On the Characterization of Product-Form Multiclass Queueing Models with Probabilistic Disciplines

Simonetta Balsamo and Andrea Marin

Università Ca' Foscari di Venezia
Dipartimento di Informatica
via Torino 155, Venezia

Abstract. Probabilistic queueing disciplines are used for modeling several system behaviors. In particular, under a set of assumptions, it has been proved that if the choice of the customer to serve after a job completion is uniform among the queue population, then the model has a BCMP-like product-form solution. In this paper we address the problem of characterizing the probabilistic queueing disciplines that can be embedded in a BCMP queueing network maintaining the product-form property. We base our result on Muntz's property $M \Rightarrow M$ and prove that the RANDOM is the only non-preemptive, non-priority, probabilistic discipline that fulfils the $M \Rightarrow M$ property with a class independent exponential server. Then we observe that the FCFS and RANDOM discipline share the same product-form conditions and a set of relevant performance indices when embedded in a BCMP queueing network. We use a simulator to explore the similarities of these disciplines in non-product-form contexts, i.e., under various non-Poisson arrival processes.

1 Introduction

Queueing models have a pivotal importance for performance evaluation purposes. They have been widely used to model various types of systems, ranging from computer hardware and software architectures, to telephony systems and communication networks. Informally, a queueing model has a set of resources that serves a finite or infinite set of customers. Customers arrive to the model according to a stochastic process (the arrival process) and then they possibly wait for the service in the queue. The service requires a time that is usually modeled by a random variable with a known distribution. After being served, the customers leave the queueing model. The analysis of queueing models, which is usually based on the definition and solution of the underlying stochastic process, provides a set of performance indices including the steady state distribution of the number of customers in the system and some average performance measures, such as the throughput and the mean response time.

In this paper we focus on multiclass queueing models with probabilistic scheduling disciplines. In the simplest queueing models, we usually assume that

K. Al-Begain, D. Fiems, and G. Horváth (Eds.): ASMTA 2009, LNCS 5513, pp. 204–218, 2009.

all the customers are identical, i.e., they all arrive according to the same arrival process and are served according to the same service time distribution. However, for many practical purposes these limitations are unrealistic. In multiclass queueing models, the customers are clustered into classes, and each class is characterized by an arrival process and a service time distribution. The scheduling discipline may also depend on the customer class (e.g., scheduling with priority). According to the literature, in the following we use *RANDOM* to denote the probabilistic queueing discipline in which every customer in the queue has the same probability to enter in service immediately after a job completion, regardless to its arrival time or class. From the performance modeling viewpoint, note that using general probabilistic disciplines allows us to model systems in which the class of a customer may influence the probability of entering in service, thus modeling a sort of mild priority mechanism. For instance this technique is applied to the analysis of the performance of the *Differentiated Services* architecture (RFC 2475) for Internet, as describe in [10]. Among probabilistic disciplines the RANDOM queueing stations have been used to model several resource contention systems, such as shared bus contention systems, as described in [1]. Multiclass queueing models can be embedded with some restrictions in product-form queueing networks (QNs), maintaining the product-form property of the model. Roughly speaking, a QN is a set of interconnected queueing systems that serve a set of customers. A QN is in product-form if its steady sate distribution can be expressed as product of functions whose arguments depend only on the state of one station. The definition of these functions depend on the structure of the network (e.g., the customer routing among the queueing nodes), on the average arrival and service rates and on the queueing discipline. QNs satisfying the well-known BCMP theorem on product-form [3] are multiclass, open, closed and mixed QNs, with probabilistic routing and consisting of stations with four possible scheduling disciplines: First Come First Served (FCFS), Last Come First Served with Preemptive Resume (LCFSPR), Processor Sharing (PS) and Infinite Servers (IS). The service time distribution is exponential for FCFS nodes and general (Coxian) for the other three types of nodes. For these models, exact and efficient analysis techniques have been defined. In [9] the author proves that all these node types fulfil a property called Markov implies Markov ($M \Rightarrow M$) and that every other node type that fulfils that property can be embedded in a BCMP QN maintaining the product-form property. Informally, the $M \Rightarrow M$ property requires that under independent Poisson arrival processes, possibly with different rates for class (class independent Poisson arrivals), a work-conserving multiclass station exhibits departure processes that are independent and Poisson distributed. The main strength of this property is that it defines a relatively easy way to decide if a station leads to a QN with product-form solution by its analysis in isolation and under Poisson arrivals. Note that it is well-known that in a QN with cycles, the arrival processes to a station may be non-Poisson, therefore $M \Rightarrow M$ allows us to study the behavior of a queueing center in a special case (independent Poisson arrivals) and then derive the product-form property of the QN. Using $M \Rightarrow M$, several extensions of the BCMP theorem have been proposed, considering various disciplines and other constraints (e.g. [1,7]).

In product-form QNs, FCFS and RANDOM disciplines share the same product-form conditions (i.e. exponential service time with the same rate for all the customer classes, i.e. class independent exponential service time [2]) and steady state queue length distribution under an appropriate aggregation of states.

In this paper we study a class of probabilistic queueing disciplines and we focus on the conditions under which multiclass queueing systems with such disciplines admit product-form solutions. We consider the queuing system in isolation, and its analysis as a building block that can be included in a product-form multiclass QN.

We address the problem of identifying possibly product-form queueing systems having a probabilistic discipline that is not RANDOM. In this paper we prove that, if we consider a general BCMP-like product-form solution, under certain conditions, the only probabilistic discipline that leads to a product-from is the RANDOM. In other words, this is a theoretical negative result that states that the extension of BCMP product-form QNs cannot include nodes with probabilistic disciplines, except for the RANDOM.

In order to discuss the effect of probabilistic disciplines on multiclass queueing systems, we present a performance comparison of two queueing systems with RANDOM and FCFS discipline, respectively, by varying the arrival processes. As a consequence of the equivalence results between FCFS and RANDOM stations under class independent Poisson arrivals, for many practical purposes, in a multiclass product-form QN, a RANDOM station can replace a FCFS station. This can be useful for those networks that although they have a BCMP-like product-form, they do not satisfy the conditions of the well-known algorithms for the exact analysis. For instance, this can happen in case of some load-dependent service rates or in case of nodes that are not strictly BCMP (e.g. Le Boudec's MSCCC [7]). In these cases, simulation may be required even for product-form QNs and using the RANDOM disciplines instead of the FCFS allows for a great simplification of the state of the model. In fact, for RANDOM stations, it is not necessary to represent the arrival order but only the number of customers in the queue for each class (therefore the simulation efficiency is improved).

We address the problem of understanding the different behaviors, in terms of stationary queue length distributions and departure processes, of the FCFS and RANDOM stations under stationary non-Poisson arrival processes. By the results of these experiments, we observe that the two queueing systems show a similar behaviors under low load conditions. Under heavy traffic conditions the FCFS and RANDOM queueing systems still have a close queue length distribution, but they exhibit some significant differences in the departure processes. When the different classes have similar arrival rates, we observe a good approximation of the steady state probabilities independently of the load factor, but the departure processes can still exhibit major differences. As a consequence we can derive the conditions under which RANDOM stations can be used to approximate FCFS stations in complex models that are not in product-form, simplifying the model state and then improving the simulation efficiency.

The paper is structured as follows. Section 2 introduces the queueing model and the definition of the underlying stochastic process. Section 3 reviews the Markov implies Markov property ($M \Rightarrow M$) and introduces the main result of this paper by proving that the RANDOM queueing discipline is the only probabilistic discipline that fulfils the $M \Rightarrow M$ property in queueing station with class independent exponential service time and without preemption or priority. Section 4 presents a performance comparison between the two queuing systems respectively with FCFS and RANDOM disciplines under various arrival processes, through a set of simulation experiments. Section 5 gives some final remarks.

2 Model Description and Notation

In this paper we consider multiclass queueing centers with probabilistic queueing disciplines, single exponential server and class independent stationary arrivals. We name this class of models PQD (Probabilistic Queueing Discipline). Let R be the number of classes, μ the class independent service rate and λ_r the arrival rate of class r customers, for $r = 1, \ldots, R$. When a customer arrives to the queuing system, its service immediately starts if the station is empty, otherwise the customer waits in the queue. At a job completion, a customer is probabilistically chosen from the queue and is immediately put in service. This choice is such that the probability of choosing a customer is non-zero and independent of its arrival time (but e.g. it may depend on its class, on its class arrival rate, on the number of customers of that class in the queue).

Since we are interested in the analysis of the product-form properties of PQD models, we assume exponential class-independent service rates. Indeed, in case of Coxian distributed service times, the station balance condition must hold [4] and this is more restrictive than $M \Rightarrow M$ (e.g. FCFS does not satisfy station balance although it is in product-form in case of class independent exponential service time). Then, in Section 4 we consider PQD models with various arrival distributions and we compare the RANDOM and the FCFS queueing systems with the same service time distribution, in terms of the steady state distributions of the queue lengths and the departure processes.

We denote the state of the model by $\mathbf{m}^{(r)}$, where $\mathbf{m}$ is a R-dimensional vector and $r = 1, \ldots, R$ denotes the class of the customer in service. If the queue is empty we use the symbol $\mathbf{m}_0^{(\epsilon)} \equiv \mathbf{m}_0$, where ϵ means that there is no customer in service. The s-th component of $\mathbf{m}^{(r)}$ is denoted by m_s and represents the number of class s customers in the queue. For instance state $\mathbf{m}^{(r)} = (m_1, \ldots, m_R)^{(r)}$ represents the station with $1 + \sum_{s=1}^{R} m_s$ customers and the customer being served has class r.

At a job completion we probabilistically choose the next customer to be served. We model this behavior by defining a set of non-negative functions $w_s(\mathbf{m}^{(r)})$ that assign a weight to each class $s = 1, \ldots, R$ for every state $\mathbf{m}^{(r)}$. Function

$w_s(\mathbf{m}^{(r)})$ assumes the value 0 if and only if $m_s = 0$. After the job completion the probability of choosing a customer of class s is then given by the following expression:

$$\frac{w_s(\mathbf{m}^{(r)})}{\sum_{t=1}^{R} w_t(\mathbf{m}^{(r)})} \tag{1}$$

It should be clear that if $w_s(\mathbf{m}^{(r)}) = m_s$ then the model is a $M/M/1/RAND$ station since every customer has the same probability of entering in service after a job completion.

In the following, $|\mathbf{m}^{(r)}|_s$ denotes the total number of class s customers in the station, i.e.:

$$|\mathbf{m}^{(r)}|_s = \begin{cases} m_s + 1 & \text{if } r = s \\ m_s & \text{otherwise} \end{cases} .$$

We define $|\mathbf{m}^{(r)}| = \sum_{s=1}^{R} |\mathbf{m}^{(r)}|_s$, i.e., the total number of customers in the station.

3 The RANDOM Queueing Discipline and the BCMP Product-Form

Since BCMP theorem [3] has been formulated, many authors tried to characterize this class of product-form models in terms of structural conditions (e.g. [4]) or conditions on the underlying continuous time Markov chain (CTMC) (e.g. [6,9]).

In particular, in this paper, we focus on the $M \Rightarrow M$ property introduced in [9]. Informally, it states that a multiclass queueing station that, under class independent Poisson arrivals, exhibits class independent Poisson departures, is in product-form. This means that such a queueing station can be embedded in a BCMP QN maintaining the product-form solution of the whole model. It is worthwhile noting that, in general, the arrival processes to the stations within a product-form QN are not Poisson. Therefore, the relevance of the $M \Rightarrow M$ is due to the possibility of studying the behavior of a station in isolation and with independent Poisson arrivals, and then derive the steady state solution when it is embedded in a BCMP-like QN. This property is defined under very general assumptions, i.e., it requires the station to be work-conserving, without priority, and with single-step transitions (see e.g. [5]). In order to prove that the underlying process of a queueing center fulfils the $M \Rightarrow M$ property it suffices to prove that [9]:

$$\forall s = 1, \dots, R, \quad \forall \gamma \in \Gamma, \qquad \sum_{\xi \in \mathcal{S}^{s+}(\gamma)} \pi(\xi) q(\xi \to \gamma) = \pi(\gamma)\lambda_s, \tag{2}$$

where: Γ is the set of reachable states, $\pi(\gamma)$ is the stationary probability of state $\gamma \in \Gamma$ and $\mathcal{S}^{s+}(\gamma) = \{\xi \in \Gamma : |\xi|_s = |\gamma|_s + 1\}$, i.e., the set of states with one customer of class s more than state γ, where $|\xi|_s$ denotes the number of class s customers in state ξ, $q(\xi \to \gamma)$ is the transition rate between states ξ and γ and λ_s is class s arrival rate.

We can rewrite condition (2) for the multiclass probabilistic queue PQD introduced in the previous section, as follows:

$$\forall s = 1, \ldots, R, \forall \mathbf{m}^{(r)} \in \Gamma \qquad \sum_{\mathbf{m}^{(s)} \in \mathcal{S}^{s+}(\mathbf{m}^{(r)})} \pi(\mathbf{m}^{(s)}) q(\mathbf{m}^{(s)} \to \mathbf{m}^{(r)}) = \pi(\mathbf{m}^{(r)}) \lambda_s, \tag{3}$$

where Γ is the (infinite) set of reachable states.

Note that, since there is a single server, if $\mathbf{m}^{(r)}$ can be reached from $\mathbf{m}^{(s)}$ in one step, and $|\mathbf{m}^{(s)}|_t = |\mathbf{m}^{(r)}|_t + 1$, then it follows that $s = t$. Intuitively, this is due to the fact that if $\mathbf{m}^{(s)}$ has a customer of class t more than $\mathbf{m}^{(r)}$ and the latter state can be reached from the former in one step, then that step corresponds to a class s job completion, hence $t = s$.

The main theoretical result of this paper is given by the following theorem. Informally it states that the RANDOM discipline is the *only* probabilistic discipline that fulfils the $M \Rightarrow M$ property for multiclass queueing centers with single server, without preemption, and with class independent exponential service rates. Note that it is well-know that the multiclass RANDOM queueing discipline fulfils the $M \Rightarrow M$ property [1], however, in this paper we prove that it is also *necessary* for those stations that satisfy the conditions of Theorem 1.

Theorem 1. *The RANDOM queue is the only PQD that fulfils the $M \Rightarrow M$ property.*

The proof that RANDOM $\Rightarrow$ ($M \Rightarrow M$) is given in [1] and the stationary distribution is derived. Hence, we have to prove that ($M \Rightarrow M$) $\Rightarrow$ RANDOM. In order to prove the theorem, we introduce some definitions and lemmas. Hereafter, for the sake of simplicity, we just write $\mathbf{m}^{(\cdot)}$ to denote a state when the class of the customer being served is not important.

Definition 1. *Let $\mathbf{m}^{(\cdot)}$,$\mathbf{p}^{(\cdot)} \in \Gamma$. We say that $\mathbf{m}^{(\cdot)} \le \mathbf{p}^{(\cdot)}$ if:*

$$\forall s = 1, \ldots, R \quad |\mathbf{m}^{(\cdot)}|_s \le |\mathbf{p}^{(\cdot)}|_s,$$

and we define $\mathrm{dist}(\mathbf{p}^{(\cdot)}, \mathbf{m}^{(\cdot)}) = |\mathbf{p}^{(\cdot)}| - |\mathbf{m}^{(\cdot)}|$.

Relation $\le$ is a partial order in Γ, and $\mathrm{dist}(\mathbf{p}^{(\cdot)}, \mathbf{m}^{(\cdot)})$ is the number of customers that $\mathbf{p}^{(\cdot)}$ has more than $\mathbf{m}^{(\cdot)}$.

Definition 2 (MM-step and MM-path). *Given states $\mathbf{p}^{(s)}$ we call MM-Step a transition to state $\mathbf{m}^{(r)}$ if $|\mathbf{p}^{(s)}|_s = |\mathbf{m}^{(r)}|_s + 1$. We denote a MM-step by: $\mathbf{p}^{(s)} \xrightarrow{r} \mathbf{m}^{(r)}$ and if $\mathbf{m}^{(r)} = \mathbf{m}_0$ we set $r = \epsilon$. A MM-path is a sequence of MM-steps and is described by a vector whose components are the ordered labels of the arrows as well as the initial state.*

In order to help the intuition let us consider a simple example for state $\mathbf{p}^{(s)} = (2, 1)^{(2)}$ of a $R = 2$ classes PQD. A possible MM-Path from $\mathbf{p}^{(s)}$ to $\mathbf{m}_0$ is:

$$(2, 1)^{(2)} \xrightarrow{1} (1, 1)^{(1)} \xrightarrow{2} (1, 0)^{(2)} \xrightarrow{1} (0, 0)^{(1)} \xrightarrow{\epsilon} \mathbf{m}_0$$

Hence the MM-path is $\alpha = (2,1)^{(2)};[1,2,1,\epsilon]$. In the following, Greek letters denote MM-pathes. Note that, in general, given a state $\mathbf{m}$, the number of possible MM-paths to $\mathbf{m}_0$ is given by the multinomial coefficient

$$\binom{\sum_{s=1}^{R} m_s}{m_1,\ldots,m_R}.$$

Definition 3 (Function Ψ). *Let $\mathbf{m}^{(\cdot)}, \mathbf{p}^{(\cdot)} \in \Gamma$, with $\mathbf{m}^{(\cdot)} < \mathbf{p}^{(\cdot)}$, and let α be a MM-path between $\mathbf{p}^{(\cdot)}$ and $\mathbf{m}^{(\cdot)}$. The function Ψ is defined as follows:*

$$\Psi(\alpha) = \begin{cases} 1 & \text{if } \alpha = \mathbf{p}^{(\cdot)};[] \vee \alpha = \mathbf{p}^{(\cdot)};[\epsilon] \\ \frac{\sum_{s=1}^{R} w_s(\mathbf{p}^{(\cdot)})}{w_r(\mathbf{p}^{(\cdot)})}\Psi(\beta) & \text{otherwise,} \end{cases} \tag{4}$$

where β is the MM-path α with the first MM-step removed.

Intuitively, function Ψ is the reciprocal of the probability that MM-path α occurs given that there has not been any arrival to the queueing center.

The following lemmas state that if the PQD model fulfils the $M \Rightarrow M$ property, then function Ψ only depends on the initial and the final states of the path, i.e., is independent of the order of the MM-steps.

Lemma 1. *If a PQD satisfies the $M \Rightarrow M$ property, then:*

1. *if ($\mathbf{p}^{(s)} \xrightarrow{r} \mathbf{m}^{(r)}$) then we can write:*

$$\frac{\pi(\mathbf{p}^{(s)})}{\pi(\mathbf{m}^{(r)})} = \frac{\lambda_s}{\mu}\frac{\sum_{t=1}^{R} w_t(\mathbf{m}^{(s)})}{w_r(\mathbf{m}^{(s)})}.$$

2. *if $\mathbf{p}^{(s)} \xrightarrow{\epsilon} \mathbf{m}_0$ then we an write:*

$$\frac{\pi(\mathbf{p}^{(s)})}{\pi(\mathbf{m}_0)} = \frac{\lambda_s}{\mu}.$$

Proof. Condition (3) holds by hypothesis, and we already noted that $\mathcal{S}^{s+}(\mathbf{m}^{(r)}) = \{\mathbf{p}^{(s)}\}$. We know that the transition rate from $\mathbf{p}^{(s)}$ to $\mathbf{m}^{(s)}$ is μ multiplied by the probability of choosing a class r customer to put in service. Then the lemma can be derived by trivial algebra. □

Lemma 2. *A PQD satisfies the $M \Rightarrow M$ property if and only if for all states $\mathbf{p}^{(s)} \in \Gamma$ and for all the MM-pathes α from $\mathbf{p}^{(s)}$ to $\mathbf{m}_0$ we have $\Psi(\alpha) = \psi(\mathbf{p}^{(s)})$. Then the stationary probability of state $\mathbf{p}^{(s)}$ can be expressed as follows:*

$$\pi(\mathbf{p}^{(s)}) = \pi(\mathbf{m}_0)\prod_{t=1}^{R}\left(\frac{\lambda_t}{\mu}\right)^{|\mathbf{p}^{(s)}|_t}\psi(\mathbf{p}^{(s)}). \tag{5}$$

Proof. Suppose that the PQD model satisfies the $M \Rightarrow M$ property, and let α be an arbitrary MM-path from $\mathbf{p}^{(r)}$ to $\mathbf{m}_0$. First, we prove by induction that

$$\pi(\mathbf{p}^{(s)}) = \pi(\mathbf{m}_0) \prod_{t=1}^{R} \left(\frac{\lambda_t}{\mu}\right)^{|\mathbf{p}^{(s)}|_t} \Psi(\alpha). \tag{6}$$

Base case: if $\mathbf{p}^{(s)} \xrightarrow{\epsilon} \mathbf{m}_0$ then Equation (6) is verified by Lemma 1 and Definition 3.

Inductive step: suppose that:

$$\alpha = \mathbf{p}^{(s)} \xrightarrow{r} \underbrace{\mathbf{m}^{(r)} \to \cdots \to \mathbf{m}_0}_{\beta},$$

and $r \neq \epsilon$. By induction we know that:

$$\pi(\mathbf{m}^{(r)}) = \pi(\mathbf{m}_0) \prod_{t=1}^{R} \left(\frac{\lambda_t}{\mu}\right)^{|\mathbf{p}^{(s)}|_t} \frac{\mu}{\lambda_s} \Psi(\beta). \tag{7}$$

We can write:

$$\frac{\pi(\mathbf{p}^{(s)})}{\pi(\mathbf{m}_0)} = \frac{\pi(\mathbf{p}^{(s)})}{\pi(\mathbf{m}^{(r)})} \frac{\pi(\mathbf{m}^{(r)})}{\pi(\mathbf{m}_0)},$$

that, by Lemma 1 combined with Equation (7) becomes:

$$\frac{\pi(\mathbf{p}^{(s)})}{\pi(\mathbf{m}_0)} = \frac{\lambda_s}{\mu} \prod_{t=1}^{R} \left(\frac{\lambda_t}{\mu}\right)^{|\mathbf{p}^{(s)}|_t} \frac{\mu}{\lambda_s} \Psi(\beta) \frac{\sum_{t=1}^{R} w_r(\mathbf{p}^{(s)})}{w_r(\mathbf{p}^{(s)})}.$$

Using definition 3 we conclude the proof by induction.

Let us consider two different MM-pathes α_1 and α_2 from $\mathbf{p}^{(s)}$ to $\mathbf{m}_0$. By the uniqueness of π for ergodic CTMCs, the following condition is satisfied:

$$\pi(\mathbf{m}_0) \prod_{t=1}^{R} \left(\frac{\lambda_t}{\mu}\right)^{|\mathbf{p}^{(s)}|_t} \Psi(\alpha_1) = \pi(\mathbf{m}_0) \prod_{t=1}^{R} \left(\frac{\lambda_t}{\mu}\right)^{|\mathbf{p}^{(s)}|_t} \Psi(\alpha_2),$$

that implies $\Psi(\alpha_1) = \Psi(\alpha_2) = \psi(\mathbf{p}^{(s)})$. Equation (5) can be obtained from Equation (6) by substitution of Ψ with ψ.

Vice versa let us assume Equation (5) and prove that the $M \Rightarrow M$ property holds. We need to prove Equation (3) rewritten as follows recalling that the transition rate from $\mathbf{p}^{(s)}$ to $\mathbf{m}^{(r)}$ is $\mu w_r(\mathbf{p}^{(s)}) / \sum_{t=1}^{R} w_t(\mathbf{p}^{(s)})$:

$$\pi(\mathbf{p}^{(s)}) \mu \frac{w_r(\mathbf{p}^{(s)})}{\sum_{t=1}^{R} w_t(\mathbf{p}^{(s)})} = \pi(\mathbf{m}^{(r)}) \lambda_s.$$

This can be easily derived by substituting function π by formula (5), in fact:

$$\prod_{t=1}^{R} \left(\frac{\lambda_t}{\mu}\right)^{|\mathbf{p}^{(s)}|_t} \psi(\mathbf{p}^{(s)}) \mu \frac{w_r(\mathbf{p}^{(s)})}{\sum_{t=1}^{R} w_t(\mathbf{p}^{(s)})} = \prod_{t=1}^{R} \left(\frac{\lambda_t}{\mu}\right)^{|\mathbf{m}^{(r)}|_t} \psi(\mathbf{m}^{(r)}) \lambda_s.$$

Noting that $|\mathbf{p}^{(s)}|_s = |\mathbf{m}^{(r)}|_s + 1$ and $|\mathbf{p}^{(s)}|_t = |\mathbf{m}^{(r)}|_t$ with $t \neq s$, we have:

$$\frac{\lambda_s}{\mu} \prod_{t=1}^{R} \left(\frac{\lambda_t}{\mu}\right)^{|\mathbf{m}^{(r)}|_t} \psi(\mathbf{p}^{(s)}) \mu \frac{w_r(\mathbf{p}^{(s)})}{\sum_{t=1}^{R} w_t(\mathbf{p}^{(s)})} = \prod_{t=1}^{R} \left(\frac{\lambda_t}{\mu}\right)^{|\mathbf{m}^{(r)}|_t} \psi(\mathbf{m}^{(r)}) \lambda_s .$$

This is true if:

$$\psi(\mathbf{p}^{(s)}) \frac{w_r(\mathbf{p}^{(s)})}{\sum_{t=1}^{R} w_t(\mathbf{p}^{(s)})} = \psi(\mathbf{m}^{(r)}).$$

Let us consider a MM-path $\alpha = \mathbf{p}^{(s)} \xrightarrow{r} \mathbf{m}^{(r)} \to \cdots \to \mathbf{m}_0$ and $\beta = \mathbf{m}^{(r)} \to \cdots \to \mathbf{m}_0$, by hypothesis $\Psi(\alpha) = \psi(\mathbf{p}^{(s)})$ and $\Psi(\beta) = \psi(\mathbf{m}^{(r)})$, then we can apply Definition 3 and conclude the proof of the lemma. □

Although Lemma 2 states that a PQD model satisfies the $M \Rightarrow M$ property if and only if for every state $\mathbf{p}^{(\cdot)} \in \Gamma$ there exists a function ψ depending only on $\mathbf{p}^{(\cdot)}$ such that $\psi(\mathbf{p}^{(\cdot)}) = \Psi(\alpha)$ for every MM-path α from $\mathbf{p}^{(\cdot)}$ to $\mathbf{m}_0$, it does not give any information about the definition of ψ. The following Lemma gives the definition for ψ from which we straightforwardly derive the proof of Theorem 1.

Lemma 3. *If a PQD model fulfils the $M \Rightarrow M$ property then we can write, for each $\mathbf{p}^{(\cdot)} \in \Gamma$:*

$$\psi(\mathbf{p}^{(\cdot)}) = \binom{\sum_{s=1}^{R} p_s}{p_1, \ldots, p_R}. \tag{8}$$

Proof. The proof is by induction on $|\mathbf{p}^{(s)}|$.

Let $A(\mathbf{p}^{(\cdot)})$ be the set of the class labels with at least one customer in the queue, defined as follows:

$$A(\mathbf{p}^{(\cdot)}) = \{r : p_r > 0\}, \text{ with } r = 1, \ldots, R$$

Note that if $|A(\mathbf{p}^{(\cdot)})| \leq 1$ then we immediately obtain $\psi(\mathbf{p}^{(\cdot)}) = 1$. Therefore if $|\mathbf{p}^{(\cdot)}| \leq 2$ the lemma is trivially satisfied.

Base case: $|\mathbf{p}^{(\cdot)}| = 3$. In this case there is a customer in service and two customers in the queue. If the latter ones belong to the same class then the base case is verified as there is obviously just one possible choice of the class to put in service. Let us consider the case that the queued customers have two different classes s and t. Then the MM-pathes from $\mathbf{p}^{(\cdot)}$ to $\mathbf{m}_0$ are:

$$\alpha : \mathbf{p}^{(\cdot)} \xrightarrow{s} \mathbf{m}^{(s)} \xrightarrow{t} \mathbf{n}^{(t)} \xrightarrow{\epsilon} \mathbf{m}_0$$
$$\beta : \mathbf{p}^{(\cdot)} \xrightarrow{t} \mathbf{m}^{(t)} \xrightarrow{s} \mathbf{n}^{(s)} \xrightarrow{\epsilon} \mathbf{m}_0,$$

hence:

$$\Psi(\alpha) = \frac{w_s(\mathbf{p}^{(\cdot)}) + w_t(\mathbf{p}^{(\cdot)})}{w_s(\mathbf{p}^{(\cdot)})}$$
$$\Psi(\beta) = \frac{w_s(\mathbf{p}^{(\cdot)}) + w_t(\mathbf{p}^{(\cdot)})}{w_t(\mathbf{p}^{(\cdot)})}$$

By Lemma 2 we have that $\psi(\mathbf{p}^{(\cdot)}) = \Psi(\alpha) = \Psi(\beta)$ because by hypothesis the station fulfils the $M \Rightarrow M$ property. Therefore, we conclude $w_s(\mathbf{p}^{(\cdot)}) = w_t(\mathbf{p}^{(\cdot)})$ and $\psi(\mathbf{p}^{(\cdot)}) = 2$.

Induction step: let us consider a state $\mathbf{p}^{(\cdot)}$ with $|\mathbf{p}^{(\cdot)}| > 3$. If $A(\mathbf{p}^{(\cdot)}) = 1$ then the result immediately follows. Let us assume $|A(\mathbf{p}^{(\cdot)})| > 1$ and let $r \in A(\mathbf{p}^{(\cdot)})$. Let us consider an MM-path α to $\mathbf{m}_0$ such that the first MM-step is $\mathbf{p}^{(\cdot)} \xrightarrow{r} \mathbf{m}^{(r)}$. Since $|\mathbf{m}^{(r)}| < |\mathbf{p}^{(\cdot)}|$, by inductive hypothesis we have that:

$$\psi(\mathbf{m}^{(r)}) = \binom{\sum_{s=1}^{R} m_s}{m_1, \ldots, m_R} = \binom{(\sum_{s=1}^{R} p_s) - 1}{p_1, \ldots, p_r - 1, \ldots, p_R}$$

that gives:

$$\Psi(\alpha) = \frac{\sum_{t \in A(\mathbf{p}^{(\cdot)})} w_t(\mathbf{p}^{(\cdot)})}{w_r(\mathbf{p}^{(\cdot)})} \psi(\mathbf{m}^{(r)}). \tag{9}$$

Let us consider $r' \in A(\mathbf{p}^{(\cdot)})$ with $r' \neq r$. In a similar manner we obtain a MM-path β and:

$$\psi(\mathbf{m}^{(r')}) = \binom{\sum_{s=1}^{R} m_s}{m_1, \ldots, m_R} = \binom{(\sum_{s=1}^{R} p_s) - 1}{p_1, \ldots, p'_r - 1, \ldots, p_R},$$

$$\Psi(\beta) = \frac{\sum_{t \in A(\mathbf{p}^{(\cdot)})} w_t(\mathbf{p}^{(\cdot)})}{w_{r'}(\mathbf{p}^{(\cdot)})} \psi(\mathbf{m}^{(r)}).$$

By Lemma 2 we have that $\Psi(\alpha) = \Psi(\beta)$. Then:

$$\frac{1}{w_r(\mathbf{p}^{(\cdot)}) p_{r'}} = \frac{1}{w_{r'}(\mathbf{p}^{(\cdot)}) p_r}.$$

that can be written as:

$$\frac{w_r(\mathbf{p}^{(\cdot)})}{w_{r'}(\mathbf{p}^{(\cdot)})} = \frac{p_r}{p_{r'}}.$$

Since this relation must hold for every couple $r, r' \in A(\mathbf{p}^{(\cdot)})$ and for each state $\mathbf{p}^{(\cdot)}$ that has at least two different classes of customers in the queue (note that when the queue is empty or all the customers belong to the same class the definition of function w is not really important) we conclude that $w_r(\mathbf{p}^{(\cdot)}) = f(\mathbf{p}^{(\cdot)}) p_r$ for every $r \in A(\mathbf{p}^{(\cdot)}$, where f is a non-negative function. By replacing this expression in Equation (9) we conclude the proof of the lemma. □

Proof (Theorem 1). Using the previous lemmas, the theorem proof is very simple. First we observe that the class r weight function w_r influences the model behavior only if there are two or more customers belonging to different classes in the queue. In this case Lemma 3 states that $w_r(\mathbf{p}^{(\cdot)}) = f(\mathbf{p}^{(\cdot)}) p_r$, with $\mathbf{p}^{(\cdot)} \in \Gamma$ and f a non-zero function. Note that f is does not affect the behavior of the PQD because the definition of the probability of entering in service is given by Expression (1). Therefore, we proved that a necessary condition for a PQD to fulfil the $M \Rightarrow M$ property is that the scheduling discipline is RANDOM. □

4 Comparison of RANDOM with FCFS Queueing Discipline

It is well-known that under class independent Poisson arrivals a queueing system with RANDOM or FCFS queueing disciplines is in product-form if the service time distribution is class independent and exponential. In particular, according to the model description given in Section 2 the steady state solution for the RANDOM discipline is:

$$\pi(\mathbf{m}) = \pi_0 \binom{\sum_{r=1}^{R} m_r}{m_1, \ldots, m_R} \prod_{r=1}^{R} \left(\frac{\lambda_r}{\mu}\right)^{m_r}, \tag{10}$$

where π_0 is the stationary probability of observing the empty station, $\mathbf{m}$ is a vector whose components m_r represent the total (in service or in queue) number of customers of class r in the station. The state of FCFS station must represent the arrival order of the customers therefore a straightforward comparison with the expression given by Formula (10) is not possible. However, if we consider an aggregation of states that just represents the number of customers in the station despite to their arrival order, Equation (10) expresses its steady state solution. Note that this equivalence holds even for multiple exponential servers as proven in [2].

In this section we compare the FCFS queueing system with the RANDOM one by assuming class independent exponential service time and under various arrival processes. We focus our attention on the following indices, that are known to be equal in case of Poisson arrivals:

- the steady state probability of observing a given number of customers of a class,
- the distribution of the interdeparture time for a class of customers in steady state. This characterizes the departure process.

By these experiments, we just aim to present an example to illustrate that in multiclass stations the queueing discipline influences the station performance indices, even if it is work-conserving and without priority, and if the service time distribution is class independent and exponential (i.e. with the memoryless property). From this observation we informally derive the conditions under which a multiclass FCFS station embedded in a large model can be approximated by a RANDOM station maintaining the overall model behavior unchanged.

4.1 Experiments

In order to obtain these estimates, we have built a simulator in Java using the combined multiple recursive generator class by L'Ecuyer [8]. Since we are interested in steady state estimates, we used Welch's procedure to define the warmup period [11], performed a set of independent replications of the simulation and constructed 90 percent confidence intervals. The validation of the model has been done with independent Poisson arrivals comparing the estimates with the exact results. We have considered uniform, hyperexponential and Erlang interarrival time distributions and different load factors.

For the sake of brevity, in this section we show the results of some experiments obtained with the following conditions:

- $R = 3$ classes of customers.
- the interarrival time distributions are Erlang r.v.s with 20 stages of service, with means $1/\lambda_1 = 1.3$, $1/\lambda_2 = 10.0$ and $1/\lambda_3 = 4.0$
- the service rate is exponentially distributed with mean $1/\mu$.

We vary the value of μ in order to change the station load factor $\rho = (\lambda_1 + \lambda_2 + \lambda_3)/\mu$. Since only the classes with the slowest arrival rates present major differences for the considered performance indices, Figures 1, 2, 3 and 4 illustrate the comparison of the two disciplines for classes 2 and 3 and different values of ρ. For all the experiments we observe that the confidence interval width is always less than 0.001, so they are not drawn in the pictures.

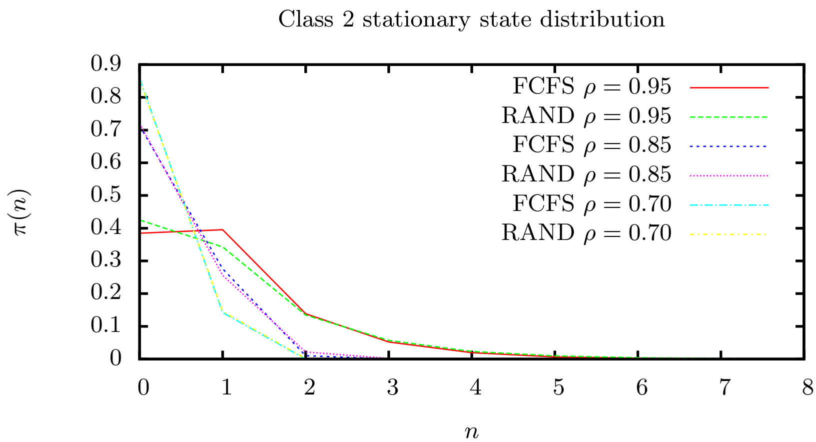

Fig. 1. Class 2 queue population for different values of ρ in steady state

The results clearly show that the queueing disciplines influence the departure processes and the queue population distributions in steady state even if the service time distribution is class independent.

In general we have observed that the FCFS and the RANDOM discipline exhibit similar behaviors in terms of queue population and departure process, if: the class arrival rates are similar, the load factor of the station is low and the interarrival rate distributions can be approximated by independent exponential random variables.

By observing figures 1, 2, 3, 4, it is possible to see the RANDOM and FCFS disciplines exhibit an almost identical behavior, in terms of the considered performance indices, when the load factor $\rho \leq 0.5$. This has an intuitive motivation,

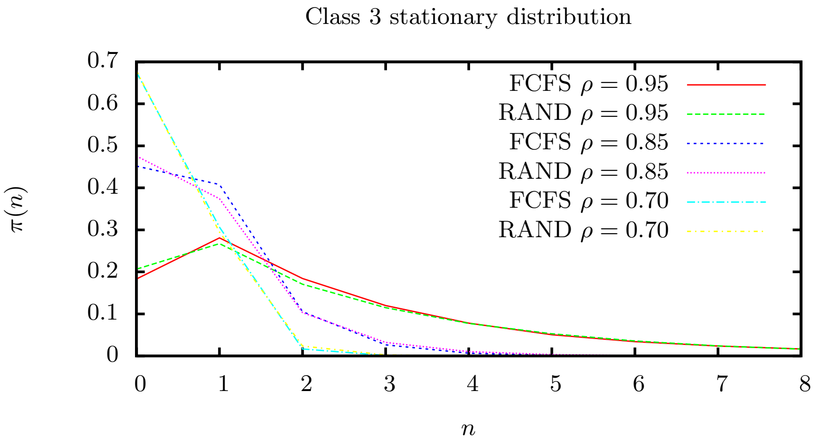

Fig. 2. Class 3 queue population for different values of ρ in steady state

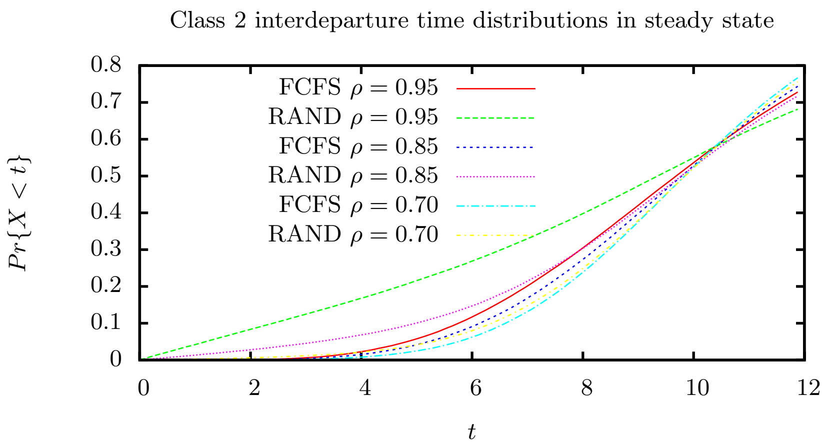

Fig. 3. Class 2 interdeparture time X accumulation function in steady state

i.e., when the traffic is low the probability of observing many customers in the queue is low, therefore the scheduling discipline tends to have a low impact on the considered performance measures. The load factor seems to be a reasonable parameter to analyze if we want to approximate the FCFS stations of a simulated model by RANDOM stations for the sake of improvement of the simulation performance using a more compact state representation.

It is worthwhile pointing out that, for the considered performance indices, the product-form property shows an insensitivity to the scheduling discipline (FCFS or RANDOM) for these multiclass models that is not true in general. Note that this is a peculiarity of multiclass models, while for single class models it is well-known the insensitivity property of work-conserving and non-priority disciplines (e.g. [5]).

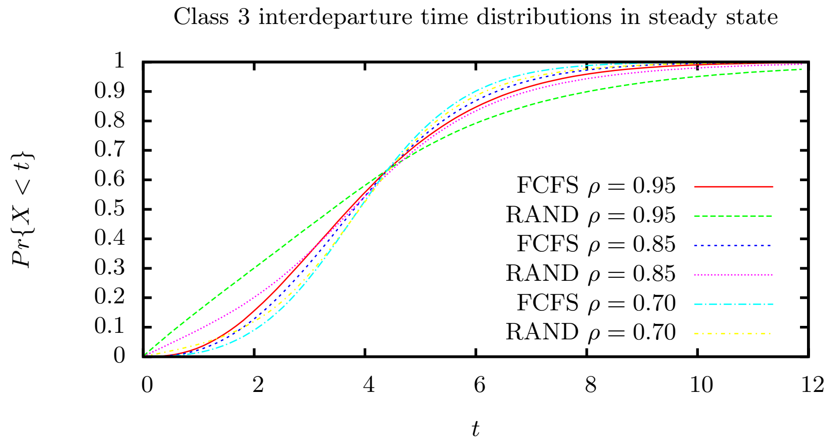

Fig. 4. Class 3 interdeparture time X accumulation function in steady state

5 Conclusions

In this paper we have presented and proved a new theoretical result that characterizes the probabilistic disciplines that fulfil the $M \Rightarrow M$ property, i.e., that can be embedded in a BCMP-like QN maintaining the product-form property of the model. Moreover, we have performed a set of simulation experiments to compare the behaviors of FCFS and RANDOM disciplines under non-Poisson arrival processes. In particular, we have shown the impact of the load-factor on the similarity of the queue population distributions and the departure processes in steady state. In the first part of the paper we have shown that assuming a class independent exponential server the only probabilistic queueing discipline that fulfils the $M \Rightarrow M$ property is the RANDOM. The importance of this result is about the impossibility of defining general non-RANDOM probabilistic disciplines (e.g. [10]) with a class independent exponential server that can be composed in a BCMP-like manner, i.e., that can be studied very efficiently by exact techniques.

In the second part of the paper we have addressed the problem of analyzing the impact on a set of relevant performance indices of the queueing discipline under some non-Poisson class independent arrival processes. The simulation results show that, in case of heavy load, the population distributions and the departure processes of the FCFS and RANDOM models differ. In practice, we explore the possibility of replacing a FCFS station by a RANDOM station (with class independent exponential server) in a non-product-form model. This can be useful because the state representation of the RANDOM station is much more compact than that of FCFS. However, although this leads to exact results in case of product-form QNs, we have observed that it is not generally true. In particular major differences on the steady state population distributions and on the departure processes are observed in case of heavy traffic.

References

1. Afshari, P.V., Bruell, S.C., Kain, R.Y.: Modeling a new technique for accessing shared buses. In: Proc. of the Computer Network Performance Symp., pp. 4–13. ACM Press, New York (1982)
2. Balsamo, S., Marin, A.: On representing multiclass M/M/k queues by generalized stochastic Petri nets. In: Proc. of ECMS/ASMTA-2007 Conf., Prague, Czech Republic, June 4-6, pp. 121–128 (2007)
3. Baskett, F., Chandy, K.M., Muntz, R.R., Palacios, F.G.: Open, closed, and mixed networks of queues with different classes of customers. J. ACM 22(2), 248–260 (1975)
4. Chandy Jr., K.M., Howard, J.H., Towsley, D.F.: Product form and local balance in queueing networks. J. ACM 24(2), 250–263 (1977)
5. Kant, K.: Introduction to Computer System Performance Evaluation. McGraw-Hill, New York (1992)
6. Kelly, F.: Reversibility and stochastic networks. Wiley, New York (1979)
7. Le Boudec, J.Y.: A BCMP extension to multiserver stations with concurrent classes of customers. In: SIGMETRICS 1986/PERFORMANCE 1986: Proc. of the 1986 ACM SIGMETRICS Int. Conf. on Computer performance modelling, measurement and evaluation, pp. 78–91. ACM Press, New York (1986)
8. L'Ecuyer, P.: Good parameters and implementations for combined multiple recursive random number generators. Operations Research 47(1), 159–164 (1999)
9. Muntz, R.R.: Poisson departure processes and queueing networks. Technical Report IBM Research Report RC4145, Yorktown Heights, New York (1972)
10. Tham, C., Yao, Q., Jiang, Y.: A multi-class probabilistic priority scheduling discipline for differentiated services networks. Computer Communications 25(17), 1487–1496 (2002)
11. Welch, P.D.: On the problem of the initial transient in steady-state simulations. Technical report, IBM Watson Research Center, Yorktown Heights, NY (1981)