Representing LCFSPR BCMP service center with Coxian service time by GSPN

Simonetta Balsamo
Dipartimento di Informatica
Universita' Ca' Foscari di Venezia
via Torino, 155, Venice, Italy
balsamo@dsi.unive.it

Andrea Marin
Dipartimento di Informatica
Universita' Ca' Foscari di Venezia
via Torino, 155, Venice, Italy
marin@dsi.unive.it

ABSTRACT

In this paper we study the relations between multi-class BCMP-like service stations with Coxian service time and generalized stochastic Petri nets (GSPN). We consider multi-class product-form service center types of QN and their representation with GSPN, in order to investigate the relation between the product-form solutions of the two models. Representing queueing discipline with GSPN models is not easy. We focus on representing multi-class queueing systems with LCFSPR scheduling discipline and Coxian distributed service time. Note that the queueing 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 a structurally finite GSPN model with product-form that does not belong to the class of known product-form GSPN. Then we show that this model is equivalent to the multi-class M/COX/k/ LCFSPR queueing system, in terms of steady state probability and average performance measures. The main idea is to define a finite GSPN model that simulates the behavior of the given queue discipline by defining some appropriate random choices. Moreover, we prove that the combination of the introduced equivalent GSPN model has a closed-form steady state probability by the $M \Rightarrow M$ property.

Keywords

GSPN, BCMP theorem, Coxian distribution, multi-class system, LCFSPR scheduling discipline

1. INTRODUCTION

Queueing networks (QNs) and Generalized Stochastic Petri Nets (GSPNs) are stochastic models that have been widely applied to evaluate system performance. In this paper we study the relations between multi-class BCMP-like service stations with Coxian service time and generalized stochastic Petri nets (GSPN). We consider multi-class product-form service center types of QN and their

representation with GSPN, in order to investigate the relation between the product-form solutions of the two models.

QNs allow representing resource sharing system. A QN model is a collection of interacting service centers representing system resources and a set of customers representing the users sharing the resources. They provide a good balance of a relative high accuracy in performance results and efficiency in model analysis and solution. Under some general assumptions QNs can be described by an associated stochastic Markov process whose solution becomes soon computationally unfeasible, due to the high cardinality of the system state space that is exponential with the QN number of components (resources and customers). The class of product-form QNs allows a closed expression for the stationary state probability and the definition of efficient polynomial algorithms to calculate the performance measures. Product-form QN stationary probability can be expressed as product of functions which depends on the state and on the type of each node. The main result is known as BCMP theorem [5] and it considers open, closed and mixed QNs which consist of multiple class of customers, probabilistic routing, Poisson arrivals and four types of service centers. Some efficient solution algorithms have been defined for BCMP QNs and some further extensions of product-form QNs have been derived under special constraints.

Generalized Stochastic Petri Nets (GSPNs) extend Petri Nets models defined to specify synchronization behavior of concurrent systems. A Petri Net consists of a set of places, a set of transitions, an input function relating places to transitions, an output function relating transition to places, and a marking function, associating to each place a nonnegative integer number. In the stochastic timed extension the firing time of a transition, that is the time taken by the activity represented by the transitions, is expressed by random variables, usually exponential. GSPNs are continuous time stochastic Petri Nets that allow both exponentially timed and untimed (or immediate) transitions. The quantitative analysis of a stochastic Petri Net is based on the identification and solution of an associated Markov process built on the basis of the net reachability graph. In the general case, GSPNs are hard to study because of the generation of its reachability set which is an NP-hard problem. The Markov process suffers of the state space explosion problem and some special classes of product-form solution nets have been recently defined. Product-form GSPNs are studied in [13, 11, 2] but even if they express the stationary probability function as product of function depending on the marking of each place, they do not overcome the problem of deciding whether a marking is reachable from the model initial

state or not. This limit influences also the applicability of the algorithms defined for product-form GSPNs [21, 11]. Some non polynomial algorithms exist for product-form stochastic Petri Net, under further structural constraints, and many approximation techniques have been defined. A different approach is used in [6] where the author defines the conditions for a product-form solution of a stochastic process defined by the competition of set of CTMCs (e.g. stochastic processes associated to GSPNs or SPNs) competing for a set of resources.

GSPNs are usually considered as stochastic model with a lower level of abstraction than QNs. In fact GSPN semantic is strictly defined in terms of transitions, places and rules of firing. On the other hand GSPNs easily allow the definition of concurrency and synchronization, such as fork-join constructions without introducing a specific semantic, so being more expressive than BCMP QNs. Extensions of classical QNs have been introduced in order to represent special system features, such as synchronization and concurrency constraints, finite capacity queues, memory constraints and simultaneous resource possession or client-server communication patterns in concurrent and/or distributed system (e.g., Extended QN and Layered QN [15, 14, 23]). However, most of such QN extensions do not yield the BCMP product-form solution and can be solved by approximate solution techniques.

In this paper we study the relations between BCMP QNs and GSPNs. At a first look the topic seems to be trivial, and actually it is a well-know result that a restricted class of GSPNs called state-machines is equivalent to single class QNs. This means that if the customers are statistically identical, state machines and QNs are equivalent even if the former model does not have the notion of queueing discipline. However, the relation between multi-class QNs with GSPNs is not trivial, and some problems arise when we try to represent BCMP QNs with multiple classes of customers with GSPNs. We observe that in this case we cannot ignore the scheduling discipline because, as shown in [5] and further studies [9, 18], it influences the performance measures. For example a First Come First Served (FCFS) station with exponential service time can be embedded in a BCMP QN only if the service time is independent on the customer class, while for an LCFSPR station this condition is not necessary.

To the best of our knowledge the only results concerning the representation of scheduling disciplines with GSPNs are presented in [3]. In [22] the authors introduce a comparison between QN models and SPN models based on the representation of multi-class features by colored Petri nets. However the differences among the various scheduling disciplines are not analyzed. Balbo et al. in [1] combine GSPN and product-form QN by replacing subsystem in a low-level model with their flow equivalents models. Still little attention is devoted to scheduling disciplines. In [3] 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 introduced in [2]. 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 [2] but it does not distinguish different queueing disciplines by

mapping everything into the PS discipline. The work [4] presents structurally finite GSPN models for all the service center types of BCMP theorem by assuming exponential service time distribution, and it is proved that they hold the $M \Rightarrow M$ property. The main idea of this results is a probabilistic model of the queue, i.e., all the customers of the same class wait in the same place. 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, there is also a probabilistic choice of the customer that has to leave the server when a new customer arrives to the system. However, in the work [4] Coxian distributed service times are not considered.

The original result of this paper with respect to [4] is the definition of a GSPN model for the LCFSPR service center with Coxian distributed service time. We define a structurally finite GSPN model with product-form and we prove that it is equivalent to the multi-class M/COX/k/LCFSPR queueing system, in terms of steady state probability and average performance measures. It is worthwhile noting that the defined GSPN model does not belong to the class of product-form GSPN defined in [2].

The GSPN model does not require the introduction of new semantic and it just uses standard exponential and immediate transitions with state-dependent weights. We prove that under an aggregation of states the stationary probabilities of the GSPN are identical to the BCMP marginal probabilities of the LCFSPR node type. The presented result is not a trivial extension of the case of exponential service time, since the GSPN model definition depends on the Coxian service distribution and we apply an appropriate aggregation on the state space. Moreover we show that the for the GSPN model, $M \Rightarrow M$ property is satisfied, in the sense specified in Section 2. Hence we can apply a compositionality property similar to the one known for product-form QNs. Specifically we can compose a set of such GSPN models as well as the ones defined in [4] so obtaining a GSPN whose solution can be expresses by the product of the solutions of the various GSPN components.

This paper, combined with [4], to the best of our knowledge, points out a new relation between BCMP QNs and GSPNs and overcomes the problem of representing the most common queueing disciplines by finite GSPNs. Moreover we define a GSPN model that, under appropriate composition rules, can be combined with similar models [4] holding a product-form stationary probability function. Finally we observe that possible practical applications of these relations concern the integration of QNs and GSPNs into a unique framework for system performance analysis. For example the presented relations could help in transforming an hybrid model that includes both QN and GSPN submodels into a unique GSPN model that can be analyzed with appropriate techniques and tools. The presented results could also allow one to obtain more efficient solution if compositionality and product-form expression introduced above hold.

The paper can be outlined as follows. Section 2 recalls some results on queueing networks and queueing systems with multiple classes of customers. Section 3 introduces the GSPN formalism we use in the paper and sketches the analysis techniques. In Section 4 we define the GSPN-COX model and we prove that it is equivalent to the multi-class LCFSPR BCMP station with multiple

constant rate servers. Moreover we prove that the model holds the $M \Rightarrow M$ property. Section 5 illustrates a simple example of BCMP QN translated into GSPN. Section 6 discusses some final remarks about the work.

2. PRODUCT-FORM QUEUEING NETWORKS WITH CLASSES OF CUSTOMERS

In this section we briefly recall some results we are using in the following concerning product-form queueing networks (PFQNs). Informally a queueing network consists of a finite set of service centers or stations $\Omega = \{1, \ldots, M\}$. Each station is characterized by a scheduling discipline, a set of class of service, and a random service time which can depend on the class. Customers are permanently clustered into chains which can be open, if they have arrivals to and departures from the chain, or closed, if their population is constant. After the service completion at station i with class (i, r) the customer immediately goes to station j and class (j, s) with probability $p_{(i,r),(j,s)}$ or it can leave the system with probability $p_{(i,r),0}$, with $\sum_{d=(j,s) \vee 0} p_{(i,r),d} = 1$ for each i, r, d.

Performance indices of general non-product-form QNs can be hard to obtain, and sometimes only approximated values, or result by simulations, can be obtained in reasonable time. BCMP QNs [5] play a special role in the field of stochastic models because they combine a good modelling flexibility with the availability of mathematical results and algorithms which consent to obtain performace indices efficiently [8, 20, 7, 12]. The main idea underlying BCMP theorem is that the QN stations can be studied in isolation, just like they had Poisson arrivals with a network-dependent parameter even if it is well-known that, in presence of feedback in the QN routing, station arrivals are not Poisson distributed. BCMP QNs consist of four possible station types identified by their scheduling discipline and service time distribution: First Come First Served (FCFS) with exponential class-independent service time, Processor Sharing (PS) and Last Come First Served with Preemptive Resume (LCFSPR) with Coxian class-dependent service time, Infinite Servers (IS) with Coxian class-dependent service time.

Several extensions to BCMP theorem have been introduced. In [18] Muntz proves an interesting property $(M \Rightarrow M)$ of stations which can be embedded in a QN keeping the product-form of the stationary probability function. We briefly review this result that we shall use in the following. A multi-class station exhibits $M \Rightarrow M$ property if under Poisson arrivals for each customer class it presents Poisson departures. A simple way to check if a station holds $M \Rightarrow M$ property is presented in [18]. Let $\mathcal{S}$ be the state space of the station and let $s \in \mathcal{S}$ be a generic state. Let us define $|s|_r$ as the number of class r customers in the station if the state is s and finally let $\mathcal{S}_r^+ = \{s' \in \mathcal{S} : |s'|_r = |s|_r + 1\}$. Then the station has the $M \Rightarrow M$ property if:

$$\forall s \in \mathcal{S}, \forall r \quad \sum_{s' \in \mathcal{S}_r^+} \pi(s')\xi(s' \to s) = \pi(s)\lambda_r, \quad (1)$$

where $\pi(s)$ is the stationary probability of state s, $\xi(s' \to s)$ is the transition rate from s' to s and finally λ_r is the arrival rate for customers of class r. Although $M \Rightarrow M$ is defined for queueing stations, in the following we are using it for characterizing GSPN models. In fact in [17] Melamed extended Muntz's results to a generic stochastic process by the use of opportune traffic processes. We point out here that because of the lower level of abstraction of the GSPN formalism with respect to QNs, the definition of those traffic processes is not immediate in the general case. However in this paper we define particular GSPN models and we know those processes because we know the semantic of what we are defining, i.e., we know which transition firings correspond to customer movements in the net.

The work [4] has introduced GSPN models whose marginal state distributions are the same of the correspondent FCFS, LCFSPR, PS, IS stations just for exponential distributed service time. The introduced models have been proved to hold the $M \Rightarrow M$ property. In order to give a complete representation of BCMP station types, we shall now consider the Coxian service time distribution. This paper focus on LCFSPR station with Coxian distributed service time.

Let us consider a M/COX/1/LCFSPR with one load-dependent server. According to the BCMP [5] model definition, we 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. The service time is modelled by a Coxian random variable with L_r exponential stages with mean $\mu_{r,l}$, where $1 \leq r \leq R$ denotes the customer class and $1 \leq l \leq L_r$ denotes the stage of service. After being served at stage l a customer of class r moves to stage $l + 1$ with probability $a_{r,l}$ and it completes the service with probability $b_{r,l}$, where $a_{r,l} + b_{r,l} = 1$ and $a_{r,L_r} = 0$ Let μ_r be the mean of the Coxian random variable for class 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 $x(1) = y_r(1) = 1$. The station state $\mathbf{q}$ is a finite but unlimited sized vector whose components are couple (r_j, l_j) with r_j representing the class of the j-th vector position customer, l_j the stage of service at which the j-th customer is in. Customers arrive to the station and they are inserted at the end of the vector. Only the customer occupying the last vector position is served and it can change its stage or leave the system. Under stability conditions, the steady state probability of this service center denoted by $\pi''(\mathbf{q})$, is given by [5]:

$$\pi''(\mathbf{q}) = \pi_0' \prod_{r=1}^{R} \lambda_r^{n_r} \prod_{j=1}^{|\mathbf{q}|} \frac{A_{r_j,l_j}}{\mu_{r_j,l_j}} \prod_{b=1}^{n} \frac{1}{x(b)} \prod_{a=1}^{n_r} \frac{1}{y_r(a)}, \quad (2)$$

where $|\mathbf{q}|$ is the dimension of vector $\mathbf{q}$ and $A_{r_j,l_j} = \prod_{i=1}^{l_j-1} a_{r_j,i}$ The marginal probability for an aggregation of states which consider just the number of customers per class $\mathbf{n} = (n_1, \ldots, n_R)$, denoted by $\pi'(\mathbf{n})$, is presented in [5], and can be expressed as follows:

$$\pi'(\mathbf{n}) = \pi_0' \frac{n!}{\prod_{i=1}^{R} n_i!} \prod_{i=1}^{R} \lambda_i^{n_i} \prod_{b=1}^{n} \frac{1}{x(b)} \prod_{r=1}^{R} \left[\left(\frac{1}{\mu_r} \right)^{n_r} \right.$$
$$\left. \prod_{a=1}^{n_r} \frac{1}{y_r(a)} \right]. \quad (3)$$

Note that in order to obtain the steady state probabilities for the multi-class systems with k servers it suffices to set appropriate capacity functions. An LCFSPR center with k load independent servers requires to set $x(n) = \frac{\min(n,k)}{n}$ and $y_r(n_r) = n_r$.

Section 4 presents a GSPN whose aggregated stationary distribution is equivalent to formula (3) and holds $M \Rightarrow M$ property.

3. GENERALIZED STOCHASTIC PETRI NETS

In this section we briefly recall the Generalized Stochastic Petri Nets (GSPNs). We consider the notation for GSPNs introduced in [16]. In order to allow marking dependent probabilities for solving conflicts among immediate transitions we use the techniques discussed in [10]. 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:

- $\mathcal{P} = \{P_1, \dots, P_M\}$ is the set of M places,

- $\mathcal{T} = \{t_1, \dots, t_N\}$ 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: $\mathbf{I}(t_i) = (i_1, \dots, i_M)$ where $i_j = I(t_i, P_j)$, $\mathbf{O}(t_i) = (o_1, \dots, o_M)$ where $o_j = O(t_i, P_j)$ and $\mathbf{H}(t_i) = (h_1, \dots, h_M)$ 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 $i = 1, \dots, M$ 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:

$$p(t_i, \mathbf{m}) = \frac{w(t_i, \mathbf{m})}{\sum_{t_j \in \mathcal{T}_\alpha} w(t_j, \mathbf{m})}. \tag{4}$$

Given a tangible marking $\mathbf{m}$ the transition with the lowest associated stochastic time fires. Sometimes it can be useful to associate a probabilistic output vector to a transition. In this case we denote a possible output bag of transition t_i by $\mathbf{O}_j(t_i)$, and its probability by $d_{i,j}$ where $\sum_j d_{i,j} = 1$. Note that this is not a real extension to the model definition. In fact the probabilistic behavior of a transition firing can be easily obtained by using immediate transitions in a trivial way. Therefore even if we introduce a model with a probabilistic behavior of the transitions firings, it can be used by tools which do not support this feature.

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 $1 \le j \le M$.

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 [16]. 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.

4. M/COX/K/LCFSPR MODEL

In this section we introduce a GSPN which can be considered equivalent to a multi-class M/COX/k queue with LCFSPR scheduling discipline, in terms of steady state probability. We provide a model for this queueing system whose structure is finite and depends only on the number of classes of customers and stages of the Coxian

service time, i.e., not on the number of servers.

DEFINITION 1 (GSPN-COX). *According to the definition given in Section 3:*

- *Places.* $\mathcal{P} = \{P_{r,l}, P_{R+r,l}, P_{r,0} : 1 \le r \le R, 1 \le l \le L_r\} \cup \{P_0, P_{2R+1}\}$

- *Transitions.* $\mathcal{T} = \mathcal{T}_s \cup \mathcal{T}_p \cup \mathcal{T}'$, *where* $\mathcal{T}_s = \{ts_{r,l} : 1 \le r \le R, 0 \le l \le L_r\}$, $\mathcal{T}_p = \{tp_{r,l} : 1 \le r \le R, 1 \le l \le L_r\}$, $\mathcal{T}' = \{T_{r,l} : 1 \le r \le R, 0 \le l \le L_r\}$. *Function* Π *is defined as follows:*

$$\Pi(\tilde{t}) = \begin{cases} 1 & if \ \tilde{t} \in \mathcal{T}_p \cup \mathcal{T}_s \\ 0 & if \ \tilde{t} \in \mathcal{T}' \end{cases}$$

- *Arcs.* *Let* $tp_{r,l} \in \mathcal{T}_p$, *then* $\mathbf{I}(tp_{r,l}) = \mathbf{e_{r,1}} + \mathbf{e_0}$, $\mathbf{H}(tp_{r,l}) = \mathbf{e_{2R+1}}$ *and* $\mathbf{O}(tp_{r,l}) = \mathbf{e_{2R+1}} + \mathbf{e_{R+r,l}}$. *Let* $ts_{r,l} \in \mathcal{T}_s$ *and* $l > 0$, *then* $\mathbf{I}(ts_{r,l}) = \mathbf{e_{R+r,l}} + \mathbf{e_{2R+1}}$, $\mathbf{H}(ts_{r,l}) = \mathbf{e_0}$ *and* $\mathbf{H}(ts_{r,l}) = \mathbf{e_{r,1}}$. *Let* $ts_{r,0} \in \mathcal{T}_s$ *then* $\mathbf{I}(ts_{r,0}) = \mathbf{e_{r,0}} + \mathbf{e_0}$, $\mathbf{H}(ts_{r,0}) = \mathbf{0}$ *and* $\mathbf{O}(ts_{r,0}) = \mathbf{e_{r,1}}$. *Let* $T_{r,l} \in \mathcal{T}'$ *then* $\mathbf{I}(T_{r,l}) = \mathbf{e_{r,1}}$, $\mathbf{H}(T_{r,l}) = 0$. *If* $l < L_r$ *then the output vector is probabilistic:* $\mathbf{O}_1(T_{r,l}) = \mathbf{e_{2R+1}}$ *with probability* $b_{r,l}$ *and* $\mathbf{O}_2(T_{r,l}) = \mathbf{e_{r,l+1}}$ *with probability* $a_{r,l}$ *with* $a_{r,l} + b_{r,l} = 1$. *If* $l = L_r$ *the output vector is deterministic,* $\mathbf{O}(T_{r,L_r}) = \mathbf{e_{2R+1}}$.

- *Weights.* *Let* $1 \le r \le R$ *and* $l > 1$ *then* $w(ts_{r,l}) = m_{R+r,l}$, $w(ts_{r,0}) = 1$, $w(tp_{r,l}) = m_{r,l}$ *and* $w(T_{r,l}) = \mu_{r,l}m_{r,l}$.

- *Initial marking.* $\mathbf{m_0} = (0, \ldots, 0, k)$.

Tokens arrive to places $P_{r,0}$ for $r = 1, \ldots, R$ and P_0.

Note that $m_0 = 1$ if and only if $m_{r,0} = 1$ for a r, and that if $m_0 > 0$ then $\mathbf{m}$ is a vanishing marking, with $1 \le r \le R$. Figure 1 illustrates a graphical model for $R = 2$, $L_1 = 3$, $L_2 = 2$. Exponential transitions $T_{0,1}$ and $T_{0,2}$ are introduced to show the arrival behaviors. The dotted lines and the solid lines represents alternative firing modes while grey lines are used for sake of clarity. We can give the following interpretation of the model: place P_3 has as many tokens as the free servers of the queuing system, place $P_{r,l}$, $1 \le r \le 2$, has as many tokens as the number of customers being served at stage l of class r. The tokens in place $P_{R+r,l}$, $1 \le r \le 2$ represent the number of customers in the queue of class r and stage l. When a customer of class r arrives to the system a token is temporally (i.e., it origins a vanishing state) stored in places $P_{r,0}$ and P_0. Immediate transition $t_{s,0}$ put the arrived token in service at stage 1, the customer which must be preempted is chosen probabilistically by the transitions tps. When a customer leaves the system by the firing of a timed transition, a preempted customer from places $P_{R+r,l}$ is chosen probabilistically by a transition tss and it resumes the service.

In the following we first introduce a lemma which gives the stationary probabilities for model GSPN-COX under Poisson arrivals and we show that the $M \Rightarrow M$ property holds. Then we aggregate the states giving the marginal probabilities in closed form. The final corollary states how the model GSPN-COX can be considered equivalent to a M/COX/k/LCFSPR station.

LEMMA 1. *Let $\mathbf{m}$ be a tangible state of model GSPN-COX. Then the steady state probability of $\mathbf{m}$ is given by:*

$$\pi(\mathbf{m}) = \pi_0 \Big[\prod_{r=1}^R \lambda_r^{m_r + m_{R+r}} \Big] \tag{5}$$

$$\cdot \Big[\frac{(\sum_{r=1}^R m_r)!}{\prod_{r=1}^R \prod_{l=1}^{L_r} m_{r,l}!} \frac{(\sum_{r=1}^R m_{R+r})!}{\prod_{r=1}^R \prod_{l=1}^{L_r} m_{R+r,l}!} \Big]$$

$$\cdot \Big[\prod_{r=1}^R \prod_{l=1}^{L_R} \Big(\frac{A_{r,l}}{\mu_{r,l}} \Big)^{m_{R+r,l} + m_{r,l}} \Big] \Big[\prod_{a=1}^m \frac{1}{\min\{n, k\}} \Big],$$

where $m_r = \sum_{l=1}^{L_r} m_{r,l}$, $m_{R+r} = \sum_{l=1}^{L_r} m_{R+r,l}$, $m = \sum_{r=1}^R m_r + m_{R+r}$ *and* π_0 *is the normalizing constant. Moreover model GSPN-COX holds $M \Rightarrow M$ property.*

PROOF. The proof is based on verifying that formula (5) satisfies the set of global balance equations (GBEs) of the Markov process associated to the model. We work by cases: first we consider (a) the case of $m_{R+1} = 0$ and $m_{R+r,l} > 0$ for some $r = 1, \ldots, R$ and $1 \le l \le L_r$, then (b)the case $m_{R+1} = 0$ and $m_{R+r,l} = 0$ for all $r = 1, \ldots, R$ and $1 \le l \le L_r$, then (c) the case $1 \le m_{R+1} < k$, and finally (d) the case $m_{R+1} = k$. The $M \Rightarrow M$ property, that is equation (1) is immediately verified by the chosen partial balance.

Case (a). Consider tangible state $\mathbf{m}$ where $m_{R+1} = 0$ and $m_{R+r,l} > 0$ for some $r = 1, \ldots, R$. State $\mathbf{m}$ can be reached from the following set of states:

- $\mathcal{A} = \{\mathbf{m}' : \mathbf{m}' = \mathbf{m} + \mathbf{e_{r,l-1}} - \mathbf{e_{r,l}}, m_{r,l} > 0, 1 \le r \le R, 1 < l \le L_r\}$ with rate $\mu_{r,l-1}a_{r,l-1}(m_{r,l-1} + 1)$.

- $\mathcal{B} = \{\mathbf{m}' : \mathbf{m}' = \mathbf{m} - \mathbf{e_{s,1}} + \mathbf{e_{r,l}} - \mathbf{e_{R+r,l}}, 1 \le r, s \le R, m_{s,1} > 0, m_{R+r,l} > 0\}$ with rate $\lambda_s(m_{r,l} + 1)/(\sum_{r'=1}^R \sum_{l'=1}^{L_{r'}} m_{r',l'})$.

- $\mathcal{C} = \{\mathbf{m}' = \mathbf{m} + \mathbf{e_{s,\ell}} + \mathbf{e_{R+r,l}} - \mathbf{e_{r,l}}, 1 \le r, s \le R, 1 \le l \le L_r, m_{r,l} > 0$ with rate $\mu_{s,m}b_{s,m}(m_{s,m} + 1)(m_{R+r,l} + 1)/(1 + \sum_{r'=1}^R \sum_{l'=1}^{L_{r'}} m_{R+r',l'})$.

The leaving rate from state $\mathbf{m}$ is:

$$\sum_{r=1}^R \lambda_r + \sum_{r=1}^R \sum_{l=1}^{L_r} m_{r,l}\mu_{r,l}, \tag{6}$$

so we have to prove that:

$$\sum_{\mathbf{m}' \in \mathcal{A} \cup \mathcal{B} \cup \mathcal{C}} \pi(\mathbf{m}')\xi(\mathbf{m}' \to \mathbf{m})$$

$$= \pi(\mathbf{m}) \Big[\sum_{r=1}^R \lambda_r + \sum_{r=1}^R \sum_{l=1}^{L_r} m_{r,l}\mu_{r,l} \Big], \tag{7}$$

where $\xi(\mathbf{m}' \to \mathbf{m})$ denotes the transition rate from state $\mathbf{m}'$ to state $\mathbf{m}$. In the following we write $\pi(\mathbf{m}')$ as product of $\pi(\mathbf{m})$ and an opportune factor.

We verify the GBEs considering the effective arrival rates from states belonging to different sets $\mathcal{A}, \mathcal{B}, \mathcal{C}$ separately. Let $\mathbf{m}' \in \mathcal{A}$, thus $m_{r,l} > 0$ and $l > 1$, then we can write the effective arrival rate to state $\mathbf{m}$ as follows:

$$\pi(\mathbf{m}) \Big[\frac{m_{r,l}}{1 + m_{r,l-1}} \frac{\mu_{r,l}}{A_{r,l}} \frac{A_{r,l-1}}{\mu_{r,l-1}} \Big] \mu_{r,l-1}a_{r,l-1}(1 + m_{r,l-1})$$

$$= \pi(\mathbf{m})[m_{r,l}\mu_{r,l}]. \tag{8}$$

Let $\mathbf{m}' \in \mathcal{B}$ thus $m_{s,1} > 0$, and let us define $Y = \{(r,l)|1 \le r \le R, 1 \le l \le L_r, m_{R+e,l} > 0\}$. Then we
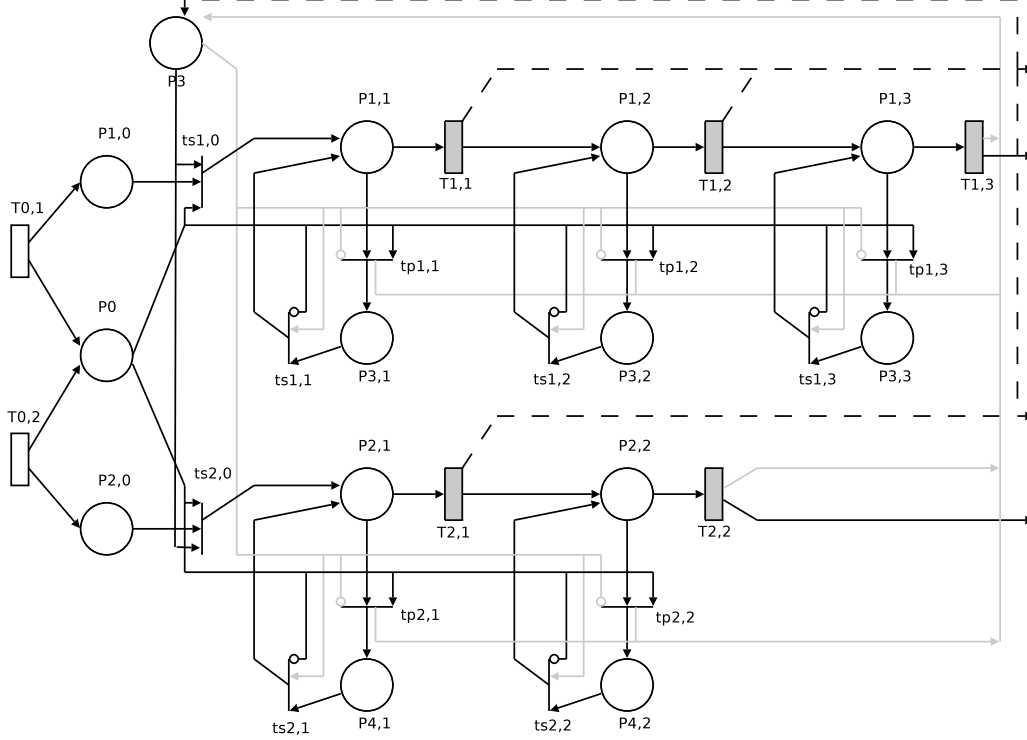
Figure 1: GSPN-Cox for $R = 2$ classes, $L_1 = 3$ and $L_2 = 2$ stages.

can write the effective arrival rate to state $\mathbf{m}$ as follows:

$$\pi(\mathbf{m})\Big[\sum_{(r,l)\in Y}\frac{1}{\lambda_s}\frac{m_{s,1}}{1+m_{r,l}}\frac{m_{R+r,l}}{\sum_{r'=1}^{R}\sum_{l'=L_{r'}}m_{R+r',l'}}\Big]$$

$$\cdot\Big[\frac{\mu_{s,1}}{A_{s,1}}k\lambda_s\frac{1+m_{r,l}}{\sum_{r'=1}^{R}\sum_{l'=1}^{L_{r'}}m_{r',l'}}\Big]$$

$$= \pi(\mathbf{m})\Big[m_{s,1}\mu_{s,1}\frac{1}{\sum_{r'=1}^{R}\sum_{l'=1}^{L_{r'}}m_{r',l'}}\sum_{(r,l)\in Y}m_{R+r}\Big]$$

$$= \pi(\mathbf{m})[m_{s,1}\mu_{s,1}]. \tag{9}$$

Consider now $\mathbf{m}' \in \mathcal{C}$. The effective arrival rate to state $\mathbf{m}$ can be written as follows:

$$\pi(\mathbf{m})\Big[\sum_{\ell=1}^{L_s}\sum_{(r,l)\in Y}\lambda_s\frac{m_{r,l}}{m_{s,\ell}+1}\frac{1+\sum_{r'=1}^{R}\sum_{l'=1}^{L_{r'}}m_{R+r',l'}}{1+m_{R+r,l}}$$

$$\frac{A_{s,\ell}}{\mu_{s,\ell}}\frac{1}{k}\mu_{s,\ell}b_{s,\ell}(m_{s,\ell}+1)\frac{1+m_{R+r,l}}{1+\sum_{r'=1}^{R}\sum_{l'=1}^{L_{r'}}m_{R+r',l'}}\Big]$$

$$= \pi(\mathbf{m})[\lambda_s\sum_{\ell=1}^{L_s}A_{s,\ell}b_{s,\ell}\sum_{(r,l)\in Y}\frac{m_{r,l}}{k}]$$

$$= \pi(\mathbf{m})\lambda_s. \tag{10}$$

Note that summing over all the possible $\mathbf{m}' \in \mathcal{A}\cup\mathcal{B}\cup\mathcal{C}$ equations (8), (9), (10) we obtain the total arrival rate to state $\mathbf{m}$ which equates the effective leaving rate from state $\mathbf{m}$ given by expression (6).

Case (b). Consider the tangible state $\mathbf{m}$ where $m_{R+1} = 0$ and $m_{R+r,l} = 0$ for all $r = 1 \ldots, R$ and $1 \le l \le L_r$. The only difference with respect to case (a) is that set $\mathcal{B}$ has to be redefined as: $\mathcal{B} = \{\mathbf{m}'|\mathbf{m}' = \mathbf{m} - \mathbf{e_{s,1}} + \mathbf{e_{R+1}}, 1 \le r, s \le R, m_{s,1} > 0\}$. Hence the effective arrival rate to

state $\mathbf{m}$ from a state in $\mathcal{B}$ can be written as follows:

$$\pi(\mathbf{m})[\frac{1}{\lambda_s}\frac{m_{s,1}}{\sum_{r'=1}^{R}\sum_{l'=1}^{L_{r'}}m_{r',l'}}\mu_{s,1}k\lambda_s] = \pi(\mathbf{m})[\mu_{s,1}m_{s,1}]. \tag{11}$$

Noting that the right hand side of equation (11) is equal to the right hand side of equation (9) and summing over all the possible states $\mathbf{m}'$ we verify the GBEs.

Case (c). Consider the tangible state $\mathbf{m}$ where $1 \le \sum_{r=1}^{R}\sum_{l=1}^{L_r}m_{r,l} < k$ thus we have $\sum_{r=1}^{R}\sum_{l=1}^{L_r}m_{R+r,l} = 0$. We partition the set of states from which state $\mathbf{m}$ can be reached as follows:

- $\mathcal{A}$ is defined as done in case (a).
- $\mathcal{B} = \{\mathbf{m}'|\mathbf{m}' = \mathbf{m} - \mathbf{e_{s,1}}, 1 \le r, s \le R, m_{s,1} > 0\}$ with rate λ_s.
- $\mathcal{C} = \{\mathbf{m}'|\mathbf{m}' = \mathbf{m} + \mathbf{e_{s,\ell}} - \mathbf{e_{R+1}}, 1 \le s \le R, 1 \le \ell \le L_s\}$ with rate $\mu_{s,\ell}b_{s,\ell}(m_{s,\ell}+1)$.

The calculations for states in $\mathcal{A}\cup\mathcal{B}$ are the same as the previous case. Consider $\mathbf{m}' \in \mathcal{C}$, the effective arrival rate to $\mathbf{m}$ can be written as follows:

$$\pi(\mathbf{m})\Big[\sum_{\ell=1}^{L_s}\lambda_s\frac{1+\sum_{r'=1}^{R}\sum_{l'=1}^{L_{r'}}m_{r',l'}}{1+m_{s,\ell}}\frac{A_{s,\ell}}{\mu_{s,\ell}}\mu_{s,\ell}b_{s,\ell}$$

$$\cdot(m_{s,\ell}+1)\frac{1}{1+\sum_{r'=1}^{R}\sum_{l'=1}^{L_{r'}}m_{r',l'}}\Big] = \pi(\mathbf{m})\lambda_s. \tag{12}$$

Noting that the right hand side of equation (12) is equal to the right hand side of equation (10) and summing over all $\mathbf{m}' \in \mathcal{A}\cup\mathcal{B}\cup\mathcal{C}$ we verify the the GBEs.

Case (d). The proof is trivial. $\square$

The following lemma states a closed form expression for the state probability of the GSPN-COX by considering

an exact aggregation of states. The basic idea is to obtain an aggregated state which does not represent the stage at which a token is, but it just distinguishes between a token being in service and being in queue. The actual relevance of the lemma is to show a result which is equivalent to the well-known result of BCMP theorem, that is, the marginal distribution for a Coxian service time LCF-SPR station is equal to the one of the same station with an exponential service time distribution with the same mean.

LEMMA 2. *Let us define the aggregate tangible state as follows:* $\mathbf{n} = (n_1, \ldots, n_R, n_{R+1}, \ldots, n_{2R}, n_{2R+1})$, *where* $n_r = m_r = \sum_{l=1}^{L_r} m_{r,l}$ *and* $n_{R+r} = m_{R+r} = \sum_{l=1}^{L_r} m_{R+r,l}$ *and* $n_{2R+1} = m_{R+1}$, *with* $1 \le r \le R$ *and* $\mathbf{m}$ *is a tangible state of model GSPN-COX. The steady state probabilities for GSPN-COX can be written as follows:*

$$\pi_a(\mathbf{n}) = \pi_0 \prod_{r=1}^{R} \lambda_r^{n_r + n_{R+r}} \frac{(\sum_{r=1}^{R} n_r)!}{\prod_{r=1}^{R} n_r!} \frac{(\sum_{r=1}^{R} n_{R+r})!}{\prod_{r=1}^{R} n_{R+r}!}$$
$$\cdot \prod_{r=1}^{R} \left(\frac{1}{\mu_r}\right)^{n_r + n_{R+r}} \prod_{a=1}^{n} \frac{1}{\min(n, k)}, \quad (13)$$

where $1/\mu_r$ *is the mean service time, i.e.,* $\sum_{l=1}^{L_r} A_{r,l}/\mu_{r,l}$, $n = \sum_{r=1}^{R} n_r = m$.

PROOF. In order to prove Lemma 2 we calculate the aggregation as follows. Given $\mathbf{n}$, the aggregation is obtained in $2R$ steps. We first sum over $\mathbf{m}$ such that $m_1 = n_1$, that is, we aggregate class 1 customers in service, obtaining an intermediate state $\mathbf{m}^{\alpha_1}$, where $\pi^{\alpha_1}(\mathbf{m}^{\alpha_1})$ is given by the sum:

$$\pi^{\alpha_1}(\mathbf{m}^{\alpha_1}) = \sum_{\mathbf{m}:m_1=n_1} \pi(\mathbf{m}). \quad (14)$$

The following step aggregates class 1 customers in the queue, obtaining the intermediate state $\mathbf{m}^{\beta_1}$, whose stationary probability is given by:

$$\pi^{\beta_1}(\mathbf{m}^{\beta_1}) = \sum_{\mathbf{m}^{\alpha_1}:m_{R+1}^{\alpha_1}=n_{R+1}} \pi^{\alpha_1}(\mathbf{m}^{\alpha_1}). \quad (15)$$

Proceeding with the aggregation for all classes $1 \ldots R$ we have that: $\mathbf{m}^{\beta_R} = \mathbf{n}$ and $\pi^{\beta_R}(\mathbf{m}^{\beta_R}) = \pi_a(\mathbf{n})$.

In order to simplify the notation, in the following we write $\sum_{\alpha_r}$ and $\sum_{\beta_r}$ to denote the sums which give intermediate states $\mathbf{m}^{\alpha_r}$ and $\mathbf{m}^{\beta_r}$ respectively.

Hence we can simplify what we have to prove as follows:

$$\sum_{\beta_R} \sum_{\alpha_R} \cdots \sum_{\beta_1} \sum_{\alpha_1} \left[\frac{(\sum_{r=1}^{R} m_r)!}{\prod_{r=1}^{R} \prod_{\ell=1}^{L_r} m_{r,\ell}!} \right.$$
$$\left. \cdot \frac{(\sum_{r=1}^{R} m_{R+r})!}{\prod_{r=1}^{R} \prod_{\ell=1}^{L_r} m_{R+r,\ell}!} \right] \cdot \left[\prod_{r=1}^{R} \prod_{\ell=1}^{L_R} \left(\frac{A_{r,\ell}}{\mu_{r,\ell}}\right)^{m_{R+r,\ell}+m_{r,\ell}} \right]$$
$$= \frac{(\sum_{r=1}^{R} n_r)!}{\prod_{r=1}^{R} n_r!} \frac{(\sum_{r=1}^{R} n_{R+r})!}{\prod_{r=1}^{R} n_{R+r}!} \prod_{r=1}^{R} \left(\frac{1}{\mu_r}\right)^{n_r + n_{R+r}}. \quad (16)$$

Let us consider the inner summatory of the left hand side

of equation (16):

$$\sum_{\alpha_1} \left[\frac{(\sum_{r=1}^{R} m_r)!}{\prod_{r=1}^{R} \prod_{\ell=1}^{L_r} m_{r,\ell}} \frac{(\sum_{r=1}^{R} m_{R+r})!}{\prod_{r=1}^{R} \prod_{\ell=1}^{L_R} m_{R+r,\ell}} \right.$$
$$\left. \cdot \left[\prod_{r=1}^{R} \prod_{\ell=1}^{L_r} \left(\frac{A_{r,\ell}}{\mu_{r,\ell}}\right)^{m_{R+r,\ell}+m_{r,\ell}} \right] \right]$$
$$= \frac{(\sum_{r=1}^{R} m_r)!}{\prod_{r=2}^{R} \prod_{\ell=1}^{L_r} m_{r,\ell}} \frac{(\sum_{r=1}^{R} m_{R+r})!}{\prod_{r=1}^{R} \prod_{\ell=1}^{L_R} m_{R+r,\ell}}$$
$$\cdot \prod_{\ell=1}^{L_1} \left(\frac{A_{1,l}}{\mu_{1,\ell}}\right)^{m_{R+1,\ell}} \prod_{r=2}^{R} \prod_{\ell=1}^{L_r} \left(\frac{A_{r,\ell}}{\mu_{r,\ell}}\right)^{m_{R+r,\ell}+m_{r,\ell}}$$
$$\cdot \sum_{\alpha_1} \left[\frac{1}{\prod_{\ell=1}^{L_1} m_{1,\ell}} \prod_{\ell=1}^{L_1} \left(\frac{A_{1,\ell}}{\mu_{1,\ell}}\right)^{m_{1,\ell}} \right]$$
$$= \frac{(\sum_{r=1}^{R} m_r)!}{\prod_{r=2}^{R} \prod_{\ell=1}^{L_r} m_{r,\ell}} \frac{(\sum_{r=1}^{R} m_{R+r})!}{\prod_{r=1}^{R} \prod_{\ell=1}^{L_R} m_{R+r,\ell}}$$
$$\cdot \prod_{\ell=1}^{L_1} \left(\frac{A_{1,l}}{\mu_{1,\ell}}\right)^{m_{R+1,\ell}} \prod_{r=2}^{R} \prod_{\ell=1}^{L_r} \left(\frac{A_{r,\ell}}{\mu_{r,\ell}}\right)^{m_{R+r,\ell}+m_{r,\ell}} \frac{1}{m_1!}$$
$$\cdot \sum_{\alpha_1} \left[\binom{m_1}{m_{1,1} \ldots m_{1,L_1}} \prod_{\ell=1}^{L_1} \left(\frac{A_{1,\ell}}{\mu_{1,\ell}}\right)^{m_{1,\ell}} \right]$$

that by applying the binomial theorem, and noting that

$$\sum_{\ell=1}^{L_1} \left(\frac{A_{1,\ell}}{\mu_{1,\ell}}\right)$$

is the mean of the Coxian distributed service time, i.e., $1/\mu_1$, can be rewritten as follows:

$$= \left[\frac{(\sum_{r=1}^{R} m_r)!}{m_1! \prod_{r=2}^{R} \prod_{\ell=1}^{L_r} m_{r,\ell}} \frac{(\sum_{r=1}^{R} m_{R+r})!}{\prod_{r=1}^{R} \prod_{\ell=1}^{L_R} m_{R+r,\ell}} \right]$$
$$\cdot \left[\prod_{r=2}^{R} \prod_{\ell=1}^{L_r} \left(\frac{A_{r,\ell}}{\mu_{r,\ell}}\right)^{m_{R+r,\ell}+m_{r,\ell}} \prod_{\ell=1}^{L_1} \left(\frac{A_{1,\ell}}{\mu_{1,\ell}}\right)^{m_{R+1,\ell}} \right.$$
$$\left. \cdot \left(\frac{1}{\mu_1}\right)^{m_1} \right] \quad (17)$$

Summing expression (17) to obtain $\mathbf{m}^{\beta_1}$, by similar calculations, gives:

$$\left[\frac{(\sum_{r=1}^{R} m_r)!}{m_1! \prod_{r=2}^{R} \prod_{\ell=1}^{L_r} m_{r,\ell}} \frac{(\sum_{r=1}^{R} m_{R+r})!}{m_{R+1}! \prod_{r=2}^{R} \prod_{\ell=1}^{L_R} m_{R+r,\ell}} \right]$$
$$\cdot \left[\prod_{r=2}^{R} \prod_{\ell=1}^{L_r} \left(\frac{A_{r,\ell}}{\mu_{r,\ell}}\right)^{m_{R+r,\ell}+m_{r,\ell}} \left(\frac{1}{\mu_1}\right)^{m_1+m_{R+1}} \right] \quad (18)$$

Expression (18) can be similarly summed over α_2 and then β_2, and so on. Noting that $m_r = n_r$ and $m_{R+r} = n_{R+r}$ for $r = 1, \ldots, R$ we have proved equation (16) and the lemma. □

The following theorem states the equivalence of the BCMP LCFSPR marginal probabilities and the GSPN-COX ones.

THEOREM 1. *Consider model GSPN-COX, let* $\mathbf{m}$ *be a tangible state. Define the aggregated state* $\mathbf{u}$ *as follows:* $u_r = \sum_{l=1}^{L_r} m_{r,l} + \sum_{l=1}^{L_r} m_{R+r,l}$ *for* $r = 1, \ldots, R$, *i.e.,* u_r *represents the number of customers of class* r *in the system, either in service or in queue. Then we can write:*

$$\pi_A(\mathbf{u}) = \pi_0 \frac{u!}{\prod_{r=1}^{R} u_r!} \prod_{r=1}^{R} \lambda_r^{u_r} \left(\frac{1}{\mu_r}\right)^{u_r} \prod_{a=1}^{u} \frac{1}{\min\{k, a\}},$$
$$\quad (19)$$

where $u = \sum_{r=1}^{R} u_r$.

PROOF. We have to prove that:

$$\pi_A(\mathbf{u}) = \sum_{\mathbf{m}: \sum_{l=1}^{L_r} (m_{r,l} + m_{R+r,l}) = u_r \atop r=1,\ldots,R} \pi(\mathbf{m}). \qquad (20)$$

By Lemma 2 we obtain a closed form for the aggregation of the stages. However, from the queueing model point of view, the aggregated state still distinguish the number of customers in service and the ones in queue. Moreover we can see the marginal probability distribution as the probability distribution of GSPN-COX model with $L_r = 1$ for every $r = 1, \ldots, R$, and $1/\mu_r = \sum_{l=1}^{L_r} (A_{r,l}/\mu_{r,l})$. In [4] it is proved that Theorem 1 holds for models with $L_r = 1$, i.e., exponential servers. $\square$

The following corollary is very simple an it immediately follows by Theorem 1 and BCMP theorem:

COROLLARY 1. *The M/COX/k/LCFSPR queueing system with R customer classes, arrival rates λ_r, Coxian distributed service times with mean μ_r, $1 \leq r \leq R$ and steady state probability $\pi'(\mathbf{n})$, is equivalent to model GSPN-COX in terms of steady state probability, i.e., $\pi_A(\mathbf{n}) = \pi'(\mathbf{n})$ for all $\mathbf{n} \in \mathbb{N}^R$ under condition that $\sum_{l=1}^{L_r} A_{r,l}/\mu_{r,l} = \mu_r$ for all $r = 1, \ldots, R$.*

5. EXAMPLE

In this section we present an example of representation of a simple multi-class QN by a GSPN. Consider the QN
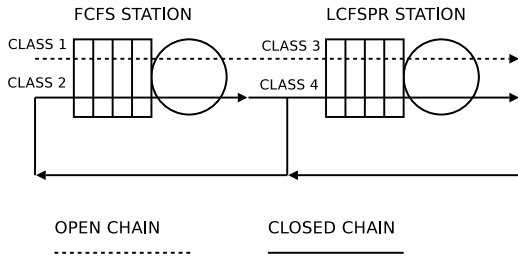

Figure 2: Example of a QN with two stations.

shown in Figure 2 that consists of two stations with different scheduling disciplines, FCFS and LCFSPR, respectively. We assume that the FCFS station has 2 identical servers and a service time exponentially distributed and independent on the class of the customer being served. The LCFSPR station consists of 3 identical servers and Coxian distributed service times that depend on the class of the customer being served. For customers of class 1, the service time is modelled by a 3-stage Coxian random variable, while for customers of class 2, the service time is modelled by a 2-stage Coxian random variable. The routing is illustrated in Figure 2 and it is probabilistic: after the service completion, a customer of class 2 can be served either by class 4 with probability $p_{2,4}$ or by class 2 with probability $p_{2,2}$, with $p_{2,4} + p_{2,2} = 1$. Note that we are using class concept in local sense [19]. The QN is mixed because there is an open chain and a closed one. The QN holds the BCMP theorem for product-form.

Figure 3 shows the equivalent GSPN model of the entire QN, where the second service center with LCFSPR and Coxian service distribution is defined according to Definition 1 of Section 4, and the first service center is defined as in [4].

Dotted lines in Figure 3 represent an alternative firing mode of a transition, while grey lines are used just to improve the readability of the figure. For example transition transition T_b has two possible output vectors according to the probabilistic routing of the closed chain customers after being served at FCFS station. Hence $O_1(T_b, P^*) = 1$, $O_1(T_b, P_{bq}) = 1$ (corresponding to the dotted line) and $O_2(T_b, P^*) = 1$, $O_2(T_b, P_0) = 1$, $O_2(T_b, P_{2,0}) = 1$. The probability associated to the output vectors are $p_{2,2}$ for firing mode 1 and $p_{2,4}$ for firing mode 2 (corresponding to the solid line). The mean of the timed transitions of the GSPN-COX block associated to the LCFSPR station are set according to the Coxian service time of the QN station. Note that if the example QN has product-form, the BCMP conditions for the QN require that the average service rate of FCFS station does not depend on the customer class. This condition corresponds the condition $\mu_a = \mu_b$ in the GSPN, where μ_a and μ_b are the firing rates of exponential transitions T_a and T_b.

Note that the QN has feedback, thus the internal flows are not Poisson distributed. Even if we have proved that the equivalence between the GSPN-COX model and the multi-class LCFSPR BCMP station holds (for marginal distributions) under Poisson arrivals, $M \Rightarrow M$ property assures that the equivalence can be straightforward extended to networks of blocks each of which holds the $M \Rightarrow M$ property. In fact Lemma 1 proves that model GSPN-COX satisfies $M \Rightarrow M$ property, and in [4] it is proved that the same holds for the GSPN block we are using for FCFS station. This ensures the equivalence between the QN of Figure 2 and the GSPN of Figure 3 in terms of marginal stationary probabilities.

As consequence the whole GSPN equivalent to the QN holds the $M \Rightarrow M$ property, hence the departure process from the GSPN (and the QN) is a Poisson process [18].

6. FINAL REMARKS

In this paper we have shown how to represent multi-class single queueing systems by structurally finite GSPN for LCFSPR queueing disciplines with Coxian distributed service time. We pointed out this investigation led to the definition of a model which can be combined with similar GSPN models [4] giving a product-form solution which does not belong to any one the well-known classes [13, 11, 2, 6].

We observe that simulating the behavior of the queueing discipline such that the GSPN Markov chain is isomorphic to the one associated to the correspondent QN leads to an infinite structure for open networks, or a structure depending on the number of customers in the system for closed ones. In order to overcome this problem and define a finite GSPN we have simulated the queueing discipline with a probabilistic behavior, so defining the GSPN-COX model whose Markov chain is *not* isomorphic to the associated queueing station Markov chain. However we have proved the equivalence between the two models for meaningful aggregations. In fact for model GSPN-COX we give the stationary distribution for three level of details, each showing a closed form expression:

- $\pi(\mathbf{m})$ represents the stationary distribution for detailed state, that specifies how many customers for each class are being served and are in queue for every stage of the service.

- $\pi_a(\mathbf{n})$ allows one to determine the stationary probability of a state considering how many customers
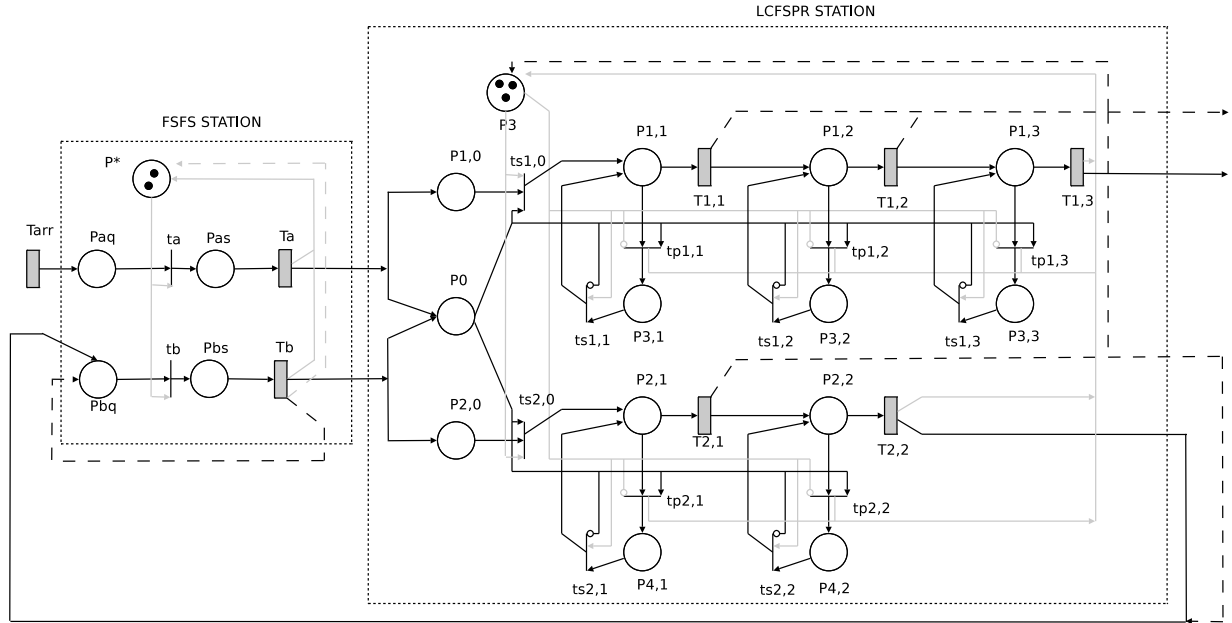
Figure 3: GSPN equivalent to the QN of Figure 2.

are being served and in the queue for every class.

- $\pi_A(\mathbf{u})$ represents the stationary distribution for the less detailed state, that gives the number of customers for every class in the net, regardless to their position, in queue or in service.

We have proved, in Section 4, that the GSPN-COX model holds the $M \Rightarrow M$ property that allows us to state that the model can be used as a building block in representing a BCMP QN with GSPN, maintaining the same product-form for the marginal state distribution. Hence each GSPN block presented in [4] and in this paper, that represent various types of QN service centers, can be composed by using the $M \Rightarrow M$ property. In order to compose the QN centers one can define an appropriate GSPN structure that represents the QN routing probabilities, as simply illustrated by an example in the previous section.

It is worthwhile noting that the product-form is a property defined on the Markov chain associated to a stochastic model. Hence further research could try to identify structurally finite GSPN models whose stochastic processes are isomorphic to the correspondent queueing center ones, by considering a significant level of detail. These kinds of model could have a stronger equivalence with BCMP QNs and not just for average performance measures.

7. REFERENCES

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

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

[3] Balbo, G., Bruell, S. C., and Sereno, M. 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.* (2003), 103–112.

[4] Balsamo, S., and Marin, A. On representing multiclass M/M/k queues by generalized stochastic Petri nets. In *Proc. of ECMS/ASMTA-2007 conference. To appear.*

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

[6] Boucherie, R. J. A characterisation of independence for competing Markov chains with applications to stochastic Petri nets. Tech. Rep. RR-1880.

[7] Bruell, S. C., G.Balbo, and Afshari, P. V. Mean Value Analysis of mixed, multiple class BCMP networks with load dependent service stations. *Performance Evaluation 4* (1984), 241–260.

[8] Buzen, J. P. Computational algorithms for closed queueing networks with exponential servers. *Commun. ACM 16*, 9 (1973), 527–531.

[9] Chandy, K. M., John H. Howard, J., and Towsley, D. F. Product form and local balance in queueing networks. *J. ACM 24*, 2 (1977), 250–263.

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

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

[12] Conway, A. E., and Georganas, N. D. Recal - a new efficient algorithm for the exact analysis of

multiple-chain closed queuing networks. *J. ACM 33*, 4 (1986), 768–791.

[13] HENDERSON, W., LUCIC, D., AND TAYLOR, P. G. A net level performance analysis of Stochastic Petri Nets. *J. Austral. Math. Soc. Ser. B 31* (1989), 176–187.

[14] KANT, K. *Introduction to Computer System Performance Evaluation*. McGraw-Hill, 1992.

[15] LAZOWSKA, E. D., ZAHORJAN, J. L., GRAHAM, G. S., AND SEVCICK, K. C. *Quantitative system performance: computer system analysis using queueing network models*. Prentice Hall, Englewood Cliffs, NJ, 1984.

[16] MARSAN, M. A., BALBO, G., CONTE, G., DONATELLI, S., AND FRANCESCHINIS, G. *Modelling with generalized stochastic Petri nets*. Wiley, 1995.

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

[18] MUNTZ, R. R. Poisson departure processes and queueing networks. Tech. Rep. IBM Research Report RC4145, Yorktown Heights, New York, 1972.

[19] REISER, M., AND SAUER, C. H. Queueing network models: Methods of solution and their program implementation. In *Current Trends in Programming Methodology*, (K. M. Chandy and R. T. Yeh), Ed. Prentice-Hall Inc., 1978.

[20] RESISER, M., AND LAVENBERG, S. S. Mean Value Analysis of closed multichain queueing network. *J. ACM 27*, 2 (1980), 313–320.

[21] SERENO, M., AND BALBO, G. Mean Value Analysis of stochastic Petri nets. *Performance Evaluation*, 29 (1997), 35–62.

[22] VERNON, M., ZAHORJAN, J., AND LAZOWSKA, E. D. A comparsion of performance Petri Nets and queueing network models. *Proc. 3rd Intern. Workshop on Modelling Techniques and Performance Evaluation* (1987), 181–192.

[23] WOODSIDE, C., NEILSON, J., PETRIU, S., AND MJUMDAR, S. The stochastic rendezvous network model for performance of synchronous client-server-like distributed software. *IEEE Transaction on Computer 44* (1995), 20–34.