ON REPRESENTING MULTICLASS M/M/k QUEUES BY GENERALIZED STOCHASTIC PETRI NETS

Simonetta Balsamo
Andrea Marin
Computer Science Department
Università Ca' Foscari di Venezia
Via Torino 155, 30172 Venezia Mestre, Italy
{balsamo,marin}@dsi.unive.it

KEYWORDS

GSPN, multiclass queuing systems.

ABSTRACT

In this paper we study the relations between multiclass BCMP-like service stations and generalized stochastic Petri nets (GSPN). Representing queuing discipline with GSPN models is not easy. We focus on representing multi-class queuing systems with different queuing disciplines by defining appropriate finite GSPN models. Note that queuing discipline in general affects performance measures in multi-class systems. For example, BCMP-like service centers with First Come First Served (FCFS) and with Last Come First Served with Preemptive Resume (LCFSPr) have a (different) product-form solution under different hypotheses. We define structurally finite GSPNs equivalent to the multi-class M/M/k queuing system with FCFS, LCFSPR, Processor Sharing (PS) and Infinite Servers (IS). Equivalence holds in terms of steady state probability function and average performance measure. The main idea is to define a finite GSPN model that simulates the behavior of a given queue discipline with some appropriate random choice. Moreover, we prove that the combination of the introduced equivalent models has a closed-form steady state probability by the $M \Rightarrow M$ property. We consider queuing systems with both a single server with load dependent service rate, and multiple servers with constant service rate.

1 INTRODUCTION

Queuing theory and (Generalized) Stochastic Petri Nets are important classes of stochastic models used to evaluate system performances. Queuing systems have been widely applied to represent resource contention systems where a set of customers competes for resource usage. Queuing networks (QN) extend and combine various queuing systems to represent more complex systems. Generalized Stochastic Petri nets (GSPN) can be naturally used to represent systems with synchronization and concurrency and to perform both qualitative and quantitative analysis. Under some exponential and independence assumptions these models can be studied through the associated continuous-time Markov chain (CTMC). In order to improve the algorithmic analysis efficiency for restricted classes of QN and SPNs, product-form theorems have been introduced. BCMP theorem (Baskett, Chandy, Muntz, & Palacios, 1975) is the main result for QN while the papers of Henderson et al. (Henderson, Lucic, & Taylor, 1989; Coleman, Henderson, & Taylor, 1996) present the main results for SPN and they have been extended to GSPN in the paper (Balbo, Bruell, & Sereno, 2002).

In this paper we investigate the relations between the BCMP queuing centers and GSPN models. The problem is trivial when all the customers in the QN are statistically identical (single class QN) thanks to the insensitivity property, i.e., the performance measures depend on the average of the service time and not on the queuing disciplines. Some difficulties arise when the QN customers are clustered in different classes which have different behaviors for each service center. In (Baskett et al., 1975) it is proved that in this case the queuing disciplines influence the QN performance measures. Thus if we try to represent a QN station by a GSPN we expect to have different models according to the queueing disciplines. Most of the works in this field devote a little attention to this problem. To the best of our knowledge, representing scheduling disciplines in multiclass models with finite GSPNs is still an open problem. In (Vernon, Zahorjan, & Lazowska, 1987) the authors introduce a comparison between QN models and SPN models based on the representation of multiclass features by colored Petri nets. However the differences between different scheduling disciplines are not analyzed. Balbo et al. in (Balbo, Bruell, & Ghanta, 1998) combine GSPN and product-form QN by replacing subsystem in a low-level model with their flow equivalent models. Still little attention is devoted to scheduling disciplines. In (Balbo, Bruell, & Sereno, 2003) the authors observe how they can map each service station of a BCMP QN to a complex GSPN which does not hold the GSPN product-form conditions of (Balbo et al., 2002). The GSPN model depends on the scheduling disciplines but it has an infinite number of places and transitions for the FCFS and LCFSPR stations. Then they give a finite and remarkably compact representation by a GSPN equivalent to the detailed model. The compact representation holds the product-form conditions for GSPN showed in (Balbo et al., 2002) but it does not distinguish different queuing disciplines by mapping everything in the PS discipline.

In this paper we present an equivalence result between two types of stochastic models. We propose a finite GSPN representation of a set of queuing systems with various scheduling disciplines. According to the BCMP-type service centers we analyze First Come First Served (FCFS), Last Come First Served with preemptive resume (LCFSPR), Processor Sharing (PS) and Infinite Servers (IS) scheduling disciplines. The main idea behind these results is a probabilistic model of the queue, i.e., all the customers of the same class wait in the same place and when a server becomes free the customer which gets the service is chosen in a probabilistic way similarly to what happens with the random queuing discipline. In the LCFSPR discipline, we also choose probabilistically the customer that looses the server when a new customer arrives to the system.

The advantage of having a finite representation which is different for the various scheduling disciplines is twofold: first it makes the analysis easier and it can be used for practical purposes. Second it does not require the definition of new semantic for the GSPN according to the queuing disciplines. Thus existing analysis or simulation tools can be used with the GSPN nets defined in this work. The proposed results are interesting because they allow the representation of an M/M/k queue with various queuing disciplines by a compact GSPN, which is equivalent to the queuing system in term of steady state queue length distribution. A practical consequence can be that it can extend a GSPN simulator or analyzer for analyzing multiclass queue systems. The only requirement is that the tool is able to model state-dependent firing rates for timed transitions and state-dependent weights for immediate transitions. There is no need to support the colored model extension to represent different classes.

The paper is structured as follows. Section 2 briefly reviews the GSPN models recalling formalism we chose, Section 3 reviews some results of the queuing systems theory used later in the paper. In Sections 4, 5 we introduce the GSPNs respectively equivalent to the FCFS and LCFSPR multiclass M/M/k queue. Section 6 discuss the GSPN models for both PS scheduling and IS systems. Section 7 uses $M \Rightarrow M$ property to state some considerations on the form of the steady state probability for some combinations of the GSPN models. Finally, Section 8 provides some concluding remarks.

2 GENERALIZED STOCHASTIC PETRI NETS

In this section we briefly recall the Generalized Stochastic Petri Nets (GSPN). We consider the notation for GSPN introduced in (Marsan, Balbo, Conte, Donatelli, & Franceschinis, 1995). In order to allow marking dependent probabilities for solving conflicts among immediate transitions we use the techniques discussed in (Chiola, Marsan, Balbo, & Conte, 1993). Let us define a marked Stochastic Petri Net which consists of a 8-tuple as follows:

$$GSPN = (\mathcal{P}, \mathcal{T}, I(\cdot,\cdot), O(\cdot,\cdot), H(\cdot,\cdot), \Pi(\cdot), w(\cdot,\cdot), \mathbf{m_0})$$

where:

- $\mathcal{P} = \{P_1, \ldots, P_M\}$ is the set of M places,
- $\mathcal{T} = \{t_1, \ldots, t_N\}$ is the set of N transitions (both immediate and timed),
- $I(t_i, p_j) : \mathcal{T} \times \mathcal{P} \to \mathbb{N}$ is the input function, $1 \le i \le N$, $1 \le j \le M$,
- $O(t_i, p_j) : \mathcal{T} \times \mathcal{P} \to \mathbb{N}$ is the output function, $1 \le i \le N$, $1 \le j \le M$,
- $H(t_i, p_j) : \mathcal{T} \times \mathcal{P} \to \mathbb{N}$ is the inhibition function, $1 \le i \le N$, $1 \le j \le M$,
- $\Pi(t_i) : \mathcal{T} \to \mathbf{N}$ is a function that specifies the priority of transition t_i, $1 \le i \le N$,
- $\mathbf{m} \in \mathbb{N}^M$ denotes a marking or state of the net, where m_i represents the number of tokens in place P_i, $1 \le i \le N$,
- $w(t_i, \mathbf{m}) : \mathcal{T} \times \mathbb{N}^M \to \mathbb{R}$ is the function which specifies for each timed transition t_i and each marking $\mathbf{m}$ a state dependent firing rate, and for immediate transitions a state dependent weight,
- $\mathbf{m_0} \in \mathbf{N}^M$ represents the initial state of the GSPN, i.e. the number of tokens in each place at the initial state.

We consider ordinary nets, i.e., functions I, O and H take values in $\{0, 1\}$. For each transition t_i let us define the input vector $\mathbf{I}(t_i)$, the output vector $\mathbf{O}(t_i)$ and the inhibition vector $\mathbf{H}(t_i)$ as follows: $\mathbf{I}(t_i) = (i_1, \ldots, i_M)$ where $i_j = I(t_i, P_j)$, $\mathbf{O}(t_i) = (o_1, \ldots, o_M)$ where $o_j = O(t_i, P_j)$ and $\mathbf{H}(t_i) = (h_1, \ldots, h_M)$ where $h_j = H(t_i, P_j)$. Function $\Pi(t_i)$ associates a priority to transition t_i. If $\Pi(t_i) = 0$ then t_i is a timed transition, i.e., it fires after an exponentially distributed firing time with mean $1/w(t_i, \mathbf{m})$, where $\mathbf{m}$ is the marking of the net. If $\Pi(t_i) > 0$ then t_i is an immediate transition and its firing time is zero. We say that transition t_a is enabled by marking $\mathbf{m}$ if $m_i \ge I(t_a, p_i)$ and $m_i < H(t_a, p_i)$ for each $i = 1, \ldots, M$ and no other transition of higher priority is enabled. We consider just two priority levels, 0 and 1. Hence when an immediate transition is enabled all the timed ones are disabled. The firing of transition t_i changes the state of the net from $\mathbf{m}$ to $\mathbf{m} - \mathbf{I}(t_i) + \mathbf{O}(t_i)$. The reachability set $RS(\mathbf{m_0})$ of the net is defined as the set of all markings that can be reached in zero or more firings from $\mathbf{m_0}$. We say that marking $\mathbf{m}$ is tangible if it enables only timed transitions and it is vanishing otherwise. For a vanishing marking $\mathbf{m}$ let $\mathcal{T}_\alpha$ be the set of enabled immediate transitions. Then the firing probability for any transition $t_i \in \mathcal{T}_\alpha$ and any state $\mathbf{m}$ is denoted by $p(t_i, \mathbf{m})$ and it is defined as follows:

$$p(t_i, \mathbf{m}) = \frac{w(t_i, \mathbf{m})}{\sum_{t_j \in \mathcal{T}_\alpha} w(t_j, \mathbf{m})}. \qquad (1)$$

Given a tangible marking $\mathbf{m}$ the transition with the lowest associated stochastic time fires.

A GSPN is represented by a graph with the following conventions: timed transitions are white filled boxes, immediate transitions are black filled boxes, places are circles, if $I(t_i, p_j) > 0$ we draw an arrow from p_j to t_i labelled with $I(t_i, p_j)$, if $O(t_i, p_j) > 0$ we draw an arrow from t_i to p_j labelled with $O(t_i, p_j)$, if $H(t_i, p_j) > 0$ we draw an circle ending line from p_j to t_i labelled with the value of $H(t_i, p_j)$, the marking $\mathbf{m}$ is represented by a set of m_j filled circles representing the tokens in place p_j for each $j = 1, \ldots, M$.

For ordinary nets we do not use labels for the arrows.

GSPN analysis consists in finding the steady state probability for each tangible marking of the reachability set. Some analysis techniques are presented in (Marsan et al., 1995). Under general assumptions, the stochastic process generated by the dynamic behavior of a standard SPN is a CTMC process. Mean state sojourn times are computed from the mean transition delays of the net. For GSPNs the distribution of the sojourn time in any marking can be expressed as a negative exponential and deterministically zero distributions for tangible and vanishing markings, respectively. Thus the marking process can be studied as a semi-Markov random process.

The GSPN models introduced in this paper present marking processes which allow us to easily reduce the semi-Markov process to a CTMC. In fact whenever a vanishing marking is reached, the next marking is tangible. Thus we can simply obtain a CTMC whose states are the tangible states of the original process and the transition rates are computed weighting the transitions rates of the original process with the firing probabilities of the immediate transitions.

Finally let us introduce some other notations: let $\mathbf{e_i}$ be an M-dimensional vector with all zero components but the i-th which is 1. We use the lower case t to name immediate transitions, the upper case T to name timed transitions, $\tilde{t}$ to name a generic timed or immediate transition.

3 SINGLE QUEUING SYSTEMS WITH DIFFERENT CLASSES OF CUSTOMERS

In this section we briefly recall single queuing systems with different classes of customers classifying them on the number of servers and scheduling disciplines. Let us consider an open queuing system with external arrivals, a queue, a set of identical servers and a set of R customer classes. Customers of class r arrive at the system according to a Poisson process with rate λ_r and require an exponentially distributed random service time with parameter μ_r, $r = 1, \ldots, R$. The system has a set of independent servers, possibly infinite.

We consider the following disciplines: First Come First Server (FCFS), Last Come First Server with Preemptive Resume (LCFSPR), Processor Sharing (PS). Let's start by considering a multiclass PS queuing system with a single server with load dependent service rate. Following the BCMP (Baskett et al., 1975) conventions, assume that the service rate can be expressed by a combination of a capacity function $x(n)$ depending on the total number of customers n at the station, and a class dependent capacity function $y_r(n_r)$, where n_r is the number of customers of class r at the station and a constant class dependent service rate μ_r. So the effective service rate for a customer of class r is given by the product $x(n)y_r(n_r)\mu_r$. Note that $x(1) = y_r(1) = 1$. Under stability conditions, the steady state probability of this service center is given by:

$$\pi'(\mathbf{n}) = \pi'_0 \frac{n!}{\prod_{i=1}^{R} n_i!} \prod_{i=1}^{R} \lambda_i^{n_i} \prod_{b=1}^{n} \frac{1}{x(b)} \prod_{r=1}^{R} \Big[\Big(\frac{1}{\mu_r} \Big)^{n_r} \prod_{a=1}^{n_r} \frac{1}{y_r(a)} \Big]. \quad (2)$$

Formula (2) holds also for single server LCFSPR with load dependent service rate. In order to obtain the steady state probabilities for M/M/k mutliclass system it suffices to set appropriate capacity function. An LCFSPR or PS center with k load independent servers requires to set $x(n) = \frac{\min(n,k)}{n}$ and $y_r(n_r) = n_r$.

Formula (2) still holds for FCFS scheduling discipline if the service rate is class independent, i.e., $\mu_r = \mu$ and $y_r(n_r) = 1$ for $1 \leq r \leq R$ and $n_r \geq 1$. In order to study the M/M/k/FCFS queuing system we have to set $x(n) = \min\{n, k\}$, and formula (2) becomes:

$$\pi'(\mathbf{n}) = \pi'_0 \frac{n!}{\prod_{i=1}^{R} n_i!} \prod_{i=1}^{R} \lambda_i^{n_i} \prod_{a=1}^{n} \frac{1}{\mu(a)}, \quad (3)$$

where $\mu(a) = x(a)\mu$.

4 REPRESENTING M/M/k/FCFS QUEUE BY GSPN

In this section we define a GSPN that represents an R-multiclass M/M/k/FCFS queue. Then we prove that the GSPN model is equivalent to the queuing system in terms of the steady state probability. Given the M/M/k/FCFS models defined as in Section 3 let us define the model called *GSPN-1.*

Definition 1 (GSPN-1). *According to GSPN definition given in Section 2:*

- $\mathcal{P} = \mathcal{P}_q \cup \mathcal{P}_s \cup \{P_{2R+1}\}$ *with* $\mathcal{P}_q = \{P_1, \ldots, P_R\}$ *and* $\mathcal{P}_s = \{P_{R+1}, \ldots, P_{2R}\}$,
- $\mathcal{T} = \mathcal{T}_w \cup \mathcal{T}_q$ *where* $\mathcal{T}_q = \{t_1, \ldots, t_R\}$ *and* $\mathcal{T}_w = \{T_{R+1}, \ldots, T_{2R}\}$,

– *function Π defined as follows:*

$$\Pi(\tilde{t_i}) = \begin{cases} 0 & \text{if } R+1 \leq i \leq 2R \\ 1 & \text{if } 1 \leq i \leq R \end{cases},$$

– *input and output vectors for transition t_i, $1 \leq i \leq R$: $\mathbf{I}(t_i) = \mathbf{e_i} + \mathbf{e_{2R+i}}$ and $\mathbf{O}(t_i) = \mathbf{e_{R+i}}$. Input and output vector for transition T_{R+i}: $\mathbf{I}(T_{R+i}) = \mathbf{e_{R+i}}$ and $\mathbf{O}(T_{R+i}) = \mathbf{e_{2R+1}}$,*
– *$\mathbf{H}(t_i) = (0, \ldots, 0)$ for all $t_i \in \mathcal{T}$,*
– *$w(T_{R+i}, \mathbf{m}) = m_{R+i}\mu$ for $1 \leq i \leq R$ and $w(t_i, \mathbf{m}) = m_i$ for $1 \leq i \leq R$,*
– *$\mathbf{m_0} = (0, \ldots, 0, k)$.*

Tokens arrive to places P_i, $1 \leq i \leq R$ according to Poisson stochastic processes.

Figure 1 illustrates the graphical representation of GSPN-1 model where $t_1, \ldots, t_R$ are immediate transitions and $T_{R+1}, \ldots, T_{2R}$ are exponential transitions.

Let $\mathbf{m}$ be a valid vanishing state of the GSPN-1, and let $\mathcal{T}_a \subseteq \mathcal{T}_q$ be the set of immediate transitions enabled by $\mathbf{m}$, then the probability of firing of $t_i \in \mathcal{T}_a$ can be written as:

$$p(t_i, \mathbf{m}) = p_i(\mathbf{m}) = \frac{m_i}{\sum_{j \in \{j | t_j \in \mathcal{T}_a\}} m_j} \qquad (4)$$

We shall now derive a closed form solution for the steady state probability of GSPN-1 model by considering the set of reachable markings $\mathbf{m} = (m_1, \ldots, m_{2R+1})$. This is given by Lemma 1. Then we introduce a state aggregation by defining the aggregate state $\mathbf{n} = (n_1, \ldots, n_R)$ where $n_i = m_i + m_{R+i}$, $1 \leq i \leq R$. This state corresponds to the number of customers of class i in the queuing model. Theorem 1 provides the closed form solution for model GSPN-1 in terms of aggregated stationary probability of state $\mathbf{n}$. Finally the GSPN-1 model is shown to be equivalent to the M/M/k FCFS multiclass queuing system in terms of stationary probability.

Lemma 1. *Let $\mathbf{m} = (m_1, \ldots, m_{2R+1})$ be a reachable tangible state of the GSPN-1. Then if the stability condition holds, the stationary state probability can be written as follows:*

$$\pi(\mathbf{m}) = \pi_0 \prod_{i=1}^{R} \lambda_i^{m_i + m_{R+i}} \frac{(\sum_{i=R+1}^{2R} m_i)!}{\prod_{i=R+1}^{2R} m_i!} \cdot \frac{(\sum_{i=1}^{R} m_i)!}{\prod_{i=1}^{R} m_i!} \prod_{j=1}^{\sum_{i=1}^{2R} m_i} \frac{1}{\mu(j)}. \qquad (5)$$

where π_0 is a normalizing constant, $\mu(j) = x(j)\mu$ following the BCMP conventions.

The proof is given in appendix in the technical report (Balsamo & Marin, 2007) and is based on verifying the set of the CTMC global balance equations.

Theorem 1. *Consider model GSPN-1 and let $n_i = m_i + m_{R+i}$, $1 \leq i \leq R$ and $\mathbf{n} = (n_1 \ldots, n_R)$ be an aggregated state. Let $\pi_a(\mathbf{n})$ be the steady state probability of n_i for $i = 1, \ldots, R$. Then we can write:*

$$\pi_a(\mathbf{n}) = \pi_0 \frac{(\sum_{i=1}^{R} n_i)!}{\prod_{i=1}^{R} n_i!} \prod_{i=1}^{R} \lambda_i^{n_i} \prod_{i=1}^{\sum_{i=1}^{R} n_i} \frac{1}{\mu(i)} \quad \forall \mathbf{n} \in \mathbb{N}^R. \qquad (6)$$

The proof is based on a convolution formula for binomial coefficients and can be found in (Balsamo & Marin, 2007).

Corollary 1. *The M/M/k queuing system with FCFS discipline, R customer classes, arrival rates λ_i, $1 \leq i \leq R$, single server rate μ and steady state probability $\pi'(\mathbf{n})$ is equivalent to the GSPN-1 in terms of steady state probability, i.e., $\pi_a(\mathbf{n}) = \pi'(\mathbf{n})$ for all $\mathbf{n} \in \mathbb{N}^R$ where $\pi_a(\mathbf{n})$ is the aggregated probability of GSPN given by formula (6).*

Proof. It follows immediately from equation (3) and Theorem 1. □

Note it can be shown by trivial counterexamples that GSPN-1 does not hold the steady state distribution (2) when the service rate is class dependent.

GSPN-1 can as well simulate a single server FCFS service station with an BCMP-like load dependent service rate as proved in the technical report (Balsamo & Marin, 2007). The net structure complexity is linear on R, the number of customer classes.

5 REPRESENTING M/M/k/LCFSPR QUEUE BY GSPN

In this section we introduce a GSPN which can be considered equivalent, for steady state probability, to a multiclass M/M/k queue with LCFS with preemptive resume scheduling discipline. As we consider just exponentially distributed service times, we do not consider the problem of representing the resume. We provide a model for this queue system whose structure is finite and depends only on the number of classes of customers, i.e., not on the number of servers.

Definition 2 (GSPN-2). *According to GSPN definition given in Section 2:*

– *$\mathcal{P} = \mathcal{P}_q \cup \mathcal{P}_w \cup \mathcal{P}_a \cup \{P_{3R+1}\}$ where $\mathcal{P}_q = \{P_1, \ldots, P_R\}$ and $\mathcal{P}_w = \{P_{R+1}, \ldots, P_{2R}\}$ and $\mathcal{P}_a = \{P_{2R+1}, \ldots, P_{3R}\}$,*
– *$\mathcal{T} = \mathcal{T}_q \cup \mathcal{T}_w \cup \mathcal{T}_f \cup \mathcal{T}_g$ where $\mathcal{T}_q = \{t_1, \ldots, t_R\}$ and $\mathcal{T}_w = \{T_{R+1}, \ldots, T_{2R}\}$ and $\mathcal{T}_f = \{t_{2R+1}, \ldots, t_{3R}\}$ and $\mathcal{T}_g = \{t_{ij},\ 1 \leq i, j \leq R\}$,*
– *function Π is defined as follows:*

$$\Pi(\tilde{t}) = \begin{cases} 1 & \text{if } \tilde{t} \in \mathcal{T}_q \cup \mathcal{T}_f \cup \mathcal{T}_g \\ 0 & \text{if } \tilde{t} \in \mathcal{T}_w \end{cases},$$

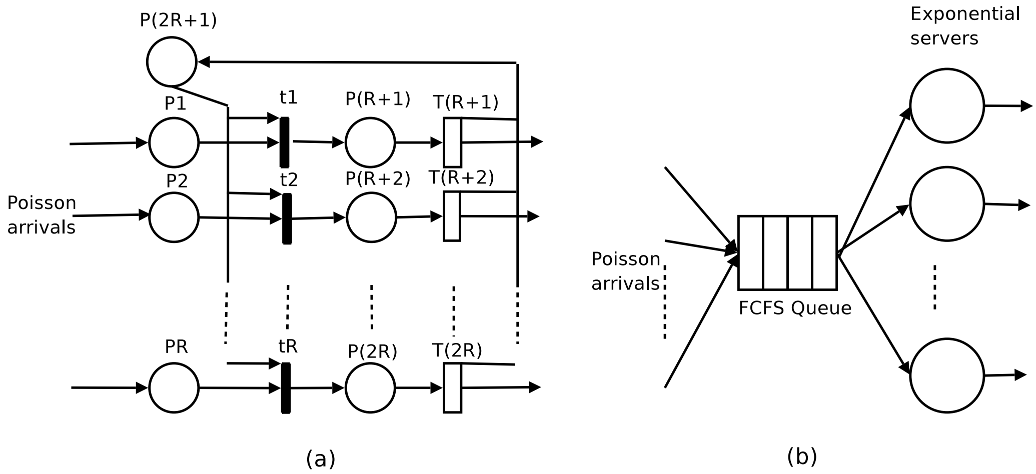

Fig. 1. (a) Graphical representation of model GSPN-1. (b) Queuing station associated

- *Let* $1 \leq i, j \leq R$. *The input and output vectors of* $t_i \in T_q$: $\mathbf{I}(t_i) = \mathbf{e}_i + \mathbf{e}_{3R+1}$ *and* $\mathbf{O}(t_i) = \mathbf{e}_{R+i}$. *The input and output vectors for* $T_{R+i} \in \mathcal{T}_w$: $\mathbf{I}(t_{R+i}) = \mathbf{e}_{R+i}$ *and* $\mathbf{O}(t_{R+i}) = \mathbf{e}_{3R+1}$. *The input and output vectors for* $t_{2R+i} \in \mathcal{T}_f$: $\mathbf{I}(t_{2R+i}) = \mathbf{e}_{2R+i} + \mathbf{e}_{3R+1}$ *and* $\mathbf{O}(t_{2R+i}) = \mathbf{e}_{R+i}$. *The input and output vectors for* $t_{ij} \in \mathcal{T}_g$: $\mathbf{I}(t_{ij}) = \mathbf{e}_{2R+i} + \mathbf{e}_{R+j}$ *and* $\mathbf{e}_j + \mathbf{e}_{R+i}$,
- $\mathbf{H}(t_i) = (0, \ldots, 0)$ *for* $t_i \in \mathcal{T}_q \cup \mathcal{T}_w \cup \mathcal{T}_f$ *and* $\mathbf{H}(t_{ij}) = \mathbf{e}_{3R+1}$ *for* $t_{ij} \in \mathcal{T}_g$,
- *for* $1 \leq i, j \leq R$ *let* $w(T_{R+i}, \mathbf{m}) = m_{R+i}\mu_i$, $w(t_i, \mathbf{m}) = m_i$, $w(t_{2R+i}, \mathbf{m}) = 1$ *and* $w(t_{ij}, \mathbf{m}) = m_{R+j}$,
- $\mathbf{m_0} = (0, \ldots, 0, k)$.

Tokens arrive to places P_{2R+i}, $1 \leq i \leq R$, *according to Poisson stochastic processes.*

Figure 2 shows a graphical model for $R = 2$ classes LCFSPR queue where dotted lines are introduced for the sake of readability and they do not have ant particular meaning. Note that when a token arrives to the place P_{2R+i} it is temporally (i.e. the state is vanishing) stored in P_{2R+i} and we have two cases:

- there is at least one free server, i.e. $m_{3R+1} > 0$, thus the customer goes immediately in service. This is modelled by the immediate transition set $\mathcal{T}_f$
- all the servers are busy, i.e. $m_{3R+1} = 0$, so a customer is preempted and put in queue and the new customer goes in service. This is modelled by R^2 transitions, $\mathcal{T}_g$. The inhibitor arcs are needed to avoid pre-emption when there is at least one free server.

By the structure analysis of the network we can solve the conflicts on immediate transitions introducing just one simple function. When one or more transitions of $\mathcal{T}_q$ are enabled, the probability of firing for the i-th transition is:

$$p(t_i, \mathbf{m}) = p_i(\mathbf{m}) = \frac{m_i}{\sum_{l=1}^{R} m_l}. \tag{7}$$

When one or more transitions of $\mathcal{T}_g$ are enabled, the probability of firing is:

$$p(t_{ij}, \mathbf{m}) = p_{ij}(\mathbf{m}) = \frac{m_{R+j}}{\sum_{l=1}^{R} m_{R+l}}. \tag{8}$$

Now we can state a main lemma for model GSPN-2 representation:

Lemma 2. *Let* $\mathbf{m} = (m_1, \ldots, m_{3R+1})$ *be a reachable tangible marking of GSPN-2 model. Then if the stability condition holds, the stationary state probability can be written as follows:*

$$\pi(\mathbf{m}) = \pi_0 \prod_{i=1}^{R} \lambda_i^{m_i + m_{R+i}} \frac{(\sum_{i=1}^{R} m_i)!}{\prod_{i=1}^{R} m_i!} \frac{(\sum_{i=1}^{R} m_{R+i})!}{\prod_{i=1}^{R} m_{R+i}!} \cdot \prod_{i=1}^{R} \left(\frac{1}{\mu_i}\right)^{m_i + m_{R+i}} \prod_{j=1}^{\sum_{i=1}^{2R} m_i} \frac{1}{\min(j, k)}. \tag{9}$$

where μ_i *is the average service rate for one customer of class* i *when there are no other customers in the system,* k *is the number of servers,* π_0 *is a normalizing constant.*

The proof is given in the technical report (Balsamo & Marin, 2007).

Theorem 2. *Consider model GSPN-2 and let* $n_i = m_i + m_{R+i}$, $1 \leq i \leq R$ *and* $\mathbf{n} = (n_1, \ldots, n_R)$ *be an aggregated state. Let* $\pi_a(\mathbf{n})$ *be the steady state probability of* n_i *for* $i = 1, \ldots, R$. *Then we can write:*

$$\pi_a(\mathbf{n}) = \pi_0 \frac{(\sum_{i=1}^{R} n_i)!}{\prod_{i=1}^{R} n_i!} \prod_{i=1}^{R} \lambda_i^{n_i} \prod_{i=1}^{R} \left(\frac{1}{\mu_i}\right)^{n_i} \cdot \prod_{i=1}^{\sum_{i=1}^{R} n_i} \frac{1}{\min(k, i)} \quad \forall \mathbf{n} \in \mathbf{N}^R. \tag{10}$$

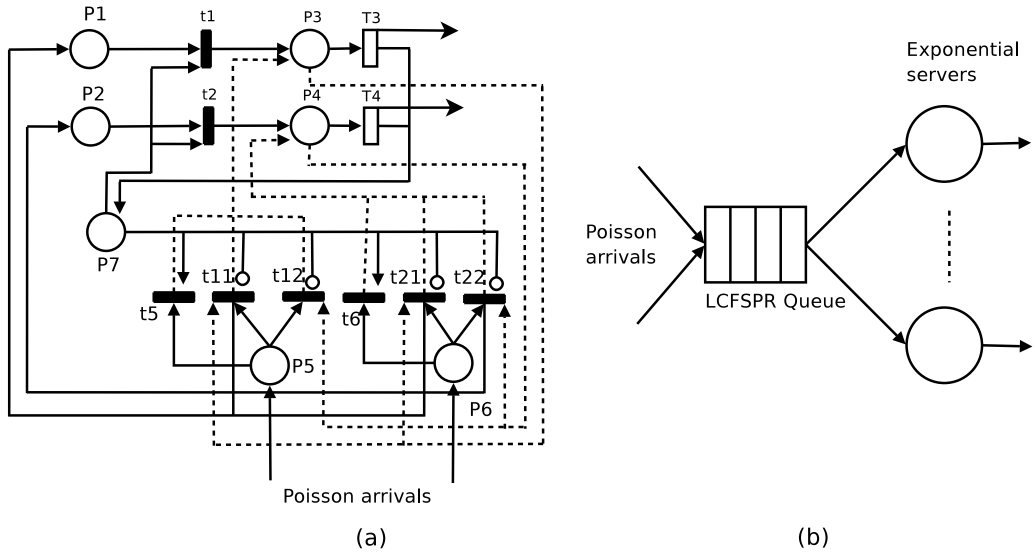

Fig. 2. (a) Graphical representation of model GSPN-2 for two classes of customers. (b) Queuing station associated.

The proof is based on the Vandermonde formula and it is similar the one given for Theorem 1 in (Balsamo & Marin, 2007).

Corollary 2. *The M/M/k queuing system with LCFSPR discipline, R customer classes, arrival rates λ_i, single server rate μ_i for class i customers and steady state probability $\pi'(\mathbf{n})$ is equivalent to model GSPN-2 in terms of steady state probability, i.e., $\pi_a(\mathbf{n}) = \pi'(\mathbf{n})$ for all $\mathbf{n} \in \mathbf{N}^R$, where $\pi_a(\mathbf{n})$ is the aggregated probability of GSPN given by formula (10). The normalizing constant is $\pi_0 = \pi(0, \ldots, 0, k) = \pi'(0, \ldots, 0, k)$.*

Proof. It follows immediately from queuing theory and Theorem 2. □

The net GSPN-2 can as well simulate a single server LCFSPR service station with a BCMP-like load dependent service rate as proved in technical report (Balsamo & Marin, 2007).

By defining $n_i = m_i + m_{R+i}$ we can aggregate the states and we can prove that the steady state probability π_a of the aggregated CTMC is identical to probability π' defined by equation (2). For what concerns the net structure complexity, the number of places grows as $\mathcal{O}(R)$ and the number of transitions grows as $\mathcal{O}(R^2)$.

6 REPRESENTING M/M/k/PS QUEUE AND M/M/∞ QUEUE BY GSPN

The processor sharing discipline can be easily represented considering that the k processors are shared among the users in the system. Different classes of users can have different average time services, but all modelled by exponentially distributed random variables. We can think that the k servers are shared among the R classes in proportion to the number of customers of the classes.

Definition 3 (GSPN-3). *Let us define the model GSPN-3 as follows:*

- $\mathcal{P} = \{P_1, \ldots, P_R\}$,
- $\mathcal{T} = \{T_1, \ldots, T_R\}$,
- $\Pi(T_i) = 1$ *for each* $T_i \in \mathcal{T}$,
- $\mathbf{I}(T_i) = \mathbf{e}_i$ *and* $\mathbf{O}(T_i) = (0, \ldots, 0)$ *for each* $T_i \in \mathcal{T}$,
- $\mathbf{H}(T_i) = (0, \ldots, 0)$ *for each* $T_i \in \mathcal{T}$,
- $w(T_i, \mathbf{m}) = \frac{m_i}{m} \min(k, m)$ *where* $m = \sum_{j=1}^{R} m_j$ *for each* $T_i \in \mathcal{T}$,
- $\mathbf{m_0} = (0, \ldots, 0)$.

Note that this model is equivalent to a queuing system with PS discipline and one server with load-dependent exponential service time to simulate the multi-server feature. Therefore it immediately follows the theorem:

Theorem 3. *Consider model GSPN-3. Then if stability condition holds the stationary state probability can be written as follows:*

$$\pi(\mathbf{m}) = \pi_0 \frac{(\sum_{i=1}^{R} m_i)!}{\prod_{i=1}^{R} m_i!} \prod_{i=1}^{R} \lambda_i^{m_i} \prod_{i=1}^{R} \left(\frac{1}{\mu_i}\right)^{m_i} \cdot \prod_{i=1}^{\sum_{i=1}^{R} m_i} \frac{1}{\min(k, i)}, \quad (11)$$

where μ_i is the average service rate for one customer of class i when there are no other customers in the system, k is the number of servers, π_0 is a normalizing constant.

This model is similar to the compact model introduced in (Balbo et al., 2003), the only difference is that we allow a whole state dependent firing rate thus we don't need a place whose tokens represent the total number of customers in the system.

Model GSPN-3 can easily represent also the IS center. It suffices to set the firing rates of each transition T_i as $m_i\mu_i$, $1 \leq i \leq R$.

7 M ⇒ M PROPERTY ON THE GSPN REPRESENTATION

Markov implies Markov property is introduced and studied by Muntz (Muntz, 1972). In that paper he shows that if a queuing system with Poisson arrivals presents departures according to a Poisson process (M⇒M property) then a combination of service centers of this type in a queuing network has a product-form solution. As we are considering GSPNs we will prove that a combination of GSPN-1, GSPN-2 and GSPN-3 models still holds a closed-form steady state probability by defining appropriate traffic processes over the CTMC associated to each of the models and using the results given in (Melamed, 1979) which generalize Muntz's work. We now briefly review Melamed's results limited to a CTMC in steady state. Consider an ergodic CTMC with state space Γ and a set of traffic transitions denoted by Θ_1, $\ldots$, Θ_R, where $\Theta_i \subseteq \Gamma \times \Gamma$, $\Theta_i \neq \emptyset$. Let us define $K_i(t)$ as the process which counts the number of transitions $(\alpha, \beta) \in \Theta_i$ up to t. Let $m_i = \sum_{\gamma\in\Gamma}\sum_{\eta\in\Theta_i(\cdot,\gamma)} \pi(\eta)\xi(\eta \to \gamma)$ and for each state $\gamma \in \Gamma$ let $m_i(\gamma) = \sum_{\eta\in\Theta_i(\cdot,\gamma)} \pi(\eta)\xi(\eta \to \gamma)$ where $\Theta_i(\cdot, \gamma) = \{\beta | (\beta, \gamma) \in \Theta_i\}$ and $\xi(\eta \to \gamma)$ is the transition rate between states η and γ.
Then we can state that $K_i(t)$ are mutually independent Poisson processes if and only if the following equation holds:

$$\forall\gamma \in \Gamma, \quad \sum_{i=1}^{R} m_i(\gamma) = \pi(\gamma) \sum_{i=1}^{R} m_i \qquad (12)$$

We aim to study the departure traffic processes from our models. Take for example model GSPN-1, we can define R traffic processes as follows:

$$\Theta_i = \{(\mathbf{m}', \mathbf{m}) : |\mathbf{m}'|_i = |\mathbf{m}|_i + 1\}, \ i = 1, \ldots, R, \qquad (13)$$

where $|\mathbf{m}|_i = m_i + m_{R+i}$. In our case, in order to prove that $K_i(t)$ are independent Poisson processes when there are Poisson arrivals, it suffices to prove that:

$$\forall\gamma \in \Gamma, \qquad \sum_{\eta\in\Theta_i(\cdot,\gamma)} \pi(\eta)\xi(\eta \to \gamma) = \lambda_i\pi(\gamma), \qquad (14)$$

In (Balsamo & Marin, 2007) we prove that this condition holds for GSPN-1, GSPN-2 and GSPN-3 models by defining appropriate traffic processes. As observed in (Melamed, 1979) this property of the CTMC is equivalent to the $M \Rightarrow M$ given by Muntz thus it assures that a BCMP-like composition of these GSPN models holds a closed-form steady state probability function. Random switches between the blocks and user class switches can be easily modelled by immediate transitions.

8 FINAL REMARKS

In this paper we have shown how to represent multiclass single queuing systems by structurally finite GSPN for various queuing disciplines. For each of the BCMP center types we have introduced a GSPN model whose steady state probability, aggregating on the number of customers in the system for each class, is equal to the correspondent single queue service center. Hence the two models are equivalent in terms of steady state distribution and average performance indexes. The main advantages of our representation are the following.

- We define a finite GSPN model. The abstraction level of the GSPN model allows the representation of the queuing behavior without introducing an high level of details in the state specification. We distinguish the customers waiting in the queue from those being served without taking in account the arrival order. This leads to a steady state probability which is less detailed than the one proposed in (Balbo et al., 2003) which considers the single station detailed representation with the order of the customers in the queue, similarly to the BCMP paper (Baskett et al., 1975). On the other side the models we introduce are more detailed than those which just consider the total number of customers in a center as the compact models of (Balbo et al., 2003).
- The FCFS and the LCFSPR (or PS) scheduling disciplines have different GSPN representations. The GSPN models simulate the corresponding queuing system even if their semantic is different.

The main idea of the definition of the GSPN models is a probabilistic choice of the customer to serve when a server is avaiable and of the customer to preempt when there is an arrival to a LCFSPR station.

The M⇒M property allows us to state that a combination of GSPN-1, GSPN-2 and GSPN-3 models similar to the service centers combination in BCMP networks, has a simple closed form steady state probability. In (Afshari, Bruell, & Kain, 1982) authors define a queuing center isomorphic to GSPN-1 and show how it can be embedded in a BCMP queuing network so that the steady state probability function of the network does not change. In the GSPN formalism probabilistic routing can be easily simulated by introducing a block with a place and an immediate transition for each possible route just after the timed transitions of the models.

Further research deals with the extension of the proposed LCFSPR model to Coxian service time distributions and the definition of algorithms to identify GSPN which are compositions of models GSPN-1, GSPN-2, GSPN-3 and in order to obtain efficiently a set of significant performace indexes.

REFERENCES

Afshari, P. V., Bruell, S. C., & Kain, R. Y. (1982). Modeling a new technique for accessing shared buses. In *Proc. of the computer network performance symposium* (pp. 4–13). New York, NY, USA: ACM Press.

Balbo, G., Bruell, S. C., & Ghanta, S. (1998). Combining queueing network and generalized stochastic Petri nets for the solution of complex models of system behavior. *IEEE Trans. on Computers*, *37*, 1251–1268.

Balbo, G., Bruell, S. C., & Sereno, M. (2002). Product form solution for Generalized Stochastic Petri Nets. *IEEE Trans. on Software Eng.*, *28*, 915–932.

Balbo, G., Bruell, S. C., & Sereno, M. (2003). On the relations between BCMP Queueing Networks and Product Form Solution Stochastic Petri Nets. *Proc. of 10th International Workshop on Petri Nets and Performance Models, 2003.*, 103–112.

Balsamo, S., & Marin, A. (2007). *On representing multiclass M/M/k queues by generalized stochastic Petri nets* (Tech. Rep. No. CS-2007-1). via Torino, 155 Venice.

Baskett, F., Chandy, K. M., Muntz, R. R., & Palacios, F. G. (1975). Open, closed, and mixed networks of queues with different classes of customers. *J. ACM*, *22*(2), 248–260.

Chiola, G., Marsan, M. A., Balbo, G., & Conte, G. (1993). Generalized stochastic Petri nets: a definition at the net level and its implications. *IEEE Trans. on Software Eng.*, *19*(2), 89–107.

Coleman, J. L., Henderson, W., & Taylor, P. G. (1996). Product form equilibrium distributions and a convolution algorithm for Stochastic Petri nets. *Performance Evaluation*, *26*, 159–180.

Henderson, W., Lucic, D., & Taylor, P. G. (1989). A net level performance analysis of Stochastic Petri Nets. *J. Austral. Math. Soc. Ser. B*, *31*, 176–187.

Marsan, M. A., Balbo, G., Conte, G., Donatelli, S., & Franceschinis, G. (1995). *Modelling with generalized stochastic Petri nets.* Wiley.

Melamed, B. (1979). On Poisson traffic processes in discrete-state markovian systems with applications to queueing theory. *Advances in Applied Prob.*(11), 218–239.

Muntz, R. R. (1972). *Poisson departure processes and queueing networks* (Tech. Rep. No. IBM Research Report RC4145). Yorktown Heights, New York.

Vernon, M., Zahorjan, J., & Lazowska, E. D. (1987). A comparsion of performance Petri Nets and queueing network models. *Proc. 3rd Intern. Workshop on Modelling Techniques and Performance Evaluation*, 181–192.

AUTHOR BIOGRAPHIES

SIMONETTA BALSAMO is a full professor of Computer Science at the university Ca' Foscari of Venice, Italy. Her research interests include performance and reliability modeling and analysis of computer and communication systems, parallel and distributed processing, distributed simulation, quantitative analysis of software architectures and integration of specification languages and performance models. She has published several papers in international journal, conference proceeding, books and special editions. She has served as general chair, program chair and program committee member for several internatinal conference, is associated editor of Performance Evaluation Journal. She is a member of ACM.

ANDREA MARIN was born in Venice, Italy and went to the Ca' Foscari University of Venice, where he studied computer science and obtained his degree in 2005. He is now a PhD student at Ca' Foscari university and he is interested in stochastic models for performance evaluation, and models for biological systems. His email address is: marin@dsi.unive.it and his Web-page can be found at http://www.dsi.unive.it/ marin/.