1989-92

SERIE RESEARCH MEMORANDA

INTERCONNECTED NETWORKS OF QUEUES WITH RANDOMIZED ARRIVAL AND DEPARTURE BLOCKING

by

Peter de Waal
and
Nico M. van Dijk

Research Memorandum 1989-92

December 1989

VRIJE UNIVERSITEIT
Faculteit der Economische Wetenschappen en Econometrie
A M S T E R D A M

Interconnected Networks of Queues with Randomized Arrival and Departure Blocking

Peter de Waal
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

Nico van Dijk
Department of Econometrics, Free University
P.O.Box 7161, 1007 MC Amsterdam, The Netherlands

November 10, 1989

Abstract

In this paper we study two interconnected multiclass non-exponential queueing networks. Jobs can jump from one cluster to another, but subject to randomized blocking depending on the class occupancies. Such systems naturally arise in communication networks, like e.g. Metropolitan Area Networks. We present sufficient conditions for the existence of a product form equilibrium distribution under both the recirculate and the stop blocking protocol. A number of examples are given.

1980 Mathematics Subject Classification: 60K25, 90B22
Keywords & phrases: queueing networks, insensitivity, product form, blocking

1 Introduction

Product form queueing networks were first introduced in JACKSON [11]. He considered single-chain networks with Poisson input and exponential service time distributions. Gordon and Newell established product form results for closed queueing networks with exponential service (c.f. GORDON AND NEWELL [7]). Closely related to the existence of a product form equilibrium distribution is the *insensitivity property*. By insensitivity is meant that the equilibrium distribution depends only on the means of the service distributions and not on higher moments. It appeared that both the product form and the insensitivity could be explained by the notion of *local* or *partial balance* (c.f. BASKETT *et.al.* [1], CHANDY *et.al.* [3], CHANDY AND MARTIN [4], HORDIJK AND VAN DIJK [8], KELLY [13], SCHASSBERGER [19], WHITTLE [23, 24]). Roughly speaking local balance means that there exists balance in probability flows on subsets of the statespace.

Although the queueing network models as described in BASKETT *et.al.* [1] have been used extensively in the context of telecommunications, computer performance modelling and flexible manufacturing, they have some serious drawbacks that limits application in a large class of real life systems. In general blocking is not allowed and, in open networks, arrival rates may depend on the network occupancy only in an elementary way. It is for example not allowed to let the arrival rate for a customer class depend on the occupancy of another class. A first extension in this direction was presented in LAM [15]. He provided sufficient conditions for an open queueing network to prohibit arrivals or departures depending on the momentary multiclass network occupancy. However, inputs were assumed to originate from Poisson sources and arrivals and departures could only be rejected or admitted without randomization. In KAUFMAN [12] and FOSCHINI AND GOPINATH [6] a single station with Poisson input is allowed to restrict the set of feasible states to a coordinate convex set by rejecting arriving customers.

In BURMAN *et.el.* [2], HORDIJK AND VAN DIJK [9], KELLY [13], PITTEL [17] Poissonian arrivals can be blocked due to finite capacity constraints in single stations under the restrictive condition of a reversible routing. In HORDIJK AND VAN DIJK [8] a number of extensions to some special non-reversible examples are given. Blocking depending on the total configuration of a group of stations, however, was hereby not involved. Extensions in this direction can be found in VAN DIJK [5], KRZESINSKI [14], SERFOZO [20] and TOWSLEY [21]. The results in [14] and [21] though concern blocking within a cluster and not upon entering and leaving a cluster (cf. [5, Section 4.3 (iii)]). The results in [5] and [20] are limited to single class exponential structures.

In *Metropolitan Area Networks (MANs)* (c.f RUBIN AND LEE [18]) in contrast one network area may feed the other where each area itself will generally have a non-reversible and non-exponential structure, while admission for connections depends upon the total multiclass network occupancies.

This paper will study multiclass non-exponential networks with both arrival and departure blocking depending on the total customer class occupancies and with arrivals generated by another non-exponential network. Specifically, MAN applications are hereby involved.

An extension of the standard insensitivity product form results for BCMP-networks will be obtained. Particularly also, for communication purposes, such as in MAN, the result will be proven not only for the *recirculate* or *repeat blocking protocol* as standardly used in literature (e.g. LAM [15]) but also for the more realistic *stop communication protocol.* Some further extension, such as to open and mixed open-closed networks, class changes and network dependent service disciplines will be briefly discussed.

The organization of this paper is as follows: in Section 2 the queueing network model is introduced and a sufficient condition for a product form is given. Examples are presented to illustrate potential applications. The main theorem for closed networks is stated. In Section 3 we present some extensions to the model of Section 2.

2 Main results

2.1 Description of the model

In this section we give the description of the queueing network model. We let N denote the set of natural numbers and $\mathsf{N}_k = \{1, \dots, k\}$. Consider a network of N stations. The stations are grouped into two clusters named C_1 and C_2. Stations in cluster C_1 are numbered $n = 1, \dots, N_1$ and stations in C_2 are numbered $n = N_1 + 1, \dots, N$. We let $\mathcal{N}$ denote the set of stations $C_1 \cup C_2$. An example of a network with two clusters is depicted in Figure 1.

Following KELLY [13] the service disciplines at the different nodes are described by three functions f_n, ϕ_n and δ_n, that have the following interpretation.

$f_n(k)$ the speed of the server at the n-th station when k customers are present,

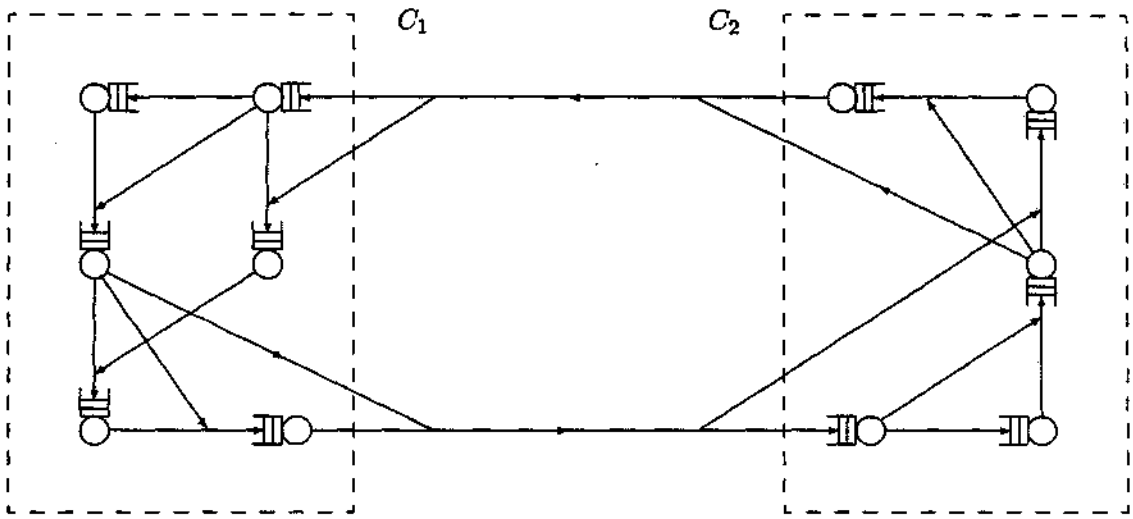

Figure 1: A network with two clusters of queueing stations.

$\phi_n(k,i)$ the fraction of the service capacity that is awarded to the customer in the i-th position at station n when k customers are present,

$\delta_n(k,i)$ the probability that a customer arriving at station n is placed at position i when k customers are present.

We assume that the shift protocol is used, i.e. if a new customer is placed at position i, then the customers at positions $i,\ldots,k$ shift to positions $i+1,\ldots,k+1$, and if a customer departs from position i, then these customers shift to positions $i,\ldots,k-1$. We also assume the buffers at all stations to be infinite. Note that from the definition of ϕ and δ we have

$$\sum_{n=1}^{k} \phi_n(k,i) = \sum_{n=1}^{k} \delta_n(k-1,i) = 1. \tag{1}$$

DEFINITION 2.1 (SYMMETRIC QUEUES) *Queue i is symmetric if and only if*

$$\phi_n(k,i) = \delta_n(k-1,i) \tag{2}$$

for all i,k. Let S denote the set of symmetric queues.

If a queue is not symmetric, then we assume that it operates under the *First Come First Served (FCFS)* queueing discipline, i.e. $\phi(k,i) = 1$ iff $i = 1$ and $\delta(k,i) = 1$ iff $i = k+1$.

In the network we distinguish K different classes of customers. The network is closed and the total number of class k customers is equal to M^k, $k = 1, \ldots, K$. On leaving queue m, a customer of class k moves to a queue n with probability R^k_{mn}. Note that

$$\sum_{n\in\mathcal{N}} R^k_{mn} = 1, \quad \forall m \in \mathcal{N}. \tag{3}$$

We assume the routing matrices $[R^k_{mn}]$ to be irreducible for each class. The routing from a station in one cluster to a station in another cluster is required to take a special form. We assume that for stations $m \in C_1$ and $n \in C_2$ the routing probability R^k_{mn} can be written as

$$R^k_{mn} = R^k_{m0} R^k_{0n} \tag{4}$$

where

$$R^k_{m0} = 1 - \sum_{l=1}^{N_1} R^k_{ml} \tag{5}$$

and

$$1 = \sum_{n=N_1+1}^{N} R^k_{0n} \tag{6}$$

and similarly for a transition from C_2 to C_1. In words this means that routing of a customer from one cluster into another is independent of the originating station. Define the *visiting ratio* θ^k_n of class k to station n as the solution of the equation

$$\theta^k_n = \sum_{m=1}^{N} \theta^k_m R^k_{mn}, \tag{7}$$

with the normalisation condition $\sum_{n=1}^{N} \theta^k_n = 1$, $k = 1, \ldots, K$. Since the routing matrices are assumed to be irreducible, the solution to this equation exists, and is unique due to the normalisation condition. By restriction (4) on the routing probabilities from one cluster to another, we have for all $k = 1, \ldots, K$

$$\sum_{n=1}^{N_1} \theta^k_n R^k_{n0} = \sum_{n=N_1+1}^{N} \theta^k_n R^k_{n0}. \tag{8}$$

This result is not surprising, since it simply states that the number of jumps from C_1 to C_2 in a routing cycle through both clusters must equal the number of jumps from C_2 to C_1. One can derive (8) quite easily from (7) by summing over $n = 1, \ldots, N_1$:

$$\sum_{n=1}^{N_1} \theta_n^k - \sum_{n=1}^{N_1}\sum_{m=1}^{N_1} \theta_m^k R_{mn}^k = \sum_{n=1}^{N_1}\sum_{m=N_1+1}^{N} \theta_m^k R_{m0}^k R_{0n}^k \Leftrightarrow \tag{9}$$

$$\sum_{m=1}^{N_1} \theta_m^k (1 - \sum_{n=1}^{N_1} R_{mn}^k) = \sum_{m=N_1+1}^{N} \theta_m^k R_{m0}^k \sum_{n=1}^{N} R_{0n}^k \Leftrightarrow \tag{10}$$

$$\sum_{m=1}^{N_1} \theta_m^k R_{m0}^k = \sum_{m=N_1+1}^{N} \theta_m^k R_{m0}^k \tag{11}$$

While travelling through the network customers pass through the stations where they request service. We assume that the service demands of customers of class k at station $n \in \mathcal{S}$ are independent and drawn from a distribution with distribution function $B_n^k(.)$. Without loss of generality we assume that these destributions are absolute continuous with density functions $\beta_n^k(.)$ and mean service rates μ_n^k. For each FCFS queueing station $n \in \mathcal{N} - \mathcal{S}$, the service requirements are required to be exponentially distributed and independent of the customerclass with mean service rate μ_n.

A state of the network is represented by a vector $\mathbf{n} = ((x_{ns}), n \in \mathcal{N}, s = 1, \ldots, k_n)$, where k_n denotes the number of customers present at queue n and x_{ns} the class of the customer at position s in station n. Define a *microstate* $\mathbf{X} = ((x_{ns}, r_{ns}), n \in \mathcal{N}, s = 1, \ldots, k_n)$, where r_{ns} is the residual service requirement of the customer at the s-th position in station n. In addition we will also use the following notations for states:

$\mathbf{X} - (n, s)$ The state that results if we remove the customer at position s in station n from state $\mathbf{X}$.

$\mathbf{X} - (n_1, s_1) + (n_2, s_2, y)$ State $\mathbf{X}$ with the customer at position s_1 in station n_1 removed and put into position s_2 at station n_2 with a residual service requirement y.

2.2 The invariance condition

A *population vector* is defined as an element of N^K. For a given state $\mathbf{X}$ we define $M_1 = (M_1^1, M_1^2, \ldots, M_1^K)$ and $M_2 = (M_2^1, M_2^2, \ldots, M_2^K)$ by

$$M_i^k = \sum_{n \in C_i} \sum_{s=1}^{k_n} 1_{(x_{ns}=k)} \quad , i = 1, 2; \;\; k = 1, ..., K. \tag{12}$$

In words the k-th component M_i^k of a population vector M_i is the total number of class k customers in cluster C_i. Since the network is closed, the sum of M_1 and M_2 must be constant. Denote this sum as $\mathcal{M} = \{M^1, ..., M^K\}$.

Next we introduce the *arrival- and departure probability functions* that describe when jumps from one cluster to another are allowed.

DEFINITION 2.2 (RECIRCULATE BLOCKING PROTOCOL)

$A^k(M_2)$ *The probability that an arrival of a class k customer, $k = 1, ..., K$, coming from C_1, is accepted at C_2, if the C_2-population is M_2. If the customer's arrival is accepted he will route into C_2 according to the normal routing probability R_{0n}^k, $n \in C_2$, otherwise he will be rerouted into C_1 according to the routing probabilities R_{0n}^k, $n \in C_1$.*

$D^k(M_2)$ *The probability that a departure of a class k customer, $k = 1, ..., K$, coming from C_2, is accepted into C_1, if the C_2-population is M_2. If the customer's departure is accepted, then he will route into C_1 according to the routing probabilities R_{0n}^k, $n \in C_1$, otherwise he will be rerouted into C_2 according to the probabilities R_{0n}^k, $n \in C_2$.*

With these functions we can disable certain jumps from one cluster to another, by setting either A or D for certain populationvectors equal to zero. With this we restrict the queueing network without controlled arrivals and departures to a subset of its original statespace. Without loss of generality we assume that this restricted statespace is irreducibel. In our analysis we will impose assumptions on A and D.

ASSUMPTION 2.3 (LAM'S CONDITION) *The arrival and departure functions A^k and D^k must satisfy for all $k = 1, ..., K$*

$$A^k(M_2) = 0 \Leftrightarrow D^k(M_2 + \mathbf{e}_k) = 0, \tag{13}$$

where $\mathbf{e}_k$ is the k-th unit vector.

This condition is identical to the one found in LAM [15, page 373]. The arrival- and departureprobabilities must thus be constructed that if an arrival of a class k customer is prohibited for a certain populationvector $\mathbf{M}$, then departures for the same class k must also be prohibited for the population vector $\mathbf{M} + \mathbf{e}_k$. This leads us to the definition of *paths*.

Definition 2.4 (Paths) *Let* $\mathbf{p}$ *and* $\mathbf{q}$ *be elements of* N^K. *If there exist an* $l \in \mathsf{N}$ *and a sequence* $\mathbf{v} : \mathsf{N}_l \to \mathsf{N}^K$, *such that either* $\mathbf{v}_{n+1} - \mathbf{v}_n = \mathbf{e}_k$ *or* $\mathbf{v}_{n+1} - \mathbf{v}_n = -\mathbf{e}_k$ *for some* $k = 1, ..., K$, *with* $\mathbf{v}_0 = \mathbf{p}$, $\mathbf{v}_l = \mathbf{q}$ *and*

$$\prod_{n=0}^{l-1} F(\mathbf{v}_n, \mathbf{v}_{n+1}) > 0, \tag{14}$$

where

$$F(\mathbf{v}_n, \mathbf{v}_{n+1}) = \begin{cases} A^k(\mathbf{v}_n) & \text{if } \mathbf{v}_{n+1} = \mathbf{v}_n + \mathbf{e}_k \\ D^k(\mathbf{v}_{n+1} & \text{if } \mathbf{v}_{n+1} = \mathbf{v}_n - \mathbf{e}_k \end{cases} \tag{15}$$

then the sequence $\mathbf{v}$ *is called a* path *from* $\mathbf{p}$ *to* $\mathbf{q}$ *and* $\mathbf{q}$ *is said to be reachable from* $\mathbf{p}$.

With the definitions of the arrival- and departureprobabilities, there exists a path from $\mathbf{p}$ to $\mathbf{q}$ when it is possible to construct with positive probability a realization of departures and arrivals between C_1 and C_2 such that the initial C_2-population is $\mathbf{p}$ and the terminating population is $\mathbf{q}$ (provided the routing probabilities admit this construction). Expression (14) is exactly the probability of this particular realization. An example of a path for a queueing network with two customerclasses is depicted in Figure 2. The solid points represent the set of possible population vectors in C_2 and the solid lines represents the arrival- and departure transitions that are allowed with positive probability. The dashed lines show an example of a path from $\mathbf{p}$ to $\mathbf{q}$.

One can easily show that the reachability as defined in Definition 2.4 is an equivalence relation (the symmetric property follows from Assumption 2.3). Since we have assumed that the restriction of arrivals and departures induces a restricted statespace that is irreducible, we have exactly one equivalence class denoted as $\mathcal{E}$. We can now state the second assumption we need for the arrival- and departure probability functions.

Assumption 2.5 *Let* $M_0 \in \mathcal{E}$ *be the initial* C_2*-population vector, i.e. the population of* C_2 *at time* $t = 0$. *For all* $\mathbf{m} \in \mathcal{E}$ *and for all paths* $\mathbf{v}$ *from* M_0 *to* $\mathbf{m}$ *the product*

$$G(\mathbf{m}) = \prod_{n=0}^{l-1} H(\mathbf{v}_n, \mathbf{v}_{n+1}) \tag{16}$$

where

$$H(\mathbf{v}_n, \mathbf{v}_{n+1}) = \begin{cases} \dfrac{A^k(\mathbf{v}_n)}{D^k(\mathbf{v}_{n+1})} & \text{if} \quad \mathbf{v}_{n+1} - \mathbf{v}_n = \mathbf{e}_k \\[2ex] \dfrac{D^k(\mathbf{v}_{n+1})}{A^k(\mathbf{v}_n)} & \text{if} \quad \mathbf{v}_{n+1} - v_n = -\mathbf{e}_k \end{cases} \tag{17}$$

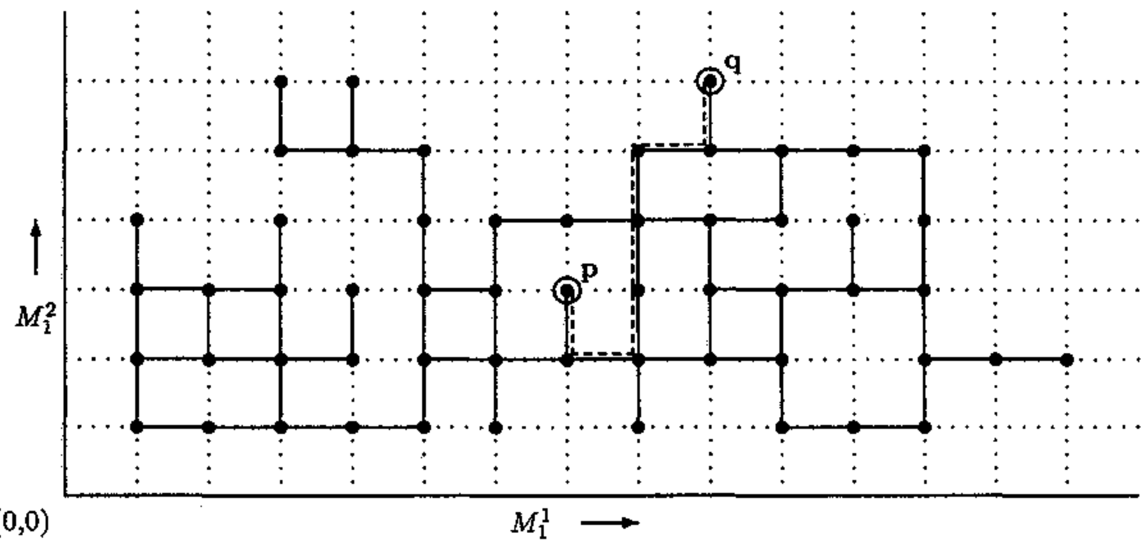

Figure 2: An example of a set of reachable population vectors and a path.

must be independent of $\mathbf{v}$.

Note that (16) is well-defined due to Assumption 2.3. In words Assumption 2.5 says that the probability to make a path from $\mathbf{p}$ to $\mathbf{q}$ must be independent of the chosen path. The expression for $G(.)$ and $H(.,.)$ is fairly complicated to allow for an irregular structure of the arrival- and departurefunctions $A(.)$ and $D(.)$ like for example in Figure 2. This complexity can be avoided if there exists a *minimal state*, i.e. a state from which a path can be made to all other reachable states only by moving customers from C_1 to C_2, i.e. $\mathbf{v}_{n+1} - \mathbf{v}_n = +\mathbf{e}_k$ for all n. The product G then becomes less complicated by choosing this minimal state as the inital state M_0. A similar statement can be made if there exists a *maximal state*, a state from which all other states can be reached by a path with $\mathbf{v}_{n+1} - \mathbf{v}_n = -\mathbf{e}_k$. Examples of networks with a minimal and a maximal state are depicted in Figure 3.

2.3 Examples

In this section we present some examples of the broad class of networks that can be modelled as the network in Section 2.1 and that satisfy the assumptions of Section 2.2. These examples and their corresponding insensitive product forms are new due to the non-exponential network input

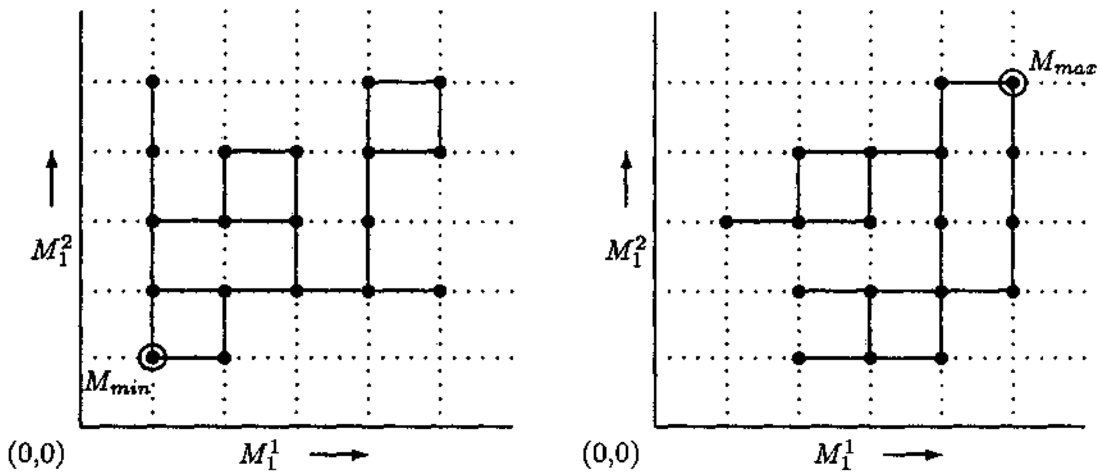

Figure 3: Examples with minimal and maximal states.

and randomized blocking. Related references for simpler cases will be mentioned as we proceed.

EXAMPLE 2.6 (COORDINATE CONVEX SETS)

Let M_{CC} be a coordinate convex subset of N^K, i.e. $\mathbf{m} = (m_1, \ldots, m_K) \in M_{CC}$ implies for all $k = 1, ..., K$, $(\mathbf{m} - \mathbf{e}_k)^+ \in M_{CC}$, where $(\mathbf{m})^+ = ((m_1)^+, \ldots, (m_K)^+)$ and $(m_k)^+ = \max(m_k, 0)$. If we define

$$A^k(\mathbf{m}) = 1_{(\mathbf{m}+\mathbf{e}_k \in M_{CC})}, \tag{18}$$

$$D^k(\mathbf{m}) = 1_{(\mathbf{m} \in M_{CC})}, \tag{19}$$

then one can show that for $M_0 = \mathbf{0}$, $\mathcal{E} = M_{CC}$.

For an example of a coordinate convex set and a non-coordinate convex set see Figure 4. If we thus start with an initially empty C_2-cluster, the C_2-population vectors will be restricted to M_{CC}. One can show that for this model we can even take $D^k \equiv 1$, since due to the definition of coordinate convexity and the definitions of the A^k's the only positive recurrent states $\mathbf{X}$ of this network have a C_2-population vector in M_{CC}. This model was introduced in LAM [15] though he considered only Poisson arrival processes. He suggested applications in store-and-forward nodes, network flow control and multiprogramming computer systems.

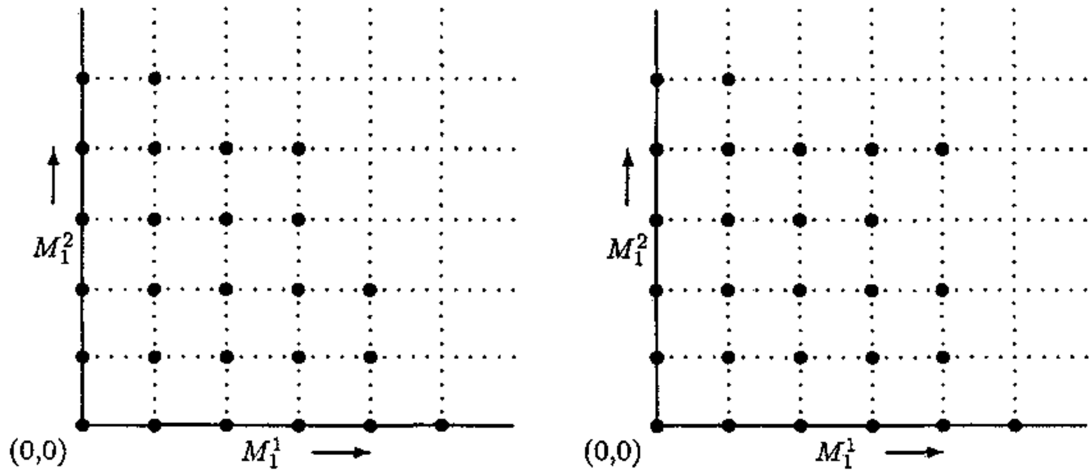

Figure 4: Example of a coordinate- and a non-coordinate convex set.

Other applications of coordinate convex sets are Metropolitan Area Network protocols as presented in RUBIN AND LEE [18]. Briefy, in interconnected MAN's we distinguish two groups of subscribers, say a group 1 and 2 with K_1 and K_2 subscribers respectively, where each group represents a local or metropolitan area network. Both within a group and in between two groups connections can be set up from one subscriber to another. If we number all subscribers $1, \ldots, K_1 + K_2$ and identify each possible connection from a source subscriber m to a destination subscriber n as a job (m, n), we can distinguish three classes of jobs, viz. local connections within group 1 and within group 2 and long-distance connections in between groups 1 and 2. The above description fits into the framework of Sections 2.1 and 2.2 where we consider a connection busy or idle depending on whether its corresponding job is in cluster C_2 or C_1. The allocation policies of RUBIN AND LEE [18] can be characterized by coordinate convex sets. An example is the *dedicated circuit allocation* policy, where L_1, L_2 and S separate circuits are available for connections within group 1, within group 2 and in between groups 1 and 2 respectively. This policy can be characterized by the coordinate convex set

$$M_{CC} = \{(m_1, m_2, m_{1-2}) | m_1 \leq L_1, m_2 \leq L_2, m_{1-2} \leq S\}$$

where m_1, m_2 and m_{1-2} represent the total number of ongoing communications within network 1, network 2 and in between networks 1 and 2. The *shared* circuit allocation policy, where the long-distance circuits can also be used for local communications, can be characterized by

$$M_{CC} = \{(m_1, m_2, m_{1-2}) | m_1 \leq L_1 + S, m_2 \leq L_2 + S, m_{1-2} \leq S - (m_1 - L_1)^+ - (m_2 - l_2)^+\},$$

where $(y)^+ = \max\{y, 0\}$. A shared allocation policy, where each long-distance connection requires a local circuit within each local area, is represented by

$$M_{CC} = \{(m_1, m_2, m_{1-2}) | m_1 + m_{1-2} \leq L_1, m_2 + m_{1-2} \leq L_2, m_{1-2} \leq S\}.$$

Extensions in this spirit are directly possible such as to introduce constraints that any subscriber m can have no more than O_m outgoing connections take place at the same time. The examples remain valid with the additional restriction to M_{CC} of

$$\sum_n 1_{\{(m,n) \in C_2\}} \leq O_m.$$

Similarly, input constraints, say I_n for subscriber n are realized by

$$\sum_m 1_{\{(m,n) \in C_2\}} \leq I_n.$$

Exclusion of busy connections (m, n) and (n, m) at the same time, reflecting one-way communication, can be put in the framework of coordinate convex sets by the constraint

$$1_{((m,n) \in C_2)} + 1_{((n,m) \in C_2)} \leq 1.$$

Note that in the interconnection networks of RUBIN AND LEE [18] both idle and busy times for connections were assumed to be exponentially distributed, whereas in our framework this condition can be relaxed.

EXAMPLE 2.7

Consider the network of Section 2.1 with two customerclasses. Let $a, b \in \mathsf{N}$, $0 < a < b$ and consider the functions

$$D^2((m^1, m^2)) = 1_{(m^1 \leq a)} + 1_{(a < m^1 \leq b)} \left(\frac{m^1 - a}{b - a} \right), \tag{20}$$

$$A^2((m^1, m^2)) = 1_{(m^1 \leq b)} \tag{21}$$

and $D^1 \equiv A^1 \equiv 1$.

In this example class 2 customers are served normally when the number of class 1 customers does not exceed a. If the number of class 1 customers exceeds b then the servicing of class 2 customers is stopped completely. Between these two levels the service rate for class 2 is gradually decreased. One can view class 1 as having a higher priority over class 2.

EXAMPLE 2.8 (ERLANG'S IDEAL GRADING)

Consider a queueing network where C_2 consists of N_2 servers, that can service at most one customer at any time. Each time a customer wants to jump from cluster C_1 to C_2, he hunts over R randomly chosen servers for a free server, where $R \leq N_2$ is a fixed integer. If it finds a free server, the customer will receive his service, otherwise it recirculates to C_1. The times spent in clusters C_1 and C_2 are usually referred to as think- and busy times. This model can be parametrized by

$$A^k(\mathbf{m}) = \left[1 - \begin{pmatrix} \|\mathbf{m}\| \\ R \end{pmatrix} / \begin{pmatrix} N_2 \\ R \end{pmatrix}\right], \tag{22}$$

where $\begin{pmatrix} n \\ m \end{pmatrix} = 0$ for $n < m$ and $\|\mathbf{m}\| = \|(m^1, ..., m^K)\| = \sum_{k=0}^{K} m^k$. When C_1 consists solely of N_1 sources with exponentially distributed think times, the model reduces to *Erlang's ideal grading.* To satisfy Assumption 2.3 it is sufficient to set $D^k(\mathbf{m}) = 0$ for $\|\mathbf{m}\| > N_2$.

2.4 The productform equilibrium distribution

In this section we present the main result of the paper. We assume without loss of generality that the Markov process corresponding to the queueing network as described in the preceding sections, has a unique equilibrium density function, that is continuously differentiable in all its arguments.

THEOREM 2.9

Let M_0 be the C_2-population vector at time 0, then the equilibrium probability $\pi(\mathbf{X})$ of being in state $\mathbf{X}$, that has a C_2-population vector $M_2 \in \mathcal{E}$ is

$$\pi(\mathbf{X}) = CG(M_2) \prod_{n=1}^{N} \prod_{s=1}^{k_n} \frac{\theta_n^{x_{ns}}}{f_n(s)} (1 - B_n^{x_{ns}}(r_{ns})) \tag{23}$$

where C is a normalisation constant.

PROOF. We prove that (23) satisfies the forward Kolmogorov differential equations. These state that the global probability flow out of a each state should equal the flow into the state. They can be written as

$$0 = \sum_n \sum_s \Xi(\mathbf{X}, n, s) \tag{24}$$

where $\Xi(\mathbf{X}, n, s)$ is a local probability flow:

$$\Xi(\mathbf{X}, n, s) := \frac{\partial \pi(\mathbf{X})}{\partial r_{ns}} f_n(k_n)\phi_n(k_n, s) \tag{25}$$

$$+ \sum_{\substack{m \in C_1 \\ m \neq n}} \sum_{t=1}^{k_m^X+1} \pi(\mathbf{X} - (n,s) + (m,t,0^+))\phi_m(k_m^X+1,t) f_m(k_m^X+1)\delta_n(k_n-1,s)\beta_n^{x_{ns}}(r_{ns}) *$$

$$*\{R_{mn}^{x_{ns}} + R_{m0}^{x_{ns}}(1 - A^{x_{ns}}(M_2))R_{0n}^{x_{ns}}\} \tag{26}$$

$$+ \sum_{t=1}^{k_n} \pi(\mathbf{X} - (n,s) + (n,t,0^+))\phi_n(k_n,t) f_n(k_n)\delta_n(k_n-1,s)\beta_n^{x_{ns}}(r_{ns}) *$$

$$*\{R_{nn}^{x_{ns}} + R_{n0}^{x_{ns}}(1 - A^{x_{ns}}(M_2))R_{0n}^{x_{ns}}\} \tag{27}$$

$$+ \sum_{m \in C_2} \sum_{t=1}^{k_m^X+1} \pi(\mathbf{X} - (n,s) + (m,t,0^+))\phi_m(k_m^X+1,t) f_m(k_m^X+1)\delta_n(k_n-1,s)\beta_n^{x_{ns}}(r_{ns}) *$$

$$*\{R_{m0}D^{x_{ns}}(M_2 + \mathrm{e}_{x_{ns}})R_{0n}^{x_{ns}}\} \tag{28}$$

for $n \in C_1$ and analogously for $n \in C_2$.

PROOF FOR SYMMETRIC QUEUES. First we prove that if n is a symmetric queue, then $\pi(.)$ as in (23) satisfies the so-called *Job Local Balance* equations. These equations state that there should be a balance of the loss in probability density of state $\mathbf{X}$ due to a departure of *one* job against the gain in probability density due to an arrival of *the same* job. From the definition of Ξ it is clear that this corresponds to the equation $\Xi(\mathbf{X}, n, s) = 0$, since in this definition (25) corresponds to flow out of state $\mathbf{X}$ due to a change in the residual service time r_{ns} of the job at position (n, s) and (26), (27) and (28) are flows into state $\mathbf{X}$ due to an arrival of a customer from queue $m \in C_1$, $m \neq n$, from queue n and from queue $m \in C_2$ respectively.

Assume that $n \in C_1$ is a symmetric queue. Note that according to (23)

$$\pi(\mathbf{X}-(n,s)+(m,t,0^+)) = \begin{cases} \dfrac{\pi(\mathbf{X}) f_n(k_n)\theta_m^{x_{ns}}}{(1-B_n^{x_{ns}}) f_m(k_m^X+1)\theta_n^{x_{ns}}} & , m \in C_1, m \neq n \\ \dfrac{\pi(\mathbf{X})}{(1-B_n^{x_{ns}})} & , m = n \\ \dfrac{\pi(\mathbf{X}) f_n(k_n)\theta_m^{x_{ns}} G(M_2+\mathbf{e}_{x_{ns}})}{(1-B_n^{x_{ns}}) f_m(k_m^X+1)\theta_n^{x_{ns}} G(M_2)} & , m \in C_2 \end{cases} \tag{29}$$

and

$$\frac{\partial \pi(\mathbf{X})}{\partial r_{ns}} = -\frac{\pi(\mathbf{X})}{1-B_n^{x_{ns}}(r_{ns})} \beta_n^{x_{ns}}(r_{ns}) \tag{30}$$

To show that $\Xi(\mathbf{X}, n, s) = 0$, we distinguish three cases:

1. $M_2 + \mathbf{e}_{x_{ns}} \notin \mathcal{E}$. (Non-admissable population vector)
2. $M_2 + \mathbf{e}_{x_{ns}} \in \mathcal{E}$ and $A^{x_{ns}}(M_2) = 0$. (Not directly admissable population vector)
3. $M_2 + \mathbf{e}_{x_{ns}} \in \mathcal{E}$ and $A^{x_{ns}}(M_2) > 0$. (Directly admissable population vector)

Since we are considering only flows due to changes in job (n, s), there should be no confusion when we omit the index x_{ns} from $A, B, \beta, \theta, \mathbf{e}$ and R. To improve the notational clarity we will also omit the index $\mathbf{X}$ from k_n and k_m.

1. $(M_2 + \mathbf{e} \notin \mathcal{E})$

Since $M_2 + \mathbf{e} \notin \mathcal{E}$, we must have $A(M_2) = 0$ and according to Assumption 2.3 also $D(M_2 + \mathbf{e}) = 0$. Consequently term (28) vanishes. With (29) the equation $\Xi(\mathbf{X}, n, s) = 0$ becomes equivalent to

$$\begin{aligned} 0 &= \phi_n(k_n, s)\theta_n \\ &+ \sum_{\substack{m \in C_1 \\ m \neq n}} \sum_{t=1}^{k_m+1} \phi_m(k_m+1, t)\theta_m \delta_n(k_n-1, s)\{R_{mn} + R_{m0}R_{0n}\} \\ &+ \sum_{t=1}^{k_n} \phi_n(k_n, t)\theta_n \delta_n(k_n-1, s)\{R_{nn} + R_{n0}R_{0n}\} \end{aligned} \tag{31}$$

Since $\delta_n(k_n - 1, s) = \phi_n(k_n, s)$ (N is symmetric) and $\sum_{t=1}^{k} \phi(k, t) = 1$ a sufficient condition for (31) is

$$\theta_n = \sum_{m \in C_1} \theta_m \{R_{mn} + R_{m0}R_{0n}\}. \tag{32}$$

Using the traffic equations (7) and the observation (8) this condition is verified.

2. ($M_2 + \mathbf{e} \in \mathcal{E}$ and $A(M_2) = 0$)

Due to Assumption 2.3, $A(M_2) = 0$ is equivalent to $D(M_2+\mathbf{e}) = 0$. Substituting this in $\Xi(\mathbf{X}, n, s) = 0$, again yields (31).

3. ($M_2 + \mathbf{e} \in \mathcal{E}$ and $A(M_2) > 0$)

Due to Assumption 2.3 we have $D(M_2 + \mathbf{e}) > 0$, so there exists a path from M_2 to $M_2 + \mathbf{e}$. Since $\mathbf{v}_0 := M_2$ and $\mathbf{v}_1 := M_2 + \mathbf{e}$ is such a path, we have from (16)

$$G(M_2 + \mathbf{e}) = G(M_2)\frac{A(M_2)}{D(M_2 + \mathbf{e})}. \tag{33}$$

The equation $\Xi(\mathbf{X}, n, s) = 0$ then yields

$$\begin{aligned} &\theta_n \phi_n(k_n, s) &(34)\\ &= \sum_{\substack{m \in C_1 \\ m \neq n}} \sum_{t=1}^{k_m+1} \theta_m \phi_m(k_m + 1, t)\delta_n(k_n - 1, s)\{R_{mn} + R_{m0}(1 - A(M_2))R_{0n}\} \\ &+ \sum_{t=1}^{k_n} \theta_n \phi_n(k_n, t)\delta_n(k_n - 1, s)\{R_{nn} + R_{n0}(1 - A(M_2))R_{0n}\} \\ &+ \sum_{m \in C_2} \sum_{t=1}^{k_m+1} \theta_m \phi_m(k_m + 1, t)\delta_n(k_n - 1, s)\{R_{m0}A(M_2)R_{0n}\} \end{aligned}$$

Using the same arguments as in Case 1 this equation is implied by

$$\theta_n = \sum_{m \in C_1} \{R_{nn} + R_{n0}(1 - A(M_2))R_{0n}\} + \sum_{m \in C_2} R_{m0}A(M_2)R_{0n} \tag{35}$$

or equivalently, using $R_{m0}R_{0n} = R_{mn}$ for $n \in C_1$, $m \in C_2$,

$$\theta_n = \sum_{m \in C_1 \cup C_2} \theta_m R_{mn} + \{\sum_{m \in C_1} \theta_m R_{m0} - \sum_{m \in C_2} \theta_m R_{m0}\}(1 - A(M_2))R_{0n}. \tag{36}$$

This equation is true due to (7) and (8).

The proof of the Job Local Balance equation for a symmetric station in C_2 proceeds in a similar manner.

Proof for FCFS queues. Now assume that $n \in C_1$ and n is a FCFS queue. It is obvious that we cannot have Job Local Balance in this case, since all the probability flow from departures is due to the customer in the first position and arrivals bring in customers in the last position. For FCFS queues we have, however, ***Station Balance***, i.e. a balance in the gain of probability density of state

$\mathbf{X}$ due to arrivals in the queue against the loss in probability density due to departures from the queue, or equivalently $\sum_{s=1}^{k_n} \Xi(\mathbf{X}, n, s) = 0$. Since $\delta_n(k_n - 1, s) = 1$ iff $s = k_n$, and $\phi(k_n, s) = 1$ iff $s = 1$, we have $\Xi(\mathbf{X}, n, s) = 0$ for $2 = 1, \ldots, k_n - 1$. This yields

$$\sum_{s=1}^{k_n} \Xi(\mathbf{X}, n, s) = \frac{\partial \pi(\mathbf{X})}{\partial r_{n1}} \beta_n(r_{n1}) f_n(k_n)$$

$$+ \sum_{m \in C_2} \sum_{t=1}^{k_m+1} \pi(\mathbf{X} - (n, k_n) + (m, t, 0^+)) f_m(k_m + 1) \phi_m(k_m + 1, t) \beta_n(r_{nk_n}) \{R_{m0} D(M_2 + \mathbf{e}) R_{0n}\}$$

$$+ \qquad \pi(\mathbf{X} - (n, k_n) + (n, 1, 0^+)) f_n(k_n) \beta_n(r_{nk_n}) \{R_{nn} + R_{n0}(1 - A(M_2)) R_{0n}\} \tag{37}$$

$$+ \sum_{\substack{m \in C_1 \\ m \neq n}} \sum_{t=1}^{k_m+1} \pi(\mathbf{X} - (n, k_n) + (m, t, 0^+)) f_m(k_m + 1) \phi_m(k_m + 1, t) *$$

$$* \beta_n(r_{nk_n}) \{R_{mn} + R_{m0}(1 - A(M_2)) R_{0n}\}$$

Where the functions A, D and the routing probabilities R should be read with index x_{nk_n}. Since $B_n(.)$ and $\beta_n(.)$ are independent of the customerclass and $B_n(r) = 1 - e^{-\mu_n r}$, we have $\beta_n(r)/(1 - B_n(r)) = \mu_n$ and (37) reduces to (32) for Case 1 and 2, and (35) for Case 3. □

COROLLARY 2.10 *For each state* $\mathbf{n}$ *the equilibrium distribution is given by*

$$\pi(\mathbf{n}) = CG(M_2) \prod_{n=1}^{N} \prod_{s=1}^{k_n} \frac{\theta_n^{x_{ns}}}{\mu_n^{x_{ns}} f_n(s)} \tag{38}$$

PROOF. By integrating equation (23) from zero to infinity for each r_{ns}, $s = 1, \ldots, k_n$, $n = 1, \ldots, N$. □

3 Extensions

In this section we discuss some extensions for the queueing network model introduced in Section 2. We present an alternative for the protocol that is used when a customer's jump is rejected. We also introduce class changes, open and mixed networks and network dependent service disciplines.

3.1 Stop protocol

In the queueing network that was introduced in Section 2, customers who were prohibited to jump from one cluster to another, were rerouted back into the originating cluster. This protocol, called the *recirculate protocol*, can be viewed as an extension of the *repeat protocol*. The latter, that is widely used in telecommunication, requires that customers repeat their service in the same station when a jump is prohibited. The product form results in the literature are all established under the repeat or, in the exponential case, recirculate protocol (cf. HORDIJK AND VAN DIJK [9], KAUFMAN [12], KELLY [13], LAM [15], SERFOZO [20], WALRAND [22]), YAO AND BUZACOTT [25]. In practice, however, a total service may comprise a number of exponential stages and is no longer of exponential form. Particularly in communication, though, the recirculate protocol, which requires all service stages to be repeated, seems unrealistic.

Alternatively, the *stop protocol* is more realistic, where service of a customer is interrupted when he is blocked. In the case of exponential service times the equivalence between the stop and recirculate (repeat) protocol is intuitively obvious and has been shown (e.g. ONVURAL AND PERROS [16]). For non-exponential situations, however, this is far from obvious and generally not true.

DEFINITION 3.1 (STOP BLOCKING PROTOCOL) *Consider the network as described in Section 2. When the occupancy of cluster C_2 is equal to M_2, then for all classes k, the amount of service acquired by class k customers in C_1 and C_2 is reduced by a factor $A^k(M_2)$ and $D^k(M_2)$ respectively. For an example, a customer in position s at station $n \in C_1$ will now receive service at a rate $f_n(k_n)\phi_n(k_n, s)A^{x_{ns}}(M_2)$.*

THEOREM 3.2 *Assume that all stations are symmetric. Let M_0 be the C_2-population vector at time 0, then also under the stop protocol the equilibrium probability $\pi(\mathbf{X})$ of being in state $\mathbf{X}$ is given by (23).*

PROOF. We have to prove that (23) satisfies the forward Kolmogorov equations. We prove this in a similar way as in Theorem 2.9. Note that the local probability flow $\Xi(\mathbf{X}, n, s)$ (25) under the stop protocol is given by

$$\Xi(\mathbf{X}, n, s) = \frac{\partial \pi(\mathbf{X})}{\partial r_{ns}} f_n(k_n)\phi_n(k_n, s)A^{x_{ns}}(M_2)$$

$$
\begin{aligned}
&+ \sum_{\substack{m\in C_1\\ m\neq n}} \sum_{t=1}^{k_m^X+1} \pi(\mathbf{X}-(n,s)+(m,t,0^+))\phi_m(k_m^X+1,t)f_m(k_m^X+1)\delta_n(k_n-1,s)\beta_n^{x_{ns}}(\tau_{ns}) * \\
&\qquad * \{R_{mn}^{x_{ns}} A^{x_{ns}}(M_2)\} \\
&+ \sum_{t=1}^{k_n} \pi(\mathbf{X}-(n,s)+(n,t,0^+))\phi_n(k_n,t)f_n(k_n)\delta_n(k_n-1,s)\beta_n^{x_{ns}}(\tau_{ns}) * \\
&\qquad * \{R_{nn}^{x_{ns}} + A^{x_{ns}}(M_2)\} \\
&+ \sum_{m\in C_2} \sum_{t=1}^{k_m^X+1} \pi(\mathbf{X}-(n,s)+(m,t,0^+))\phi_m(k_m^X+1,t)f_m(k_m^X+1)\delta_n(k_n-1,s)\beta_n^{x_{ns}}(\tau_{ns}) * \\
&\qquad * \{R_{m0} D^{x_{ns}}(M_2+\mathbf{e}_{x_{ns}}) R_{0n}^{x_{ns}}\}
\end{aligned}
$$

If we have $A^{x_{ns}}(M_2) = 0$ (Case 1 and 2 in the proof of Theorem 2.9), then $D^{x_{ns}}(M_2 + \mathbf{e}_{x_{ns}}) = 0$ and $\Xi(\mathbf{X}, n, s) = 0$. If $A^{x_{ns}}(M_2) > 0$, then $D^{x_{ns}}(M_2 + \mathbf{e}_{x_{ns}}) > 0$ and $\Xi(\mathbf{X}, n, s) = 0$ is equivalent to

$$
\begin{aligned}
&\theta_n\phi_n(k_n, s) \\
&= \sum_{\substack{m\in C_1\\ m\neq n}} \sum_{t=1}^{k_m+1} \theta_m\phi_m(k_m+1,t)\delta_n(k_n-1,s)R_{mn}A(M_2) && (39) \\
&+ \sum_{t=1}^{k_n} \theta_n\phi_n(k_n,t)\delta_n(k_n,s)R_{nn}A(M_2) && (40) \\
&+ \sum_{m\in C_2} \sum_{t=1}^{k_m+1} \theta_m\phi_m(k_m+1,t)\delta_n(k_n-1,s)R_{m0}A(M_2)R_{0n} && (41)
\end{aligned}
$$

This equation is again implied by the traffic equations (7). □

3.2 Type changes

Like the queueing networks in BASKETT *et.al.* [1] and LAM [15] we can extend our networks by allowing customers to change class when jumping. Suppose that a customer of class k who completes service at station n jumps to station m and becomes a customer of class l with probability $R_{nk;ml}$. Assume that the routing matrix $[R_{nk;ml}]$ defines a Markov chain with states (n,k) that can be decomposed into K ergodic routing subchains $\mathcal{R}_1, \ldots, \mathcal{R}_K$. We denote *type* as a membership of one of the subchains and we assume the number of customers of each type to be fixed. Define population vectors $\mathcal{M}, M_1$ and M_2, where the k-th component now denotes the number of customers of type k, and define the arrival- and departure functions as $A^k(.)$ and $D^k(.)$ for each *type* k. Service time distributions in symmetric stations are still allowed to be dependent of customerclass. For this

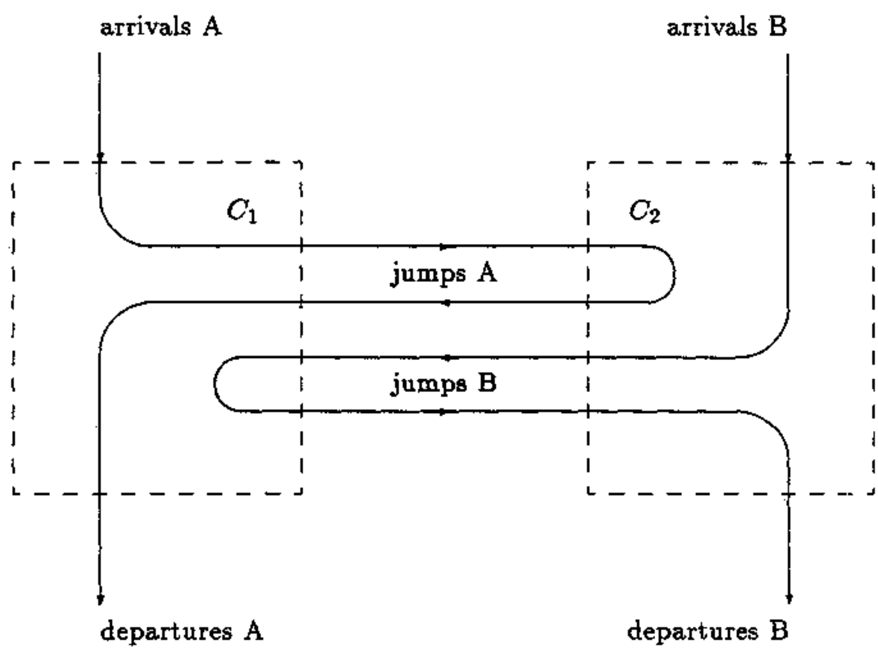

Figure 5: Mixed open and closed networks

network Theorem 2.9 still holds, but with θ_n^k obtained from slightly more detailed traffic equations (cf. LAM [15]).

3.3 Mixed open and closed networks

Extensions can be given where externally arriving jobs, generated by Poisson sources, can enter a cluster, provided they can depart the system only from that cluster (cf. Figure 5).

With the technique as described in HORDIJK AND VAN DIJK [10, pages 433–435] stations in the network can be replaced by infinite sources, thus extending the results to open and mixed open-closed networks.

3.4 Network dependent service disciplines

As in CHANDY AND MARTIN [4] the service discipline functions δ and ϕ can be made totally network dependent, provided the symmetry at non-exponential stations is not violated. As in VAN DIJK [5] and SERFOZO [20] the service capacity function $f_i(\mathbf{k})$ at station i when the state of the network is

$\mathbf{k}$, can be taken to be of the form

$$f_i(\mathbf{k}) = \frac{\chi(\mathbf{k} - \mathbf{e}_i)}{\psi(\mathbf{k})}.$$

Here $\chi(.)$ and $\psi(.)$ are arbitrary strictly positive functions and $\mathbf{k} = (k_1, \ldots, k_{N_1+N_2})$ is the station occupancy vector with k_n the number of customers at station n. These state dependent service rates lead to the stationary probability of being in state $\mathbf{X}$

$$\pi(\mathbf{X}) = CG(M_2)\psi(\mathbf{k}) \prod_{n=1}^{N} \prod_{s=1}^{k_n} \theta_n^{x_{ns}}(1 - B_n^{x_{ns}}(r_{ns}))$$

where $\mathbf{k}$ is the station occupancy vector when the microstate is $\mathbf{X}$.

References

[1] F. Baskett, K.M. Chandy, R.R. Muntz, and F.G. Palacios (1975). Open, Closed, and Mixed Networks of Queues with Different Classes of Customers. *J. Assoc. Comput. Mach.*, 22, 248–260.

[2] D.Y. Burman, J.P. Lehoczky, and Y. Lin (1982). Insensitivity of Blocking Probabilities in a Circuit Switching Network. *J. Appl. Prob.*, 21, 850–859.

[3] K.M. Chandy, J.H. Howard, and D.F. Towsley (1977). Product Form and Local Balance in Queueing Networks. *J. Assoc. Comput. Mach.*, 24, 250–263.

[4] K.M. Chandy and J. Martin (1983). A Characterization of Product Form Queueing Networks. *J. Assoc. Comput. Mach.*, 30, 250–263.

[5] N.M. Van Dijk (1989). *Product Froms for Queueing Networks with Limited Clusters.* Research Memorandum 1989-29, Free University, Amsterdam.

[6] G.J. Foschini and B. Gopinath (1983). Sharing Memory Optimally. *IEEE Trans. Comm.*, COM-31, 352–359.

[7] W.J. Gordon and G.F. Newell (1967). Closed Queueing Systems with Exponential Servers. *Oper. Res.*, 15, 254–265.

[8] A. HORDIJK AND N. VAN DIJK (1983). Networks of Queues. In *Modelling and Performance Evaluation Methodology*, F. Baccelli and G. Fayolle (eds.), 151–205, Springer-Verlag, Berlin.

[9] A. HORDIJK AND N. VAN DIJK (1981). Networks of Queues with Blocking. In *Performance '81*, F.J. Kylstra (ed.), 51–65, North-Holland.

[10] A. HORDIJK AND N. VAN DIJK (1981). Stationary Probabilities for Networks of Queues. In *Applied Probability - Computer Science, The Interface, vol. II*, R.L. Disney and T.J. Ott (eds.), 423–451, Birkhauser, Boston.

[11] J.R. JACKSON (1957). Networks of Waiting Lines. *Oper. Res.*, 5, 518–521.

[12] J.S. KAUFMAN (1981). Blocking in a Shared Resource Environment. *IEEE Trans. Comm.*, COM-29, 1474–1481.

[13] F.P. KELLY (1979). *Reversibility and Stochastic Networks.* Wiley, New York.

[14] A.E. KRZESINSKI (1987). Multiclass Queueing Networks with State-Dependent Routing. *Perf. Eval.*, 7, 125–143.

[15] S.S. LAM (1977). Queueing Networks with Population Size Constraints. *IBM Journal of Research and Development*, 21, 370–378.

[16] R.O. ONVURAL AND H.C. PERROS (1986). On Equivalencies of Blocking Mechanisms in Queueing Networks with Blocking. *Oper. Res. Letters*, 5, 293–297.

[17] B. PITTEL (1979). Closed Exponential Networks of Queues with Saturation. The Jackson-type Stationary Distribution and its Asymptotic Analysis. *Math. Oper. Res.*, 4, 357–378.

[18] I. RUBIN AND J.K. LEE (1988). Performance Analysis of Interconnected Metropolitan Area Circuit-Switched Telecommunications Networks. *IEEE Trans. Comm.*, COM-36, 171–185.

[19] R. SCHASSBERGER (1978). Insensitivity of Steady State Distributions of Generalised Semi-Markov Processes with Speeds. *Adv. Appl. Prob.*, 10, 836–851.

[20] R.F. SERFOZO (1988). *Markovian Network Processes with System-Dependent Transition Rates.* Research Report, Georgia Institute of Technology, Atlanta.