Analysis of Blocking Networks with Temporal Dependence

Vittoria De Nitto Personè[a], Giuliano Casale[b], Evgenia Smirni[c]

[a]*University of Rome Tor Vergata, Italy*
[b]*Imperial College London, U.K.*
[c]*College of William and Mary, Virginia, U.S.A.*

***Abstract*—In this paper we estend the class of MAP queueing networks to include blocking models. We consider two different blocking mechanisms: Repetitive Service-Random Destination and Blocking After Service. We analyze the Markov process underlying the MAP queueing network and propose a methodology based on a partition of the state space into "marginal state spaces". By using this partition, we prove a set of "partial" balance equations that relates blocking performance indexes. The proposed methodology can be a sound framework to define approximate solution methods for MAP queueing networks with blocking.**

1. Introduction

Blocking queueing networks are powerful tools to deal with more representative models of real-life systems. Their solutions are difficult and, despite of their importance, there is a general lack of methodologies aimed to include blocking phenomenon into solution tools.

MAP queueing networks are shown to be the first analytical methodology to describe and predict accurately the performance of complex systems operating under temporal dependencies [CaMiSmi08, CaMiSmi10]. MAP queueing networks include servers whose service times are Markovian Arrival Processes (MAPs), a class of Markov-modulated point processes that can model general distributions as well as the main features of nonrenewal workloads, such as autocorrelation in service times and burstiness [Neu89].

In this paper, we extend the class of MAP queueing network to include blocking mechanism. We consider two different blocking models: the Repetitive Service-Random Destination (RS-RD) mechanism and the Blocking After Service mechanism (BAS). These two different models are extensively studied in the literature and are representative of telecommunication systems and computer systems with limited shared resources.

For these two mechanisms we analyze the Markov process underlying the MAP queueing network. Because of the state space explosion, the queueing network's equilibrium behavior cannot be determined exactly, but we argue that it can still be bounded accurately by describing the system with "reduced" state spaces, which we call *marginal state spaces*. Marginal state spaces capture the behavior of the network conditioned on a given queue being active, blocked or idle.

In the following section, we present the considered blocking models and the related literature. In the section 3, we define the MAP blocking queueing networks and we prove the marginal balance for BAS blocking. In the section 4, we extend the methodology and the results to the RS-RD blocking.

Finally, the section 5 concludes the paper.

2. Blocking models

Queueing network models with finite capacity queues and blocking have been introduced and applied as more realistic models of systems with finite capacity resources and population constraints. In these models, when a queue reaches its maximum capacity, the flow of customers entering the service node is stopped, both from other service nodes[1] and from external sources in case of open networks. This phenomenon is referred to as blocking. In particular, each blocking mechanism defines the service node blocking time, the behaviours of arriving customers to full capacity service center and the servers' activity in the network.

Queueing networks with blocking can be applied to telecommunication systems and computer systems with limited shared resources, such as interconnecting links or store and forward buffers, as well as in production systems with the finite storage buffers. Different blocking mechanisms representing distinct behaviours of real systems with limited resources have been defined and analyzed in the literature. The interested reader can refer to [BalDenOnv01, Onv90, Onv93, Per84, Per89, Per94] for an extensive bibliography.

More recent results can be found in several research application areas such as computer systems [DeKe00], communication systems and networks [AwYaWo06, DaHo08, LiToLe07], manufacturing systems [YaMiYaMa09], software architectures [BaDenIn03] and also in the emerging area of "health systems" [KoKuSmi05].

Since the first results on product forms [Aky87, Aky88, Aky89, BalCl98, BalDen91, BouVan97, Cl98, Onv89, Ser99], and equivalence and monotonicity properties [AdVdw89, AmmGer89, BalDenIa87, BalDon89, DaLiTo94, DaTo91, Den94, OnPe89, ShaYa89, VanTi86], few effort has been devoted to develop general methodologies to deal with blocking queueing networks. More recently, some new solution approaches have been proposed [Be&alii07, OsBi09].

Bounding analysis is a one of the most attractive methodologies to estimate performance measures with a limited cost. Tandem networks with general service time distribution and Blocking Before Service mechanism or BAS blocking were considered in [Nak00]. The author defines bounds for the expected cycle time. The presented results are good for very limited size of the finite buffers. Open general topologies networks with multiserver exponential queues and RS-RD blocking were considered in [KuSriKu98]. The model includes multiclass population. The authors defined bounds to the queues throughput and the blocking probabilities.

[1] Throughout this paper we use the terms queue, node and service center interchangeably.

Our methodology combines the blocking characteristic with the temporal dependencies characteristic in the service process.

As introduced above, we consider two different blocking models that are extensively used in the literature.

Let us consider a closed queueing network, with routing matrix $P=|p_{ij}|$, that is when $p_{ij}>0$ queue i is connected to queue j. We consider finite capacity queues, with F_i the queue i capacity. Note that F_i includes the jobs in service. If n_i denotes the instantaneous population at queue i, when $n_i = F_i$ queue i is *full* and any incoming job cannot be accommodated until a departure from queue i takes place. Each blocking model defines the behaviour of the "sending" queue and the behaviours of arriving customers to full capacity queue. The considered blocking mechanisms are defined in the following.

Repetitive Service-Random Destination (RS-RD)

A queue i, if not empty, processes a job regardless of the job population at its destination j ($p_{ij}>0$). When node i completes, if node j is full, the completed job is rerouted at node i where, according to node i scheduling, it will receive a new service. During the new service, the job will select a new destination independently from the previous one.

Note that according to RS-RD blocking a node will never be actually blocked, but it "wastes" its service, it could have to repeat it. In this case, we define **blocking** as the time the node is working for a full destination node.

RS-RD blocking is used to model mainly telecommunications system. Recently, this blocking mechanism was used to model congestion control in the internet [AwYaWo06].

Blocking After Service mechanism (BAS)

A queue i, if not empty, processes a job regardless of the job population at its destination j ($p_{ij}>0$). When node i completes, if node j is full, node i stops its activity (it is blocked) and the completed job waits until a departure will occur from node j. At that moment two simultaneous transitions take place: the completed job from i to j and the job from j.

In a general network topology where more than one queue can complete job towards a full queue j, an "unblocking" policy has to be defined. Usually, the First Blocked First Unblocked (FBFU) policy is considered the fairest policy: first unblock the job was blocked first. In the following we assume FBFU policy.

BAS blocking has been used mainly to model production systems and disk I/O subsystems. Recently, this blocking was studied for two stages tandem queues with MAP and phase-type distributions [GoMar06].

3. MAP Blocking Queueing Networks

We introduce the class of MAP queueing networks supporting nonrenewal service which is studied in the rest of the paper. A summary of the main notation is given in Table I. In this section, we present the case of BAS blocking mechanism. The RS-RD blocking mechanism case is presented in Section 4.

3.1 Model Definition

We consider a closed network with finite single-server queues, which serve jobs according to a MAP service time process and under work-conserving FCFS scheduling. Each queue has finite capacità F_i and the same blocking mechanism. The service process is independent of both the job allocation across the queues and the state of other service processes. The network is composed by M queues and populated by N statistically indistinguishable jobs (single class model), which proceed through the queues according to a state-independent routing scheme. That is, upon departure from a queue i, a job joins queue j with fixed probability p_{ij}.

TABLE I
SUMMARY OF MAIN NOTATION

h, k, u	phase indexes
i, j, m	queue indexes
$\vec{k}$	phase vector, i.e., active phases
k_i	active phase at queue i in $\vec{k}$
K_i	number of phases in queue i's MAP
K_{max}	maximum K_i, $1 \leq i \leq M$
M	number of queues in the network
μ_i	mean service rate of queue i
$\mu_i^{k,h}$	completion rate of queue i, phase change $k \rightarrow h$
$v_i^{k,h}$	background trans. rate of queue i, phase $k \rightarrow h$
N	number of jobs in the network
n_i	number of jobs at queue i
p_{ij}	routing prob. from queue i to queue j
$\pi(n_i, h, n_j, k)$	prob. of n_i jobs in queue i in phase h and n_j jobs in queue j in phase k
$q_{i,j}^{k,h}$	rate for a queue i job completion towards queue j, phase change $k \rightarrow h$
Q_i	mean queue-length at queue i
Q_i^k	mean queue-length at queue i in phase k
U_i	mean utilization of queue i
U_i^k	mean utilization of queue i in phase k
Uef_i	mean effective utilization of queue i

Uef_i^k mean effective utilization of queue i in phase k

The service process at queue i is modeled by a MAP with $K_i \geq 1$ phases. General service can be approximated accurately by a MAP [ZhaCaSmi08]. If $K_i = 1$, then the MAP reduces to an exponential distribution, otherwise it generates service time samples that are phase-type (PH) distributed [Neu89]. That is, hyperexponential, hypoexponential, Erlang, and Coxian are all allowed service time distributions; nonrenewal service is also supported, e.g., Markov Modulated Poisson Process (MMPP), Interrupted Poisson Process (IPP) [FiMe93]. It should be nevertheless remarked that MAP fitting can be still a challenging problem if the data has an irregular temporal dependence structure, see [HorTe02] for a review. We point to [ZhaCaSmi08] for a new technique, called Kronecker Product Composition (KPC), that can provide MAP fitting of higher-order moments and temporal dependence structure of arbitrary processes.

The transition from phase k to phase h for the MAP service process of queue i has rate $\phi_i^{k,h}$ and produces a service completion with probability $t_i^{k,h}$; if $h = k$ then $t_i^{k,k} = 1$ according to the MAP definition. We define $\mu_i^{k,h} = t_i^{k,h}\phi_i^{k,h}$ to be the rate of job completions in phase k that leaves the MAP in phase h; $v_i^{k,h} = (1-t_i^{k,h})\phi_i^{k,h}$, $k \neq h$ is the complementary rate of transitions not associated with job completions that only change the MAP active phase (background transitions). It is worth noting that if a queue is blocked it completely stops its activity. As a consequence, a phase transition cannot occur during the blocking time. Note that this holds for BAS blocking, but not for RS-RD where a queue is never effectively blocked. In this representation of queue i's MAP, $\mu_i^{k,h}$ is the element in row k and column h of the D_1 matrix; $v_i^{k,h}$ is in row k and column h of D_0. We point the reader to [HorTe02] and references therein for background on MAPs and MAP fitting.

3.2. Underlying Markov Process for BAS blocking

General MAP service requires to maintain information at the process level on the current service phase at each queue. A feasible network state in the queueing network underlying Markov process is a tuple $\mathbf{s}= (\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_M)$, where $\mathbf{s}_i$ is the state of queue i with $\mathbf{s}_i = (n_i, b_i, \boldsymbol{m}_i, k)$ defined as follows: n_i is the queue population (including the job in service); b_i is the state of node i (1=blocked, 0=active); $\boldsymbol{m}_i$ is the list of queues blocked on queue i; $k \in K_i$ is the phase of queue i. As stated in section 2., the First Blocked First Unblocked order is assumed. Head($\boldsymbol{m}_i$) is the queue that will be unblocked by a departure from queue i in state $\mathbf{s}$.

In the following, for the sake of simplicity, we assume to omit b_i=0 (queue i is blocked) and $\boldsymbol{m}_i$ when this is empty (there are no queues blocked on the considered queue). Finally, let E_{BAS} be the state space of the queueing network when all queues behave according to BAS blocking.

According to this space, the Markov process transitions have rate $q_{i,j}^{k,h}$ from state $\mathbf{s}=(\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_M)$ to $\mathbf{s}'=(\mathbf{s}'_1, \mathbf{s}'_2, ..., \mathbf{s}'_M)$; the rate is given by

$$q_{i,j}^{k,h} = \begin{cases} p_{ij}\mu_i^{k,h}, & i \neq j, \\ v_i^{k,h} + p_{ij}\mu_i^{k,h}, & i = j \text{ and } k \neq h. \end{cases} \quad (1)$$

The state $\mathbf{s}'$ is defined in Table II according to the blocking and unblocking rules and the simultaneous transitions characterizing BAS blocking. In particular, the first row corresponds to the normal transition between *i* and *j* when queue *j* is not full and queue i does not have blocked queues on it. The second row corresponds to the case when queue *i* is full with queue *m* blocked on it and queue j is not full (for the aim of simplicity, we omit the case of multiple simultaneous transitions. The interested reader can refer [BalDenOnv01]). Finally the third row corresponds to the case of blocking, since queue *j* is full. Note that the remaining components of $\mathbf{s}'$ are unchanged in respect of $\mathbf{s}$. All these transitions have rate $q_{i,j}^{k,h}$.

TABLE II
TRANSITION RELATED STATES

$\mathbf{s}'=(\mathbf{s}'_1, \mathbf{s}'_2, ..., \mathbf{s}'_M)$	Conditions on $\mathbf{s}=(\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_M)$
$\mathbf{s}'_i = (n_i\text{-}1,\ b'_i,\ \boldsymbol{m}'_i,\ h) \wedge \mathbf{s}'_j = (n_j\text{+}1,\ b'_j,\ \boldsymbol{m}'_j,\ k'_j)$	$b_i=0,\ n_j < F_j,\ \boldsymbol{m}_i = \varnothing$
$\mathbf{s}'_i = (n_i,\ b'_i,\ \boldsymbol{m}'_i,\ h) \wedge \mathbf{s}'_j = (n_j\text{+}1,\ b'_j,\ \boldsymbol{m}'_j,\ k'_j) \wedge$ $\mathbf{s}'_m = (n_m\text{-}1,\ b'_m,\ \boldsymbol{m}'_m,\ k'_m)$	$b_i=0,\ n_i = F_i,\ n_j < F_j,\ \boldsymbol{m}_i \neq \varnothing$, head($\boldsymbol{m}_i$)=m, $\boldsymbol{m}_m = \varnothing$
$\mathbf{s}'_i = (n_i,\ 1,\ \boldsymbol{m}'_i,\ h) \wedge \mathbf{s}'_j = (F_j,\ b'_j,\ \boldsymbol{m}'_j,\ k'_j)$	$b_i=0,\ n_j = F_j$

In (1), $q_{i,j}^{k,h}$ is for $i \neq$ j the rate of departures from *i* to *j* triggering a phase transition in *i*'s service process from phase *k* to *h*; otherwise it accounts for the background transitions $v_i^{k,h}$ and the rate of the self-looping jobs $p_{ij}\mu_i^{k,h}$. Note that the case for $i = j$ and $k = h$ is not explicitly accounted since it corresponds to the diagonal of the infinitesimal generator of the Markov process. This diagonal is computed to make each row sum to zero. The size of the infinitesimal generator corresponds to the cardinality of the related global balance equations. By considering only the population components n_i, the state space of a blocking network is a subset of the state space of the same network but with infinite capacity queues. In particular, all states with $n_i > F_i$ are cutted. On the other hand, all the different components b_i, $\boldsymbol{m}_i$, increase the state space cardinality of the cutted subspace. To the best of our knowledge, a formula does not exist to compute the state space cardinality for blocking networks.

The state subspace of queue *i* in phase *k* can be partitioned in three subspaces: an *idle* subspace in which queue *i* is idle I_i^k, a *blocked* subspace in which queue *i* is blocked B_i^k, an *active* subspace in which queue *i* is neither empty nor blocked A_i^k, that is queue *i* is active.

Definition 1 (Marginal State Spaces)

The *idle condition subspace* $I_i^k = \{\mathbf{s} : \mathrm{s_i}(0,0,\varnothing,k)\}$ is the set of states of the MAP network where queue i is empty (n_i=0), active (b_i=0), with no node blocked on it ($\mathbf{m_i}=\varnothing$) and in phase k∈K_i.

The *active condition subspace* $A_i^k = \{\mathbf{s} : \mathrm{s_i}(n_i,0,\mathbf{m}_i,k)\}$ is the set of states of the MAP network where queue i is not empty (n_i>0), active (b_i=0) and in phase k∈K_i. Note that if queue i is full (n_i=F_i), it is also possible that some node is blocked on it ($\mathbf{m_i}\neq\varnothing$).

The *blocked condition subspace* $B_i^k = \{\mathbf{s} : \mathrm{s_i}(n_i,1,\mathbf{m}_i,k)\}$ is the set of states of the MAP network where queue i is not empty (n_i>0), blocked (b_i=1) and in phase k∈K_i. Note that if queue i is full (n_i=F_i), it is also possible that some node is blocked on it ($\mathbf{m_i}\neq\varnothing$).

As a summarizing example, the MAP network in Figure 1 with routing probabilities p_{11}, p_{12}, p_{13} = 1 − p_{11}− p_{12} at the first queue and p_{21} = 1, p_{31} = 1, at the remaining queues has underlying Markov process as shown in Figure 2. In the last figure, each queue *i* has finite capacity F_i = 2 and the total number of jobs circulating in the network is *N*=3. Two queues are exponential with rates $\mu_1 \equiv \mu_1^{1,1}$ and $\mu_2 \equiv \mu_2^{1,1}$, respectively; the third queue is a MAP with K_3 = 2 phases having $\mu_3^{k,h} = 0$ for $k \neq h$, that is a MMPP(2) process.

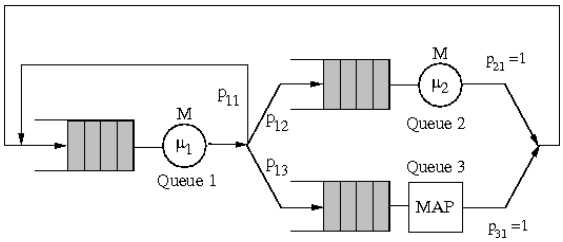

Fig. 1. Example network composed by two exponential queues and a MAP queue.

In figure 2, we show the Markov subspace related to queue 3 in phase 1, for the network of Figure 1 with finite capacity F_i = 2 ∀ *i*, *i*= 1, 2, 3, and *N*=3 jobs in the network. An analogous graph could be shown for the queue 3 in phase 2. For the sake of readability the phase change transitions are omitted. In the figure, the partition in the active-subspace, the idle-subspace and the blocked-subspace is also shown. The states in the active set are the only states that contribute to the transitions out from a queue *i*. According to the defined state notation, the state ((2 [3]) 0 (1,1), 1) in the blocked-subspace, denotes the case queue 3 is blocked (b_3=1) since it completed a job for queue 1 that is full. As soon as queue 1 completes a job, two simultaneous transitions will take place. The

process transition will be into the state (2 1 0, 1) in the idle-subspace if the completed job is destined to queue 2, while it will be into the state (2 0 1, 1) in the active-subspace if the completed job is destined to queue 3.

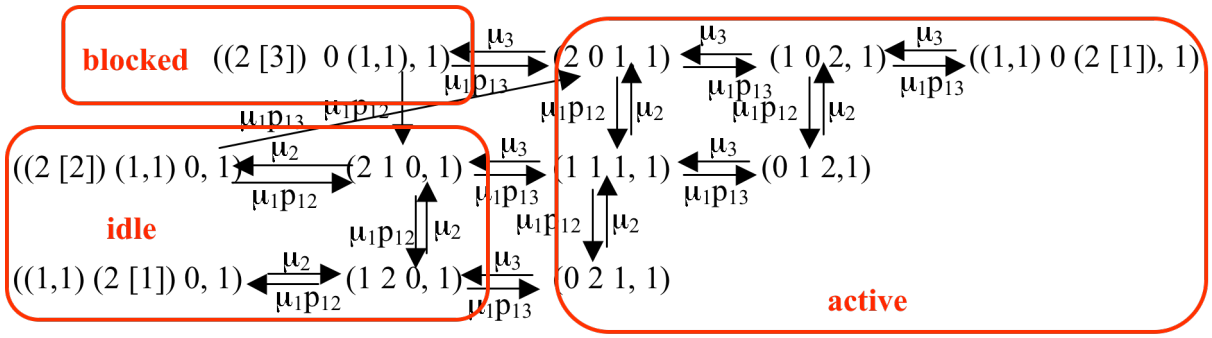

Fig. 2. Underlying Markov process of the network in Figure 1 with BAS and N=3 (subspace for queue 3 in phase 1).

In the following figures 3.a and 3.b the active subspaces are shown for queues 1 and 2 respectively. The blocked and empty subspaces can be easily derived.

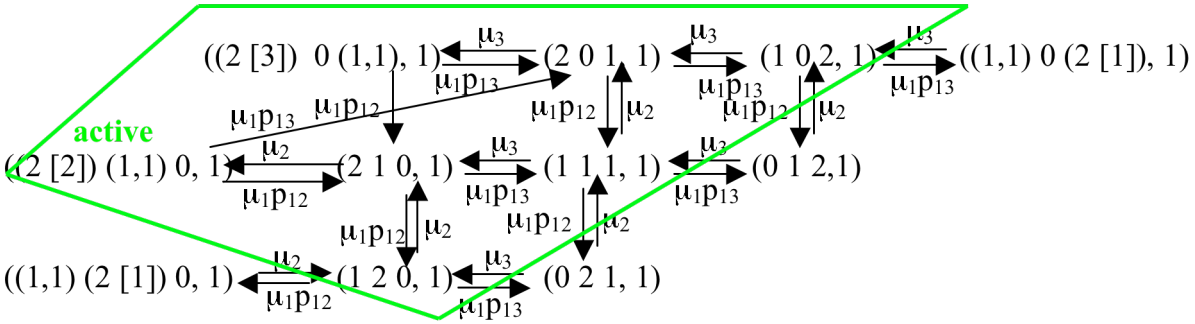

Fig. 3.b Subspace for queue 1 in phase 1.

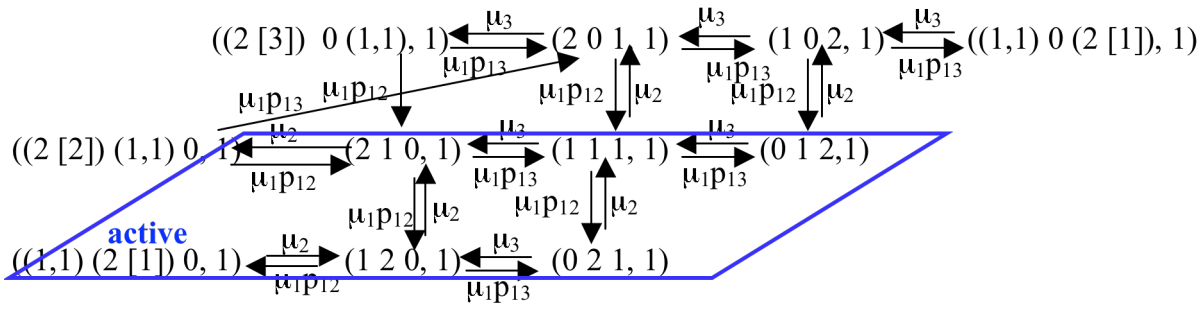

Fig. 3.c Subspace for queue 2 in phase 1.

In the following section, we present a "partial" balance analysis. Indeed, we prove partial balance equations that relate the performance indexes of the queues.

3.3. Marginal balance and performance indexes for BAS blocking

In this Section, we define a sort of "partial" balance, although the class of models considered in this paper in non-product-form. The first step is the definition of a marginal steady state probability that relates two queues of the network.

Let us define the marginal probability function that relates two queues in the network:

$$\boldsymbol{\pi}\left(n_i,k,n_j,u\right) = \sum_{\substack{\forall \mathbf{s}' \in E_{BAS}: \boldsymbol{s}'_i=(n'_i,b'_i,\boldsymbol{m}'_i,k'_i): n'_i=n_i,k'_i=k, \\ \boldsymbol{s}'_j=(n'_j,b'_j,\boldsymbol{m}'_j,k'_j): n'_j=n_j,k'_j=u}} \boldsymbol{\pi}(\mathbf{s}') \qquad (2)$$

where E_{BAS} is the state space, that is $E_{BAS} = \{ (\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_M) \mid \mathbf{s}_i = (n_i, b_i, \boldsymbol{m_i}, k_i), 0 \le n_i \le F_i, \sum_{i=1}^{M} n_i = N \}$.

The function represents the joint steady state probability to have queue *i* in phase *k* with n_i jobs and queue *j* in phase *u* with n_j jobs.

We also define a *subspace* Y_i^k *marginal probability function* $\boldsymbol{\pi}_i^k\left(n_i,k,n_j,u\right)$ as the marginal probability function defined in (2), but restricted to the states in the subspace Y_i^k, where $Y_i^k \in \{ I_i^k, B_i^k, A_i^k \}$, that is

$$\boldsymbol{\pi}_i^k\left(n_i,k,n_j,u\right) = \sum_{\substack{\forall \mathbf{s}' \in Y_i^k: \boldsymbol{s}'_i=(n'_i,b'_i,\boldsymbol{m}'_i,k'_i): n'_i=n_i,k'_i=k, \\ \boldsymbol{s}'_j=(n'_j,b'_j,\boldsymbol{m}'_j,k'_j): n'_j=n_j,k'_j=u}} \boldsymbol{\pi}(\mathbf{s}') \qquad (3)$$

Let us define the performance indexes in terms of the marginal probability function.

The mean queue length of queue *i* is $Q_i = \sum_{k=1}^{K_i} Q_i^k$ with

$$Q_i^k = \sum_{n_i=1}^{\min\{N,F_i\}} \sum_{n_j=0}^{\min\{N,F_j\}} \sum_{h=1}^{K_j} n_i \pi\left(n_i,k,n_j,h\right) \qquad (4)$$

with respect to any queue j, $1 \le j \le M$.

Let us define the *classical utilization* of queue *i* as

$$U_i = \sum_{k=1}^{K_i} U_i^k \qquad (5)$$

where we denote by U_i^k the classical utilization of queue *i* in phase *k* and it can be computed as follows:

$$U_i^k = \sum_{n_i=1}^{\min\{N,F_i\}} \sum_{n_j=0}^{\min\{N,F_j\}} \sum_{h=1}^{K_j} \pi(n_i,k,n_j,h) \quad (6)$$

with respect to any queue j, 1≤ j ≤M.

According to the blocking theory, the classical utilization is a measure of the occupancy degree of the queue, but this does not correspond to the queue productivity, since a not-idle queue could even be blocked if its destination is full.

For the aim to measure the effective productivity of a queue, we have to define an "effective utilization". The *effective utilization* of queue *i* is defined as follows:

$$Uef_i = \sum_{k=1}^{K_i} Uef_i^k \quad (7)$$

where we denote by Uef_i^k the effective utilization of queue *i* in phase *k* and it can be computed as follows:

$$Uef_i^k = \sum_{\mathbf{s}' \in A_i^k : n'_i = n_i, k'_i = k} \boldsymbol{\pi}(n_i,k,n_i,k) \quad (8)$$

As introduced above, the effective utilization takes into account the "useful" utilization of a queue, that is the period of time the queue is busy and it is not blocked, so it can produce useful work.

Let us define the mean queue length of *i* $C_j^k(i)$ in the active and blocked subspace, $A_j^k \cup B_j^k$, of queue j in phase k:

$$C_j^k(i) = \sum_{n_j=1}^{\min\{N,F_j\}} \sum_{n_i=1}^{\min\{N,F_i\}} \sum_{h=1}^{K_i} n_i \pi(n_i,h,n_j,k) \quad (9)$$

As a consequence, $C_j^k(j) = Q_j^k$. The following theorem relates classical utilization with mean queue length.

Theorem 1

In the state subspace where queue j is not-idle and in phase k, the $C_j^k(i)$ sum to NU_j^k, i.e.

$$\sum_{i=1}^{M} C_j^k(i) = NU_j^k \quad (10)$$

$$1 \le k \le K_j$$

Proof

By using (6) and the population constraint, we have

$$NU_j^k = N \sum_{n_j=1}^{\min\{N,F_j\}} \sum_{n_i=0}^{\min\{N,F_i\}} \sum_{h=1}^{K_i} \pi\left(n_j,k,n_i,h\right) = \sum_{z=1}^{M} n_z \sum_{n_j=1}^{\min\{N,F_j\}} \sum_{n_i=0}^{\min\{N,F_i\}} \sum_{h=1}^{K_i} \pi\left(n_j,k,n_i,h\right) =$$

$$= \sum_{z=1}^{M} \sum_{n_j=1}^{\min\{N,F_j\}} \sum_{n_i=0}^{\min\{N,F_i\}} \sum_{h=1}^{K_i} n_z \pi\left(n_j,k,n_i,h\right)$$

and by the definition (6) one can choose any *i*, with $1 \le i \le M$. So by choosing i=z , the following holds:

$$NU_j^k = \sum_{z=1}^{M} \sum_{n_j=1}^{\min\{N,F_j\}} \sum_{n_z=0}^{\min\{N,F_z\}} \sum_{h=1}^{K_z} n_z \pi\left(n_j,k,n_z,h\right) = \sum_{z=1}^{M} C_j^k(z)$$

Note that in the theorem 1 the classical utilization is considered, since we are interested in the computation of mean queue length. So we have to include the "blocking period".
In the following theorem we derive a balance between the effective utilization of queue i in all its phases. Indeed, for the BAS blocking the transitions are associated only to the non-blocking states. As a consequence, since the Theorem 2 represents a balance between transitions, it can only consider the effective utilization.

Theorem 2

The utilizations of queue *i* in its K_i phases are in equilibrium, i. e.,

$$\sum_{j=1}^{M} \sum_{\substack{h=1 \\ h\neq k \text{ if } j=i}}^{K_i} q_{i,j}^{k,h}(n_i) Uef_i^k = \sum_{j=1}^{M} \sum_{\substack{h=1 \\ h\neq k \text{ if } j=i}}^{K_i} q_{i,j}^{h,k}(n_i) Uef_i^h \qquad (11)$$

Proof

Let δ_m be a binary variable that is one if and only if queue *m* is not-idle in state $(\vec{n}, \vec{k})$, i.e. $n_m \ge 1$.

Let us consider the global balance equation for state $\left(\bar{n}, \bar{k}\right)$:

$$\sum_{m=1}^{M} \sum_{j=1}^{M} O_{m,j}\left(\bar{n}, \bar{k}\right) = \sum_{m=1}^{M} \sum_{j=1}^{M} I_{m,j}\left(\bar{n}, \bar{k}\right)$$

where for *j=m*

$$O_{m,m}\left(\bar{n}, \bar{k}\right) = \delta_m \sum_{\substack{h=1 \\ h\neq k}}^{K_m} q_{m,m}^{k,h} \pi\left(\bar{n}, \bar{k}\right)$$

$$I_{m,m}\left(\bar{n},\bar{k}\right)=\delta_m\sum_{\substack{h=1\\h\neq k}}^{K_m} q_{m,m}^{h,k}\pi\left(\bar{n},\bar{k}+\left(h-k\right)\bar{e}_m\right)$$

and for $j\neq m$

$$O_{m,j}\left(\bar{n},\bar{k}\right)=\delta_m\sum_{\substack{h=1\\h\neq k,\left(\bar{n},\bar{k}\right):\bar{n}_j<F_j}}^{K_m} q_{m,j}^{k,h}\pi\left(\bar{n},\bar{k}\right)$$

$$I_{m,j}\left(\bar{n},\bar{k}\right)=\delta_j\sum_{\substack{h=1\\h\neq k}}^{K_m} q_{m,j}^{h,k}\pi\left(\bar{n}-\bar{e}_j+\bar{e}_m,\bar{k}+\left(h-k\right)\bar{e}_m\right)$$

By considering all global balance equations of states in which queue i is in phase k, we evaluate the following identity relation:

$$\sum_{(\vec{n},\vec{k}):k_i=k}\ \sum_{m=1}^{M}\sum_{j=1}^{M}\left(O_{m,j}\left(\bar{n},\bar{k}\right)-I_{m,j}\left(\bar{n},\bar{k}\right)\right)=0.$$

For $m=i$ and $j=i$

$$\sum_{(\vec{n},\vec{k}):k_i=k}\delta_i\sum_{\substack{h=1\\h\neq k}}^{K_i}\left(q_{i,i}^{k,h}\pi\left(\bar{n},\bar{k}\right)-q_{i,i}^{h,k}\pi\left(\bar{n},\bar{k}+\left(h-k\right)\bar{e}_i\right)\right)=$$

$$=\sum_{n_i=1}^{\min\{N,F_i\}}\left(\sum_{\substack{h=1\\h\neq k}}^{K_i}\left(q_{i,i}^{k,h}\pi\left(n_i,k,n_i,k\right)-q_{i,i}^{h,k}\pi\left(n_i,h,n_i,h\right)\right)\right)=$$

$$=\sum_{\substack{h=1\\h\neq k}}^{K_i}\left(q_{i,i}^{k,h}U_i^k-q_{i,i}^{h,k}U_i^h\right)$$

for $m=i$ and $j\neq i$

$$\sum_{(\vec{n},\vec{k}):k_i=k}\ \sum_{\substack{j=1\\j\neq i}}^{M}\ \sum_{\substack{h=1\\h\neq k,\left(\bar{n},\bar{k}\right):\bar{n}_j<F_j}}^{K_i}\left(\delta_i q_{i,j}^{k,h}\pi\left(\bar{n},\bar{k}\right)-\delta_j q_{i,j}^{h,k}\pi\left(\bar{n}-\bar{e}_j+\bar{e}_i,\bar{k}+\left(h-k\right)\bar{e}_i\right)\right)=$$

$$=\sum_{\substack{h=1\\h\neq k}}^{K_i}\left(q_{i,j}^{k,h}Uef_i^k-q_{i,j}^{h,k}Uef_i^h\right)$$

for $m\neq i$ and $1\leq j\leq \mathrm{M}$

$$\sum_{(\vec{n},\vec{k}):k_i=k}\ \sum_{\substack{m=1\\m\neq i}}^{M}\sum_{j=1}^{M}\ \sum_{\substack{h=1\\h\neq k,\left(\bar{n},\bar{k}\right):\bar{n}_j<F_j}}^{K_m}\left(\delta_m q_{m,j}^{u,h}\pi\left(\bar{n},\bar{k}\right)-\delta_j q_{m,j}^{h,u}\pi\left(\bar{n}-\bar{e}_j+\bar{e}_m,\bar{k}+\left(h-k\right)\bar{e}_m\right)\right)=0$$

$$\sum_{(\vec{n},\vec{k}):k_i=k}\ \sum_{\substack{m=1\\ m\neq i}}^{M}\left(\sum_{\substack{j=1\\ j\neq m}}^{M}\ \sum_{\substack{h=1\\ h\neq u}}^{K_m}\left(q_{m,j}^{u,h}Uef_m^u - q_{m,j}^{h,u}Uef_m^h\right)+\sum_{\substack{h=1\\ h\neq u}}^{K_m}\left(q_{m,m}^{u,h}U_m^u - q_{m,m}^{h,u}U_m^h\right)\right)=0$$

In the following theorem, we prove that a "marginal balance" holds between the marginal probabilities similarly to the global balance between the steady state probabilities.

Theorem 3

The arrival rate at queue i when its queue-length is n_i jobs, $0< n_i \leq min\{N, F_i\}$-1, is balanced by the rate of departures when the queue-length is n_i +1, i.e.

$$\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{k=1}^{K_j}\ \sum_{\forall \mathbf{s}\in A_j^k:\overline{\mathbf{m}}_j=\emptyset \vee \mathrm{head}\left(\overline{\mathbf{m}}_j\right)\neq i}\ \sum_{h=1}^{K_j}\sum_{u=1}^{K_i}q_{j,i}^{k,h}\pi_j^k\left(n_i,u,n_j,k\right) \qquad (12)$$

$$=\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{n_j=0}^{\min\{N,F_j\}-1}\sum_{u=1}^{K_j}\ \sum_{k=1}^{K_i}\ \sum_{\mathbf{s}'\in A_i^k:n'_i=n_i+1,\overline{\boldsymbol{m}}_i=\emptyset,n'_j=n_j,k'_j=u}\ \sum_{h=1}^{K_i}q_{i,j}^{k,h}\pi_i^k\left(n_i+1,k,n_j,u\right)+$$

$$+\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{n_j=1}^{\min\{N,F_j\}}\sum_{\substack{l=1\\ l\neq i}}^{M}\ \sum_{n_l=0}^{\min\{N,F_l\}-1}\ \sum_{u=1}^{K_l}\ \sum_{k=1}^{K_j}\ \sum_{\mathbf{s}'\in A_j^k:n'_i=n_i+1,head\left(\overline{\boldsymbol{m}}_j\right)=i}\ \sum_{h=1}^{K_j}q_{j,l}^{k,h}\pi_j^k\left(n_j,k,n_l,u\right)$$

for all $1 \leq i \leq$ M. In the case $n_i = 0$ the marginal balance specializes to the more informative relation

$$\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{k=1}^{K_j}\ \sum_{\forall \mathbf{s}\in A_j^k:\overline{\mathbf{m}}_j=\emptyset \vee \mathrm{head}\left(\overline{\mathbf{m}}_j\right)\neq i}\ \sum_{h=1}^{K_j}q_{j,i}^{k,h}\pi_j^k\left(n_i=0,u,n_j,k\right) \qquad (13)$$

$$=\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{n_j=0}^{\min\{N,F_j\}-1}\sum_{h=1}^{K_j}\ \sum_{k=1}^{K_i}\ \sum_{\mathbf{s}'\in A_i^k:n'_i=n_i+1,\boldsymbol{m}'_i=\emptyset,n'_j=n_j,k'_j=u}q_{i,j}^{k,u}\pi_i^k\left(n_i=1,k,n_j,h\right)+$$

$$+\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{n_j=1}^{\min\{N,F_j\}}\sum_{\substack{l=1\\ l\neq i}}^{M}\ \sum_{n_l=0}^{\min\{N,F_l\}-1}\ \sum_{w=1}^{K_l}\ \sum_{k=1}^{K_j}\ \sum_{\mathbf{s}'\in A_j^k:n'_i=1,k'_i=u,head\left(\overline{\boldsymbol{m}}_j\right)=i}\ \sum_{h=1}^{K_j}q_{j,l}^{k,h}\pi_j^k\left(n_j,k,n_l,w\right)$$

which holds for each phase u, $1 \leq u \leq K_i$, with $1 \leq i \leq$ M.

Proof

The left side of the equation considers all the departures from queue j towards queue i: j has to be in the active subspace and i cannot be full. All these departures yield the population in i to become n_i+1, except for the case in which i is unblocked by the departure from j (head($\boldsymbol{m}_j$)=i, that is i was waiting for free space in j, and by effect of the simultaneous transitions the population in i would remain n_i.

The right side of the equation considers all the departures from queue *i*, when the population in *i* is n_i+1. All these departures yield the population in *i* to become n_i. These departures are:
a. the transitions from *i* towards *j*: *i* has to be active and *j* non-full. Moreover, queue *i* doesn't have blocked qeues on it ($\boldsymbol{m}_i=\varnothing$), otherwise its population would still remain n_i +1;
b. the transitions between any pair of nodes *j* and *l*, with *j*, *l* ≠ *i*, when the queue *j* is full and the queue *i* is the first blocked node in the *j* list $\boldsymbol{m}_j$, that is head($\boldsymbol{m}_j$)=*i*. These transitions trigger a simultaneous one from queue *i*, thus decreasing its population to n_i.

Let $S(k, n_i)\equiv\{ \mathbf{s}= (\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_M) \mid \mathbf{s_i} = (n_i', b_i', \overline{m}_i, k_i') : n_i' \le n_i, k_i' = k \}$, since the theorem requires $n_i \le min\{N, F_i\}$-1 there always exists the related set $\overline{S}(k, n_i)\equiv\{ \mathbf{s}= (\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_M) \mid \mathbf{s_i} = (n_i', b_i', \overline{m}_i, k_i') : n_i' \ge n_i + 1, k_i' = k \}$.

The equilibrium probability flux exchanged by $\cup_{k=1}^{K_i} S(k, n_i)$ and $\cup_{k=1}^{K_i} \overline{S}(k, n_i)$ must be in balance because their union is the entire state space. We seek for a representation of the exchanged probability flux using the marginal probabilities. The flux F from $\cup_{k=1}^{K_i} \overline{S}(k, n_i)$ to $\cup_{k=1}^{K_i} S(k, n_i)$ needs to decrease the queue-length of queue *i* to n_i . By considering that batch completions are not allowed, these transitions are the two cases a. and b. described above. Therefore F is the following flux of job completions:

$$F \equiv \sum_{\substack{j=1 \\ j\neq i}}^{M} \sum_{k=1}^{K_i} \sum_{\mathbf{s}'\in A_i^k : n'_i = n_i+1, \boldsymbol{m}_i=\varnothing, n'_j<F_j} \sum_{h=1}^{K_i} q_{i,j}^{k,h} \pi(\mathbf{s}') +$$

$$+ \sum_{\substack{j=1 \\ j\neq i}}^{M} \sum_{\substack{l=1 \\ l\neq i}}^{M} \sum_{k=1}^{K_j} \sum_{\mathbf{s}'\in A_j^k : n'_i = n_i+1, head(\overline{\boldsymbol{m}}_j)=i, n'_l<F_l} \sum_{h=1}^{K_j} q_{j,l}^{k,h} \pi(\mathbf{s}')$$

which excludes the self-routed jobs (case *j* = *i*) that do not decrease n_i +1 to n_i. Note that the job transition towards queue *x* is possible only if *x* is not full ($n_x < F_x$), *x*=*j*, *l*. The opposite flux G needs to increase the queue-length of queue *i* to n_i+1. The transitions towards states where *i* has n_i+1 jobs are allowed from states where the following conditions hold: the sending queue *j* is not blocked, $\text{head}(\boldsymbol{m}_j) \neq i$ such that a simultaneous transition doesn't occur (otherwise the *i* population doesn't change), queue *i* is not full ($n_i < F_i$). Therefore G is the following flux of job completions:

$$G \equiv \sum_{\substack{j=1 \\ j\neq i}}^{M} \sum_{k=1}^{K_j} \sum_{h=1}^{K_j} q_{j,i}^{k,h} \sum_{\mathbf{s}'\in A_j^k : n'_i = n_i, \overline{\boldsymbol{m}}_j=\varnothing \vee \text{head}(\overline{\boldsymbol{m}}_j)\neq i} \pi(\mathbf{s}')$$

and describes all possible transitions that bring a job from queue *j* ≠ *i*.

However by the given definition (3)

$$\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{k=1}^{K_i}\sum_{\mathbf{s}'\in A_i^k:n'_i=n_i+1,\boldsymbol{m}_i=\varnothing,n'_j<F_j}\sum_{h=1}^{K_i}q_{i,j}^{k,h}\pi(\mathbf{s}')+$$

$$+\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{\substack{l=1\\ l\neq i}}^{M}\sum_{k=1}^{K_j}\sum_{\mathbf{s}'\in A_j^k:n'_i=n_i+1,head(\overline{\boldsymbol{m}}_j)=i,n'_l<F_l}\sum_{h=1}^{K_j}q_{j,l}^{k,h}\pi(\mathbf{s}')=$$

$$=\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{n_j=0}^{\min\{N,F_j\}-1}\sum_{u=1}^{K_j}\sum_{k=1}^{K_i}\sum_{\mathbf{s}'\in A_i^k:n'_i=n_i+1,\overline{\boldsymbol{m}}_i=\varnothing,n'_j=n_j,k'_j=u}\sum_{h=1}^{K_i}q_{i,j}^{k,h}\boldsymbol{\pi}_i^k(n_i+1,k,n_j,u)+$$

$$+\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{n_j=1}^{\min\{N,F_j\}}\sum_{\substack{l=1\\ l\neq i}}^{M}\sum_{n_l=0}^{\min\{N,F_l\}-1}\sum_{u=1}^{K_l}\sum_{k=1}^{K_j}\sum_{\mathbf{s}'\in A_j^k:n'_i=n_i+1,head(\overline{\boldsymbol{m}}_j)=i}\sum_{h=1}^{K_j}q_{j,l}^{k,h}\pi_j^k(n_j,k,n_l,u)$$

and

$$\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{k=1}^{K_j}\sum_{h=1}^{K_j}q_{j,i}^{k,h}\sum_{\mathbf{s}'\in A_j^k:n'_i=n_i,\overline{\boldsymbol{m}}_j=\varnothing\vee head(\overline{\boldsymbol{m}}_j)\neq i}\pi(\mathbf{s}')=$$

$$=\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{k=1}^{K_j}\sum_{\forall\mathbf{s}\in A_j^k:\overline{\mathbf{m}}_j=\varnothing\vee head(\overline{\mathbf{m}}_j)\neq i}\sum_{h=1}^{K_j}\sum_{u=1}^{K_i}q_{j,i}^{k,h}\pi_j^k(n_i,u,n_j,k)$$

and by substituting them

$$F=\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{n_j=0}^{\min\{N,F_j\}-1}\sum_{u=1}^{K_j}\sum_{k=1}^{K_i}\sum_{\mathbf{s}'\in A_i^k:n'_i=n_i+1,\overline{\boldsymbol{m}}_i=\varnothing,n'_j=n_j,k'_j=u}\sum_{h=1}^{K_i}q_{i,j}^{k,h}\pi_i^k(n_i+1,k,n_j,u)+$$

$$+\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{n_j=1}^{\min\{N,F_j\}}\sum_{\substack{l=1\\ l\neq i}}^{M}\sum_{n_l=0}^{\min\{N,F_l\}-1}\sum_{u=1}^{K_l}\sum_{k=1}^{K_j}\sum_{\mathbf{s}'\in A_j^k:n'_i=n_i+1,head(\overline{\boldsymbol{m}}_j)=i}\sum_{h=1}^{K_j}q_{j,l}^{k,h}\pi_j^k(n_j,k,n_l,u)$$

$$G=\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{k=1}^{K_j}\sum_{\forall\mathbf{s}\in A_j^k:\overline{\mathbf{m}}_j=\varnothing\vee head(\overline{\mathbf{m}}_j)\neq i}\sum_{h=1}^{K_j}\sum_{u=1}^{K_i}q_{j,i}^{k,h}\pi_j^k(n_i,u,n_j,k)$$

and by imposing the equilibrium balance $F=G$ for $n_i\geq 1$ we find immediately (12).

Note that (12) would hold also for $n_i = 0$; nevertheless, in this case we can give the more detailed condition (13) by recalling that if $n_i = 0$ phase transitions in i are not posssible, hence the balance $F = G$ splits into a set of disjoint probability flux balances, one for each phase u of i. The proof in this case is almost identical by considering the interface between the sets $S(k, n_i=0)\equiv\{\ \mathbf{s}=(\mathbf{s}_1,\mathbf{s}_2,...,\mathbf{s}_M)\ |\ \mathbf{s_i}=(n_i',b_i',\overline{m}_i,k_i'):n_i'\leq 0,k_i'=u\}$, and $\cup_{k=1}^{K_i}\overline{S}(k, n_i=1)$.

The following theorem can be seen as an extension of the Theorem 1. Indeed Theorem 4 defines a relation between the sum of mean queue-lengths of all queues and the classical utilization of any queue *i* when a queue *j* is in phase *k*, for any *j* and *k*.

Theorem 4

The sum of mean queue-lengths of all queues *t*, *t*=1, …, *M*, when queue *j* is in phase *k*, satisfies

$$\sum_{t=1}^{M}\left(\sum_{n_t=1}^{\min\{N,F_t\}}\sum_{n_j=0}^{\min\{N,F_j\}}\sum_{h=1}^{K_t} n_t\pi\left(n_t,h,n_j,k\right)\right) \geq N\sum_{h=1}^{K_i}\sum_{n_i=1}^{\min\{N,F_i\}}\sum_{n_j=0}^{\min\{N,F_j\}}\pi\left(n_i,h,n_j,k\right) \qquad (16)$$

for all $1\leq i \leq M$, $1\leq j \leq M$, $1\leq k \leq K_j$.

Proof

Letting $\sum_{\mathbf{s}_j=\left(n_j,b_j,\mathbf{m}_j,k_j\right):k_j=k} \equiv \sum_{\mathbf{s}\in A_j^k\cup B_j^k\cup I_j^k:\mathbf{s}_j=\left(n_j,b_j,\mathbf{m}_j,k_j\right):k_j=k}$, we have

$$N\sum_{\mathbf{s}_j=\left(n_j,b_j,\mathbf{m}_j,k_j\right):k_j=k}\pi(\mathbf{s}) = \sum_{t=1}^{M}\ \sum_{\mathbf{s}_j=\left(n_j,b_j,\mathbf{m}_j,k_j\right):k_j=k} n_t\pi(\mathbf{s}) =$$

$$=\sum_{t=1}^{M}\left(\sum_{n_t=1}^{\min\{N,F_t\}}\sum_{n_j=0}^{\min\{N,F_j\}}\sum_{h=1}^{K_t} n_t\pi\left(n_t,h,n_j,k\right)\right)$$

On the other hand

$$N\sum_{\mathbf{s}_j=\left(n_j,b_j,\mathbf{m}_j,k_j\right):k_j=k}\pi(\mathbf{s}) \geq N\sum_{n_i=1}^{\min\{N,F_i\}}\sum_{n_j=0}^{\min\{N,F_j\}}\pi\left(n_i,h,n_j,k\right)$$

since in the first member there are also states with n_i=0.

Finally in the following theorem, all the performance indexes of queue *i* in all its phase are related.

Theorem 5

The performance indexes of queue *i* in phase *k* and in phase *h* are related by the following equation

$$\sum_{\substack{h=1\\h\neq k}}^{K_i}\ \sum_{j=1}^{M}\sum_{n_j=0}^{min\{N,F_j\}}\sum_{u=1}^{K_j}\ \sum_{\mathbf{s}'\in A_i^k:n'_i=n_i,k'_i=k,n'_j=n_j,k'_j=u} q_{i,j}^{k,h}\,n_i\boldsymbol{\pi}\left(n_i,k,n_j,u\right)+$$

$$+\sum_{\substack{j=1\\j\neq i}}^{M}\sum_{h=1}^{K_i}\sum_{n_j=0}^{min\{N,F_j\}}\sum_{u=1}^{K_j}\ \sum_{\mathbf{s}'\in A_i^k:n'_i=n_i,k'_i=h,n'_j=n_j,k'_j=u} q_{i,j}^{h,k}\boldsymbol{\pi}\left(n_i,h,n_j,u\right)= \qquad (17)$$

$$=\sum_{\substack{j=1\\j\neq i}}^{M}\sum_{h=1}^{K_j}\sum_{u=1}^{K_j}\ \sum_{\substack{\mathbf{s}'\in A_j^h:n'_i=n_i,k'_i=k,n'_j=n_j,k'_j=h,\\ m_j=\varnothing \vee top(m_j)=i}}\ \sum_{n_i=0}^{min\{N,F_i\}} q_{j,i}^{h,u}\boldsymbol{\pi}\left(n_i,k,n_j,h\right)+$$

$$+\sum_{\substack{h=1\\h\neq k}}^{K_i}\ \sum_{j=1}^{M}\sum_{n_j=0}^{min\{N,F_j\}}\sum_{u=1}^{K_j}\ \sum_{\mathbf{s}'\in A_i^h:n'_i=n_i,k'_i=h,n'_j=n_j,k'_j=u} q_{i,j}^{h,k}\,n_i\boldsymbol{\pi}\left(n_i,h,n_j,u\right)$$

Proof

Let us consider the weighted sum of all global balance equations of states in which queue *i* is in phase *k*:

$$\sum_{A_i^k\cup B_i^k\cup I_i^k} n_i\sum_{m=1}^{M}\sum_{j=1}^{M}\left(O_{m,j}(\mathbf{s})-I_{m,j}(\mathbf{s})\right)=0$$

The proof follows a similar high-level structure than the proof of theorem 2.

4. MAP Queueing Networks with RS-RD blocking

In this section we present MAP queueing networks with RS-RD blocking. In this case, since the given mechanism definition, a simple state notation is enough to characterize unambiguously the network behaviour. A feasible network state in the queueing network underlying Markov process is a tuple $(\vec{n},\vec{k})$, where $\vec{n}=(n_1, n_2, \ldots, n_M)$, $0\le n_i\le F_i$, $\sum_{i=1}^{M} n_i=N$, describes the number of jobs in each queue, and $\vec{k}=(k_1, k_2, \ldots, k_M)$, $1\le k_i\le K_i$, specifies the active phase for each service process. Note that this state description coincides with the one used for the non-blocking case. According to this space, the Markov process transitions have rate $q_{i,j}^{k,h}$ from state $(\vec{n},\vec{k})$ to $(\vec{n}-\vec{e}_i+\vec{e}_j,\ \vec{k}')$, $k_i = k$, $k_i^{'} = h$, where $\vec{e}_t$ is a vector of zeros with a one in the t-th position; the rate is given by (1). Finally, let $E_{RS\text{-}RD}$ be the state space of the queueing network when all queues behave according to RS-RD blocking.

As stated above for the BAS case, the state space of a blocking network is a subset of the state space of the same network but with infinite capacity queues. In particular, for RS-RD, the state space can be simply obtained by cutting all states with $n_i > F_i$. To the best of our knowledge, a

formula does not exist to compute the state space cardinality for blocking networks. The interested reader can refer to [BalDenOnv01] for a recursive expression to compute the state space cardinality for a queueing network in which all queues have the same capacity and RS-RD blocking.
By considering the network of Figure 1, the underlying Markov process is shown in Figure 4.

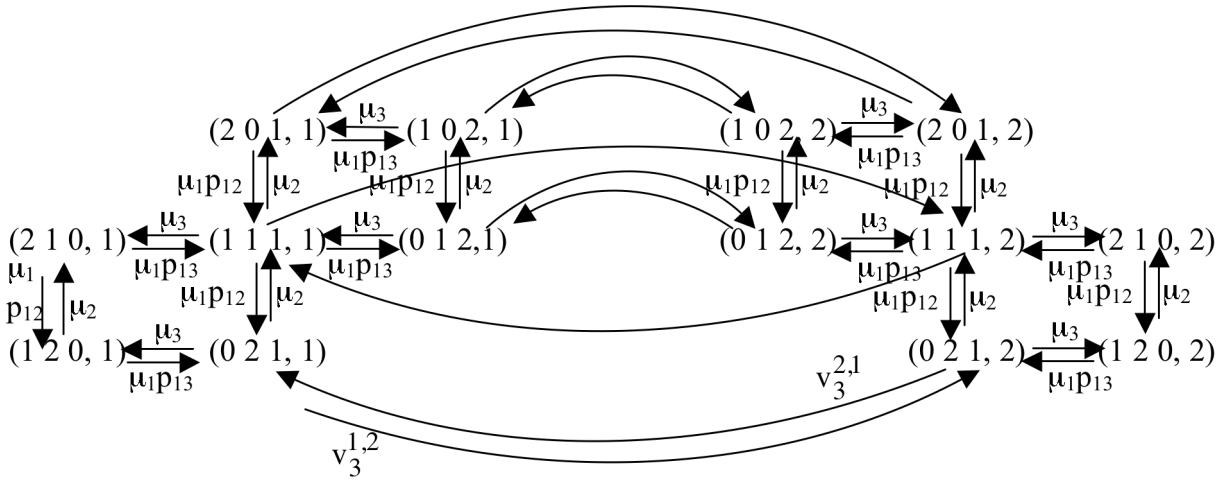

Fig. 4. Underlying Markov process of the network in Figure 1 with RS-RD and N=3.

The notation, e.g., (102, 1) indicates that the exponential queue 2 is idle, the exponential queue 1 has 1 job in service, and the MAP queue has two jobs and is in phase 1. According to RS-RD mechanism, in this state queue 1 is "blocked" with probability $0 < p_{13} < 1$ (it is working for a full queue). In (210, 2), the phase 2 is the phase left active by the last served job and queue 2 is "blocked" with probability $p_{21} = 1$.
In the following section, we present the "partial" balance analysis for the case of RS-RD blocking.

4.1. Marginal balance and performance indexes for RS-RD blocking

Let us define the marginal probability function that relates two queues of the network:

$$\pi\left(n_i,k,n_j,u\right) = \sum_{\left(\bar{n}',\bar{k}'\right)\in E_{RS-RD}\,:n'_i=n_i,k'_i=k,n'_j=n_j,k'_j=u} \pi\left(\bar{n}',\bar{k}'\right) \qquad (18)$$

where $E_{RS\text{-}RD}$ is the state space, that is $E_{RS\text{-}RD}=\{(\vec{n},\vec{k}) \mid \vec{n}=(n_1, n_2, \ldots, n_M),\ 0 \le n_i \le F_i,\ \sum_{i=1}^{M} n_i = N\}$.

Similarly to equation (2), this represents the joint steady state probability to have queue i in phase k with n_i jobs and queue j in phase u with n_j jobs.
For the sake of brevity, we do not repeat the definitions that are identical to those defined for the BAS blocking. As a consequence, the *mean queue length* Q_i and the *classical utilization* U_i for

queue *i* with RS-RD blocking, are defined by using the marginal probability function (18), instead of (2), in the equations (4), (5) and (6).

To measure the effective utilization for RS-RD blocking, we have to exclude the useless work destined to full queues. Indeed, according to the blocking definition, this work will have to be repeated, therefore it is considered useless work. The effective utilization Uef_i^k of queue *i* in phase *k* can be computed as follows:

$$Uef_i^k = \sum_{n_i=1}^{min\{N,F_i\}} \left(\pi(n_i,k,n_i,k) - \sum_{\substack{j=1 \\ j\neq i,\, p_{ij}>0}}^{M} \sum_{h=1}^{K_j} p_{ij}\pi\left(n_i,k,F_j,h\right) \right) \qquad (19)$$

The *effective utilization* Uef_i of queue *i*, can be simply obtained by substituting (19) into equation (7).

Note that the mean queue length $C_j^k(i)$, defined in (9) for BAS corresponds to the mean queue length of *i* when queue *j* is not idle and it is in phase *k* for RS-RD. As a consequence, the Theorem 1 holds with the same proof.

In the following theorem we derive a balance between the utilization of queue *i* in all its phases, both classical and effective. Note that this is different from the Theorem 2 for the BAS case. Indeed, for the RS-RD mechanism, queue blocking doesn't really occur. As a consequence, since this Theorem represents a balance between transitions, it includes also the classical utilization.

Theorem 2'

The utilizations of queue *i* in its K_i phases are in equilibrium, i. e.,

$$\sum_{\substack{j=1 \\ j\neq i}}^{M} \sum_{h=1}^{K_i} q_{i,j}^{k,h} Uef_i^k + \sum_{\substack{h=1 \\ h\neq k}}^{K_i} q_{i,i}^{k,h} U_i^k = \sum_{\substack{j=1 \\ j\neq i}}^{M} \sum_{h=1}^{K_i} q_{i,j}^{h,k} Uef_i^h + \sum_{\substack{h=1 \\ h\neq k}}^{K_i} q_{i,i}^{h,k} U_i^h \qquad (20)$$

Proof

Let δ_m be a binary variable that is one if and only if queue *m* is not-idle in state $(\vec{n},\vec{k})$, i.e. $n_m \geq 1$.

Let us consider the global balance equation for state $\left(\bar{n},\bar{k}\right)$:

$$\sum_{m=1}^{M}\sum_{j=1}^{M} O_{m,j}\left(\bar{n},\bar{k}\right) = \sum_{m=1}^{M}\sum_{j=1}^{M} I_{m,j}\left(\bar{n},\bar{k}\right)$$

where for *j=m*

$$O_{m,m}\left(\bar{n},\bar{k}\right) = \delta_m \sum_{\substack{h=1 \\ h\neq k}}^{K_m} q_{m,m}^{k,h} \pi\left(\bar{n},\bar{k}\right)$$

$$I_{m,m}\left(\bar{n},\bar{k}\right)=\delta_m\sum_{\substack{h=1\\h\neq k}}^{K_m} q_{m,m}^{h,k}\pi\left(\bar{n},\bar{k}+\left(h-k\right)\bar{e}_m\right)$$

and for $j\neq m$

$$O_{m,j}\left(\bar{n},\bar{k}\right)=\delta_m\sum_{\substack{h=1\\h\neq k,\left(\bar{n},\bar{k}\right):\bar{n}_j<F_j}}^{K_m} q_{m,j}^{k,h}\pi\left(\bar{n},\bar{k}\right)$$

$$I_{m,j}\left(\bar{n},\bar{k}\right)=\delta_j\sum_{\substack{h=1\\h\neq k}}^{K_m} q_{m,j}^{h,k}\pi\left(\bar{n}-\bar{e}_j+\bar{e}_m,\bar{k}+\left(h-k\right)\bar{e}_m\right)$$

By considering all global balance equations of states in which queue i is in phase k, we evaluate the following identity relation:

$$\sum_{(\vec{n},\vec{k}):k_i=k}\ \sum_{m=1}^{M}\sum_{j=1}^{M}\left(O_{m,j}\left(\bar{n},\bar{k}\right)-I_{m,j}\left(\bar{n},\bar{k}\right)\right)=0.$$

For $m=i$ and $j=i$

$$\sum_{(\vec{n},\vec{k}):k_i=k}\delta_i\sum_{\substack{h=1\\h\neq k}}^{K_i}\left(q_{i,i}^{k,h}\pi\left(\bar{n},\bar{k}\right)-q_{i,i}^{h,k}\pi\left(\bar{n},\bar{k}+\left(h-k\right)\bar{e}_i\right)\right)=$$

$$=\sum_{n_i=1}^{\min\{N,F_i\}}\left(\sum_{\substack{h=1\\h\neq k}}^{K_i}\left(q_{i,i}^{k,h}\pi\left(n_i,k,n_i,k\right)-q_{i,i}^{h,k}\pi\left(n_i,h,n_i,h\right)\right)\right)=$$

$$=\sum_{\substack{h=1\\h\neq k}}^{K_i}\left(q_{i,i}^{k,h}U_i^k-q_{i,i}^{h,k}U_i^h\right)$$

for $m=i$ and $j\neq i$

$$\sum_{(\vec{n},\vec{k}):k_i=k}\ \sum_{\substack{j=1\\j\neq i}}^{M}\ \sum_{\substack{h=1\\h\neq k,\left(\bar{n},\bar{k}\right):\bar{n}_j<F_j}}^{K_i}\left(\delta_i q_{i,j}^{k,h}\pi\left(\bar{n},\bar{k}\right)-\delta_j q_{i,j}^{h,k}\pi\left(\bar{n}-\bar{e}_j+\bar{e}_i,\bar{k}+\left(h-k\right)\bar{e}_i\right)\right)=$$

$$=\sum_{\substack{h=1\\h\neq k}}^{K_i}\left(q_{i,j}^{k,h}Uef_i^k-q_{i,j}^{h,k}Uef_i^h\right)$$

for $m\neq i$ and $1\leq j\leq \mathrm{M}$

$$\sum_{(\vec{n},\vec{k}):k_i=k}\ \sum_{\substack{m=1\\m\neq i}}^{M}\sum_{j=1}^{M}\ \sum_{\substack{h=1\\h\neq k,\left(\bar{n},\bar{k}\right):\bar{n}_j<F_j}}^{K_m}\left(\delta_m q_{m,j}^{u,h}\pi\left(\bar{n},\bar{k}\right)-\delta_j q_{m,j}^{h,u}\pi\left(\bar{n}-\bar{e}_j+\bar{e}_m,\bar{k}+\left(h-k\right)\bar{e}_m\right)\right)=0$$

$$\sum_{(\vec{n},\vec{k}):k_i=k}\ \sum_{\substack{m=1\\ m\neq i}}^{M}\left(\sum_{\substack{j=1\\ j\neq m}}^{M}\ \sum_{\substack{h=1\\ h\neq u}}^{K_m}\left(q_{m,j}^{u,h}Uef_m^u - q_{m,j}^{h,u}Uef_m^h\right)+\sum_{\substack{h=1\\ h\neq u}}^{K_m}\left(q_{m,m}^{u,h}U_m^u - q_{m,m}^{h,u}U_m^h\right)\right)=0$$

In the following theorem, we prove that a "marginal balance" holds between the marginal probabilities similarly to the global balance between the steady state probabilities.

Theorem 3'

The arrival rate at queue i when its queue-length is n_i jobs, $0< n_i \leq min\{N, F_i\}-1$, is balanced by the rate of departures when the queue-length is n_i+1, i.e.

$$\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{n_j=1}^{\min\{N,F_j\}}\ \sum_{k=1}^{K_j}\sum_{h=1}^{K_j}\sum_{u=1}^{K_i} q_{j,i}^{k,h}\pi\left(n_i,u,n_j,k\right)$$

$$=\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{n_j=0}^{min\{N,F_j\}-1}\sum_{u=1}^{K_j}\ \sum_{k=1}^{K_i}\sum_{h=1}^{K_i} q_{i,j}^{k,h}\boldsymbol{\pi}\left(n_i+1,k,n_j,u\right) \quad (21)$$

for all $1 \leq i \leq M$. In the case $n_i = 0$ the marginal balance specializes to the more informative relation

$$\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{n_j=1}^{\min\{N,F_j\}}\ \sum_{k=1}^{K_j}\sum_{h=1}^{K_j} q_{j,i}^{k,h}\pi\left(n_i=0,u,n_j,k\right)$$

$$=\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{n_j=0}^{min\{N,F_j\}-1}\sum_{h=1}^{K_j}\ \sum_{k=1}^{K_i} q_{i,j}^{k,u}\boldsymbol{\pi}\left(n_i=1,k,n_j,h\right) \quad (22)$$

which holds for each phase u, $1 \leq u \leq K_i$, with $1 \leq i \leq$ M.

Proof

Let $S(k, n_i)\equiv\{\left(\bar{n}',\bar{k}'\right): n_i' \leq n_i, k_i' = k\}$, since the theorem requires $n_i \leq min\{N, F_i\}-1$ there always exists the related set $\bar{S}(k, n_i)\equiv\{\left(\bar{n}',\bar{k}'\right): n_i' \geq n_i+1, k_i' = k\}$.

The equilibrium probability flux exchanged by $\cup_{k=1}^{K_i} S(k, n_i)$ and $\cup_{k=1}^{K_i} \bar{S}(k, n_i)$ must be in balance because their union is the entire state space. We seek for a representation of the exchanged probability flux using the marginal probabilities. The flux F from $\cup_{k=1}^{K_i} \bar{S}(k, n_i)$ to $\cup_{k=1}^{K_i} S(k, n_i)$ needs to decrease the queue-length of queue i to n_i. Only states where i has n_i+1 jobs can have transitions to states where i has n_i jobs (batch completions are not allowed); therefore F is the following flux of job completions

$$F \equiv \sum_{\substack{j=1 \\ j\neq i}}^{M} \sum_{n_j=1}^{\min\{N,F_j\}-1} \sum_{k=1}^{K_i} \sum_{h=1}^{K_i} q_{i,j}^{k,h} \sum_{(\bar{n}',\bar{k}')\in\{\bar{S}(k,n_i):n'_i=n_i+1,n'_j=n_j\}} \pi(\bar{n}',\bar{k}')$$

which excludes the self-routed jobs (case $j = i$) that do not decrease $n_i + 1$ to n_i. Note that the job transition towards queue j is possible only if j is not full ($n_j < F_j$). The opposite flux is

$$G \equiv \sum_{\substack{j=1 \\ j\neq i}}^{M} \sum_{n_j=1}^{\min\{N,F_j\}} \sum_{k=1}^{K_j} \sum_{h=1}^{K_j} q_{j,i}^{k,h} \sum_{u=1}^{K_i} \sum_{(\bar{n}',\bar{k}')\in\{S(u,n_i):n'_i=n_i,n'_j=n_j\}} \pi(\bar{n}',\bar{k}')$$

and describes all possible transitions that bring a job from queue $j \neq i$ for all possible phases u of i in $\vec{k}$. To account for the single step behavior, we have imposed that the population in i is of n_i jobs. However by the given definition (18)

$$\sum_{(\bar{n}',\bar{k}')\in\{\bar{S}(k,n_i):n'_i=n_i+1,n'_j=n_j\}} \pi(\bar{n}',\bar{k}') = \sum_{u=1}^{K_j} \pi(n_i+1,k,n_j,u)$$

$$\sum_{(\bar{n}',\bar{k}')\in\{S(u,n_i):n'_i=n_i,n'_j=n_j\}} \pi(\bar{n}',\bar{k}') = \sum_{k=1}^{K_j} \pi(n_i,u,n_j,k)$$

and by substituting them

$$F \equiv \sum_{\substack{j=1 \\ j\neq i}}^{M} \sum_{n_j=1}^{\min\{N,F_j\}-1} \sum_{k=1}^{K_i} \sum_{h=1}^{K_i} q_{i,j}^{k,h} \sum_{u=1}^{K_j} \pi(n_i+1,k,n_j,u)$$

$$G \equiv \sum_{\substack{j=1 \\ j\neq i}}^{M} \sum_{n_j=1}^{\min\{N,F_j\}} \sum_{k=1}^{K_j} \sum_{h=1}^{K_j} q_{j,i}^{k,h} \sum_{u=1}^{K_i} \sum_{k=1}^{K_j} \pi(n_i,u,n_j,k)$$

and by imposing the equilibrium balance $F=G$ for $n_i \geq 1$ we find immediately (21).
The proof for the more detailed condition (22) can be simply derived as for the BAS blocking.

Analogously to what stated above for the Theorem 1, Theorem 4 still holds for the RS-RD case. The proof follows the same steps by considering that the subspace $A_j^k \cup B_j^k \cup I_j^k$ is simply the subspace of queue j in phase k for RS-RD.

Finally, as for the BAS case, in the following theorem all the performance indexes of queue i in all its phases are related.

Theorem 5'

The performance indexes of queue i in phase k and in phase h are related by the following equation

$$\sum_{\substack{h=1 \\ h\neq k}}^{K_i} \sum_{\substack{j=1 \\ j\neq i}}^{M} \sum_{n_j=0}^{min\{N,F_j\}-1} \sum_{u=1}^{K_j} \sum_{n_i=1}^{min\{N,F_i\}} q_{i,j}^{k,h} n_i \boldsymbol{\pi}\left(n_i,k,n_j,u\right) +$$

$$+\sum_{\substack{h=1 \\ h\neq k}}^{K_i} \sum_{n_i=1}^{min\{N,F_i\}} q_{i,i}^{k,h} n_i \boldsymbol{\pi}\left(n_i,k,n_i,k\right) +$$

$$+\sum_{\substack{j=1 \\ j\neq i}}^{M} \sum_{h=1}^{K_i} \sum_{n_j=0}^{min\{N,F_j\}-1} \sum_{u=1}^{K_j} \sum_{n_i=1}^{min\{N,F_i\}} q_{i,j}^{h,k} \boldsymbol{\pi}\left(n_i,h,n_j,u\right) = \qquad (23)$$

$$= \sum_{\substack{j=1 \\ j\neq i}}^{M} \sum_{h=1}^{K_j} \sum_{u=1}^{K_j} \sum_{n_j=1}^{min\{N,F_j\}} q_{j,i}^{h,u} \sum_{n_i=0}^{min\{N,F_i\}-1} \boldsymbol{\pi}\left(n_i,k,n_j,h\right) +$$

$$+\sum_{\substack{h=1 \\ h\neq k}}^{K_i} \sum_{\substack{j=1 \\ j\neq i}}^{M} \sum_{n_j=0}^{min\{N,F_j\}-1} \sum_{u=1}^{K_j} \sum_{n_i=1}^{min\{N,F_i\}} q_{i,j}^{h,k} n_i \boldsymbol{\pi}\left(n_i,h,n_j,u\right) +$$

$$+\sum_{\substack{h=1 \\ h\neq k}}^{K_i} \sum_{n_i=1}^{min\{N,F_i\}} q_{i,i}^{h,k} n_i \boldsymbol{\pi}\left(n_i,h,n_i,h\right)$$

Proof

Let us consider the weighted sum of all global balance equations of states in which queue *i* is in phase *k*:

$$\sum_{(\vec{n},\vec{k}):k_i=k} n_i \sum_{m=1}^{M} \sum_{j=1}^{M} \left(O_{m,j}\left(\bar{n},\bar{k}\right) - I_{m,j}\left(\bar{n},\bar{k}\right)\right)=0$$

The proof follows a similar high-level structure than the proof of theorem 2'.

5. Conclusions

In this paper, we consider the class of closed queueing networks with blocking and MAP service times. We consider two of the most used blocking mechanisms: BAS and RS-RD blocking.
We analyze the underlying Markov processes and we prove a set of marginal balance equations that relates the blocking performance indexes. We argue that the marginal balance can be a sound framework for approximate solution techniques. A bounding analysis will be presented in a forthcoming paper.

References

[AdVdw89] Adan, I., and J. Van Der Wal "Monotonicity of the throughput in single server production and assembly networks with respect to the buffer size" in Queueing Networks with Blocking (H.G. Perros and T. Altiok Eds.), Elsevier, 1989, 325-344.

[Aky87] Akyildiz, I.F. "Exact Product Form Solutions for Queueing Networks with Blocking" IEEE Trans. on Computers, Vol. 1 (1987) 121-126.

[Aky88] Akyildiz, I.F. "Mean Value Analysis for Blocking Queueing Networks" IEEE Trans. on Software Engineering, Vol. 14 (1988) 418-429.

[Aky89] Akyildiz, I.F. "Product Form Approximations for Queueing Networks with Multiple Servers and Blocking" IEEE Trans. Computers, Vol. 38 (1989) 99-115.

[AmmGer89] Ammar, M.H., and S.B. Gershwin "Equivalence Relations in Queueing Models of Fork/Join Networks with Blocking" Performance Evaluation, Vol. 10 (1989) 233-245.

[AwYaWo06] Awan I., Yar A., Woodward M.E., "Analysis of Queueing Networks with Blocking under Active Queue Management Scheme", 12th International Conference on Parallel and Distributed Systems, ICPADS'06

[BalCl98] Balsamo, S., C. Clò "A Convolution Algorithm for Product Form Queueing Networks with Blocking" Annals of Operations Research, Vol. 79 (1998) 97-117.

[BalDen91] Balsamo, S., and V. De Nitto Personè "Closed queueing networks with finite capacities: blocking types, product-form solution and performance indices" Performance Evaluation, Vol. 12, 4 (1991) 85-102.

[BalDenIa87] Balsamo, S., V. De Nitto, G. Iazeolla "Identity and Reducibility Properties of Some Blocking and Non-Blocking Mechanisms in Congested Networks" in Flow Control of Congested Networks, (A.R. Odoni, L. Bianco, G. Szego Eds.), NATO ASI Series, Comp. and System Science, Vol.F38, Springer-Verlag, 1987.

[BaDenIn03] Balsamo S., de Nitto Personè V., Inverardi P., "A review on Queueing Network Models with finite capacity queues for Software Architectures performance prediction", Performance Evaluation, Vol 51/2-4 (2003) pp 269 - 288.

[BalDenOnv01] S. Balsamo, V. de Nitto Personè, R. Onvural, "Analysis of Queueing Networks with Blocking", Kluwer Academic Publishers, ISBN 0-7923-7996-9, 2001

[BalDon89] Balsamo, S., and L. Donatiello "Two-stage Queueing Networks with Blocking: Cycle Time Distribution and Equivalence Properties", in Modelling Techniques and Tools for Computer Performance Evaluation (R. Puigjaner, D. Potier Eds.) Plenum Press, 1989.

[Be&alii07] Begin T., Brandwajn A., Baynat B., Wolfinger B.E., Fdida S., "High-level approach to modeling of observed system behavior", ACM SIGMETRICS Performance Evaluation Review, Volume 35 , Issue 3 (December 2007)

[BouVan97] Boucherie, R., and N. Van Dijk "On the arrival theorem for product form queueing networks with blocking" Performance Evaluation, 29 (1997) 155-176.

[CaMiSmi08] G. Casale, N. Mi, E. Smirni. Bound Analysis of Closed Queueing Networks with Workload Burstiness. in Proc. of ACM SIGMETRICS 2008, 13-24, Annapolis, MD, ACM Press, June 2008.

[CaMiSmi10] G. Casale, N. Mi, and E.Smirni. "Model-Driven System Capacity Planning Under Workload Burstiness" IEEE Transactions on Computers, 59(1):66-80, Jan 2010.

[Cl98] Clò, C. "MVA for Product-Form Cyclic Queueing Networks with RS Blocking" Annals of Operations Research, Vol. 79 (1998).

[DaHo08] Daduna H., Holst M., "Customer Oriented Performance Measures for Packet Transmission in a Ring Network with Blocking", 14th GI/ITG Conf. On Measurement, Modeling and Evaluation of Computer and Communication System (MMB 2008)

[DaLiTo94] Dallery, Y., Z. Liu, and D.F. Towsley "Equivalence, reversibility, symmetry and concavity properties in fork/join queueing networks with blocking" Techn, Report, MASI.90.32, Université Pierre et Marie Curie, France, June, 1990 and J. of the ACM, Vol. 41 (1994) 903-942.

[DaTo91] Dallery, Y., and D.F. Towsley "Symmetry property of the throughput in closed tandem queueing networks with finite buffers" Op. Res. Letters, Vol. 10 (1991) 541-547.

[Den94] De Nitto Personè, V. "Topology related index for performance comparison of blocking symmetrical networks" European J. of Oper. Res., Vol. 78 (1994) 413-425.

[DeKe00] De Almeida D., Kellert P. "Markovian and analytical models for multiple bus multiprocessor systems with memory blockings", Journal of Systems Architecture, 46, (2000), pp. 455 - 477

[FiMe93] W. Fischer and K. S. Meier-Hellstern. The Markov- Modulated Poisson Process (MMPP) cookbook. Perf. Eval., 18(2):149–171, 1993.

[GoMar06] A. Gomez_Corral, M.E. Martos "Performance of two-stage tandem queues with blocking: the impact ofseveral flows of signals", Performance Evaluation, Vol. 63, Issue 9, (October 2006)

[HorTe02] A. Horv´ath and M. Telek. Markovian modeling of real data traffic: Heuristic phase type and MAP fitting of heavy tailed and fractal like samples. In Performance Evaluation of Complex Systems: Techniques and Tools, IFIP Performance 2002, LNCS Tutorial Series Vol 2459, pages 405–434, 2002.

[KoKuSmi05] Koizumi N., Kuno E., Smith T.E., "Modeling patient flows using a queuing network with blocking", Health Care Management Science, 8 (1), (2005) pp. 49-60.

[KuSriKu98] Kumar S., Srikant R., Kumar P.R., "Bounding blocking probabilities and throughput in queueing networks with buffer capacity constraints", Queueing Systems 28 (1998) 55–77 55

[LiToLe07] Zhen Liu C.H., Towsley D., Lerarge M., "Scalability of Fork/Join Queueing Networks with Blocking" SIGMETRICS'07, June 12-16, 2007 , San Diego CA

[Nak00] K. Nakade "New bounds for expected cycle times in tandem queues with blocking", European Journal of Operational Research, Volume 125, Issue 1, 16 August 2000, Pages 84-92

[Neu89] Neuts M. F. Structured Stochastic Matrices of M/G/1 Type and Their Applications. Marcel Dekker, NY, 1989.

[Onv89] Onvural, R.O. "A Note on the Product Form Solutions of Multiclass Closed Queueing Networks with Blocking" Performance Evaluation, Vol.10 (1989) 247-253.

[Onv90] Onvural, R.O. "Survey of Closed Queueing Networks with Blocking" ACM Computing Surveys, Vol. 22, 2 (1990) 83-121.

[Onv93] Onvural, R.O. Special Issue on Queueing Networks with Finite Capacity, Performance Evaluation, Vol. 17, 3 (1993).

[OnPe89] Onvural, R.O., and H.G. Perros "Some equivalencies on closed exponential queueing networks with blocking" Performance Evaluation, Vol.9 (1989) 111-118.

[OsBi09] Osorio C., Bierlaire M. "An analytic finite capacity queueing network model capturing the propagation of congestion and blocking", European Journal of Operational Research, 196 (2009) pp. 996 - 1007

[Per84] Perros, H.G. "Queueing Networks with Blocking: A Bibliography" ACM Sigmetrics, Performance Evaluation Review, Vol. 12 (1984) 8-12.

[Per89] H.G. Perros, A bibliography of papers on queueing networks with finite capacity queues, Performance Evaluation, vol. 10, n.3 (1989) 225-260.

[Per94] Perros, H.G. Queueing networks with blocking. Oxford University Press, 1994.

[Ser99] Sereno, M. "Mean Value Analysis of product form solution queueing networks with repetitive service blocking" Performance Evaluation, Vol. 36-37 (1999) 19-33.

[ShaYa89] Shanthikumar, G.J., and D.D. Yao "Monotonicity Properties in Cyclic Queueing Networks with Finite Buffers" in First International Workshop on Queueing Networks with Blocking, (Perros and Altiok Eds), Elsevier Science Publishers, North Holland, 1989.

[YaMiYaMa09] Yamadaa T., Mizuharab N., Yamamotoc H., Matsuib M., "A performance evaluation of disassembly systems with reverse blocking", Computers & Industrial Engineering Intelligent Manufacturing and Logistics, Volume 56, Issue 3, April 2009, Pages 1113-1125

[VanTi86] Van Dijk, N.M., and H.C. Tijms "Insensitivity in Two Node Blocking Models with Applications" in Teletraffic Analysis and Computer Performance Evaluation (Boxma, Cohen and Tijms Eds.), Elsevier Science Publishers, North Holland, 1986, 329-340.

[ZhaCaSmi08] E.Z. Zhang, G. Casale, and E. Smirni. Interarrival Times Characterization and Fitting for Markovian Traffic Analysis. TR WM-CS-2008-02, College of William and Mary, 2008.