SERIE RESEARCH MEMORANDA

NETWORK OF QUEUES WITH
SERVICE ANTICIPATING ROUTING

N.M. van Dijk
M. Rumsewicz

Researchmemorandum 1988-5 Feb. '88

VRIJE UNIVERSITEIT
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN
EN ECONOMETRIE
AMSTERDAM

NETWORKS OF QUEUES WITH SERVICE ANTICIPATING ROUTING

Nico M. van Dijk*† and Michael Rumsewicz**

Abstract

Queueing networks are studied with general service requirements and routing depending upon the next required service time. A non-standard product form expression is obtained that averages out dependencies of routing and service times.

Keywords

NETWORKS OF QUEUES, ANTICIPATING ROUTING, PRODUCT FORM

* Free University, Amsterdam, The Netherlands
† This paper has been written during a visit to the University of Adelaide.
** University of Adelaide, South Australia

1. Introduction

Motivated by Jackson's historical product form results (c.f. Jackson (1957)), queueing networks have been extensively studied over recent decades in applications areas such as telecommunications, computer performance evaluation and flexible manufacturing (c.f. Kelly (1979), Buzacott and Yao (1986), Walrand (1988) and references therein). In particular, various sufficient, as well as some necessary, conditions for product form expressions and the related insensitivity property have been established (for example, König and Jansen (1974), Baskett *et al* (1975), Chandy *et al* (1977), Schassberger (1978), Kelly (1979), Chandy and Martin (1983), Hordijk and van Dijk (1983), Whittle (1985)). By insensitivity it is meant that the steady state distribution is determined only by the means of the service times and not their distributional forms. Roughly speaking, these product form and insensitivity results turn out to be based upon notions of partial balance.

The product form results reported, however, do not allow routing (probabilities) from one station to another to depend upon the next required service amount, as may naturally arise in various present day applications. For example, in manufacturing one may typically encounter a structure as depicted in Figure 1, where the assemblage of parts constitutes several stages with more workstations being available at some stages. Which workstation is actually to be used may depend upon the required processing time of a particular part, determined by, for instance, a random distribution or the part type. It is noted here that by employing this so-called anticipating routing one can also allow the service time of a job to depend on the station from which it has come. Anticipating routing thus also involves history dependence.

Figure 1 near here

This paper aims to generalise standard product form results to queueing networks with anticipating routing. A product form expression will be obtained that averages out the dependencies between routing and service. Though standard in form, this expression is non-standard in that the means of the service distributions are replaced by specially computed parameters. These parameters, however, have the natural interpretation of mean service times such as observed by the system in equilibrium. The computation of these parameters comes down to solving the standard traffic equations with routing probabilities averaged over the known service distributions. It is interesting to note, moreover, that while conventionally partial balance yields insensitivity results, in this paper it is used to obtain results that explicitly depend on service distributional forms. The results of this paper partially relate to recent age dependent routing results of Rumsewicz and Henderson (1987), but are in fact complementary and require a more complicated averaging procedure.

To highlight the essential feature of anticipating routing and its consequences, the presentation will first be restricted to closed systems and one job type. The extended formula for open systems, multiple job types and more general service rates is presented and briefly argued as being direct. As particular examples we show that the standard Jackson network and networks with mixed stations are special cases.

2. Model, product form result and examples

2.1 Model

Consider a closed network comprising of a set of labelled queues $\mathcal{N} = \{1, 2, \ldots, N\}$ and M customers. The state of the network can be represented by a vector $\mathbf{n} = (n_1, n_2, \ldots, n_N)$, where n_i is the number of customers present at queue i. Assume that service facility i works at a rate $\phi_i(n_i)$ when n_i customers are present and a proportion $\gamma_i(l, \mathbf{n})$ of this effort is dedicated to the customer in position l when the state is $\mathbf{n}$. If, after the arrival of a customer to queue i the state of the network is $\mathbf{n}$, the customer moves into position l with probability $\delta_i(l, \mathbf{n})$. When a customer leaves position l in queue i the customers in positions $l+1,\ldots,n_i$ move to positions $l,\ldots,n_i-1$ respectively. Similarly, when a customer moves into position l the customers in positions $l,\ldots,n_i-1$ move into positions $l+1,\ldots,n_i$. Note that

$$\sum_{l=1}^{n_i} \gamma_i(l,\mathbf{n}) = \sum_{l=1}^{n_i} \delta_i(l,\mathbf{n}) = 1. \tag{2.1}$$

Definition : Queue i is symmetric if and only if $\gamma_i(l,\mathbf{n}) = \delta_i(l,\mathbf{n})$ for all $l,\mathbf{n}$.

Let the set of symmetric queues be $\mathcal{S} \subseteq \mathcal{N}$.

On leaving queue $i \in \mathcal{N}$, a customer moves to a queue $j \in \mathcal{S}$, and requires a service $y \in (0,\infty)$ with probability density $p_{ij}(y)$. The customer moves to a queue $j \in \mathcal{N}-\mathcal{S}$ with probability p_{ij} and requires a service time drawn from a negative exponential distribution with mean μ_i^{-1}. For $i \in \mathcal{N}, j \in \mathcal{S}$, let

$$p_{ij} = \int_0^\infty p_{ij}(y)dy \tag{2.2}$$

and note that

$$\sum_{j\in\mathcal{N}} p_{ij} = 1, \quad \forall i \in \mathcal{N}. \tag{2.3}$$

Without loss of generality assume that the probability matrix $[p_{ij}]$ is irreducible. By also recalling (3), there thus exists a unique probability distribution $\{\lambda_1,\lambda_2,\ldots,\lambda_N\}$ satisfying the traffic equations

$$\lambda_i = \sum_{j\in\mathcal{N}} \lambda_j p_{ji}, \quad \forall i \in \mathcal{N}. \tag{2.4}$$

Define a microstate $\mathbf{x} = (x_{il}, i \in \mathcal{N}, l = 1,\ldots,n_i)$, where x_{il} is the residual service requirement of the customer in position l of queue i and let $\mathbf{x}-(i,l)+(j,m,y)$ be the state created from $\mathbf{x}$ by the removal of the customer in position l of queue i and the insertion of a customer with residual service time y in position m of queue j. Again without loss of generality, assume that the corresponding Markov process has a unique equilibrium density function which is continuously differentiable in all its arguments.

2.2 Product form result

Theorem 2.1

With C a normalising constant, for any feasible $\mathbf{x}$ the equilibrium density is given by

$$\pi(\mathbf{x}) = C \prod_{i \in \mathcal{N}} \prod_{l=1}^{n_i} \left(\frac{\lambda_i}{\phi_i(l)} \right) (1 - G_i(x_{il})), \tag{2.5}$$

where

$$G_i(y) = \begin{cases} 1 - \exp(-\mu_i y) & \text{for } i \in \mathcal{N} - \mathcal{S} \\ \sum_{j \in \mathcal{N}} \int_0^y \lambda_j p_{ji}(u) du / \lambda_i & \text{for } i \in \mathcal{S}. \end{cases} \tag{2.6}$$

Remark : It is interesting to notice the form that $G_i(.)$ takes for $i \in \mathcal{S}$ as the actual service time for a customer at queue i depends on the queue that it came from. When considered in equilibrium, however, queue i behaves as if it is receiving customers in a Poisson stream at rate λ_i, according to the traffic equations (2.4). The service at queue i then seems to ignore routing and source dependencies by averaging over the respective service time density functions according to the relative arrival rates $\lambda_j p_{ji}/\lambda_i$. Correspondingly, $G_i(.)$ has the interpretation that one would expect by simply averaging all of the features of the process per queue while ignoring queue dependencies.

Proof : Recalling that $\pi(\mathbf{x})$ is assumed to be continuously differentiable, the supplemented global balance equations require that

(2.7)

$$\sum_{i\in\mathcal{N}-\mathcal{S}}\Bigg[\sum_{l=1}^{n_i}\frac{\partial}{\partial x_{il}}\pi(\mathbf{x})\gamma_i(l,\mathbf{n})\phi_i(n_i) + \sum_{j\in\mathcal{N}}\sum_{m=1}^{n_j+1}\pi(\mathbf{x}-(i,l)+(j,m,0^+))\gamma_j(m,n_j+1)\phi_j(n_j+1)p_{ji}\delta_i(l,\mathbf{n})\mu_i\exp(-\mu_i x_{il})\Bigg]$$

$$+\sum_{i\in\mathcal{S}}\sum_{l=1}^{n_i}\Bigg[\frac{\partial}{\partial x_{il}}\pi(\mathbf{x})\gamma_i(l,\mathbf{n})\phi_i(n_i) + \sum_{j\in\mathcal{N}}\sum_{m=1}^{n_j+1}\pi(\mathbf{x}-(i,l)+(j,m,0^+))\gamma_j(m,n_j+1)\phi_j(n_j+1)p_{ji}(x_{il})\delta_i(l,\mathbf{n})\Bigg]$$

$$=0.$$

It is assumed that the set of equations (2.7) has a unique, non-negative solution. When expression (2.5) is substituted into the first term of equation (2.7), for each $i \in \mathcal{N}-\mathcal{S}$, the term in [ ] reduces to

(2.8)

$$\begin{aligned}\pi(\mathbf{x})\sum_{l=1}^{n_i}&\left[-\mu_i\gamma_i(l,\mathbf{n})\phi_i(n_i)+\sum_{j\in\mathcal{N}}\sum_{m=1}^{n_j+1}\frac{\lambda_j}{\lambda_i}\phi_i(n_i)\gamma_j(m,n_j+1)p_{ji}\delta_i(l,\mathbf{n})\mu_i\right]\\ &=\pi(\mathbf{x})\frac{\mu_i\phi_i(n_i)}{\lambda_i}\left[-\lambda_i+\sum_{j\in\mathcal{N}}\lambda_j p_{ji}\right],\\ &=0,\end{aligned}$$

where the first equality follows from (2.1) and the second from (2.4).

Substituting expression (2.5) into the second term of equation (2.7), for each $i \in \mathcal{S}$ and position $l = 1, 2, \ldots, n_i$, the term in [ ] reduces to

(2.9)

$$\begin{aligned}\frac{\pi(\mathbf{x})}{1-G_i(x_{il})}\Bigg[&-\frac{d}{dx_{il}}G_i(x_{il})\gamma_i(l,\mathbf{n})\phi_i(n_i)\\ &+\sum_{j\in\mathcal{N}}\sum_{m=1}^{n_j+1}\frac{\lambda_j}{\lambda_i}\phi_i(n_i)\gamma_j(m,n_j+1)p_{ji}(y)\delta_i(l,\mathbf{n})\Bigg]\\ &=\pi(\mathbf{x})\phi_i(n_i)\gamma_i(l,\mathbf{n})\left[-\frac{d}{dx_{il}}G_i(x_{il})+\sum_{j\in\mathcal{N}}\lambda_j p_{ji}(y)/\lambda_i\right]=0,\end{aligned}$$

where the first first equality follows from (2.1) and $\gamma_i(l, \mathbf{n}) = \delta_i(l, \mathbf{n})$, for $i \in \mathcal{S}$, and the second from the integral expression (2.6) for $i \in \mathcal{S}$.

Recalling the uniqueness assumption, expression (2.5) is thus the equilibrium density for any feasible $\mathbf{x}$. ■

Remark : Note that in the proof given above, we actually verified balance per queue for non-symmetric queues and per position for symmetric queues. These notions of partial balance are well-known in the literature (c.f. Chandy *et al* (1977), Kelly(1979), Hordijk and van Dijk (1983)) for proving product form results. Their applicability for the present non-standard network is therefore of interest in itself. In particular, while in standard frameworks balance per position will lead to insensitivity results (that is, results independent of the distributional forms of the service requirements except through their means, e.g. König and Jansen, (1974), Barbour (1976), Schassberger (1978)) in the present framework it is used to obtain a result which is dependent upon the service time distributions. However, the following "insensitivity analogue" can be concluded.

Corollary 2.2

For $i \in \mathcal{N}$, let μ_i^{-1} be the mean of $G_i(.)$, that is,

$$\mu_i^{-1} = \int_0^\infty (1 - G_i(y))dy = \frac{1}{\lambda_i} \int_0^\infty (\lambda_i - \sum_{j \in \mathcal{N}} \lambda_j p_{ji}(y))dy. \tag{2.10}$$

The equilibrium distribution is given by

$$\pi(\mathbf{n}) = C \prod_{i \in \mathcal{N}} \left(\frac{\lambda_i}{\mu_i}\right)^{n_i} \prod_{l=1}^{n_i} \frac{1}{\phi_i(l)}. \tag{2.11}$$

Proof : By integrating equation (2.4) from zero to infinity for each x_{il}, $l = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, N$. ■

2.3 Examples

Jackson Networks : Consider a network of the type described in Section 2.1 but where the routing probabilities for moving to queue i after completing service at queue j are fixed to be p_{ji}. At queue i, $i \in \mathcal{S}$, the service time of the customer is drawn from a general distribution $\hat{G}_i(.)$ with mean μ_i^{-1}. If $i \in \mathcal{N} - \mathcal{S}$ the service time is negative exponentially distributed with mean μ_i^{-1}.

Then, in the notation of Section 2.1, the anticipating routing densities, $p_{ji}(y)$, are given by $p_{ji}(y) = p_{ji}\frac{d}{dy}\hat{G}_i(y)$ for $i \in \mathcal{S}$. From equations (2.4) and (2.6), we now find

$$G_i(y) = \sum_{j \in \mathcal{N}} \frac{\lambda_j}{\lambda_i} p_{ji} \hat{G}_i(y) = \hat{G}_i(y). \tag{2.12}$$

The equilibrium distribution given by equation (2.11) now corresponds with the well known result on Jackson networks (see Jackson (1957), Chandy and Martin (1983)).

Mixed networks : Consider a network as described in Section 2.1. For each $j \in \mathcal{N}$, let $O_2(j) \subseteq \mathcal{S}$, be a subset of symmetric nodes and let $O_1(j) = \mathcal{N} - O_2(j)$ be its complement. Jobs completing service at queue j, move to queue $i \in O_1(j)$ with probability $\hat{p}_{ji}$ and receive a service time drawn from a general distribution $\hat{G}_i(.)$, for $i \in \mathcal{S}$, or from a negative exponential distribution, mean μ_i^{-1}, for $i \in \mathcal{N} - \mathcal{S}$. Hence with probability $[1 - \sum_{i \in O_1(j)} \hat{p}_{ji}]$ their next service time, y, is drawn from the general distribution $\bar{G}_j(.)$ and the job moves to queue $i \in O_2(j)$ with probability $\bar{p}_{ji}(y)$. That is, a node $i \in O_2(j)$ is chosen dependent upon the next service requirement. As a result, for $i \in \mathcal{S}$, it is possible to distinguish $I_1(i) = \{j | i \in O_1(j)\}$ and $I_2(i) = \mathcal{N} - I_1(i)$. $I_1(i)$ is the set of queues that send jobs to queue i according to fixed routing probabilities, while $I_2(i)$ is to be interpreted as the set of queues that use service anticipating routing before sending jobs to queue i. This routing is schematically illustrated in Figures 2 and 3. Such processes might be used to model systems where lengthy jobs may incur penalties at some queues. Note that

$$\sum_{i \in O_1(j)} \hat{p}_{ji} + \sum_{i \in O_2(j)} \int_0^\infty \bar{p}_{ji}(y) d\bar{G}_j(y) = 1, \quad \forall j \in \mathcal{N}. \tag{2.13}$$

Figures 2 and 3 near here

The anticipating routing probabilities, $p_{ji}(y)$, are given by

$$p_{ji}(y) = \begin{cases} \hat{p}_{ji}\frac{d}{dy}\hat{G}_i(y) & \text{for } j \in I_1(i) \leftrightarrow i \in O_1(j) \\ \bar{p}_{ji}(y)\frac{d}{dy}\bar{G}_j(y) & \text{for } j \in I_2(i) \leftrightarrow i \in O_2(j) \end{cases} \tag{2.14}$$

and the effective service time distribution, $G_i(.)$, is given by

$$G_i(y) = \frac{1}{\lambda_i}\left[\sum_{j\in I_1(i)} \lambda_j \hat{p}_{ji}\hat{G}_i(y) + \sum_{j\in I_2(i)} \lambda_j \int_0^\infty \bar{p}_{ji}(y)d\bar{G}_j(y)\right]. \tag{2.15}$$

3. Extensions

Section 2 has been restricted to closed networks with only one customer type and simple forms for the service effort at each queue so as not to distract from the essential features of the results being presented. However, the extension to open networks with many customer types and a more general form for the service effort is, as will be argued below, straightforward.

Consider a network which comprises a set of labelled queues $\mathcal{N} = \{1, 2, \ldots, N\}$ with customers labelled by type from the set $\mathcal{T} = \{1, 2, \ldots, T\}$. The state of the network can be represented by a vector $\mathbf{c}$ which gives the type of customer in each position of each queue. Let $n(i,t)$ be the number of type t customers in queue i. Define $\mathbf{n} = (n(i), 1 \leq i \leq N)$, where $n(i) = \sum_{t\in\mathcal{T}} n(i,t)$, to be the "macrostate" giving the total number of customers at each node, irrespective of type and $\mathbf{e}_i$ to be an N-vector with a one in the i^{th} position and zeroes elsewhere.

Assume that service facility i works at a rate $\phi_i(\mathbf{n}) = \Phi(\mathbf{n} - \mathbf{e}_i)/\Phi(\mathbf{n})$ for arbitrary $\Phi(.)$ and a proportion $\gamma_i(l, \mathbf{n})$ of this effort is dedicated to the customer in position l when the macrostate is $\mathbf{n}$. Customers of type t arrive at queue i from outside the network in a

Poisson stream at rate ν_{it}. If, after the arrival of a customer to queue i the macrostate is $\mathbf{n}$, the customer moves into position l with probability $\delta_i(l, \mathbf{n})$. As in Section 2.1, when a customer leaves position l in queue i the customers in positions $l+1, \dots, n_i$ move to positions $l, \dots, n_i - 1$ respectively. Similarly, when a customer moves into position l the customers in positions $l, \dots, n_i - 1$ move into positions $l+1, \dots, n_i$. Again as in Section 2.1, let the set of symmetric queues be $\mathcal{S} \subseteq \mathcal{N}$.

Customers of type t arriving from outside the network to queue i, $i \in \mathcal{S}$, have service time drawn from the general distribution $\hat{G}_{it}(.)$. If $i \in \mathcal{N} - \mathcal{S}$ the service time is negative exponentially distributed with mean μ_i^{-1}. A customer of type t, on leaving queue i, for $i \in \mathcal{N}$, moves to queue j, for $j \in \mathcal{S}$, as a customer of type s with service requirement y with probability density $p_{ij;ts}(y)$. The customer moves to queue j, for $j \in \mathcal{N} - \mathcal{S}$, as a customer of type s with probability $p_{ij;ts}$ and their service time is drawn from a negative exponential distribution with mean μ_i^{-1}. For $i \in \mathcal{N}, j \in \mathcal{S}$, $s, t \in \mathcal{T}$, let

(3.1) $$p_{ij;ts} = \int_0^\infty p_{ij;ts}(y) dy.$$

Note that

(3.2) $$\sum_{j \in \mathcal{N}} \sum_{s \in \mathcal{T}} p_{ij;ts} = 1 \quad \forall (i,t) \in \mathcal{N} \times \mathcal{T}$$

and let

(3.3) $$\lambda_{it} = \nu_{it} + \sum_{j \in \mathcal{N}} \sum_{s \in \mathcal{T}} \lambda_{js} p_{ji;st}, \quad \forall (i,t) \in \mathcal{N} \times \mathcal{T},$$

in parallel with Section 2. Without loss of generality it is assumed that there exists a unique solution of the traffic equations (3.3). Similarly, assume that there exists a unique and continuously differentiable equilibrium density function for the c-process.

Theorem 3.1

With D a normalising constant, the equilibrium distribution for the network described above is given by

$$\pi(\mathbf{c}) = D\Phi(\mathbf{n}) \prod_{i \in \mathcal{N}} \prod_{t \in \mathcal{T}} \left(\frac{\lambda_{it}}{\mu_{it}}\right)^{n_{it}} \tag{3.4}$$

where

$$\mu_{it}^{-1} = \int_0^\infty (1 - G_{it}(y) dy, \quad \forall (i,t) \in \mathcal{N} \times \mathcal{T}, \tag{3.5}$$

and

(3.6)
$$G_{it}(y) = \begin{cases} 1 - \exp(-\mu_i y) & \text{for } (i,t) \in (\mathcal{N} - \mathcal{S}) \times \mathcal{T} \\ (\sum_{j \in \mathcal{N}} \sum_{s \in \mathcal{T}} \int_0^y \lambda_j p_{ji;st}(u) du + \hat{G}_{it}(y) \nu_{it}) / \lambda_{it} & \text{for } (i,t) \in \mathcal{S} \times \mathcal{T}. \end{cases}$$

Proof : The proof follows along similar lines to that given for Theorem 2.1 and is therefore not presented here. Roughly speaking, the essence of the proof of Theorem 1 is the notion of job local balance, on the one hand, and, on the other, the averaging involved in equations (2.5) and (2.6) due to anticipating routing. Provided any extensions retain the notion of balance the product form is guaranteed. As the system described without anticipating routing satisfies this notion of balance the product form is thus retained with anticipating routing. ∎

Acknowledgement

The research in this paper has been supported by an award of the Australian Research Grant Scheme.

References

BARBOUR, A. (1976) Networks of queues and the method of stages. *Adv. Appl. Prob.*, 8, 584-591.

BUZACOTT, J.A. AND YAO, D.D. (1986) On queueing network models of flexible manufacturing systems. *Queueing Syst. Theory Appl.*, **1**, 5-27.

CHANDY, K.M., HOWARD, J.H., AND TOWSLEY, D.F. (1977) Product form and local balance in queueing networks. *J. Assoc. Comp. Mach.*, **24**, 250-263.

CHANDY, K.M. AND MARTIN, J. (1983) A characterization of product form queueing networks. *J. Assoc. Comp. Mach.*, **30**, 250-263.

HORDIJK, A., AND VAN DIJK, N. (1983) Networks of queues. Part I : Job-local-balance and the adjoint process. Part II : General routing and service characteristics. *Lecture notes in control and information sciences,* Springer-Verlag, **60**, 158-205.

JACKSON, J. (1957) Networks of waiting lines. *Operations Research*, **5**, 518-521.

KELLY, F.P. (1979) *Reversibility and stochastic processes.* Wiley, London.

KÖNIG, D. AND JANSEN, U. (1974) Stochastic processes and properties of invariance for queueing systems with speeds and interruptions. *Trans 7th Prague Conference on Information Theory* Reidel, Dordrecht, 335-343.

RUMSEWICZ, M. AND HENDERSON W. (1987) Insensitivity with age dependent routing. Submitted.

SCHASSBERGER, R. (1978) Insensitivity of steady state distributions of generalised semi-Markov processes with speeds. *Adv. Appl. Prob.* **10**, 836-851.

WALRAND J. (1988) *Introduction to queueing networks.* Prentice-Hall, to appear.

WHITTLE, P. (1985) Partial balance and insensitivity. *J. Appl. Prob.* **22**, 168-176.

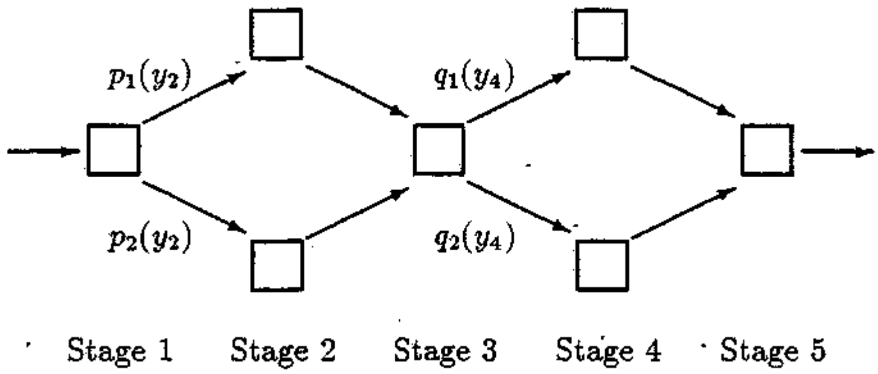

Figure 1. A parts assembly model, where y_i denotes the service requirement at stage i.

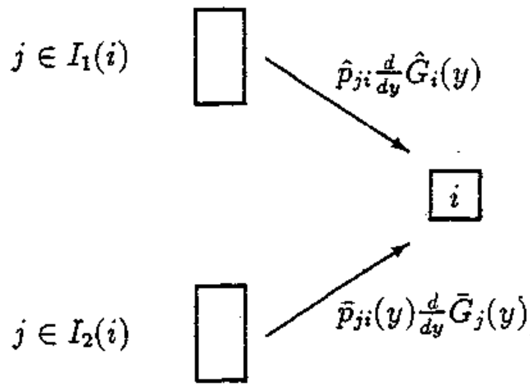

Figure 2. Schematic representation of the routing into queue i for mixed networks.

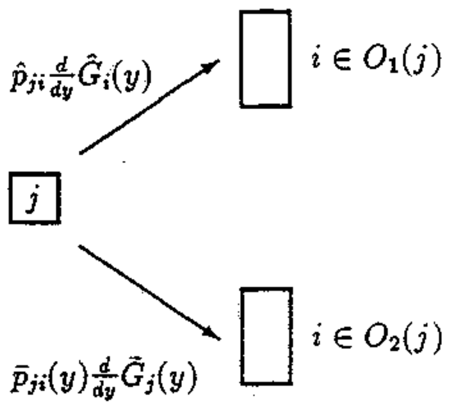

Figure 3. Schematic representation of the routing from queue j for mixed networks.