Product Form Queueing Petri Nets: A Combination of Product Form Queueing Networks and Product Form Stochastic Petri Nets

Falko Bause, Peter Buchholz

Informatik IV
Universität Dortmund
D-44221 Dortmund
Germany
e-mail:{bause,buchholz}@ls4.informatik.uni-dortmund.de

Abstract. The product form results recently published for Stochastic Petri Nets are combined with the well known product form results for Queueing Networks in the model formalism of Queueing Petri Nets yielding the class of product form Queueing Petri Nets. This model class includes Stochastic Petri Nets with product form solution and BCMP Queueing Networks as special cases. We introduce an arrival theorem for the model class and present an exact aggregation approach extending known approaches from Queueing Networks.

1 Introduction

The use of product form Queueing Networks (PQNs) as introduced in [6] for performance analysis of computer and communication systems is widespread since the product form allows an extremely efficient solution compared with any other analysis approach. More recently a class of Stochastic Petri Nets (SPNs) with product form solution (PSPNs) has been introduced in various papers [10, 11, 12]. The two model classes are different and the resulting product forms are also different. In this paper we combine both approaches in one model formalism. The basic formalism follows ideas from Queueing Petri Nets (QPNs) as described in [5]. We start with a brief description of PSPNs and traditional PQNs where queues are of the so called BCMP type extended by the flow equivalent server as defined in [8, 13]. Combination of these two model classes leads to a new model class with product form solution, named PQPNs, which includes PSPNs and PQNs as special cases in a common framework.

We show that an arrival theorem holds for this model class which is an extension of the arrival theorem in PQNs. Furthermore, an aggregation approach is presented which allows to analyse a subnet using only very limited information about the environment and to substitute the subnet by a simpler aggregate. This approach extends aggregation techniques in PQNs, although the resulting aggregates do not in general have the physical interpretation which can be given in PQNs.

The outline of the paper is as follows. The next three sections summarise very briefly PSPNs and PQNs and introduce PQPNs and the main product form result for PQPNs. Subsequently the arrival theorem for PQPNs is established. In section 6 an aggregation approach for PQPNs is presented. An example is given to clarify the concepts. The appendix contains the detailed proof of the product form result and defines some functions to compute a detailed queue state.

2 SPNs with product form solution

We start with a brief description of PSPNs following [11, 12]. For details we refer to the cited papers. Consider a live and bounded SPN in continuous time, where all transitions have exponentially distributed firing delays. Let P denote the set of places, T the set of transitions and C the set of token colours. All sets are finite. Let M be a reachable marking, $M(p)(c)$ the number of colour c tokens on

place p in marking M. $I(t)$ is the input bag for transition t. Transitions are allowed to have several output bags which are chosen according to fixed probabilities. The j-th output bag for transition t is denoted by $O_j(t)$. $r(M,t)$ is the firing rate of transition t in marking M.
A SPN of the above type has to observe the following conditions to yield a product form solution:

1. No two transitions have the same input bag, i.e., $I(t) \neq I(s)$ for all $s,t \in T$, $s \neq t$.
2. Each output bag of a transition is the input bag of some transition, i.e., $O_j(t) = I(s)$ for all $t \in T$ and some $s \in T$.
3. The marking dependent firing rate of each transition t can be expressed as $r(M,t) = \frac{\varphi(M-I(t))\chi(t)}{\Phi(M)}$, where φ, Φ and χ are arbitrary non-negative functions.

We can define the routing chain of a PSPN as a discrete time Markov chain with one step transition probabilities $p(s,t)$, where $O_j(s) = I(t)$ and $p(s,t)$ is the probability that the firing of s yields a new input bag for t. An invariant measure of the routing chain is the solution of

$$x(t) = \sum_{s \in T} x(s)p(s,t) \tag{1}$$

A set of functions F from T to $(0,\infty)$ is defined such that $f(t)\chi(t)$ is an invariant measure of the routing chain.

$$F = \{f : T \longrightarrow (0,\infty) : \chi(t)f(t) = \sum_{s \in T} \chi(s)f(s)p(s,t), \forall t \in T\} \tag{2}$$

Since the values of f are non-zero, the routing chain has to consist of positive recurrent classes of transitions only. This implies the following additional structural condition.

2'. The input bag of each transition is an output bag of some transition, i.e., $I(t) = O_j(s)$ for all $t \in T$ and $s \in T$.

In the sequel we will denote conditions 2 and 2' as condition 2.

Let $\mathcal{V} = \{V_1, \ldots, V_R\}$ be the set of recurrent classes of transitions, with each V_r called a subchain. Apart from f a function g has to exist for a product form solution such that

$$\frac{g(M + I(s))}{g(M + I(t))} = \frac{f(s)}{f(t)} \tag{3}$$

If the above conditions are observed in the SPN, then the ratio $f(s)/f(t)$ is unique for all transitions in a recurrent class, although $f(t)$ is not unique. Additionally the function g is defined uniquely up to a constant for all possible markings M. The equilibrium distribution of the PSPN has the form

$$\pi(M) = \frac{1}{G}\, g(M)\Phi(M), \text{ where } G \text{ is a normalisation constant.}$$

3 QNs with product form solution

The basic class of product form Queueing Networks (PQNs) has been defined in [6] and are often denoted as BCMP networks. Several extensions to the basic class appeared subsequently. Here we consider closed PQNs, start with the original BCMP queues and introduce queues of the flow equivalent server type with service rates depending on the state of the queue. Additional types of queues with a product form solution can be found in [8], the results presented here can be easily extended to those types.

Consider a closed QN including several classes of customers. Let Q be the set of queues and K the set of classes. A queue consists of one or several waiting lines and one or several servers. We assume that each class visits exactly one queue in the QN, such that the class index implicitly determines the queue index. Each QN can be transformed to meet this condition by introduction of additional classes. Routing of customers is described by a routing matrix $R = [R_{k,l}]$ $(k, l \in K)$.

Matrix R can be considered as the transition matrix of a discrete time Markov chain. This Markov chain can be decomposed into a set of ergodic subchains $\mathcal{V} = \{V_1, \ldots, V_R\}$. For each subchain the following set of equations can be solved uniquely up to a multiplicative constant.

$$v(l) = \sum_{k \in V_r} v(k) R_{k,l} \qquad \forall V_r \in \mathcal{V},\ \forall l \in V_r$$

The $v(k)$ are often denoted as relative visit ratios of class k. Service time distributions at the various queues (as defined below) are described by Coxian distributions with class dependent parameters. The distribution for class k is characterised by the number of Coxian phases denoted by u_k, the service rate of phase l $(1 \leq l \leq u_k)$ denoted by $\mu_{k,l}$, and the probabilities of entering phase $l+1$ after l denoted by $a_{k,l}$, where $a_{k,u_k} = 0$. Let $A_{k,l} = \prod_{i=1}^{l-1} a_{k,i}$ be the probability of a class k customer to reach phase l of the corresponding service time distribution. All queues in the PQN are of one of the five following types:

1. The service discipline is first come first served (FCFS), service times of all customer classes are exponentially distributed with identical mean μ^{-1}.
2. The service discipline is processor sharing (PS), service times may be class specific and represented by Coxian distributions.
3. Each customer gets his own, exclusive server (infinite server, IS), service times may be class specific and represented by Coxian distributions.
4. The service discipline is last come first served preemptive resume (LCFS), service times may be class specific and represented by Coxian distributions.
5. From each class of customers with a non-zero population in the queue one customer is served, the remaining customers are waiting in class specific waiting rooms with FCFS service discipline, service times are exponentially distributed with mean $(\mu_k(n))^{-1}$ for class k and population $n = (n_1, \ldots, n_K)$. For the mean values the relations

$$\mu_k(n)\mu_l(n - e_k) = \mu_l(n)\mu_k(n - e_l), \text{ where } n - e_k = (n_1, \ldots, n_{k-1}, n_k - 1, n_{k+1}, \ldots, n_K) \tag{4}$$

   hold, whenever the population vectors are defined [8, 13]. This type of queue is called the flow equivalent service center (FESC).

The detailed state of a PQN is defined by a vector $x = (x_1, \ldots, x_Q)$, where x_p is the detailed state of queue p (see appendix A for a definition). Let $n = (n_1, \ldots, n_Q)$ be the marginal state of the network such that n_p is the population vector or marginal state of queue p. The state x_p of a queue depends on the type of the queue. Let $h_p(x_p)$ be a function which depends only on the queue type, the service time distributions and the relative visit ratios $v(k)$ of classes which visit p, then

$$\pi(x) = \frac{1}{G} \prod_{p=1}^{Q} h_p(x_p) \text{ , where } G \text{ is a normalisation constant.} \tag{5}$$

The functions $h_p(x_p)$ can be found in [6] and in appendix A. The marginal distribution of the QN which considers only the populations at the queues and not the detailed states can be computed as

$$\pi(n) = \frac{1}{G} \prod_{p=1}^{Q} d_p(n_p) \tag{6}$$

where $d_p(n_p)$ is a function which depends only on the queue type, the relative visit ratios and the mean of the service time distribution (cf. appendix A). Define furthermore

$$S_p(x_p) = h_p(x_p)/d_p(n_p)\ , \text{ where detailed state } x_p \text{ belongs to marginal state } n_p. \qquad (7)$$

$S_p(x_p)$ denotes the conditional probability of the detailed state x_p when the marginal state is n_p. According to the population n_p in queue p we can define a mean departure rate of class k customers as

$$\lambda(n_p, k) = \begin{cases} \mu n_p(k)/\sum_{l\in K} n_p(l) & \text{for type 1 queues} \\ \mu_k n_p(k)/\sum_{l\in K} n_p(l) & \text{for type 2 and 4 queues} \\ \mu_k n_p(k) & \text{for type 3 queues} \\ \mu_k(n_p) & \text{for type 5 queues} \end{cases} \qquad (8)$$

where $k \in K$ and $\frac{1}{\mu_k}$ is the mean service time of class k customers.

All types of queues observe the conditions which have been defined for the firing rates of PSPNs in the previous section. This can be proved by setting $\Phi = \varphi$ as shown in [11].

For later use, $\lambda_{ph}(x, k, l)$ denotes the rate of class k customers leaving phase l in state x and entering phase $l + 1$ of their service and let $\lambda_{ex}(x, k, l)$ denote the rate of class k customers in phase l leaving the queue in state x (see appendix A).

4 QPNs with product form solution

In this section we establish a product form solution for Queueing Petri Nets, thus combining the former approaches. PQNs and PSPNs are special cases of this class of Queueing Petri Nets with product form solutions.

Queueing Petri Nets (QPNs) [4, 5] combine Coloured Generalized Stochastic Petri Nets (CGSPNs) [1, 9] with Queueing Networks (QNs) by hiding queues in special places of a CGSPN which are called queueing places or timed places.

Figure 1 i) depicts such a queueing place and its pictorial representation is given in Fig. 1 ii). A queueing place consists of two components, a queue and a depository for tokens having completed their service at the queue. Tokens, when fired onto a queueing place by any of its input transitions, are inserted into the queue. Tokens within a queue are not available for any QPN transition. After completion of its service a token moves to the depository. The tokens are now available for all output transitions of the queueing place. An enabled timed transition will fire after a certain exponentially distributed delay according to a race policy as in GSPNs. Enabled immediate transitions will fire according to relative firing frequencies. The firings of immediate transitions have priority over those of timed transitions. A QPN describes a stochastic process which can be analysed by Markovian analysis techniques [5].

For our purposes concerning product form QPNs, a QPN can be defined by a triple $(CGSPN, P_1, P_2)$ where $CGSPN$ is a coloured GSPN, P_1 is the set of queueing places and P_2 is the set of ordinary places where no queue is integrated. $P_1 \cup P_2$ constitutes the set of places P of the $CGSPN$.

To obtain a direct correspondence of a $CGSPN$ and a PSPN, let us assume that each transition has only one colour, i.e. it fires only in one mode. This can easily be achieved by unfolding all transitions. In the following we will denote such transitions as uncoloured.

Thus a $CGSPN$ here is a 4-tuple (CPN, T_1, T_2, W) where CPN is the underlying Coloured Petri Net, $T_1 \subseteq T$ is the set of timed transitions, $T_2 \subseteq T$ is the set of immediate transitions and $T = T_1 \cup T_2$ is the set of the CPN's transitions. $W = (w_1, \ldots, w_{|T|})$ is an array whose entry w_i

- is a (possibly marking dependent) rate $\in \mathbb{R}^+$ of a negative exponential distribution specifying the firing delay of t_i if $t_i \in T_1$ or

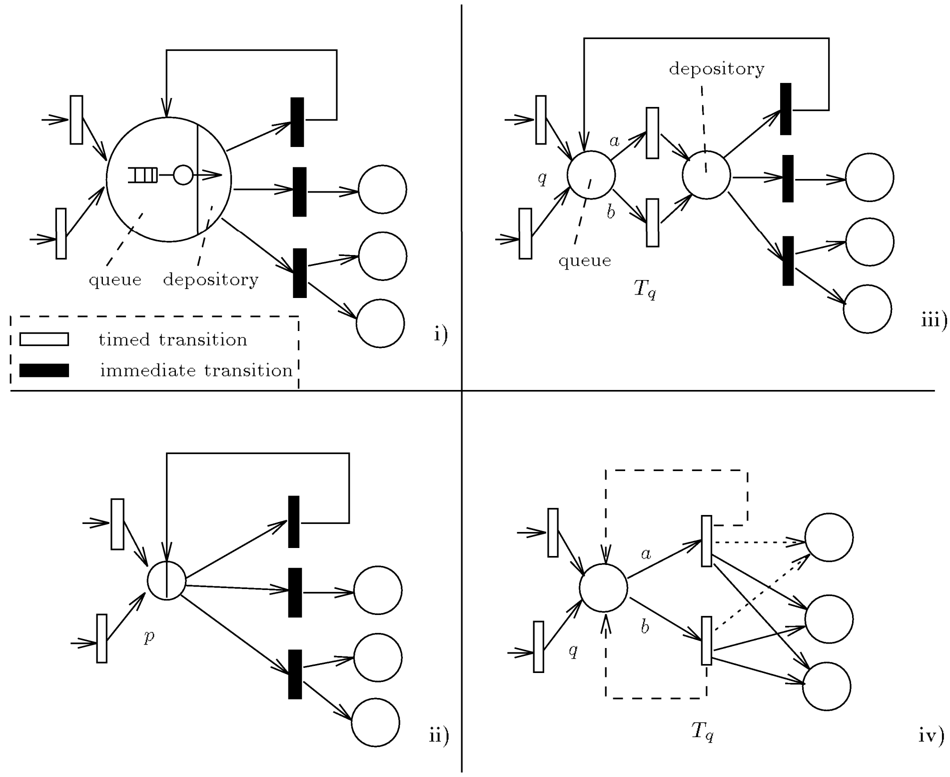

Fig. 1. i) Example queueing place of a QPN, ii) its pictorial representation, iii) its simplified interpretation, and iv) its SPN-skeleton

– is a weight $\in \mathbb{R}^+$ specifying the relative firing frequency of t_i if $t_i \in T_2$.

The QPNs for which we establish a product form solution are restricted as follows:

Definition 1. A *product form QPN (PQPN)* is a QPN with $P_1 \cup T_1 \neq \emptyset$, satisfying the following restrictions:

1. All transitions are uncoloured.
2. The input bags of immediate transitions are disjoint or equal and do not intersect with the input bags of timed transitions, i.e.

$$\forall s, t \in T_2 : I(s) \cap I(t) = \emptyset \text{ or } I(s) = I(t) \quad \text{and} \quad \forall s \in T_1, t \in T_2 : I(s) \cap I(t) = \emptyset$$

3. The relative firing frequencies of immediate transitions do not depend on the marking of the QPN.
4. Output places of immediate transitions are not input places of immediate transitions, i.e. $\forall s, t \in T_2 : O(s) \cap I(t) = \emptyset$.
5. No two timed transitions have the same input bag, i.e. $I(t) \neq I(s), \forall s, t \in T_1$.
6. The input bag of each transition is the output bag of some other transition and vice versa, i.e. $\forall t \in T, \exists s \in T : I(t) = O(s)$ and $\forall t \in T, \exists s \in T : O(t) = I(s)$.
7. The marking dependent firing rate of each timed transition $t \in T_1$ can be expressed as $r(M,t) = \frac{\varphi(M - I(t))\chi(t)}{\Phi(M)}$, where φ, Φ and χ are arbitrary non-negative functions.
8. Input transitions of queueing places have no further output places, output transitions of queueing places have no further input places and only single tokens arrive at and leave from a queue[1], i.e.
   (a) $\forall p \in P_1, t \in T : \sum_{c \in C(p)} I(t)(p)(c) \leq 1$ and $\sum_{c \in C(p)} O(t)(p)(c) \leq 1$
   where $C(p)$ denotes the set of colours of place p.

[1] Note that in our approach customers of one queue do not necessarily correspond to customers in a different queue. In particular, queues serve input bags of transitions which, for simplicity, have been transformed to a single token.

(b) $\forall p \in P_1, t \in T$:
$$O(t)(p) \neq 0 \Longrightarrow O(t)(p') = 0, \forall p' \in P \setminus \{p\}$$
$$I(t)(p) \neq 0 \Longrightarrow I(t)(p') = 0, \forall p' \in P \setminus \{p\}$$

9. All queues in queueing places $p \in P_1$ are of PQN-type (cf. Sec. 3).

Restrictions 2-4 allow us to eliminate immediate transitions on the net level by defining probabilistic output bags for timed transitions whose output bag coincides with the input bag of immediate transitions. Restrictions 5-7 are the ones for PSPNs (see Sec. 2).

Each queueing place can be replaced by its simplified interpretation (see. Fig. 1 iii), where timed transitions ($\in T_q$) connect the place q representing the former queue and the depository. Each such timed transition corresponds to a distinct colour of its input place modelling the service of tokens of that colour. The (marking dependent) rates of those transitions are the average departure rates for tokens of the corresponding colour in the queue given by (8). If, e.g., the queue in Fig. 1 serves two customer classes (token colours) a and b, then the set T_q for that queue comprises two uncoloured transitions: one for a and one for b. Substituting all queueing places by their simplified interpretation yields a CGSPN where the marking of q corresponds to a marginal state of the former queue.

Since restriction 8 holds and since the original QPN is a PQPN, this CGSPN satisfies restrictions 1-7. After eliminating immediate transitions by defining probabilistic output bags for timed transitions, we arrive at a net, called the *SPN-skeleton* of the QPN.

Definition 2. Given a PQPN and let $CGSPN = ((P, T, C, I, O), T_1, T_2, W)$ be obtained by the simplified representation of all queueing places. Then the *SPN-skeleton* of the PQPN is defined as $CSPN = ((\tilde{P}, \tilde{T}, \tilde{C}, \tilde{I}, \tilde{O}), \tilde{W})$ where

* $\tilde{P} = P \setminus \{p \in P | \exists t \in T_2 : I(t)(p) \neq 0\}$
* $\tilde{T} = T \setminus T_2$
* $\tilde{C} = C$
* $\tilde{I}(t) = I(t), \forall t \in \tilde{T}$
* The function $\tilde{O}$ now denotes several (probabilistic) output bags for each transition $t \in \tilde{T}$:
$$\tilde{O}(t) = \begin{cases} \{O(s) | s \in T_2 : O(t) = I(s)\} & \text{if } \exists t' \in T_2 : O(t) = I(t') \\ \{O(t)\} & \text{otherwise} \end{cases}$$
* $\tilde{W} = (\tilde{w}_1, \ldots, \tilde{w}_{|\tilde{T}|})$ where $\tilde{w}_i = w_i$ if $t_i \in T_1$.

The probabilities $p(s, t)$ for choosing a specific output bag of a transition $s \in \tilde{T}$ which corresponds to the input bag of $t \in \tilde{T}$ are given by

$$p(s,t) = \begin{cases} \dfrac{w_j}{\displaystyle\sum_{k : t_k \in \{t \in T_2 | I(t) = I(t_j)\}} w_k} & \text{if } \exists t_j \in T_2 : O(s) = I(t_j) \text{ and } O(t_j) = I(t) \\ 1.0 & \text{otherwise} \end{cases}$$

Of course, $p(s, t) = 0$ if $\tilde{I}(t) \notin \tilde{O}(s)$. Obviously the SPN-skeleton satisfies the conditions for product form SPNs (cf. Sec. 2).

Fig. 1 iv) shows an example of a part of a SPN-skeleton where immediate transitions have been removed (cf. Fig. 1 iii). The various probabilistic output bags are here depicted by different type of lines.

During construction of a PQPN's SPN-skeleton we have eliminated all immediate transitions. Since markings of a PQPN, in which immediate transitions are enabled, are vanishing (cf. [1, 5]), they are of no interest for a steady state analysis. Thus we can also employ the SPN-skeleton for specifying the tangible markings of the PQPN: Restrictions 2 and 6 imply a zero probability for a non-empty marking of an input place/depository of an immediate transition. Therefore we do not have to consider such components in the state descriptor and can define the state of a PQPN directly on the basis of its SPN-skeleton. Consider a place q of the SPN-skeleton obtained by the simplified interpretation of a queueing place (cf. Fig. 1 iv) and let P_Q denote the set of such places. For each $q \in P_Q$ let T_q denote the set of timed transitions of the simplified interpretation of that queue, modelling the service of tokens in this queue (cf. Fig. 1 iii and iv) and let T be the set of transitions of the QPN's SPN-skeleton. Together with the transitions from T_q, place q represents the behaviour of the queue if we assume that the marking of q is described by the detailed state of that queue. For notational convenience suppose that the first $|P_1|$ places are places representing the former queues. The state of a PQPN is given by a pair (x, M) where $x = (x_1, \ldots, x_{|P_1|})$ and x_i is the detailed state of the queue in queueing place p_i (cf. Sec. 3 and appendix A) and M is the marking of the corresponding SPN-skeleton. The marking $M(q)$ of a place in P_Q, representing the former queue, denotes the marginal state of a detailed state x_q, thus the pair (x, M) contains some redundant information. As in Sect. 3 we assume that the token colour uniquely determines the transition, place or queue being considered. This assumption simplifies the notation of the global balance equations given in the following.

Theorem 3. *Let QPN be a PQPN and f, g functions for the SPN-skeleton satisfying (2) and (3) resp. Then the steady state distribution of the QPN is given by*

$$\pi(x, M) = \pi((x_1, \ldots, x_{|P_1|}), M) = \frac{1}{G} g(M) \Phi(M) \prod_{p \in P_Q} S_p(x_p) \tag{9}$$

where S_p is given by (7) and G is a normalisation constant.

Proof. Define the set of a queue's input transitions by $T_i^q = \{s \in T | \exists t \in T_q : p(s,t) > 0\}$ and its set of output transitions by $T_o^q = \{s \in T | \exists t \in T_q : p(t,s) > 0\}$. Since each transition uniquely corresponds to a token colour/customer class, we will use transitions to also denote such token colours/customer classes.
Define $T_Q = \cup_{q \in P_Q} T_q, T_I = \cup_{q \in P_Q} T_i^q$ and $T_O = \cup_{q \in P_Q} T_o^q$.
Furthermore we use the following convention to denote the states of a queue:

$x - (s, l)$ is the state just before the "arrival" of a token of colour s in phase l, leading to state x.
$x + (s, l)$ is the state just before the "departure" of a token of colour s in phase l, leading to state x.

E.g. if in state $x + (s, l) - (s, l+1)$ a token of colour s leaves its l-th phase and enters the next phase, we obtain state x and, e.g., if $s \in C(p_j), p_j \in P_Q$, then $x - (s, l) = (x_1, \ldots, x_{j-1}, x_j - (s, l), x_{j+1}, \ldots, x_{|P_1|})$ and $x_j - (s, l)$ is the state of the queue in p_j just before "arrival" of a token of colour s in phase l, leading to state x_j of the queue in p_j. For simplifying notation also define
$\lambda_{ph}(x, s, l) = \begin{cases} \lambda_{ph}(x_q, s, l) & \text{if } s \in C(q) \\ 0 & \text{otherwise} \end{cases}$ and $\lambda_{ex}(x, s, l) = \begin{cases} \lambda_{ex}(x_q, s, l) & \text{if } s \in C(q) \\ 0 & \text{otherwise} \end{cases}$
With that the global balance equations of the PQPN can be written as

$$\pi(x,M)\ (\sum_{t\in T\setminus T_Q} r(M,t) + \sum_{s\in T_Q}\sum_{l=1}^{u_s}(\lambda_{ph}(x,s,l)+\lambda_{ex}(x,s,l)))$$

$$= \sum_{t\in T\setminus(T_O\cup T_Q)}\ \sum_{s\in T\setminus T_Q} \pi(x,M-I(t)+I(s))r(M-I(t)+I(s),s)p(s,t)$$

$$+ \sum_{t\in T_O\setminus T_Q} (\sum_{s\in T\setminus T_Q} \pi(x,M-I(t)+I(s))r(M-I(t)+I(s),s)p(s,t) + \sum_{s\in T_Q}\sum_{l=1}^{u_s}\pi(x+(s,l),M-I(t)+I(s))\lambda_{ex}(x+(s,l),s,l)p(s,t))$$

$$+ \sum_{s\in T_Q}(\sum_{t\in T_I\setminus T_Q} \pi(x-(s,1),M-I(s)+I(t))r(M-I(s)+I(t),t)p(t,s) + \sum_{t\in T_I\cap T_Q}\sum_{l=1}^{u_t}\pi(x+(t,l)-(s,1),M-I(s)+I(t))\lambda_{ex}(x+(t,l)-(s,1),t,l)p(t,s))$$

$$+ \sum_{s\in T_Q}\sum_{l=1}^{u_s-1}\pi(x+(s,l)-(s,l+1),M)\lambda_{ph}(x+(s,l)-(s,l+1),s,l)$$

Using (9) the global balance equations all reduce to the defining equations for function f (cf. (2)). More details are given in appendix C. □

Obviously PQNs and PSPNs are special cases of PQPNs. For PQNs, function f is equivalent to the visit ratios and (2) are the well-known traffic equations.

5 The arrival theorem for PQPNs

In PQNs a well-known arrival theorem is the basis for algorithms calculating important performance measures, like e.g. MVA [15], starting with an empty net and successively increasing the customer population up to a desired value.

The arrival theorem states that the steady state probability for a state of the net upon arrival of a customer is the same as the steady state probability for the corresponding state without this customer. The next theorem shows that this relationship holds provided $\Phi = \varphi$.

Denote by $RS(QPN, M_0)$ the reachability set of the PQPN and by $RS(SPN, M_0)$ the reachability set of the corresponding SPN-skeleton.

Theorem 4. *Given a PQPN with $\Phi = \varphi$. Let $p \in P_Q$ and $s \in T_p$. Assume that a function f satisfying (2), that a function g for $RS(SPN, M_0)$ and a function g' for $RS(SPN, M_0 + I(s))$ both satisfying (3) exist for the QPN's SPN-skeleton. Then*

$$PA_s[x+(s,1), M+I(s)] = \pi(x,M)$$

where $PA_s[y,M']$ denotes the probability of being in state (y,M') immediately after arrival of an input bag $I(s)$ at place p.

Proof. We have to show that the flow according to the arrival of an input bag $I(s)$ leading to state $(x+(s,1), M+I(s))$ divided by the sum of all flows concerning the arrival of an input bag $I(s)$ equals $\pi(x,M)$.

Employing the corresponding part of the global balance equations of theorem 3 (see also appendix C), the arrival theorem holds if

$$\begin{aligned}
&[\sum_{t \in T_I \setminus T_Q} \tilde{\pi}(x, M + I(t)) r(M + I(t), t) p(t, s) + \\
&\sum_{t \in T_I \cap T_Q} \sum_{l=1}^{u_t} \tilde{\pi}(x + (t, l), M + I(t)) \lambda_{ex}(x + (t, l), t, l) p(t, s)] \\
/ &[\sum_{(x', M') \in RS(QPN, M_0)} (\sum_{t \in T_I \setminus T_Q} \tilde{\pi}(x', M' + I(t)) r(M' + I(t), t) p(t, s) + \\
&\sum_{t \in T_I \cap T_Q} \sum_{l=1}^{u_t} \tilde{\pi}(x' + (t, l), M' + I(t)) \lambda_{ex}(x + (t, l), t, l) p(t, s))] \\
= &\pi(x, M)
\end{aligned} \tag{10}$$

where the left part denotes the probability PA_s and $\tilde{\pi}$ is the steady state distribution for the PQPN with the additional customer/token of class/colour s.

For a marking $M + I(t)$ with $M \in RS(SPN, M_0)$ we have

$$g'(M + I(t)) = Cf(t)g(M) \tag{11}$$

for some constant $C \in \mathbb{R}^+$, because (3) defines a function up to a constant. Let $S(x) = \prod_{p \in P_Q} S_p(x_p)$

Since local balance holds and the terms of the sums (22) and (26) are equal (cf. appendix C) and for all queues of QPN-type

$$S(x + (s, 1))(\lambda_{ph}(x + (s, 1), s, 1) + \lambda_{ex}(x + (s, 1), s, 1)) = S(x) r(M + I(s), s)$$

the left part of (10) is equal to

$$\begin{aligned}
&\frac{S(x) r(M + I(s), s) f(s) g(M) \Phi(M + I(s))}{\sum_{(x', M') \in RS(QPN, M_0)} S(x') r(M' + I(s), s) f(s) g(M') \Phi(M' + I(s))} \\
= &\frac{S(x) g(M) \varphi(M) \chi(s)}{\sum_{(x', M') \in RS(QPN, M_0)} S(x') g(M') \varphi(M') \chi(s)} \quad (\text{“}q\Phi = \varphi\chi\text{“}) \\
= &\frac{S(x) g(M) \Phi(M)}{\sum_{(x', M') \in RS(QPN, M_0)} S(x') g(M') \Phi(M')} \quad (\text{“}\varphi = \Phi\text{“}) \\
= &\pi(x, M)
\end{aligned}$$

which completes the proof. □

A similar theorem also holds for the transitions of the PQPN at the arrival instant of an input bag.

Theorem 5. *Given a PQPN with $\Phi = \varphi$. Let $s \notin T_Q$. Assume that a function f satisfying (2), that a function g for $RS(SPN, M_0)$ and a function g' for $RS(SPN, M_0 + I(s))$ both satisfying (3) exist for the QPN's SPN-skeleton. Then*

$$PA_s[x, M + I(s)] = \pi(x, M)$$

where $PA_s[x, M + I(s)]$ now denotes the probability of being in state $(x, M + I(s))$ immediately after arrival of an input bag $I(s)$ at the input places of s.

Proof. The proof is similar to the proof of theorem 4. □

The restriction $\Phi = \varphi$ is, e.g., satisfied in PQNs. We will meet this restriction also in the next section concerning exact aggregation in PQPNs.

6 Aggregation in PQPNs

An aggregation technique allowing the substitution of some queues of a PQN by a single flow equivalent substitute queue has been introduced in [7] and extended to multiple classes and more general net structures in [2, 3, 16]. The results of the aggregation approach are exact in PQNs, but there is no gain concerning solution effort if only one solution run is compared. Nevertheless, aggregation is very important since it can be used in parametric studies, where several solution runs are performed modifying one or several parameters in the model. In this case queues with parameters that are constant in all runs can be aggregated and the parametric study performed with the smaller aggregated net. Another important area is the development of approximation algorithms for non product form QNs which are often based on aggregation approaches, first examples are given in [2].

We extend the approaches from [2, 3, 16] for PQPNs and also for pure PSPNs. However, for this purpose we have to restrict the model class appropriately since aggregation implicitly introduces certain independence assumptions which are observed in PQNs but not generally in PSPNs. We consider the decomposition of a PSPN into two subnets, the aggregation of subnets and the analysis of an aggregated net. Since it has been shown that the analysis of a PQPN can be decomposed into the analysis of the underlying SPN-skeleton and the isolated computation of the detailed state probabilities of queues it is sufficient to present the approach for PSPNs, the results obviously hold if queues are included.

We start with the SPN-skeleton of a PQPN and assume that the conditions for a product form solution are observed. Let $T_{\mathcal{A}}$ be a subset of the set of skeleton transitions T and let $T_{\mathcal{R}} = T \setminus T_{\mathcal{A}}$. The set of transitions $T_{\mathcal{A}}$ forms the subnet $\mathcal{A}$ to be aggregated, the set $T_{\mathcal{R}}$ describes the rest of the net which forms the environment of $\mathcal{A}$ and is also a subnet (which can also be aggregated). The following conditions have to be observed for aggregation:

1. The input bags of subnet and environment have to be disjoint, i.e.,

   $$\cup_{t \in T_{\mathcal{A}}} I(t) \cap \cup_{t \in T_{\mathcal{R}}} I(t) = \emptyset$$

   A marking M of the complete net can be decomposed into a part $M_{\mathcal{A}}$ including those tokens which are in the input bag for some $t \in T_{\mathcal{A}}$ and $M_{\mathcal{R}}$ including those tokens which are in the input bag for some $t \in T_{\mathcal{R}}$. Assume that places are reordered such that all places of $\mathcal{A}$ occur first, thus $M = (M_{\mathcal{A}}, M_{\mathcal{R}})$. We denote $M_{\mathcal{A}}$ as the local marking of $\mathcal{A}$ and $M_{\mathcal{R}}$ as the non-local part of a marking.
2. Rates of transitions from $T_{\mathcal{A}}$ ($T_{\mathcal{R}}$) are allowed to depend only on $M_{\mathcal{A}}$ ($M_{\mathcal{R}}$), i.e.,

   $$r(M,t) = \begin{cases} r((M_{\mathcal{A}}, 0), t) \text{ for } t \in T_{\mathcal{A}} \\ \\ r((0, M_{\mathcal{R}}), t) \text{ for } t \in T_{\mathcal{R}} \end{cases}$$

   where 0 denotes the empty marking.

Let $RS(SPN, M_0)$ be the reachability set of the PSPN. We define

$$D_A(M_{\mathcal{R}}) = \{M_{\mathcal{A}} | (M_{\mathcal{A}}, M_{\mathcal{R}}) \in RS(SPN, M_0)\} \text{ and } D_{\mathcal{R}}(M_{\mathcal{A}}) = \{M_{\mathcal{R}} | (M_{\mathcal{A}}, M_{\mathcal{R}}) \in RS(SPN, M_0)\}$$

i.e., the set of all markings which are reachable in $\mathcal{A}$ when the marking of the rest is $M_{\mathcal{R}}$ and the set of all markings in $\mathcal{R}$ which are reachable when $M_{\mathcal{A}}$ is the marking of $\mathcal{A}$. The independence of transition rates from non-local markings implies $\Phi = \varphi$ and the following decomposition of the function Φ

$$\Phi((M_{\mathcal{A}}, M_{\mathcal{R}})) = \Phi_{\mathcal{A}}(M_{\mathcal{A}}) \Phi_{\mathcal{R}}(M_{\mathcal{R}})$$

$\Phi_\mathcal{A}$ and $\Phi_\mathcal{R}$ are functions defined on the markings of the subnets $\mathcal{A}$ and $\mathcal{R}$ and are used to determine the firing rates for transitions in a subnet.

$$r((M_\mathcal{A}, M_\mathcal{R}), t) = \begin{cases} \frac{\Phi_\mathcal{A}(M_\mathcal{A} - I(t))}{\Phi_\mathcal{A}(M_\mathcal{A})}\chi(t) \text{ for } t \in T_\mathcal{A} \\ \\ \frac{\Phi_\mathcal{R}(M_\mathcal{R} - I(t))}{\Phi_\mathcal{R}(M_\mathcal{R})}\chi(t) \text{ for } t \in T_\mathcal{R} \end{cases} \tag{12}$$

Like for the analysis of a PSPN as a whole, appropriate functions have to be found for the subnets. The above decomposition of Φ implies that firing rates of a transition in a subnet are independent of the environment. The next step is to decompose the invariant measures of the routing chain and the function g. We define $f_\mathcal{A}(t) = f(t)$ for $t \in T_\mathcal{A}$ and $f_\mathcal{R}(s) = f(s)$ for $s \in T_\mathcal{R}$. Below it will be shown how these measures can be computed locally for $\mathcal{A}$ and $\mathcal{R}$. Now let

$$g(M_\mathcal{A}, M_\mathcal{R}) = g_\mathcal{A}(M_\mathcal{A}) g_\mathcal{R}(M_\mathcal{R})$$

where

$$g_\mathcal{A}(M_\mathcal{A}) = \begin{cases} g((M_\mathcal{A}, 0) - I(t) + I(s))\frac{f_\mathcal{A}(t)}{f_\mathcal{A}(s)} \text{ for } s, t \in T_\mathcal{A} \text{ with } p(s,t) > 0 \\ \\ g((M_\mathcal{A}, 0) - I(t)) f_\mathcal{A}(t) \quad \text{for } t \in T_\mathcal{A} \text{ and some } s \in T_\mathcal{R} \text{ with } p(s,t) > 0 \\ \\ g((M_\mathcal{A}, 0) + I(s))\frac{1}{f_\mathcal{A}(s)} \quad \text{for } s \in T_\mathcal{A} \text{ and some } t \in T_\mathcal{R} \text{ with } p(s,t) > 0 \end{cases} \tag{13}$$

and $g_\mathcal{R}(M_\mathcal{R})$ defined accordingly.

The values can be computed recursively starting with $g(M_0) = g((M_{0,\mathcal{A}}, M_{0,\mathcal{R}})) = g_\mathcal{A}(M_{0,\mathcal{A}}) = g_\mathcal{R}(M_{0,\mathcal{R}}) = 1.0$ and computing all possible successors $M + I(s) - I(t)$ with $p(s,t) > 0$ (see [12]). With this decomposition the stationary solution can be expressed as

$$\pi((M_\mathcal{A}, M_\mathcal{R})) = \frac{1}{G} g_\mathcal{R}(M_\mathcal{R}) \Phi_\mathcal{R}(M_R) g_\mathcal{A}(M_\mathcal{R}) \Phi_\mathcal{A}(M_\mathcal{A})$$

Theorem 6. *Let* $\pi_\mathcal{R}(M_\mathcal{R}) = \sum_{M_\mathcal{A} \in D_\mathcal{A}(M_\mathcal{R})} \pi((M_A, M_R))$ *be the marginal distribution of the net neglecting the detailed marking of subnet* $\mathcal{A}$. $\pi_\mathcal{R}(M_\mathcal{R})$ *can be computed as*

$$\pi_\mathcal{R}(M_\mathcal{R}) = \tfrac{1}{G} g_\mathcal{R}(M_\mathcal{R}) \Phi_\mathcal{R}(M_\mathcal{R}) \Delta_\mathcal{A}(M_\mathcal{R})$$

$$\text{where } \Delta_\mathcal{A}(M_\mathcal{R}) = \sum_{M_\mathcal{A} \in D_\mathcal{A}(M_\mathcal{R})} g_\mathcal{A}(M_\mathcal{A}) \Phi_\mathcal{A}(M_\mathcal{A})$$

Proof. The proof is straightforward substituting the definition of $\Delta_\mathcal{A}(M_\mathcal{R})$ into the equation. □

Since $\Delta_\mathcal{A}(M_\mathcal{R})$ is independent of $r(M,t)$ for $t \in T_\mathcal{R}$, the value can be computed once and used afterwards for parametric studies modifying transition rates in $\mathcal{R}$. However, the above result is not really based on aggregation, since $\mathcal{A}$ is not analysed in isolation. For an isolated analysis we start with the computation of the invariant measures of the routing subchains $f_\mathcal{A}(t)$ ($t \in T_\mathcal{A}$). If a subchain V_r visits only $\mathcal{A}$, then $f_\mathcal{A}(s)$ can be computed for all transitions s in this subchain using only information about subnet $\mathcal{A}$. In the same way we can compute $f_\mathcal{R}(t)$ for transitions from subchains which visit only $\mathcal{R}$ using only information about subnet $\mathcal{R}$. Thus it remains to consider subchains V_r including transitions from $T_\mathcal{A}$ and $T_\mathcal{R}$. We define the following matrices: $H^r_\mathcal{A}$ including all probabilities $p(s,t)$ for $s, t \in T_\mathcal{A} \cap V_r$, $H^r_{\mathcal{A},\mathcal{R}}$ including probabilities $p(s,t)$ for $s \in T_\mathcal{A} \cap V_r$, $t \in T_\mathcal{R} \cap V_r$, $H^r_\mathcal{R}$ including probabilities $p(s,t)$ for $s, t \in T_\mathcal{R} \cap V_r$ and $H^r_{\mathcal{R},\mathcal{A}}$ including probabilities $p(s,t)$ for $s \in T_\mathcal{R} \cap V_r$, $t \in T_\mathcal{A} \cap V_r$.

Let $x^r_{\mathcal{A}}$ be a vector of invariant measures for $t \in T_{\mathcal{A}} \cap V_r$ (cf. (1)). The vector is computed up to a constant as the solution of

$$x^r_{\mathcal{A}} = x^r_{\mathcal{A}}(H^r_{\mathcal{A}} + H^r_{\mathcal{A},\mathcal{R}}(I - H^r_{\mathcal{R}})^{-1} H^r_{\mathcal{R},\mathcal{A}}) = x^r_{\mathcal{A}} \bar{H}^r_{\mathcal{A}} \ .$$

It is easy to show that the vector $x^r_{\mathcal{A}}$ equals the corresponding subvector of the invariant measure x as computed in equation (1) up to a constant. Thus the values $f_{\mathcal{A}}(t)$ can be determined up to a constant using the above matrices. The computation of $x^r_{\mathcal{A}}$ depends on the rest of the net due to matrices including an index $\mathcal{R}$. There are some situations where the analysis is independent of $\mathcal{R}$, e.g., if $H^r_{\mathcal{R},\mathcal{A}}$ contains one element in each column. The latter case corresponds to PQN submodels with a single input port per routing chain. Thus parametric studies can also include modifications of the probabilities $p(s,t)$ in $\mathcal{R}$, if the above matrix is not modified. The matrix $\bar{H}^r_{\mathcal{A}}$ describes the subnet in "short-circuit" for routing subchain r, which is realised by setting firing rates of transitions from V_r in $\mathcal{R}$ to infinity. We denote by $\bar{p}(s,t)$ the corresponding probabilities, i.e., transition s fires and generates an input bag for t as the next transition with finite rate with probability $\bar{p}(s,t)$, transitions from $T_{\mathcal{R}}$ are skipped in zero time. Accordingly matrices $\bar{H}^r_{\mathcal{R}}$ can be defined for the "short-circuited" environment.

Thus we have for subnet $\mathcal{A}$ the values $f_{\mathcal{A}}(t)$ and $\Phi_{\mathcal{A}}(M_{\mathcal{A}})$. However, the set of reachable markings $M_{\mathcal{A}}$ depends on $RS(SPN, M_0)$, the reachability set of the complete net. In PQNs aggregation of a subnet can be performed by short-circuiting the subnet assuming zero service times for all queues of the environment, analysing the short-circuited subnet for all possible population vectors and substituting the subnet by a flow equivalent service center. A similar approach will now be developed for PQPNs. In contrast to PQNs, routing chains in PQPNs are not in general disjoint. Therefore the population of a PQPN can change and the aggregation approach from PQNs is not directly usable.

In what follows an approach is described which allows to compute the reachability set of a subnet without computing the reachability set of the complete net. We present two different steps. The first consists of finding sets of independent subchains of the routing chain such that the overall reachability set can be expressed as the cross product over subspaces. The second step shows that the set of markings reachable in the subnet is the same for all environment markings in a local reachability set computed from the short-circuited subnet. Using the second step, macro markings can be defined which correspond to the subnet population in PQNs, but might be much more complex in PQPNs or PSPNs since the population is not unique. Afterwards conditional probabilities for micro markings belonging to one macro marking are computed locally. The combination of the conditional probabilities of micro markings in a macro marking together with the probability for the macro marking (as computed from an aggregated model) yields the exact solution of a PQPN. This approach is efficient since we do not have to enumerate the complete reachability set during analysis.

We start with the decomposition of the routing chain into independent subchains. Two subchains $V_{r1}, V_{r2} \in \mathcal{V}$ are disjoint, if for all $t_1 \in V_{r1}, t_2 \in V_{r2} : I(t_1) \cap I(t_2) = \emptyset$. A subset $\mathcal{V}' \subseteq \mathcal{V}$ is closed if each $V_r \in \mathcal{V}'$ and each $V_s \in \mathcal{V} \setminus \mathcal{V}'$ are disjoint and for each pair $V_0, V_r \in \mathcal{V}'$ a sequence $V_1, \ldots, V_{r-1} \in \mathcal{V}'$ exists such that V_j and V_{j+1} $(0 \leq j < r)$ are not disjoint.

Theorem 7. *Let $\mathcal{V}$ be the set of routing subchains of a PSPN, decomposed into closed subsets $\{\mathcal{V}_1, \ldots, \mathcal{V}_R\}$ and let $r(M,t) > 0$ for $M - I(t) \geq 0$. Let M_0 be the initial marking which can be uniquely decomposed into $(M_0^1, \ldots, M_0^R\}$ such that M_0^r includes all places and colours which belong to some $I(t)$ for $t \in \mathcal{V}_r$. Let $RS^r(SPN, M_0^r)$ be the reachability graph of PSPN including only transitions from $\mathcal{V}_r$ with initial marking M_0^r, then*

$$RS(SPN, M_0) = \times_{r=1}^{R} RS^r(SPN, M_0^r) \ .$$

Proof. Since the input bags of transitions from different closed subsets are disjoint, the decomposition of markings according to the subsets is unique. Furthermore, the enabling condition of a transition $t \in \mathcal{V}^r$ depends only on M^r. The firing rate $r(M,t)$ is allowed to depend on the complete marking, however, as assumed $r(M,t) > 0$ whenever t is enabled. Thus firing of transitions in the closed subsets is completely independent yielding the above decomposition of the reachability set. □

Since the SPN-skeleton of a PQPN is a PSPN we can use the above theorem to compute the possible population vectors of the queues; detailed queue states can be determined from this information and from the type and parameters of the queue. In pure PQNs each subchain forms its own closed subset with the consequence that the population in a subchain is constant. In PSPNs the computation of the set $D_{\mathcal{A}}(M_{\mathcal{R}})$ can be decomposed into computations of $D^r_{\mathcal{A}}(M^r_{\mathcal{R}})$. If the closed subset $\mathcal{V}^r$ includes only one subchain, then the behaviour is similar to PQNs with constant population now expressed in terms of input bags for transitions from $\mathcal{V}^r$, therefore the detailed state of $\mathcal{R}$ is not needed for the computation of markings in $\mathcal{A}$, only the population in $\mathcal{R}$ is necessary.

Theorem 7 describes a way to decompose the computation of the reachability set in the computation of a couple of subsets, however, it does not allow the isolated analysis of these subsets. In the sequel we introduce another decomposition of the PQPN which decomposes the analysis of the PQPN into the analysis of isolated subnets and an aggregated net describing interactions between subnets. The key question is, for which markings $M_{\mathcal{R}}, M'_{\mathcal{R}}$ holds $D_{\mathcal{A}}(M_{\mathcal{R}}) = D_{\mathcal{A}}(M'_{\mathcal{R}})$, or, in other words, can we reduce the reachability set of the environment when computing the reachability set of subnet $\mathcal{A}$ (and vice versa).

Due to the input/output relation of the transitions several flows of tokens in a PSPN can be defined without considering the specific marking. For a marking $M_{\mathcal{A}}$ of subnet $\mathcal{A}$ we define $RS(\mathcal{A}, M_{\mathcal{A}})$ as the reachability set of $\mathcal{A}$ where all routing subchains are short-circuited by setting transition rates in $\mathcal{R}$ to infinity and the initial marking is $M_{\mathcal{A}}$. Accordingly $RS(\mathcal{R}, M_{\mathcal{R}})$ is defined for $\mathcal{R}$. We assume in the sequel that $r(M,t) > 0$ for $M - I(t) \geq 0$.

Theorem 8. *For a reachability set $RS(\mathcal{A}, M_{\mathcal{A}})$ the following relations hold:*

1. *$RS(\mathcal{A}, M_{\mathcal{A}})$ is irreducible for each $M_{\mathcal{A}}$.*
2. *If $(M_{\mathcal{A}}, M_{\mathcal{R}}) \in RS(SPN, M_0)$, then $(M'_{\mathcal{A}}, M'_{\mathcal{R}}) \in RS(SPN, M_0)$ for each $M'_{\mathcal{A}} \in RS(\mathcal{A}, M_{\mathcal{A}})$ and each $M'_{\mathcal{R}} \in RS(\mathcal{R}, M_{\mathcal{R}})$.*
3. *$D_{\mathcal{A}}(M_{\mathcal{R}}) = D_{\mathcal{A}}(M'_{\mathcal{R}})$ for $M'_{\mathcal{R}} \in RS(\mathcal{R}, M_{\mathcal{R}})$ and $D_{\mathcal{R}}(M_{\mathcal{A}}) = D_{\mathcal{R}}(M'_{\mathcal{A}})$ for $M'_{\mathcal{A}} \in RS(\mathcal{A}, M_{\mathcal{A}})$.*

Proof.

1. Follows immediately from the irreducibility of the routing subchains, since $f(t) > 0$.
2. Since $M'_{\mathcal{A}} \in RS(\mathcal{A}, M_{\mathcal{A}})$, a sequence of transitions exists in the short-circuited subnet $\mathcal{A}$ such that $M_{\mathcal{A}}$ is transformed into $M'_{\mathcal{A}}$. Since such a sequence exists, there also exists a sequence in the complete net which possibly contains certain transitions from $\mathcal{R}$. However, the firing of these transitions from $\mathcal{R}$ uses only one input bag which is "routed through" $\mathcal{R}$ and afterwards generates a new input bag in $\mathcal{A}$. This sequence of transitions does not depend on the marking of $\mathcal{R}$ and does not influence the marking of $\mathcal{R}$. Thus marking $(M_{\mathcal{A}}, M_{\mathcal{R}})$ can be transformed into $(M'_{\mathcal{A}}, M_{\mathcal{R}})$ which can be transformed into $(M'_{\mathcal{A}}, M'_{\mathcal{R}})$ using the above arguments for $\mathcal{R}$ instead of $\mathcal{A}$.
3. Follows from 2). □

Referring to the short-circuited subnet we can define an aggregation of markings, i.e., $M_{\mathcal{A}}$ and $M'_{\mathcal{A}}$ are aggregated, if $M'_{\mathcal{A}} \in RS(\mathcal{A}, M_{\mathcal{A}})$. Markings of the aggregated reachability set are denoted by small letters and $\Omega_{\mathcal{A}}(m_{\mathcal{A}})$ denotes the set of markings which is represented by $m_{\mathcal{A}}$. The same quantities can be defined for $\mathcal{R}$. Let $rs(agg, (m_{0,\mathcal{A}}, m_{0,\mathcal{R}}))$, with $M_{0,\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{0,\mathcal{A}})$ and $M_{0,\mathcal{R}} \in \Omega_{\mathcal{R}}(m_{0,\mathcal{R}})$, be the aggregated reachability set. For notational convenience we use in the sequel the notations $M_{\mathcal{A}} \pm I(t)$ and $M_{\mathcal{R}} \pm I(s)$ instead of $(M_{\mathcal{A}}, 0) \pm I(t)$ and $(0, M_{\mathcal{R}}) \pm I(s)$ for $t \in T_{\mathcal{A}}$ and $s \in T_{\mathcal{R}}$.

Theorem 9. *Let $M'_{\mathcal{A}} \in RS(\mathcal{A}, M_{\mathcal{A}})$ and $t, t' \in V_r \cap T_{\mathcal{A}}$, then $M'_{\mathcal{A}} + I(t') \in RS(\mathcal{A}, M_{\mathcal{A}} + I(t))$.*

Proof. There exists a sequence of transitions transforming $M_{\mathcal{A}}$ into $M'_{\mathcal{A}}$ without using tokens from the newly arrived input bags $I(t)$ or $I(t')$. Since t and t' belong to the same routing subchain there exists a sequence of transitions transforming input bag $I(t)$ into input bag of $I(t')$ without using any additional token. □

Unfortunately we cannot conclude in general that $M'_{\mathcal{A}} - I(t') \in RS(\mathcal{A}, M_{\mathcal{A}} - I(t))$ for all $M'_{\mathcal{A}} \in RS(\mathcal{A}, M_{\mathcal{A}})$ with $M_{\mathcal{A}} - I(t) \geq 0$ and $M'_{\mathcal{A}} - I(t) \geq 0$. In the sequel of this section we restrict the class of PSPNs such that

$$M'_{\mathcal{A}} \in RS(\mathcal{A}, M_{\mathcal{A}}) \Longrightarrow M'_{\mathcal{A}} - I(t') \in RS(\mathcal{A}, M_{\mathcal{A}} - I(t))$$

for $M_{\mathcal{A}} - I(t) \geq 0$ and $M'_{\mathcal{A}} - I(t') \geq 0$ and $t, t' \in V_r \cap T_{\mathcal{A}}$. This assumption implies that we can remove from a given marking $M_{\mathcal{A}}$ an input bag for some transition from a fixed routing chain and the resulting aggregated marking is independent from the specific transition. The assumption helps to compute the aggregated reachability set, although the aggregation approach presented below does not really rely on it. We define the following notation

$$\Omega_{\mathcal{A}}(m_{\mathcal{A}} + I(t^r_{\mathcal{A}})) = RS(\mathcal{A}, M_{\mathcal{A}} + I(t))$$

for some $M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})$ and $t \in T_{\mathcal{A}} \cap V_r$. In a similar way $\Omega_{\mathcal{A}}(m_{\mathcal{A}} - I(t^r_{\mathcal{A}}))$ and the corresponding quantities for $\mathcal{R}$ can be defined. Since the reachability sets $RS(\mathcal{A}, M_{\mathcal{A}})$ and $RS(\mathcal{R}, M_{\mathcal{R}})$ are irreducible, the same sets are generated for each $M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})$ and $M_{\mathcal{R}} \in \Omega_{\mathcal{R}}(m_{\mathcal{R}})$ and each $t, t' \in V_r$. Under the above conditions the reachability graph of the complete PSPN can be expressed as

$$RS(SPN, M_0) = \bigcup_{(m_{\mathcal{A}}, m_{\mathcal{R}}) \in rs(agg, m_0)} \Omega_{\mathcal{A}}(m_{\mathcal{A}}) \times \Omega_{\mathcal{R}}(m_{\mathcal{R}})$$

The aggregative analysis is performed in two steps. First, the subnets $\mathcal{A}$ and $\mathcal{R}$ are analysed in isolation for each $\Omega_{\mathcal{A}}(m_{\mathcal{A}})$ and $\Omega_{\mathcal{R}}(m_{\mathcal{R}})$ by assuming infinite transition rates for transitions that are not in the subnet to be analysed (short-circuit analysis). Subsequently, an aggregated model is analysed to determine the distribution among aggregated markings.

We start with the short-circuit pre-analysis of a subnet and show here the results for subnet $\mathcal{A}$. Of course, similar results hold for $\mathcal{R}$. Assume that short-circuit analysis is performed for macro marking $m_{\mathcal{A}}$ and let $M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})$ be one micro marking included in the macro marking which serves as initial marking. The short-circuited subnet is characterised by the functions $\Phi_{\mathcal{A}}$ and χ to express transition rates, $f_{\mathcal{A}}$ calculated from the invariant measure of the routing chain in the short-circuited subnet and $\bar{p}(s, t)$ describing the probability that a firing of transition s generates an input bag for t in the short-circuited subnet. Subnet $\mathcal{A}$ analysed in short-circuit yields a function $\bar{g}_{\mathcal{A}}$ recursively computed as

$$\bar{g}_{\mathcal{A}}(M_{\mathcal{A}} + I(s) - I(t)) = \bar{g}_{\mathcal{A}}(M_{\mathcal{A}}) \frac{f(s)}{f(t)} \tag{14}$$

for $s, t \in T_{\mathcal{A}}$ and $\bar{p}(s, t)$. The above relation allows to compute $\bar{g}_{\mathcal{A}}$ uniquely up to a constant. The next theorem relates $\bar{g}_{\mathcal{A}}$ and $g_{\mathcal{A}}$.

Theorem 10. *Let $\bar{g}_{\mathcal{A}}$ be computed from (14) and let $g_{\mathcal{A}}$ be computed via (13), then*

$$\frac{\bar{g}_{\mathcal{A}}(M_{\mathcal{A}})}{\bar{g}_{\mathcal{A}}(M'_{\mathcal{A}})} = \frac{g_{\mathcal{A}}(M_{\mathcal{A}})}{g_{\mathcal{A}}(M'_{\mathcal{A}})} \text{ for all } M'_{\mathcal{A}} \in RS(\mathcal{A}, M_{\mathcal{A}}).$$

Proof. Assume that $g_{\mathcal{A}}$ is known, then we can choose $\bar{g}_{\mathcal{A}}(M_{\mathcal{A}}) = g_{\mathcal{A}}(M_{\mathcal{A}})$. Now consider $M'_{\mathcal{A}} = M_{\mathcal{A}} + I(s) - I(t)$ for $s, t \in T_{\mathcal{A}}$.

For $p(s,t) > 0$ we have

$$g_{\mathcal{A}}(M'_{\mathcal{A}}) = g_{\mathcal{A}}(M_{\mathcal{A}})\frac{f(s)}{f(t)} = \bar{g}_{\mathcal{A}}(M_{\mathcal{A}})\frac{f(s)}{f(t)} = \bar{g}_{\mathcal{A}}(M'_{\mathcal{A}}) \ .$$

Now let $p(s,t) = 0$ and $\bar{p}(s,t) > 0$. A sequence of transitions $t_1, \ldots, t_n \in T_{\mathcal{R}}$ exists with $p(s,t_1) > 0$, $p(t_i, t_{i+1}) > 0$ and $p(t_n, t) > 0$ which transforms $M_{\mathcal{A}}$ into $M'_{\mathcal{A}}$ and leaves the marking of $\mathcal{R}$ unchanged. This sequence yields

$$g_{\mathcal{A}}(M'_{\mathcal{A}}) = g_{\mathcal{A}}(M_{\mathcal{A}})\frac{f(s)}{f(t_1)}\frac{f(t_1)}{f(t_2)}\cdots\frac{f(t_{n-1})}{f(t_n)}\frac{f(t_n)}{f(t)} = \bar{g}_{\mathcal{A}}(M_{\mathcal{A}})\frac{f(s)}{f(t)} = \bar{g}_{\mathcal{A}}(M'_{\mathcal{A}}) \ .$$

By induction we can conclude that the result holds for all $M'_{\mathcal{A}} \in RS(\mathcal{A}, M_{\mathcal{A}})$. □

The proof of the above theorem also shows that $\bar{g}_{\mathcal{A}}$ exists whenever $g_{\mathcal{A}}$ exists. Similarly we can compute $\bar{g}_{\mathcal{R}}(M_{\mathcal{R}})$. The steady state distribution of micro markings in a macro marking in a short-circuited subnet is given by

$$\bar{\pi}_{\mathcal{A}}(M_{\mathcal{A}}) = \frac{\bar{g}_{\mathcal{A}}(M_{\mathcal{A}})\Phi_{\mathcal{A}}(M_{\mathcal{A}})}{\bar{G}_{\mathcal{A}}(m_{\mathcal{A}})} \qquad \bar{\pi}_{\mathcal{R}}(M_{\mathcal{R}}) = \frac{\bar{g}_{\mathcal{R}}(M_{\mathcal{R}})\Phi_{\mathcal{R}}(M_{\mathcal{R}})}{\bar{G}_{\mathcal{R}}(m_{\mathcal{R}})} \tag{15}$$

where $M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})$, $M_{\mathcal{R}} \in \Omega_{\mathcal{R}}(m_{\mathcal{R}})$ and $\bar{G}_{\mathcal{A}}(m_{\mathcal{A}})$, $\bar{G}_{\mathcal{R}}(m_{\mathcal{R}})$ are appropriate normalisation constants.

Since theorem 10 holds, the above values also represent the conditional probabilities of markings in a subset of markings in the complete net, i.e.,

$$\bar{\pi}(M_{\mathcal{A}}) = \frac{\sum_{M'_{\mathcal{R}} \in \Omega_{\mathcal{R}}(m_{\mathcal{R}})} \pi((M_{\mathcal{A}}, M'_{\mathcal{R}}))}{\sum_{M'_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})} \sum_{M'_{\mathcal{R}} \in \Omega_{\mathcal{R}}(m_{\mathcal{R}})} \pi((M'_{\mathcal{A}}, M'_{\mathcal{R}}))}$$

for some $(m_{\mathcal{A}}, m_{\mathcal{R}}) \in rs(agg, m_0)$, $M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})$ and $M_{\mathcal{R}} \in \Omega_{\mathcal{A}}(m_{\mathcal{R}})$.

We have not defined a representation for macro markings $(m_{\mathcal{A}}, m_{\mathcal{R}})$. In contrast to PQNs, where macro states include the population in a subnet, there might be no straightforward interpretation of such a marking in a PSPN, although $(m_{\mathcal{A}}, m_{\mathcal{R}})$ can usually be interpreted in terms of input bags for transitions from the subchains. Under the restrictions made, transitions on the aggregated reachability set are well defined and a specification of the behaviour of a net to compute the steady state distribution of the macro markings can be given. For each routing subchain V_r which visits $\mathcal{A}$ and $\mathcal{R}$ we define transitions $t^r_{\mathcal{A}}$ and $t^r_{\mathcal{R}}$ such that

* $I(t^r_{\mathcal{A}}) = O(t^r_{\mathcal{R}})$ and $I(t^r_{\mathcal{R}}) = O(t^r_{\mathcal{A}})$.
* $p(t^r_{\mathcal{A}}, t^r_{\mathcal{R}}) = p(t^r_{\mathcal{R}}, t^r_{\mathcal{A}}) = 1$.
* $m_{\mathcal{A}} \pm I(t^r_{\mathcal{A}})) = m'_{\mathcal{A}}$ for $M_{\mathcal{A}} \pm I(t) = M'_{\mathcal{A}}$, where $M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})$, $M'_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m'_{\mathcal{A}})$ and $t \in T_{\mathcal{A}} \cap V_r$ and $m_{\mathcal{R}} \pm I(t^r_{\mathcal{R}})) = m'_{\mathcal{R}}$ for $M_{\mathcal{R}} \pm I(t) = M'_{\mathcal{R}}$, where $M_{\mathcal{R}} \in \Omega_{\mathcal{R}}(m_{\mathcal{R}})$, $M'_{\mathcal{R}} \in \Omega_{\mathcal{R}}(m'_{\mathcal{R}})$ and $t \in T_{\mathcal{R}} \cap V_r$.
* The marking dependent rates are expressed as

$$\begin{aligned} r(m_{\mathcal{A}}, t^r_{\mathcal{A}}) &= \sum_{M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})} \bar{\pi}(M_{\mathcal{A}}) \sum_{t \in T_{\mathcal{A}} \cap V_r} r(M_{\mathcal{A}}, t) \sum_{s \in T_{\mathcal{R}} \cap V_r} p(t,s) \\ r(m_{\mathcal{R}}, t^r_{\mathcal{R}}) &= \sum_{M_{\mathcal{R}} \in \Omega_{\mathcal{R}}(m_{\mathcal{R}})} \bar{\pi}(M_{\mathcal{R}}) \sum_{t \in T_{\mathcal{R}} \cap V_r} r(M_{\mathcal{R}}, t) \sum_{s \in T_{\mathcal{A}} \cap V_r} p(t,s) \end{aligned} \tag{16}$$

The conditions do not completely specify a physical representation of aggregated markings. However, a specification in terms of the number of input bags of transitions for the different routing chains is possible, but has to take into account that subchains are not disjoint. The resulting net consists of two transitions for each routing subchain V_r which visits $\mathcal{A}$ and $\mathcal{R}$. The following analysis, of course, can be performed without a physical interpretation of the aggregated marking.

Theorem 11. *Define* $\tilde{\Phi}_{\mathcal{A}}(m_{\mathcal{A}}) = \sum_{M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})} \Phi_{\mathcal{A}}(M_{\mathcal{A}}) g_{\mathcal{A}}(M_{\mathcal{A}})$ *and* $\chi(t^r_{\mathcal{A}}) = \sum_{t \in T_{\mathcal{A}} \cap V_r} x(t) \sum_{s \in T_{\mathcal{R}} \cap V_r} p(t,s)$, *then*

$$r(m_{\mathcal{A}}, t^r_{\mathcal{A}}) = \frac{\tilde{\Phi}_{\mathcal{A}}(m_{\mathcal{A}} - I(t^r_{\mathcal{A}}))}{\tilde{\Phi}_{\mathcal{A}}(m_{\mathcal{A}})} \chi(t^r_{\mathcal{A}})$$

where x is the invariant measure defined in (1).

Proof. First notice that $\tilde{\Phi}_{\mathcal{A}}(m_{\mathcal{A}})$ equals the normalisation constant $\bar{G}_{\mathcal{A}}(m_{\mathcal{A}})$ used in (15). The following relations are used in the proof:

$$g(M_{\mathcal{A}}) = \tfrac{1}{f(t)} g(M_{\mathcal{A}} + I(t)) \qquad \text{for some } t \in T_{\mathcal{A}} \text{ according to (11)}$$

$$\Phi_{\mathcal{A}}(M_{\mathcal{A}}) = \Phi_{\mathcal{A}}(M_{\mathcal{A}} + I(t)) \tfrac{r(M_{\mathcal{A}}+I(t),t)}{\chi(t)} \qquad \text{for some } t \in T_{\mathcal{A}} \text{ according to (12)}$$

$$\bar{\pi}_{\mathcal{A}}(M_{\mathcal{A}}) = \tfrac{1}{\tilde{\Phi}_{\mathcal{A}}(m_{\mathcal{A}})} g_{\mathcal{A}}(M_{\mathcal{A}}) \Phi_{\mathcal{A}}(M_{\mathcal{A}}) \qquad \text{for } M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}}) \text{ according to (15)}$$

For notational convenience we prove the result for $m_{\mathcal{A}} + I(t^r_{\mathcal{A}})$:

$$\frac{\tilde{\Phi}_{\mathcal{A}}(m_{\mathcal{A}})}{\tilde{\Phi}_{\mathcal{A}}(m_{\mathcal{A}} + I(t^r_{\mathcal{A}}))} \chi(t^r_{\mathcal{A}}) = \frac{\sum_{M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})} \Phi_{\mathcal{A}}(M_{\mathcal{A}}) g_{\mathcal{A}}(M_{\mathcal{A}}) \sum_{t \in T_{\mathcal{A}} \cap V_r} x(t) \sum_{s \in T_{\mathcal{R}} \cap V_r} p(t,s)}{\tilde{\Phi}_{\mathcal{A}}(m_{\mathcal{A}} + I(t^r_{\mathcal{A}}))}$$

For some fixed $t' \in T_{\mathcal{A}} \cap V_r$ we can represent the numerator of this fraction depending on $M_{\mathcal{A}} + I(t')$ instead of $M_{\mathcal{A}}$.

$$\frac{\sum_{M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})} \Phi_{\mathcal{A}}(M_{\mathcal{A}}+I(t')) g_{\mathcal{A}}(M_{\mathcal{A}}+I(t')) r(M_{\mathcal{A}}+I(t'),t')/(f(t')\chi(t')) \sum_{t \in T_{\mathcal{A}} \cap V_r} x(t) \sum_{s \in T_{\mathcal{R}} \cap V_r} p(t,s)}{\tilde{\Phi}_{\mathcal{A}}(m_{\mathcal{A}}+I(t^r_{\mathcal{A}}))} =$$

$$\sum_{M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})} \bar{\pi}_{\mathcal{A}}(M_{\mathcal{A}} + I(t')) r(M_{\mathcal{A}} + I(t'), t')/x(t') \sum_{t \in T_{\mathcal{A}} \cap V_r} x(t) \sum_{s \in T_{\mathcal{R}} \cap V_r} p(t,s) \qquad (*)$$

Now consider for some $s, t \in T_{\mathcal{A}} \cap V_r$

$$\bar{\pi}(M_{\mathcal{A}} + I(s)) r(M_{\mathcal{A}} + I(s), s) \tfrac{x(t)}{x(s)} =$$

$$\tfrac{1}{G} g_{\mathcal{A}}(M_{\mathcal{A}} + I(s)) \Phi_{\mathcal{A}}(M_{\mathcal{A}}) \chi(s) \tfrac{x(t)}{x(s)} =$$

$$\tfrac{1}{G} g_{\mathcal{A}}(M_{\mathcal{A}} + I(t)) \Phi_{\mathcal{A}}(M_{\mathcal{A}} + I(t)) r(M_{\mathcal{A}} + I(t), t) \tfrac{\chi(s)x(t)f(s)}{\chi(t)x(s)f(t)} =$$

$$\bar{\pi}(M_{\mathcal{A}} + I(t)) r(M_{\mathcal{A}} + I(t), t)$$

Substituting this relation in $(*)$ yields

$$\sum_{M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})} \sum_{t \in T_{\mathcal{A}} \cap V_r} \bar{\pi}_{\mathcal{A}}(M_{\mathcal{A}} + I(t)) r(M_{\mathcal{A}} + I(t), t) \sum_{s \in T_{\mathcal{R}} \cap V_r} p(t,s)$$

which relates the short circuit throughput of each transition in $\mathcal{A}$ to the rate of the aggregate transition. By taking into account that $r(M_{\mathcal{A}}, t) = 0$ for $M_{\mathcal{A}} - I(t) \not\geq 0$, the above result is equivalent to the one computed for $r(m_{\mathcal{A}}, t^r_{\mathcal{A}})$ using the short-circuit pre-analysis results in (16). □

Note that the transition rate of the substitute transition has a representation which has the same form as any other transition in a PSPN. This implies that the aggregated net is a PSPN when the original net is a PSPN. Let $\tilde{\pi}(m_{\mathcal{A}}, m_{\mathcal{R}})$ be the steady state probability of marking $(m_{\mathcal{A}}, m_{\mathcal{R}})$ which can be expressed as

$$\tilde{\pi}(m_{\mathcal{A}}, m_{\mathcal{R}}) = \frac{1}{\tilde{G}} \tilde{\Phi}_{\mathcal{A}}(m_{\mathcal{A}}) \tilde{\Phi}_{\mathcal{R}}(m_{\mathcal{R}})$$

where $\tilde{G}$ is an appropriate normalisation constant. The above result follows directly from the product form, since $g(m_{\mathcal{A}}, m_{\mathcal{R}})$ can be chosen as 1 for all $(m_{\mathcal{A}}, m_{\mathcal{R}}) \in rs(agg, m_0)$.

Theorem 12. *The steady state distribution of the complete PSPN can be expressed as*

$$\pi(M_{\mathcal{A}}, M_{\mathcal{R}}) = \tilde{\pi}(m_{\mathcal{A}}, m_{\mathcal{R}}) \bar{\pi}_{\mathcal{A}}(M_{\mathcal{A}}) \bar{\pi}_{\mathcal{R}}(M_{\mathcal{R}})$$

for $(m_{\mathcal{A}}, m_{\mathcal{R}}) \in rs(agg, m_0)$, $M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})$ *and* $M_{\mathcal{R}} \in \Omega_{\mathcal{R}}(m_{\mathcal{R}})$.

Proof. We have

$$\tilde{\pi}(m_{\mathcal{A}}, m_{\mathcal{R}}) \bar{\pi}_{\mathcal{A}}(M_{\mathcal{A}}) \bar{\pi}_{\mathcal{R}}(M_{\mathcal{R}}) =$$

$$\frac{\sum\limits_{M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})} g_{\mathcal{A}}(M_{\mathcal{A}}) \Phi_{\mathcal{A}}(M_{\mathcal{A}}) \sum\limits_{M_{\mathcal{R}} \in \Omega_{\mathcal{R}}(m_{\mathcal{R}})} g_{\mathcal{R}}(M_{\mathcal{R}}) \Phi_{\mathcal{R}}(M_{\mathcal{R}})}{\sum\limits_{(m_{\mathcal{A}}, m_{\mathcal{R}}) \in rs(agg, m_0)} \sum\limits_{M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})} g_{\mathcal{A}}(M_{\mathcal{A}}) \Phi_{\mathcal{A}}(M_{\mathcal{A}}) \sum\limits_{M_{\mathcal{R}} \in \Omega_{\mathcal{R}}(m_{\mathcal{R}})} g_{\mathcal{R}}(M_{\mathcal{R}}) \Phi_{\mathcal{R}}(M_{\mathcal{R}})}$$

$$\frac{\bar{g}_{\mathcal{A}}(M_{\mathcal{A}}) \Phi_{\mathcal{A}}(M_{\mathcal{A}})}{\sum\limits_{M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})} \bar{g}_{\mathcal{A}}(M_{\mathcal{A}}) \Phi_{\mathcal{A}}(M_{\mathcal{A}})} \frac{\bar{g}_{\mathcal{R}}(M_{\mathcal{R}}) \Phi_{\mathcal{R}}(M_{\mathcal{R}})}{\sum\limits_{M_{\mathcal{R}} \in \Omega_{\mathcal{R}}(m_{\mathcal{R}})} \bar{g}_{\mathcal{R}}(M_{\mathcal{R}}) \Phi_{\mathcal{A}}(M_{\mathcal{R}})}$$

Since theorem 10 holds, we can substitute $\bar{g}$ by g in the numerator and denominator of the last two fractions giving

$$\frac{g_{\mathcal{A}}(M_{\mathcal{A}}) \Phi_{\mathcal{A}}(M_{\mathcal{A}}) g_{\mathcal{R}}(M_{\mathcal{R}}) \Phi_{\mathcal{R}}(M_{\mathcal{R}})}{\sum\limits_{(m_{\mathcal{A}}, m_{\mathcal{R}}) \in rs(agg, m_0)} \sum\limits_{M_{\mathcal{A}} \in \Omega_{\mathcal{A}}(m_{\mathcal{A}})} g_{\mathcal{A}}(M_{\mathcal{A}}) \Phi_{\mathcal{A}}(M_{\mathcal{A}}) \sum\limits_{M_{\mathcal{R}} \in \Omega_{\mathcal{R}}(m_{\mathcal{R}})} g_{\mathcal{R}}(M_{\mathcal{R}}) \Phi_{\mathcal{R}}(M_{\mathcal{R}})} =$$

$$\frac{g((M_{\mathcal{A}}, M_{\mathcal{R}})) \Phi((M_{\mathcal{A}}, M_{\mathcal{R}}))}{\sum\limits_{(M_{\mathcal{A}}, M_{\mathcal{R}}) \in RS(SPN, M_0)} g((M_{\mathcal{A}}, M_{\mathcal{R}})) \Phi((M_{\mathcal{A}}, M_{\mathcal{R}}))} =$$

$$\pi(M_{\mathcal{A}}, M_{\mathcal{R}})$$

□

Consequently the complete analysis can be decomposed into the isolated analysis of each subnet for each possible macro marking and the analysis of an aggregated net. The enumeration of the complete reachability set is not necessary. The aggregation approach can be interpreted as an extension of flow equivalent server aggregation in PQNs, where the aggregates are described by single queues with state dependent service rates. We conceive that the approach may be extended to perform approximative analysis of QPNs which slightly violate the conditions for a product form solution. Similar approaches for QNs are well known and can be adopted in the given framework. Additionally, the aggregation approach can easily be extended to decompose the net into more than two subnets, or to use aggregation hierarchically by decomposing a subnet into sub-subnets.

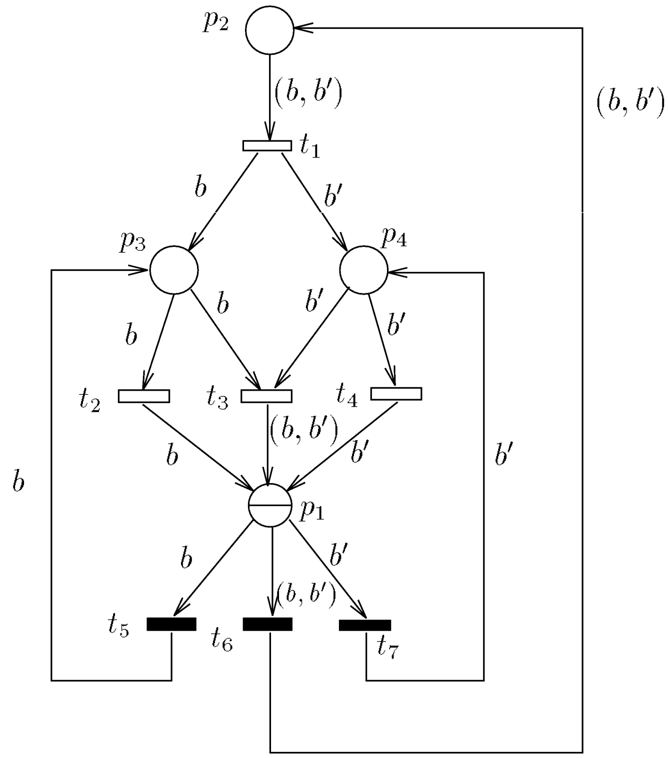

Fig. 2. Example of a PQPN

7 Example

In this example we consider processes which fork into two subprocesses. Processes and subprocesses require independently service at a queue with PS discipline. Service times at the queue are assumed to have a Coxian distribution with two phases and (sub)process specific service rates.

Figure 2 shows the PQPN for this example using the expression representation for CPNs [14]. Place p_1 is the queueing place including a queue whose service discipline is PS. Assume that we start with an initial marking of (b, b') on p_2. (b, b') denotes a process which forks into subprocesses: b and b' due to firing transition t_1. After firing of t_1 subprocess b (b') might enter the queue by firing t_2 (t_4) or the process (b, b') enters the queue by firing transition t_3. When completing its service a subprocess might reenter the queue again or joins with its counterpart by firing t_3 and the process (b, b') enters the queue. Note that, after forking, the subprocesses of a process might directly join again. Fork and join operations take some time as realised by the timed transitions t_1, t_2, t_3 and t_4.

For simplicity we assume that the firing rates of all timed transitions are independent of the current marking and are defined as $r(M,t) = \chi(t), \forall t \in T_1$. The relative firing frequencies of all immediate transitions are set to 1.0. As mentioned before, the service time distribution of a (sub)process at the queue is given by a Coxian distribution with two phases and parameters as specified in Sec. 3.

Using the simplified interpretation for queueing places gives the SPN-skeleton with the following rates:

$$r(M,t) = \chi(t),\ \forall t \in \{t_1, t_2, t_3, t_4\}$$
$$r(M,t_z) = \frac{M(p_1)(z)}{\sum_{z \in \{b,b',(b,b')\}} M(p_1)(z)} \chi(t_z),\ \forall z \in \{b, b', (b,b')\}$$

where $\chi(t_z) := [(\mu_{p_1,z,1})^{-1} + a_{p_1,z,1}(\mu_{p_1,z,2})^{-1}]^{-1}, \forall z \in \{b, b', (b,b')\}$ is the mean service rate for a class z token in the queue and $\{t_b, t_{b'}, t_{(b,b')}\} = T_q$ for queueing place p_1. Thus $\tilde{T} = \{t_1, t_2, t_3, t_4, t_b, t_{b'}, t_{(b,b')}\}$

is the set of transitions of the SPN-skeleton.
Since all output transitions of p_1 are immediate, the depository is not part of the SPN-skeleton, and we will denote the place representing the former queue also by p_1.
All probabilities $p(s,t)$ are 1.0, since there is only one enabled transition per routing subchain. Furthermore

$$\Phi(M) = \prod_{i=1}^{M(p_1)(b)} [\chi(t_b)]^{-1} \prod_{j=1}^{M(p_1)(b')} [\frac{j}{j+M(p_1)(b)}\chi(t_{b'})]^{-1} \prod_{k=1}^{M(p_1)((b,b'))} [\frac{k}{k+M(p_1)(b)+M(p_1)(b')}\chi(t_{(b,b')})]^{-1} \quad (17)$$

and defining $\Phi = \varphi$ we have $r(M,t) = \frac{\Phi(M-I(t))\chi(t)}{\Phi(M)} = \chi(t), \forall t \in \tilde{T}$.
A function f of the PQPN's SPN-skeleton satisfying (2) is e.g. $f(t) = \frac{1}{\chi(t)}, \forall t \in \tilde{T}$ and

$$g(M) = (\frac{\chi(t_3)}{\chi(t_1)})^{M(p_2)((b,b'))}(\frac{\chi(t_2)}{\chi(t_b)})^{M(p_1)(b)}(\frac{\chi(t_4)}{\chi(t_{b'})})^{M(p_1)(b')}(\frac{\chi(t_3)}{\chi(t_{(b,b')})})^{M(p_1)((b,b'))} \quad (18)$$

satisfies (3) for all reachable markings M. Since (cf. appendix B)

$$S_{p_1}(x_{p_1}) = \prod_{k=1}^{2}(y_{k,1}+y_{k,2})!\,(\mu_{p_1,k,1}+a_{p_1,k,1}\mu_{p_1,k,2})[(\frac{1}{\mu_{p_1,k,1}})^{y_{k,1}}\frac{1}{y_{k,1}!}(\frac{a_{p_1,k,1}}{\mu_{p_1,k,2}})^{y_{k,2}}\frac{1}{y_{k,2}!}] \quad (19)$$

for a state $x_{p_1} = ((y_{1,1}, y_{1,2}), (y_{2,1}, y_{2,2}))$, multiplication of (17), (18) and (19) gives the steady state distribution up to a constant (cf. (9)).
Note that if we start with several (b,b')-tokens on p_2 at the initial marking, the subprocesses being joined might be different from the subprocesses generated during the fork operation.

We now extend the example by a second subnet, which is described by a simple PQN including two queues and use here the specification as a PQPN to fit in our framework. A possible interpretation of the extended model is a client server scenario. The PQPN introduced above specifies processes running on a client, each process type, represented by token colours b, b' and (b,b'), requests service from a server or cycles in subnet 1, leaving the CPU (in place p_1) and entering places p_3, p_4. We assume that communication times between client and server are negligible and need not to be considered in the model. The server consists of two queues describing cpu (queue 2) and memory (queue 3). Each request needs several cycles between both queues before it is satisfied and returns to the client. Queue 2 is a PS queue with Cox 2 service time distribution and queue 3 is a FCFS queue with exponential service time distribution. The transitions t_5, t_6 and t_7 realise the cycling of jobs in subnet 1. We introduce new transitions t_8, t_9 and t_{10} describing the movement from subnet 1 into the new subnet 2. Newly arriving tokens entering subnet 2 enter the queue 2, which is part of place p_5. After leaving queue 2 a token either enters queue 3, embedded in place p_6, or returns to subnet 1 which means that a token of the appropriate colour is added to place p_2, p_3 or p_4, respectively. The complete model is shown in figure 3.

Let the initial marking of the PQPN be given by two tokens of colour (b,b') at place p_2. The aggregated markings of the first subnet can be expressed by the number of colour b and the number of colour b' tokens in the subnet; a token of colour (b,b') counts for both colours b and b'. The aggregated reachability set of the second subnet is described by the number of colour b, b' and (b,b') tokens, here (b,b') counts as a normal colour which is independent of b and b'. The physical interpretation of the aggregated marking results from the input/output bags for the three routing chains. In the first column of table 1 the aggregated states of the complete PSPN are listed. States are encoded as 5 tuples, the first two entries include the number of colour b and b' tokens in the first subnet, the last three values describe the number of colour b, b' and (b,b') tokens in the second subnet.

The next step is to analyse both subnets in isolation by short circuiting the environment for each routing subchain, here represented by the colours b, b' and (b,b'). This analysis step has to be performed

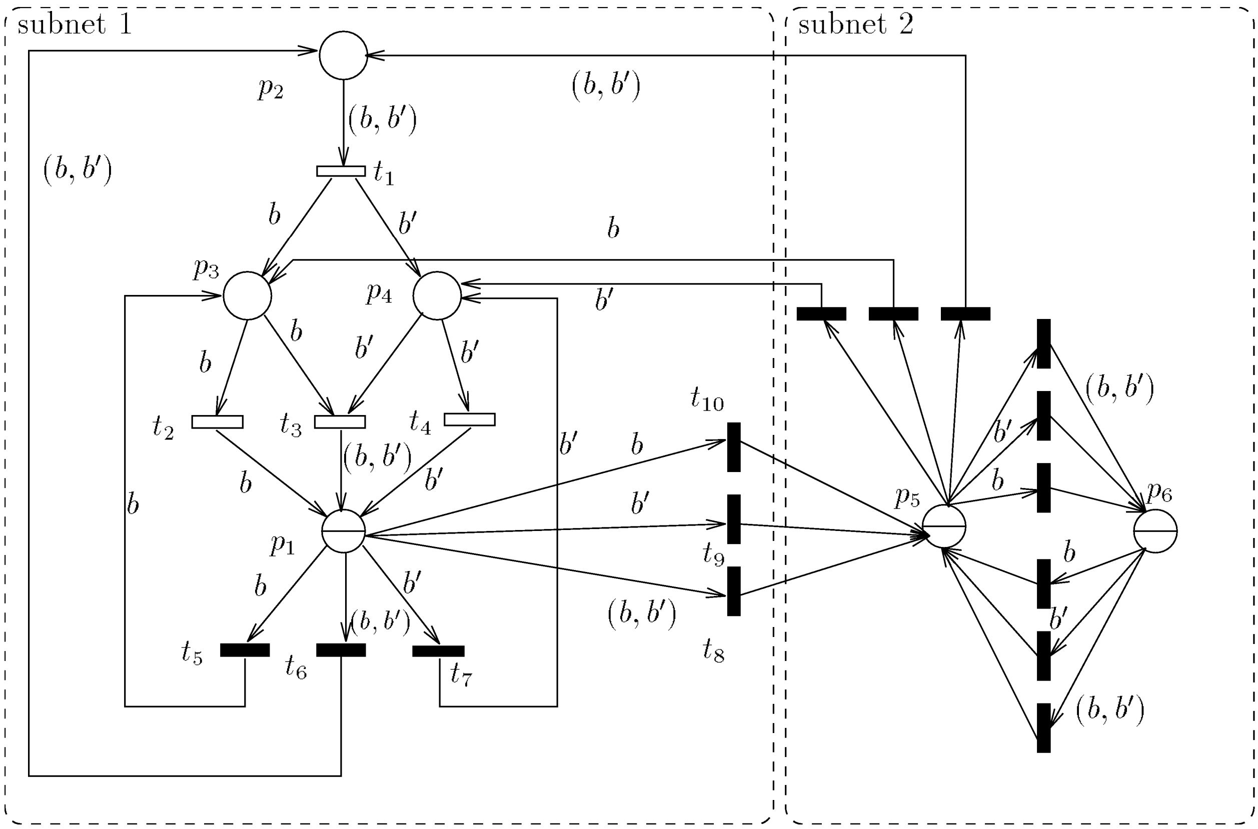

Fig. 3. Complete net with both subnets

for each aggregated state which implies 8 solution runs for subnet 1 and 13 solution runs for subnet 2. The example has been analysed by assuming firing rate 10 for transitions $t_1, \ldots, t_4$, mean service time 1 for b, b' at all queues, mean service time 2 for (b, b') at the queues in the places p_1, p_5 and mean service time 1 for (b, b') at the queue in place p_6. Weights of the immediate transitions are chosen such that colour b and b' tokens stay with probability 0.9 in a subnet and colour (b, b') tokens stay with probability 0.8 in a subnet. Since the marginal probabilities of the PQPN depend only on the means of the service times, no further assumptions about the service time distributions at the PS queues are necessary. The service time at the FCFS queue has to be exponential. The short circuit throughputs for these parameters and the different routing chains are given in table 1. These throughputs determine the firing rates of substitute transitions for the subnet.

A PSPN model where both subnets are substituted by appropriate places and transitions is presented in figure 4. The firing rate of transition t_i is given in the $i+1$-th column of table 1 and depends only on the marking of the corresponding subnet (i.e., the places p_1 and p_2 for subnet 1 and p_3, p_4 and p_5 for subnet 2). The aggregate for subnet 2 is, of course, a standard flow equivalent service center. The aggregate for subnet 1, however, is more complex: tokens of colours b and b' are joined to form a token of colour (b, b') in subnet 2. Analysis of the aggregated net yields the aggregated steady state probabilities $\tilde{\pi}(m_{\mathcal{A}}, m_{\mathcal{R}})$. Short circuit analysis yields conditional probabilities of micro markings in macro markings. From these results all steady state performance results can be computed, e.g., the probability of two tokens (or customers) of colour (b, b') in the queue included in p_5 is given by $\tilde{\pi}(0,0,0,0,2)\bar{\pi}_{subnet2}((0,0,2)(0,0,0))$, where markings of subnet 2 are denoted as a pair of triplets including the number of b, b' and (b, b') tokens in p_5 and p_6, respectively.

	subnet 1			subnet 2		
agg. marking	b	b'	(b,b')	b	b'	(b,b')
(2,2,0,0,0)	4.919e-2	4.919e-2	6.497e-3	-	-	-
(2,1,0,1,0)	6.605e-2	3.230e-2	6.401e-3	-	5.263e-2	-
(2,0,0,2,0)	9.910e-2	-	-	-	7.011e-2	-
(1,2,1,0,0)	3.230e-2	6.605e-2	6.401e-3	5.263e-2	-	-
(1,1,1,1,0)	4.846e-2	4.846e-2	8.881e-3	3.506e-2	3.506e-2	-
(1,1,0,0,1)	4.846e-2	4.846e-2	8.881e-3	-	-	1.538e-1
(1,0,1,2,0)	9.091e-2	-	-	2.627e-2	5.253e-2	-
(1,0,0,1,1)	9.091e-2	-	-	-	3.523e-2	1.030e-1
(0,2,2,0,0)	-	9.910e-2	-	7.011e-2	-	-
(0,1,2,1,0)	-	9.091e-2	-	5.253e-2	2.627e-2	-
(0,1,1,0,1)	-	9.091e-2	-	3.523e-2	-	1.030e-1
(0,0,2,2,0)	-	-	-	4.199e-2	4.199e-2	-
(0,0,1,1,1)	-	-	-	2.640e-2	2.640e-2	7.755e-2
(0,0,0,0,2)	-	-	-	-	-	2.015e-1

Table 1. Short-circuit throughputs for subnet 1 and 2

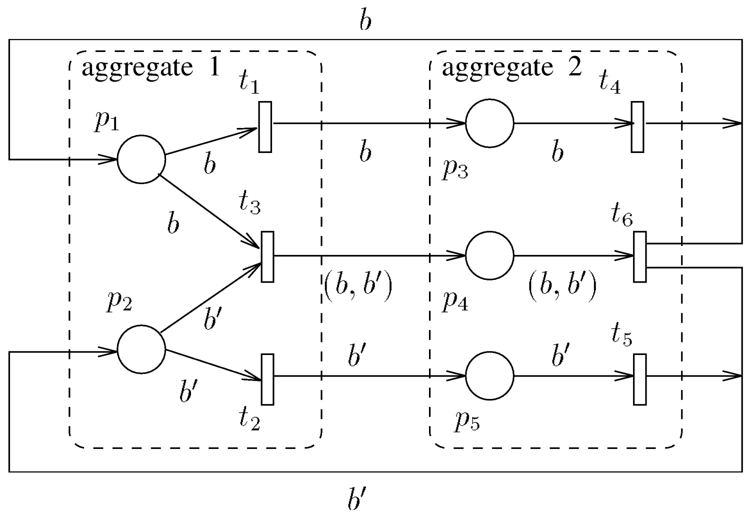

Fig. 4. Aggregated representation of both subnets

8 Conclusions

Queueing Petri Nets are a superset of Queueing Networks and (Generalized) Stochastic Petri Nets. They are suitable for modelling synchronisation and concurrency situations as well as sharing of resources which appear in most distributed systems. From the specification point of view, one great benefit consists of not being forced to model queues by ordinary (CGS)PN elements which simplifies the description of systems. Queueing places can be viewed as simple subnets of a hierarchically specified model.
Besides their descriptional power, QPNs also lead to new insight into the analysis of systems.
In this article we have introduced product form QPNs (PQPNs) which combine PSPNs and PQNs in one model formalism, namely the QPN formalism. Furthermore, we presented an arrival theorem and discussed exact aggregation in PQPNs.

Future work is directed to the development of algorithms like MVA and LBANC [15] in the context of product form Queueing Petri Nets, and to the investigation of approximative analysis techniques, based on the presented aggregation technique for QPNs, but with violation of product form conditions.

References

1. M. Ajmone-Marsan, G. Balbo, G. Conti; Performance Models of Multiprocessor Systems; *MIT Press Series in Computer Science, 1986.*
2. G. Balbo, S.C. Bruell; Computational Aspects of Aggregation in Multiple Class Queueing Networks; *Performance Evaluation (3), 1983, pp. 177-185.*
3. S. Balsamo, G. Iazeolla; An Extension of Norton's Theorem for Queueing Networks; *IEEE Trans. on Softw. Eng. (8) 4, 1982, pp. 298-305.*
4. F. Bause, H. Beilner; Eine Modellwelt zur Integration von Warteschlangen- und Petri-Netz-Modellen; *Informatik-Fachberichte, 218, 1989, pp. 190-204.*
5. F. Bause; Queueing Petri Nets: A Formalism for the Combined Qualitative and Quantitative Analysis of Systems; *Proc. of the 5th Int. Workshop on Petri Nets and Performance Models, IEEE Press 1993, pp. 14-23.*
6. F. Baskett, K.M. Chandy, R.R. Muntz, F.G. Palacois; Open, Closed and Mixed Networks of Queues with Different Classes of Customers; *JACM (22) 2, 1975, pp. 248-260.*
7. K.M. Chandy, U. Herzog, L. Woo; Parametric Analysis of Queueing Networks; *IBM J. Res. Develop. (19), 1975, pp. 36-42.*
8. K.M. Chandy, A.J. Martin; A Characterization of Product-Form Queueing Networks; *JACM (30) 2, 1983, pp. 286-299.*
9. G. Chiola, G. Bruno, T. Demaria; Introducing a Color Formalism into Generalized Stochastic Petri Nets; *Proceedings of the 9th International Workshop on Application and Theory of Petri Nets, Venice (Italy), 1988, pp. 202-215.*
10. J.L. Coleman, W. Henderson, P.G. Taylor; Product Form Equilibrium Distribution and a Convolution Algorithm for Stochastic Petri Nets; *University of Adelaide, 1993, (submitted for publication).*
11. W. Henderson, D. Lucic, P.G. Taylor; A Net Level Performance Analysis of Stochastic Petri Nets; *J. Austral. Math Soc. Ser. B (31), 1989, pp. 176-187.*
12. W. Henderson, P.G. Taylor; Embedded Processes in Stochastic Petri Nets; *IEEE Trans. on Softw. Eng. (17) 2, 1991, pp. 108-116.*
13. J.P. Ho, G. Kim; Class Dependent Queueing Disciplines with Product Form Solutions; *Proc. of the Performance 83, North Holland 1983, pp. 341-349.*
14. K. Jensen; Coloured Petri Nets; Basic Concepts, Analysis Methods and Practical Use; *EATCS Monographs on Theoretical Computer Science, Vol. 1, 1992.*
15. K. Kant; Introduction to Computer System Performance Evaluation; *McGraw-Hill, Inc., 1992.*
16. P.S. Kritzinger, S. van Wyk, A.E. Kryzesinski; A Generalization of Norton's Theorem for Multiclass Queueing Networks; *Performance Evaluation (2), 1982, pp. 98-107.*
17. M. Sereno, G. Balbo; Computational Algorithms for Product Form Solution of Stochastic Petri Nets; *Proc. of the 5th Int. Workshop on Petri Nets and Performance Models, IEEE Press 1993, pp. 98-107.*

A Detailed states in PQNs

In this appendix we describe some details of PQNs which can be found in [6, 15].

Let x be the detailed state of a queue, n its population vector, $n(k)$ the number of class k customers and n_Σ the total number of customers. μ_k^{-1} is the mean service time of class k; $\mu_{k,l}$, $a_{k,l}$ and u_k are the parameters of the Coxian distribution for class k. K denotes the total number of classes. The detailed state description depends on the queue type:

- For type 1 (FCFS) we have $x = (x_1, \ldots, x_{n_\Sigma})$, where x_i describes the class of the customer in position i. x_1 is the class of the presently served customer.
- For type 2 and 3 (PS and IS) we have $x = (y_1, \ldots, y_K)$, where $y_k = (y_{k,1}, \ldots, y_{k,u_k})$ and $y_{k,l}$ is the number of class k customers which are in the l-th phase of their service time distribution.
- For type 4 (LCFS) we have $x = (z_1, \ldots, z_{n_\Sigma})$, where $z_i = (z_i^k, z_i^l)$ and z_i^k is the class of the i-th customer in LCFS order and z_i^l is the phase of the service time distribution the customer is in.
- For type 5 (FESC) the state is completely defined by the marginal distribution of customers, i.e. $x = n$.

The function $h(x)$ used in (5) is defined as

$$h(x) = \begin{cases} (\frac{1}{\mu})^{n_\Sigma} \prod_{i=1}^{n_\Sigma} v(x_i) & \text{for type 1 queues} \\ n_\Sigma! \prod_{k=1}^{K} \prod_{l=1}^{u_k} (v(k)\frac{A_{k,l}}{\mu_{k,l}})^{y_{k,l}}(\frac{1}{y_{k,l}!}) & \text{for type 2 queues} \\ \prod_{k=1}^{K} \prod_{u=1}^{u_k} (v(k)\frac{A_{k,l}}{\mu_{k,l}})^{y_{k,l}}(\frac{1}{y_{k,l}!}) & \text{for type 3 queues} \\ \prod_{i=1}^{n_\Sigma} (v(z_i^k)\frac{A_{z_i^k,z_i^l}}{\mu_{z_i^k,z_i^l}}) & \text{for type 4 queues} \\ \nu(n) \prod_{k=1}^{K} (v(k))^{n(k)} & \text{for type 5 queues} \end{cases}$$

where $\nu(n) = \nu(n - e_k)\mu_k(n)$ and $\nu(0) = 1.0$.

The function $d(n)$ used in (6) is defined as

$$d(n) = \begin{cases} n_\Sigma!(\frac{1}{\mu})^{n_\Sigma} \prod_{k=1}^{K} \frac{1}{n(k)!}(v(k))^{n(k)} & \text{for type 1 queues} \\ n_\Sigma! \prod_{k=1}^{K} \frac{1}{n(k)!}(\frac{v(k)}{\mu_k})^{n(k)} & \text{for type 2 and 4 queues} \\ \prod_{k=1}^{K} \frac{1}{n(k)!}(\frac{v(k)}{\mu_k})^{n(k)} & \text{for type 3 queues} \\ h(x) & \text{for type 5 queues} \end{cases}$$

where n is the marginal state for the detailed state x.

The functions λ_{ph} and λ_{ex} are given by

$$\lambda_{ph}(x,k,l) = \begin{cases} \frac{y_{k,l}}{n_\Sigma}\mu_{k,l}a_{k,l} & \text{for type 2 queues} \\ y_{k,l}\mu_{k,l}a_{k,l} & \text{for type 3 queues} \\ \mu_{k,l}a_{k,l} & \text{if } x = ((k,l),\ldots) \text{ for type 4 queues} \\ 0 & \text{otherwise} \end{cases}$$

$$\lambda_{ex}(x,k,l) = \begin{cases} \mu & \text{if } x = (k,\ldots) \text{ and } l = 1 \text{ for type 1 queues} \\ \frac{y_{k,l}}{n_\Sigma}\mu_{k,l}(1 - a_{k,l}) & \text{for type 2 queues} \\ y_{k,l}\mu_{k,l}(1 - a_{k,l}) & \text{for type 3 queues} \\ \mu_{k,l}(1 - a_{k,l}) & \text{if } x = ((k,l),\ldots) \text{ for type 4 queues} \\ \mu_k(n) & \text{for type 5 queues} \\ 0 & \text{otherwise} \end{cases}$$

B The conditional probability S_p

Let $p \in P_Q$ and x denote the detailed state of the corresponding queue. The conditional probability S_p (cf. (7)) is given by

$$S_p(x) = \begin{cases} \frac{1}{n_\Sigma!} \prod_{k=1}^{K} n(k)! & \text{for type 1 queues} \\ \prod_{k=1}^{K}(n(k)!(\mu_k)^{n(k)} \prod_{l=1}^{u_k}[(\frac{A_{k,l}}{\mu_{k,l}})^{y_{k,l}} \frac{1}{y_{k,l}!}]) & \text{for type 2 and 3 queues} \\ \frac{1}{n_\Sigma!} \prod_{i=1}^{n_\Sigma}(\frac{A_{z_i^k,z_i^l}}{\mu_{z_i^k,z_i^l}}) \prod_{k=1}^{K}(n(k)!(\mu_k)^{n(k)}) & \text{for type 4 queues} \\ 1.0 & \text{for type 5 queues} \end{cases}$$

The conditional probabilities S_p for slightly different states of a queue can be easily related to each other. These relations can be exploited in the proof of the product form theorem for PQPNs, see appendix C. Define the following notations:

$x-(s,l)$ is the state just before the "arrival" of a token of colour s in phase l, leading to state x.
$x+(s,l)$ is the state just before the "departure" of a token of colour s in phase l, leading to state x.

The conditional probabilities S_p for some different states of a queue are related as follows:

type of queue	factor of $S_p(x)$ for			
	$S_p(x-(s,1))$	$S_p(x+(s,l))$	$S_p(x-(s,l)+(s,m))$	$S_p(x+(t,l)-(s,1))$
FCFS	$\frac{n_\Sigma}{n(s)}$	$\frac{n(s)+1}{n_\Sigma+1}$	1	$\frac{n(t)+1}{n(s)}$
PS,IS	$\frac{\mu_{s,1}}{n(s)\mu_s} y_{s,1}$	$(n(s)+1)\frac{\mu_s}{\mu_{s,l}}\frac{A_{s,l}}{(y_{s,l}+1)}$	$\frac{A_{s,m}}{A_{s,l}}\frac{\mu_{s,l}}{\mu_{s,m}}\frac{y_{s,l}}{(y_{s,m}+1)}$	$\frac{(n(t)+1)}{n(s)}\frac{\mu_t}{\mu_s}\frac{y_{s,1}}{(y_{t,l}+1)}\frac{\mu_{s,1}}{\mu_{t,l}}A_{t,l}$
LCFS	$\frac{n_\Sigma}{n(s)}\frac{\mu_{s,1}}{\mu_s}$	$\frac{(n(s)+1)}{(n_\Sigma+1)}\frac{\mu_s}{\mu_{s,l}}A_{s,l}$	$\frac{\mu_{s,l}}{\mu_{s,m}}\frac{A_{s,m}}{A_{s,l}}$	$\frac{(n(t)+1)}{n_s}\frac{\mu_t}{\mu_s}\frac{\mu_{s,1}}{\mu_{t,l}}A_{t,l}$

Note that for FCFS queues only $l = m = 1$ makes sense. The above given table can be read as follows. E.g., for a FCFS queue we have:

$$S_p(x-(s,1)) = \frac{n_\Sigma}{n(s)} S_p(x)$$

C Detailed proof of the product form theorem for PQPNs

With the notations defined in the proof of theorem 3 the global balance equations can be rewritten as

$$\pi(x,M)\ (\sum_{t\in T\setminus(T_O\cup T_Q)} r(M,t)+ \quad (20)$$

$$\sum_{t\in T_O\setminus T_Q} r(M,t)+ \quad (21)$$

$$\sum_{s\in T_Q}(\lambda_{ph}(x,s,1)+\lambda_{ex}(x,s,1))+ \quad (22)$$

$$\sum_{s\in T_Q}\sum_{l=2}^{u_s}(\lambda_{ph}(x,s,l)+\lambda_{ex}(x,s,l))) \quad (23)$$

$$=$$

$$\sum_{t\in T\setminus(T_O\cup T_Q)}\sum_{s\in T\setminus T_Q}\pi(x,M-I(t)+I(s))r(M-I(t)+I(s),s)p(s,t) \quad (24)$$

$$+\sum_{t\in T_O\setminus T_Q}(\sum_{s\in T\setminus T_Q}\pi(x,M-I(t)+I(s))r(M-I(t)+I(s),s)p(s,t)+$$

$$\sum_{s\in T_Q}\sum_{l=1}^{u_s}\pi(x+(s,l),M-I(t)+I(s))\lambda_{ex}(x+(s,l),s,l)p(s,t)) \quad (25)$$

$$+\sum_{s\in T_Q}(\sum_{t\in T_I\setminus T_Q}\pi(x-(s,1),M-I(s)+I(t))r(M-I(s)+I(t),t)p(t,s)+$$

$$\sum_{t\in T_I\cap T_Q}\sum_{l=1}^{u_t}\pi(x+(t,l)-(s,1),M-I(s)+I(t))\lambda_{ex}(x+(t,l)-(s,1),t,l)p(t,s)) \tag{26}$$

$$+\sum_{s\in T_Q}\sum_{l=1}^{u_s-1}\pi(x+(s,l)-(s,l+1),M)\lambda_{ph}(x+(s,l)-(s,l+1),s,l) \tag{27}$$

Terms (20) and (24) correspond to the flow not affected by queues, (21) and (25) to the flow through transitions in T_O, (22) and (26) to the flow through the first phase of service in each queue, and (23) and (27) to the flow through all other phases of service in a queue.

In the following we will show that local balance holds by equating the corresponding terms of the sums of (20) = (24), (21) = (25), (22) = (26) and (23) = (27).

Let $S(x)=\prod_{p\in P_Q}S_p(x_p)$.

ad (20) = (24): We show $\forall t\in T\setminus(T_O\cup T_Q)$:

$$\pi(x,M)r(M,t)=\sum_{s\in T\setminus T_Q}\pi(x,M-I(t)+I(s))r(M-I(t)+I(s),s)p(s,t) \tag{28}$$

Since $t\notin T_O$, we have $p(s,t)=0,\forall s\in T_Q$. Thus (28) becomes

$$\pi(x,M)r(M,t)=\sum_{s\in T}\pi(x,M-I(t)+I(s))r(M-I(t)+I(s),s)p(s,t)$$

After substituting $\pi(x,M)$ with the right part of (9) we obtain the defining equations for function f as for PSPNs (cf. (2)).

ad (21) = (25): We show $\forall t\in T_O\setminus T_Q$:

$$\begin{aligned}&\pi(x,M)(\ r(M,t))\\&\quad=\\&\quad(\sum_{s\in T\setminus T_Q}\pi(x,M-I(t)+I(s))r(M-I(t)+I(s),s)p(s,t)+\\&\qquad\sum_{s\in T_Q}\sum_{l=1}^{u_s}\pi(x+(s,l),M-I(t)+I(s))\lambda_{ex}(x+(s,l),s,l)p(s,t))\end{aligned}$$

Using (9) and (3) we obtain

$$\begin{aligned}&\Phi(M)r(M,t)=\\&\quad(\sum_{s\in T\setminus T_Q}\Phi(M-I(t)+I(s))\frac{f(s)}{f(t)}r(M-I(t)+I(s),s)p(s,t)+\\&\qquad\sum_{s\in T_Q}\frac{f(s)}{f(t)}p(s,t)\Phi(M-I(t)+I(s))\sum_{l=1}^{u_s}\frac{S(x+(s,l))}{S(x)}\lambda_{ex}(x+(s,l),s,l))\end{aligned}$$

implying

$$\begin{aligned}&\varphi(M-I(t))\chi(t)=\\&\quad(\sum_{s\in T\setminus T_Q}\varphi(M-I(t))\chi(s)\frac{f(s)}{f(t)}p(s,t)+\\&\qquad\sum_{s\in T_Q}\frac{f(s)}{f(t)}p(s,t)\frac{\varphi(M-I(t))\chi(s)}{r(M-I(t)+I(s),s)}\sum_{l=1}^{u_s}\frac{S(x+(s,l))}{S(x)}\lambda_{ex}(x+(s,l),s,l))\end{aligned}$$

Since $t \notin T_Q$, the last equation holds if

$$
\begin{aligned}
r(M - I(t) + I(s), s) &= r(M + I(s), s) \\
&= \sum_{l=1}^{u_s} \frac{S(x + (s,l))}{S(x)} \lambda_{ex}(x + (s,l), s, l)
\end{aligned}
$$

which can easily be shown to be satisfied for all queues of PQN-type.

ad (22) = (26): Here we show $\forall s \in T_Q$:

$$
\begin{aligned}
&\pi(x, M)\,(\,(\lambda_{ph}(x,s,1) + \lambda_{ex}(x,s,1))) \\
&\quad = \;(\sum_{t \in T_I \setminus T_Q} \pi(x - (s,1), M - I(s) + I(t)) r(M - I(s) + I(t), t) p(t,s) + \\
&\qquad \sum_{t \in T_I \cap T_Q} \sum_{l=1}^{u_t} \pi(x + (t,l) - (s,1), M - I(s) + I(t)) \lambda_{ex}(x + (t,l) - (s,1), t, l) p(t,s))
\end{aligned}
$$

Using (9) and dividing by $\frac{1}{G}\Phi(M) g(M) S(x)$ yields:

$$
\begin{aligned}
&(\lambda_{ph}(x,s,1) + \lambda_{ex}(x,s,1)) \\
&\quad = \;(\sum_{t \in T_I \setminus T_Q} \frac{\Phi(M - I(s) + I(t)) r(M - I(s) + I(t), t)}{\Phi(M)} \frac{S(x - (s,1))}{S(x)} \frac{f(t)}{f(s)} p(t,s) + \\
&\qquad \sum_{t \in T_I \cap T_Q} \frac{\Phi(M - I(s) + I(t))}{\Phi(M)} \frac{f(t)}{f(s)} p(t,s) \sum_{l=1}^{u_t} [\frac{S(x + (t,l) - (s,1))}{S(x)} \lambda_{ex}(x + (t,l) - (s,1), t, l)])
\end{aligned}
$$

implying

$$
\begin{aligned}
&(\lambda_{ph}(x,s,1) + \lambda_{ex}(x,s,1)) \\
&\quad = \;(\sum_{t \in T_I \setminus T_Q} \frac{\varphi(M - I(s)) \chi(t)}{\Phi(M)} \frac{S(x - (s,1))}{S(x)} \frac{f(t)}{f(s)} p(t,s) + \\
&\qquad \sum_{t \in T_I \cap T_Q} \frac{\Phi(M - I(s) + I(t))}{\Phi(M)} \frac{f(t)}{f(s)} p(t,s) \sum_{l=1}^{u_t} [\frac{S(x + (t,l) - (s,1))}{S(x)} \lambda_{ex}(x + (t,l) - (s,1), t, l)])
\end{aligned}
$$

yielding

$$
\begin{aligned}
&f(s)(\lambda_{ph}(x,s,1) + \lambda_{ex}(x,s,1)) \\
&\quad = \;\sum_{t \in T_I \setminus T_Q} \frac{r(M,s)}{\chi(s)} \frac{S(x - (s,1))}{S(x)} \chi(t) f(t) p(t,s) + \\
&\qquad \sum_{t \in T_I \cap T_Q} \frac{\Phi(M - I(s) + I(t))}{\Phi(M) \chi(t)} \sum_{l=1}^{u_t} [\frac{S(x + (t,l) - (s,1))}{S(x)} \lambda_{ex}(x + (t,l) - (s,1), t, l)] \chi(t) f(t) p(t,s)
\end{aligned}
$$

This equation is satisfied if the following hold:

a) $(\lambda_{ph}(x,s,1) + \lambda_{ex}(x,s,1)) = \Gamma_x \chi(s)$

b) $\frac{r(M,s)}{\chi(s)} \frac{S(x - (s,1))}{S(x)} = \Gamma_x$

c) $\frac{\Phi(M - I(s) + I(t))}{\Phi(M) \chi(t)} \sum_{l=1}^{u_t} [\frac{S(x + (t,l) - (s,1))}{S(x)} \lambda_{ex}(x + (t,l) - (s,1), t, l)] = \Gamma_x$

where $\Gamma_x \in \mathbb{R}^+$ is some positive value.

Fusing conditions a)+b) and b)+c) we obtain two conditions:

α) $r(M,s) = \frac{S(x)}{S(x - (s,1))} (\lambda_{ph}(x,s,1) + \lambda_{ex}(x,s,1))$

β) $\frac{r(M,s)}{\chi(s)}\frac{S(x-(s,1))}{S(x)} = \frac{\Phi(M-I(s)+I(t))}{\Phi(M)\chi(t)}\sum_{l=1}^{u_t}[\frac{S(x+(t,l)-(s,1))}{S(x)}\lambda_{ex}(x+(t,l)-(s,1),t,l)]$

$\Longleftrightarrow$

$\frac{r(M,s)\Phi(M)}{\Phi(M-I(s)+I(t))}\frac{\chi(t)}{\chi(s)} = \frac{S(x)}{S(x-(s,1))}\sum_{l=1}^{u_t}[\frac{S(x+(t,l)-(s,1))}{S(x)}\lambda_{ex}(x+(t,l)-(s,1),t,l)]$
$\Longleftrightarrow$

$\frac{\varphi(M-I(s))\chi(t)}{\Phi(M-I(s)+I(t))} = \sum_{l=1}^{u_t}[\frac{S(x+(t,l)-(s,1))}{S(x-(s,1))}\lambda_{ex}(x+(t,l)-(s,1),t,l)] \Longleftrightarrow$
$r(M-I(s)+I(t),t) = \sum_{l=1}^{u_t}[\frac{S(x+(t,l)-(s,1))}{S(x-(s,1))}\lambda_{ex}(x+(t,l)-(s,1),t,l)]$

Note that $s,t \in T_Q$ holds for this case.

Both conditions are satisfied for queues of PQN-type. Note, that for FCFS we have $u_t = 1$ and for other type of queues we have $\sum_{l=1}^{u_t} A_{t,l}(1-a_{t,l}) = 1$.

ad (23) = (27): We show $\forall s \in T_Q$:

$$\pi(x,M)\,(\sum_{l=2}^{u_s}(\lambda_{ph}(x,s,l)+\lambda_{ex}(x,s,l))) = \sum_{l=1}^{u_s-1}\pi(x+(s,l)-(s,l+1),M)\lambda_{ph}(x+(s,l)-(s,l+1),s,l)$$

For FCFS queues and the flow equivalent service center there is nothing to show, since service times are exponentially distributed ($u_s = 1$) and thus $0 = 0$ trivially holds.

So let us consider some other type of queue. Shifting the index $l \mapsto l+1$ on the left side this equation becomes

$$\pi(x,M)\,(\sum_{l=1}^{u_s-1}(\lambda_{ph}(x,s,l+1)+\lambda_{ex}(x,s,l+1))) = \sum_{l=1}^{u_s-1}\pi(x+(s,l)-(s,l+1),M)\lambda_{ph}(x+(s,l)-(s,l+1),s,l)$$

Equating the individual terms of the sums yields

$$\pi(x,M)\,(\,(\lambda_{ph}(x,s,l+1)+\lambda_{ex}(x,s,l+1))) = \pi(x+(s,l)-(s,l+1),M)\lambda_{ph}(x+(s,l)-(s,l+1),s,l)$$

Using (9) we obtain

$$S(x)\,(\,(\lambda_{ph}(x,s,l+1)+\lambda_{ex}(x,s,l+1))) = S(x+(s,l)-(s,l+1))\lambda_{ph}(x+(s,l)-(s,l+1),s,l)$$

which holds for all queues of type PS, IS or LCFS.

With that the global balance equations hold and $\pi(x,M)$ given by (9) is the steady state distribution.