On Exact Equilibrium Distributions of Stochastic Petri Nets

by

Diana Lucic
B.Sc.(Ma.)(Hons.), University of Adelaide, South Australia.

Thesis submitted for the degree
of Doctor of Philosophy
in the
Department of Applied Mathematics
University of Adelaide.

November 1990.

Contents

List of Figures

Summary

Chapter 1 introduces a PN and its time extended counterparts, the timed PN and the SPN. The problems associated with their analysis are presented and a summary of the techniques we use in an attempt to overcome these problems is given.

Chapter 2 gives a survey of the types of properties that are analysed in PNs. Time-extended PNs are then introduced and a survey of the different models used is given.

Chapter 3 describes the extended product form solutions for SPNs and gives some explanatory examples.

Chapter 4 addresses the problems related to generally distributed firing times. It contains an introduction to the topics of insensitivity and Generalised Semi-Markov processes. The concepts are then applied to SPNs, and further extensions are introduced, allowing the consideration of transition merging, marking amalgamation and another extension to the theory of insensitivity.

Chapter 4 is used as the foundation for the work in Chapter 5. Using the theory of insensitivity, we show how to aggregate and then disaggregate SPNs to yield exact equilibrium distributions.

Chapter 6 contains conclusions and some ideas for further research, arising from the work presented in this thesis.

Declaration

I declare that the contents of this thesis have not been submitted to any university for the purpose of obtaining any other degree or diploma.

Also, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text.

I consent to this thesis being made available for photocopying and loan.

Diana Lucic.

Acknowledgements

I would like to sincerely thank my supervisor Bill Henderson, for his endless patience and encouragement throughout the work toward this thesis. I would also like to thank my parents for their support and for providing me with this opportunity. Also, thanks go to Allan Bramley, for proof reading this thesis in it's later stages.

The receipt of a Commonwealth Postgraduate Research Award, and a short term Adelaide Scholarship, supplemented by a grant from the Teletraffic Research Centre at the University of Adelaide are acknowledged.

Finally, I would like to thank Nigel for his encouragement, for being so tolerant and for his astounding patience. Thank you also, for drawing my pictures and proof reading this thesis. Most importantly, thank you for all your love.

Chapter 1

Introduction

A Petri Net (PN) is a graph model, designed to include the analysis of systems which exhibit concurrent and conflicting behaviour. As a graphical tool, PNs share the same task as a flow chart, block diagram or network. In addition, tokens are used to simulate the dynamic and concurrent activities in the model.

Historically, the basic concepts of PN theory were developed from the work of Carl Adam Petri [80], in his doctoral dissertation. Further developments of the theory came from a report by Holt and Commoner [44], in which it was shown how PNs could be applied to the modelling and analysis of systems of concurrent components.

Formally, a PN is a bipartite directed graph, which can be defined by the following five-tuple,

$$N = (\mathcal{P}, \mathcal{T}, \mathbf{I}, \mathbf{O}, \mathbf{m}_0).^1$$

$\mathcal{P} = \{p_1, \ldots, p_r\}$, $r = |\mathcal{P}|$, is a finite set of places, represented by circles, and $\mathcal{T} = \{t_1, \ldots, t_s\}$, $s = |\mathcal{T}|$, is a finite set of transitions, represented by bars. The directed arcs, connect places to transitions and transitions to places. The input function, $\mathbf{I} : \mathcal{T} \to \mathbb{Z}^r$, is a mapping from a transition $t \in \mathcal{T}$ to its input bag $\mathbf{I}(t)$, which is an $r \times 1$ column vector, giving the input places (including multiplicity) of transition t. The output function, $\mathbf{O} : \mathcal{T} \to \mathbb{Z}^r$, is a mapping from a transition t to its output bag $\mathbf{O}(t)$, which is an $r \times 1$ column vector, giving the output

[1]Note that boldface characters will be used at all times to represent vectors, components of a vector will be in normal type.

places (including multiplicity) of t . These represent the pre- and post- conditions respectively. The use of bags, rather than sets, allows a place to be a multiple input or multiple output, (for a brief discussion on bag theory see Peterson [79]).

A marking, $\mathbf{m}$, is an $r \times 1$ column vector, which assigns to each place a non-negative integer. For example, the ith component, $\mathrm{m}(i)$, is the number of tokens in place $p_i \in \mathcal{P}$. The initial marking, $\mathbf{m}_0$, represents the initial token distribution.

In the modelling of PNs, it is often useful to identify places with conditions and to identify the firing of transitions with events. Before a transition can fire, or an event can occur, a pre-condition must be satisfied and is given by,

$$(\mathbf{m} - \mathbf{I}(t)) \geq \mathbf{0}.$$

That is, for every input place into a transition, there must be at least as many tokens in it, as there are arcs from the place to the transition. When this condition is satisfied, the transition is said to be enabled and may then fire. When a transition fires, tokens are removed from its input places and deposited in its output places, one token for each arc. This is executed instantaneously for PNs and executed in two phases for time extended PNs. In the first phase, the transition becomes enabled and waits for a period of time, until it fires (if it still can). In the second phase the transition starts firing, thereby removing tokens from it's input places. At the end of the firing time, the tokens are deposited into it's output places.

The state of the PN is defined by its marking. When a transition fires, there is a change in the token distribution and therefore in the marking of the net. The result of firing a transition, from a marking $\mathbf{m}$, is a new marking $\mathbf{m}'$. Successive firings of the enabled transitions in the PN, result in a set of markings, $\mathcal{M}$, which are connected through the firing sequence. The state space of the PN is called a reachability tree/graph.

The analysis of PNs, has revolved around checking correctness, and observing the behaviour, of the system modelled by the PN. Consequently, a variety of behavioural and structural properties have been observed, which have therefore induced the creation of several analysis techniques. In Chapter 2, we give a sample of the types of properties analysed and the techniques used to verify them.

The concept of time was not explicitly given in the original definition of PNs (see Petri [80]). When timing was introduced, it was in the form of time delays, associated with either the firing of the transitions, or the time spent by the tokens in the places of the net. If the time delays are given deterministically, the nets are commonly known as timed PNs and if they are given probabilistically, they are known as Stochastic PNs (SPNs).

With the advent of time-extended PNs, analysts in the field have turned their attention to the performance evaluation of the system being modelled by the time-extended PN. We continue with the same theme in this thesis, and aim to find the equilibrium distribution of SPNs in exact form. Conventional methods have relied on analysing the underlying Markov process, however, there are inherent difficulties in attempting to find the equilibrium distribution in this way, and so alternative methods are required. The techniques we present, for finding the equilibrium distribution in exact form, avoid some of these stumbling blocks.

There are two problems that make solving the global balance equations unsuitable. Firstly, using any firing time other than the negative exponentially distributed firing time, causes the underlying process to lose it's Markovian nature. Secondly, the number of markings in the Markov process, grows exponentially with the number of tokens and places in the net.

Obviously, the first problem does not arise if no other firing time distribution but the negative exponential distribution is used in the model. However, if reality is to be modelled accurately, this is a harsh restriction. In Sections 2.3, 2.4 and 2.5, we give a survey, from the literature, of the types of models that incorporate generally distributed firing times (we include deterministic firing times in this class) and the type of restrictions imposed to ensure that a solution can be easily extracted. In Chapter 4, we introduce the theory of "insensitivity" in generalised semi-Markov processes and apply this theory to SPNs. In brief, the concept of insensitivity is that the governing distributions in the process, may be replaced by other distributions, each with the same mean, without altering the equilibrium distribution of the process. Therefore, the difficulties associated with the general distributions can be reduced to a relatively simple problem, involving only negative exponential distributions, or other convenient distributions, with the same means.

Further extensions to the theory of insensitivity, by Rumsewicz and Henderson [87] discussed in Section 4.5, create the foundations for some new results presented in the remainder of Chapter 4. In Section 4.6, we present the theory on transition merging. In Section 4.7, we consider marking amalgamation of the underlying process and provide conditions for which we are able to extract exact marginal distributions for the original SPN. The procedure is to amalgamate the markings of the process, so that the resultant process is insensitive to its generally distributed firing times. In Section 4.8 an extension to the theory of insensitivity is given, which allows a set of generally distributed firing times to be enabled simultaneously.

The marking amalgamation procedure outlined above, reduces the number of markings and therefore the size of the process that must be analysed. In Chapter 3, we present a solution technique which also reduces the effect of the marking explosion problem. Note that the marking explosion problem is not unique to SPNs. A similar problem in networks of queues lead researchers in the field, to look for a method of finding the equilibrium distribution, which did not require the analysis of the state space. The result was what is called the product form solution. Standard assumptions in queueing theory are, that customers arrive and are routed through the network in a manner which may depend on the state, but not on the routeing of other customers. In a SPN this is not true, as the tokens move around in batches and the routeing of individual tokens is correlated. Consequently, the theory on product form solutions for queueing networks has not been applicable to SPNs. In Chapter 3, we adapt the work of Henderson and Taylor [41], who incorporate batch movement and correlated routeing into their queueing network. In doing this we achieve an extended product form solution for SPNs. The central feature of this technique, is to consider the transitions of the SPN to be the states of a Markov chain, called the routeing process. This is an obvious advantage, since the number of transitions in the net will, almost always, be fewer in number than the markings of the net.

In Chapter 5, we present another technique which reduces the effect of the marking explosion problem, by aggregating and disaggregating the subnets of the SPN, using the results of Chapter 4. The aim of this technique, is to amalgamate subnets of the SPN, so that the remaining "skeleton" net is insensitive to the time spent in the subnets. Once this is established, the exact marginal distribution for the original net can be obtained. We then use the results of insensitivity theory, to disaggregate the subnets, and obtain the equilibrium distribution for the original net in exact form. The benefit of this approach, is that the size of the marking process has been reduced considerably, either when extracting the marginal distribution using the skeleton net, or when disaggregating to find the equilibrium distribution for the original net.

Chapter 2

An Introduction to Petri Nets, Timed Petri Nets and Stochastic Petri Nets

PNs have no time associated with an activity in the model and are therefore not relevant in analysing time dependent performance measures, such as an equilibrium distribution. As was stated in the introduction, it is the focus of this thesis to extract such time dependent behaviour. For this reason we do not dwell on the explanation and use of PNs, rather we give a brief overview, and direct those readers who may be interested to a list of references given below. The extension of time in PNs will be addressed later in this chapter.

From 1970 to 1975 the Computation Structure Group from M.I.T. produced many reports and theses on PNs. The first book concerned with PNs was written by J.L. Peterson [79] in 1980, a second book was written by W. Reisig [86] in 1984. These two books contain references for most of the PN related works up to and including 1984.

Since then there have been numerous workshops and conferences devoted to PNs and time extended PNs. The proceedings are mostly limited to the conference participants, however, some selected papers and other articles have been published in *Advances in Petri Nets*. The 1987 volume contains the most comprehensive bibliography of papers concerned with PNs and time extended PNs, from 1962 to 1987. For an updated bibliography on where to find further publications see Murata [76].

A series of international conferences concerned mainly with timed and Stochastic PNs began in 1985 and have since been held biennially. Each of these workshops is accompanied by proceedings; all three are available from the IEEE Computer Society Press.

In the next section, we discuss the types of properties searched for in PNs. Since we are interested in the performance analysis of time extended PNs, this section will give a brief summary. For further information we suggest Peterson [79], Reisig [86] or Murata [76], amongst others.

2.1 The Behavioural and Structural Properties of Petri Nets

As mentioned in Chapter 1, PNs were initially proposed as a tool for checking correctness and observing behaviour in parallel systems. This lead to a classification of the type of behavioural and structural properties observed and therefore some methods of analysis. In this section, we list the properties, provide a motivation for their analysis and briefly summarise the techniques used to study them.

The distinction between the two types of property, behavioural and structural, lies in the dependence on the initial marking. Those properties which depend on the initial marking are classified as behavioural properties and those which do not, are classified as structural properties. We begin by discussing the behavioural properties.

2.1.1 Behavioural Properties

One of the most basic analysis problems for PNs is to determine which markings are reachable from either the initial marking or any other marking in the reachability tree. First define, for a PN, N, $R(N, \mathbf{m})$ to be a set of all the markings that are reachable from $\mathbf{m}$.

Definition 2.1: Immediately Reachable

If there exists a transition, t, enabled by a marking, $\mathbf{m}$, which when fired creates a new marking, $\mathbf{m}' = \mathbf{m} - \mathbf{I}(t) + \mathbf{O}(t)$, then these markings are said to be immediately reachable.

Definition 2.2: Reachability

Given a PN, N, and markings $\mathbf{m}_i$ and $\mathbf{m}_j$, $\mathbf{m}_j$ is said to be reachable from $\mathbf{m}_i$ if

$$\mathbf{m}_j \in R(N, \mathbf{m}_i).$$

Note that the reachability relationship is the reflexive, transitive closure of the immediately reachable relationship. Many of the other properties can be stated in terms of the reachability problem. For example, reachability analysis can be used to detect deadlock.

Definition 2.3: Deadlock

Deadlock in a PN, is a marking in which no transition can fire, halting the execution of the PN. This is of interest, for example, in resource allocation where any sequence of allocation resulting in deadlock can be identified and then eliminated. Another problem related to the reachability problem is the coverability problem.

Definition 2.4: Coverability

Given a PN, N, with initial marking $\mathbf{m}_0$ and marking $\mathbf{m}_i \in R(N, \mathbf{m}_0)$, then $\mathbf{m}_i$ is said to be coverable if there is a reachable marking $\mathbf{m}_j \in R(N, \mathbf{m}_0)$, $\mathbf{m}_j \neq \mathbf{m}_i$, such that $\mathbf{m}_j \geq \mathbf{m}_i$. The coverability problem is referred to later when we turn our attention to analysing the behavioural properties of PNs.

Definition 2.5: Conservative

A PN, N, with initial marking $\mathbf{m}_0$ is conservative, if, $\forall \mathbf{m}_j \in R(N, \mathbf{m}_0)$,

$$\sum_{p_i \in \mathcal{P}} \mathrm{m}_j(p_i) = \sum_{p_i \in \mathcal{P}} \mathrm{m}_0(p_i).$$

Definition 2.6: Boundedness

Boundedness is the property whereby each place in the PN cannot contain more than a predefined upper bound of tokens. For example, if the upper bound is defined by an integer k, then $\mathrm{m}(p) \leq k$, $\forall p \in \mathcal{P}$ and $\forall \mathbf{m} \in R(N, \mathbf{m}_0)$. The PN is termed k-bounded in this case. If $k = 1$ the PN is said to be safe. Places in PNs are often used to represent buffers and registers for storing intermediate data. By verifying boundedness, it is guaranteed that there will be no overflow in the buffers or registers.

Definition 2.7: Liveness

Liveness implies a complete absence of deadlock in the model. It is, however, a stronger property and is defined as follows. A transition, t, is potentially fireable in a marking, $\mathbf{m}$, if there exists a marking, $\mathbf{m}' \in R(N, \mathbf{m})$, such that t is enabled in $\mathbf{m}'$. A transition is live in a marking, $\mathbf{m}$, if it is potentially fireable in every marking in $R(N, \mathbf{m})$. A PN is live, if every transition is live in the initial marking, $\mathbf{m}_0$.

Definition 2.8: Persistence

A persistent PN is one in which a transition, once enabled, will remain enabled until fired. A set of conflicting transitions behave in exactly the opposite way. When one of the members in the conflict set fires, it disables the remaining transitions which must be re-enabled in order to compete for firing once again. A conflict free net is persistent.

Definition 2.9: Recoverability

A recoverable PN will return to the initial marking or some home state eventually. This property is called reversibility in the literature, however we shall reserve the term reversibility for an idea discussed further in Section 4.4.

2.1.2 Analysing the Behavioural Properties

The methods of analysing these behavioural properties are well documented (see Peterson [79], Reisig [86] and Murata [76]) and have been classified into three groups. These are the coverability tree method, the matrix equation approach and the reduction technique. We shall proceed by briefly explaining how each technique is performed.

If the PN is unbounded, that is any number of tokens may be found in some place, the reachability graph will grow infinitely large. The coverability tree approach is used to keep the tree finite in size, by employing an algorithm which alters markings by introducing a symbol, ω, into the first marking that is coverable by another marking. The alteration is made in the following way: If $\mathbf{m}_i$ is coverable by $\mathbf{m}_j$, then replace $\mathrm{m}_j(p)$ by a symbol ω, for each place, p, such that $\mathrm{m}_j(p) > \mathrm{m}_i(p)$. If such an altered marking is revisited, that branch does not have to be checked for coverability again. Figure 2.2 is the coverability tree of the unbounded PN given in Figure 2.1. The initial marking is $(1,0,0)$. If t_2 fires first, a token is removed from p_1 and a token is deposited in both p_2 and p_3, creating marking $(0,1,1)$ as indicated on the coverability tree. If t_3 then fires, a token will be removed from p_2 and p_3 and a token will be deposited in p_3, creating marking $(0,0,1)$ and so creating a deadlock. Suppose now that t_1 fires first, creating marking $(1,1,0)$. From this marking, t_1 can fire again creating $(1,2,0)$. Since $(1,2,0)$ covers $(1,1,0)$ we alter the marking by replacing the entry for m(2) by the symbol ω. That is, the first marking in that branch becomes $(1,\omega,0)$ as indicated on the coverability tree. The search down that branch is now complete. A similar argument holds if t_2 fires after t_1, creating the other branch on the left hand side. The disadvantages of this method are that it is a time consuming operation and information is invariably lost due to the lumping of the markings. Murata [76] notes that since some information has been lost, only a subclass of the behavioural properties of the PN may be analysed. The reachability problem and liveness problem cannot be solved by this method alone.

A second approach is based on the matrix equations which govern the dynamic behaviour of the PN. Consider a PN, N, with s transitions and r places. The incidence matrix, A, is an $s \times r$ matrix of integers with entries given by,

$$a_{ij} = a_{ij}^{+} - a_{ij}^{-},$$

where, a_{ij}^{+} is the number of output arcs from t_i to p_j (which may be interpreted as the number of tokens deposited in p_j after t_i fires), and a_{ij}^{-} is the number of input arcs to t_i from p_j (which may be interpreted as the number of tokens removed

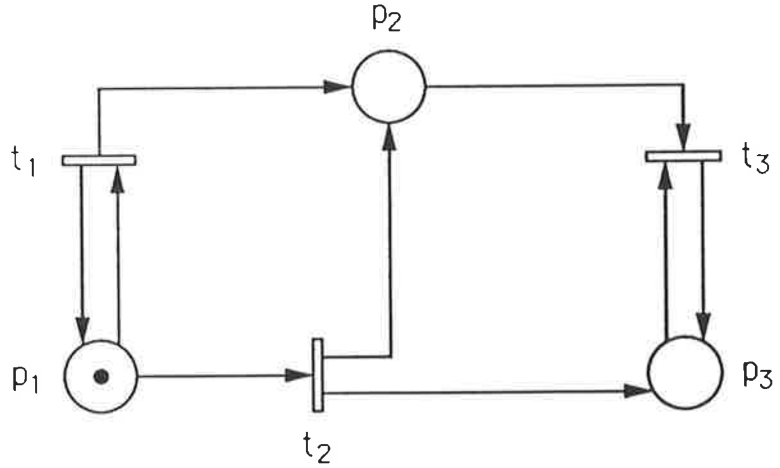

Figure 2.1: A SPN to illustrate the coverability tree approach.

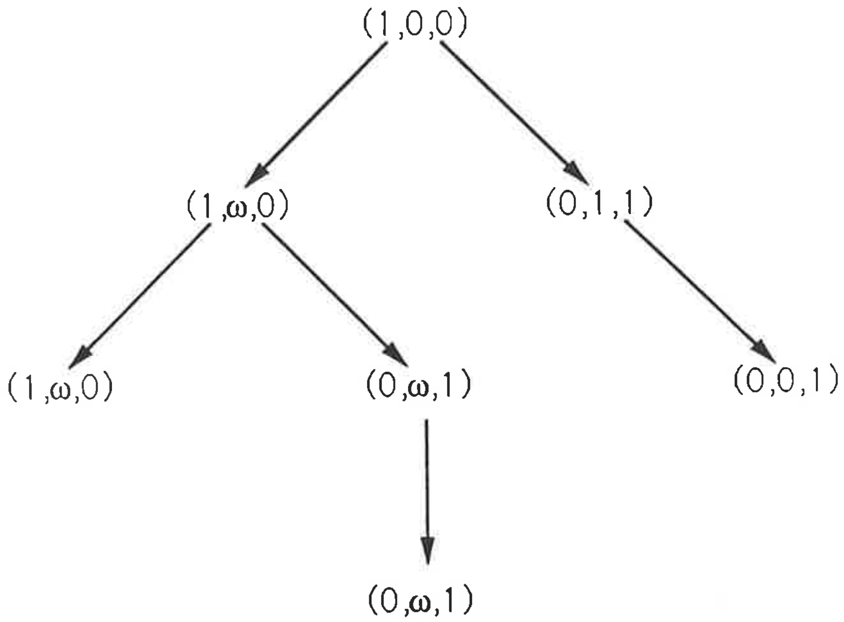

Figure 2.2: The coverability tree of the SPN in Figure 2.1.

from p_j when t_i fires). a_{ij} can therefore be interpreted as the total change in the number of tokens in p_j when t_i fires. Convention in the PN literature has A^T as the incidence matrix. However, Murata [76] points out that incidence matrices are also defined for directed graphs, but unfortunately the directed graph incidence matrix corresponds to A rather than A^T. In a small attempt to encourage consistency in the operations research literature we shall use the same notation.

Transition t_i is enabled in marking $\mathbf{m}$, iff,

$$a_{ij}^- \le \mathrm{m}(p_j), \quad 1 \le j \le r.$$

This is an expression of the pre-condition for firing t_i. Define $\mathbf{u}_n$, an $s \times 1$ column vector, to be the nth firing vector, indicating which transition fires at the nth firing in a firing sequence σ. A one, in the ith position, indicates that t_i fires at the nth firing. The remaining entries are equal to zero. Assume that the nth firing is feasible, then, from Murata [76], the state equation for N may be written as,

$$\mathbf{m}_n = \mathbf{m}_{n-1} + A^T \mathbf{u}_n, \quad n = 1, 2 \ldots \tag{2.1}$$

where $\mathbf{m}_n$ is an $r \times 1$ column vector.

Additional results follow directly from Equation (2.1). For example,

$$\mathbf{m}_n = \mathbf{m}_0 + A^T \sum_{k=1}^{n} \mathbf{u}_k, \tag{2.2}$$

which says that marking $\mathbf{m}_n$ is reachable from $\mathbf{m}_0$ through a firing sequence $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$.

A non-negative integer solution, $\mathbf{x} \neq \mathbf{0}$, of the equation,

$$A^T \mathbf{x} = \mathbf{0},$$

is called a T-invariant. For recoverable PNs, the T-invariant indicates how often, starting from some marking, each transition has to fire to reproduce that marking.

A non-negative integer solution, $\mathbf{y} \neq \mathbf{0}$, of the equation,

$$A\mathbf{y} = \mathbf{0},$$

is called an S-invariant. It is a weighting factor, such that the inner product of the S-invariant and any marking in the reachability graph, is a constant.

To gain insight into these invariants consider Equation (2.2). Let $\mathbf{x} = \sum_{k=1}^{n} \mathbf{u_k}$. $\mathbf{x}$ is an $s \times 1$ column vector of non-negative integers and is commonly called the firing count vector. The ith entry denotes the number of times that t_i must fire to transform $\mathbf{m}_0$ into $\mathbf{m}_n$. With this substitution Equation (2.2) becomes,

$$\mathbf{m}_n = \mathbf{m}_0 + A^T\mathbf{x}. \tag{2.3}$$

Therefore, a T-invariant, $\mathbf{x}$, with $A^T\mathbf{x} = \mathbf{0}$, implies that Equation (2.3) becomes,

$$\mathbf{m}_n = \mathbf{m}_0,$$

which is consistent with the definition of a T-invariant given earlier.

Now consider the inner product of $\mathbf{m}_n$, from Equation (2.3), and an S-invariant $\mathbf{y}$,

$$\mathbf{m}_n^T\mathbf{y} = \mathbf{m}_0^T\mathbf{y} + \mathbf{x}^T A\mathbf{y}. \tag{2.4}$$

Therefore, an S-invariant $\mathbf{y}$, with $A\mathbf{y} = \mathbf{0}$, implies that Equation (2.3) becomes,

$$\mathbf{m}_n^T\mathbf{y} = \mathbf{m}_0^T\mathbf{y} = \text{constant}.$$

Implying that the net is conservative with respect to the weighting factor $\mathbf{y}$.

This method is limited because of the non-deterministic nature of PNs, and because the solutions must be found as non-negative integers. a_{ij}^+ and a_{ij}^- do not cater for probabilistic routeing in the PN, therefore we are restricted to deterministic PNs when using this technique. In Section 3.3 we suggest a technique which overcomes this problem.

A third technique is called the reduction technique as it physically reduces the PN. To aid in the analysis of a large system, it is often possible to reduce the model to a simpler one. For example, an implicit place is one whose addition to (or elimination from) the net does not modify the firing sequence. Note that an implicit place has also been called a redundant place in the PN literature. A modified net can be obtained by removing such a place along with it's input and output arcs. We use the definition of an implicit place from Silva and Colom [92].

Definition 2.10: Implicit Place

A place, p_i, is an implicit place of a PN, N, with initial marking $\mathbf{m}_0$, if there exists an $r \times 1$ vector, $\mathbf{y}$, of positive rationals ($\mathbf{y} \neq \mathbf{0}$) and a rational, μ, such that for any $\mathbf{m} \in R(N, \mathbf{m}_0)$:

1. $\mathrm{m}(p_i) \geq \mathbf{m}^T\mathbf{y} + \mu, \quad ||\mathbf{y}|| \subseteq \mathcal{P} \setminus \{p_i\}$.

2. $\forall t_k \in T, \; \forall p_j \in P \setminus \{p_i\}, \quad \mathrm{m}(p_j) \geq a^-_{kj} \;\Rightarrow\; \mathrm{m}(p_i) \geq a^-_{ki}$.

$||\mathbf{y}|| = \{p \mid y(p) > 0\}$ and $||\mathbf{y}||$ is the set of implying places. The second part of condition 1 implies that $y(p_i) = 0$. Note that this definition can include p_i as an implicit place if it is a sole input place to some transition. This is contrary to the idea that the removal, or addition, of p_i should not effect the firing sequence of the PN. For this reason we exclude any places that are sole input places for any transitions.

To aid in the understanding of this definition we give our interpretation. If $m(p_i) \geq \mathbf{m}^T\mathbf{y}, \; \forall \; \mathbf{m} \in R(N, \mathbf{m}_0)$, then let $\mu = 0$ and Condition 1 is satisfied. If $m(p_i) < \mathbf{m}^T\mathbf{y}$, for some $\mathbf{m} \in R(N, \mathbf{m}_0)$, then let,

$$-\mu = \sup_{\mathbf{m} \in R(N, \mathbf{m}_0)} \left\{ \mathbf{m}^T\mathbf{y} - m(p_i) \right\},$$

and Condition 1 is satisfied. If the supremum does not exist, we cannot find μ given our choice of $\mathbf{y}$ to satisfy Condition 1 and therefore p_i is not implicit. This is a technical condition ensuring that if the weighted number of tokens in the implying places is unbounded, the difference between the number of tokens in the implicit place and the implying places is bounded.

Condition 2 is an instructive condition. For every marking it asks the question, "for every transition t_k, does place p_i obey the pre-condition for firing, given that another place satisfies the pre-condition for firing ?". If p_i does, then it is an implicit place.

Alternative reduction techniques involve fusion of a series of places, fusion of a series of transitions, fusion of parallel places and fusion of parallel transitions. For further information on these methods refer to Berthelot, Roucairol and Valk [10], Berthelot [9], Johnsonbaugh and Murata [48] and Chu [22]. Reduction is particularly relevant in SPNs, and so we apply the technique to the examples in Chapter 3.

2.1.3 Structural Properties

The invariants discussed in the last section have been shown to be powerful tools for studying, not only the behavioural, but also the structural properties of PNs (see Reisig [86] and Murata [76]). The structural properties of a PN are those that depend on the topological structure and not on any initial token distribution. Consequently they are functions of only the incidence matrix, A, which contains the information on the multiplicity of the input and output arcs of the PN. These properties are listed below with their corresponding methods of analysis.

Definition 2.11: Structurally Live

A PN is structurally live, if there exists a live initial marking. This is a difficult property to check for general PNs. However, it can be shown that a subclass of PNs, namely the marked graph, is structurally live (see Murata [76] for details).

Definition 2.12: Completely Controllable

A PN is completely controllable, if any marking is reachable from any other. If a PN is completely controllable then $\text{Rank}(A) = r$. For the proof of this result refer to Murata [76]. This property is independent of the initial marking, whereas, the property of recoverability is not. Consequently, completely controllable is a stronger property.

Definition 2.13: Structurally Bounded

A PN is structurally bounded, if for an arbitrary finite initial marking, $\mathbf{m}_0$, the PN is bounded. If there exists an $r \times 1$ column vector, $\mathbf{y} > \mathbf{0}$, of integers such that $A\mathbf{y} \leq \mathbf{0}$, then from Equation (2.4) $\mathbf{m}_n^T\mathbf{y} \leq \mathbf{m}_0^T\mathbf{y}$. Therefore $\text{m}(p)$, the number of tokens in place p is bounded by,

$$\mathrm{m}(p) \leq \frac{[\mathbf{m}_0^T \mathbf{y}]}{\mathrm{y}(p)},$$

where $\mathrm{y}(p)$ is the pth entry of $\mathbf{y}$, $\forall\, p \in \mathcal{P}$.

Definition 2.14: Repetitive

If there exists an initial marking $\mathbf{m}_0$, and a firing sequence from $\mathbf{m}_0$ where every transition fires infinitely often, the PN is repetitive. An equivalent result is that there exists an $s \times 1$ integer column vector, $\mathbf{x} > \mathbf{0}$, such that $A^T\mathbf{x} \geq \mathbf{0}$. If we substitute $A^T\mathbf{x} \geq \mathbf{0}$ into Equation (2.3), then there exist markings, $\mathbf{m}_0$ and $\mathbf{m}_n$ such that, $\mathbf{m}_n - \mathbf{m}_0 \geq \mathbf{0}$, that is $\mathbf{m}_n \geq \mathbf{m}_0$. Therefore, given that there are enough tokens in $\mathbf{m}_0$ (and therefore $\mathbf{m}_n$), so that the tokens will not be exhausted during the firing sequence, every transition can fire infinitely often in the firing sequence. Repetitive is a stronger property than coverability since it is independent of the reachability graph, whereas coverability is not.

Definition 2.15: Consistent

If there exists an initial marking $\mathbf{m}_0$, and a firing sequence which connects $\mathbf{m}_0$ back to $\mathbf{m}_0$, such that every transition fires at least once, the PN is said to be consistent. An equivalent result is that there exists an $s \times 1$ integer column vector, $\mathbf{x} > \mathbf{0}$, such that $A^T\mathbf{x} = \mathbf{0}$. Applying the same argument as above, substitution into Equation (2.3) yields $\mathbf{m}_n - \mathbf{m}_0 = \mathbf{0}$. Therefore, given that there are enough tokens in $\mathbf{m}_0$, the firing sequence can be executed. Note that if the net is consistent with this initial marking then it will also be recoverable.

Definition 2.16: Structurally Implicit

A place, p, is a structurally implicit place, if there exists an $r \times 1$ vector, $\mathbf{y} \geq 0$, where $||\mathbf{y}||$, given in Definition 2.10, obeys $||\mathbf{y}|| \subseteq \mathcal{P} \setminus \{p_i\}$, such that,

$$A\mathbf{y} \leq A(p),$$

where $A(p)$ is the column of the incidence matrix corresponding to place p. $A\mathbf{y}$ gives the total weighted change in the number of tokens in the implying places.

$A(p)$ gives the total change in the number of tokens in p. Therefore, the condition requires that more tokens are deposited into place p, than the implying places, and less tokens are removed from p, than the implying places.

Note that it is possible to make a structurally implicit place, p, implicit. This is achieved by adding tokens to p in the initial marking $\mathbf{m}_0$, so that p is implicit in the new initial marking.

2.2 Subclasses of PNs

Some subclasses of PNs include:

1. State machines: A state machine is a PN such that each transition has exactly one input and one output place. This is analogous to a queueing network.

2. Marked Graphs: A marked graph is a PN such that each place has exactly one input transition and exactly one output transition.

3. Free Choice PNs: In a free choice net any conflict situation must involve only one place.

4. Simple PNs: A simple PN requires that each transition has at most one input place that is shared with another transition. (For details see Murata [76]).

2.3 Time Extended Models

A summary of PNs with deterministic delays and related works follows in Section 2.4. A summary of SPNs follows in Section 2.5 (for a review on both types of net see Wong and Henderson [100] and Marsan [60]). Although timed PNs can be treated as SPNs, with degenerate probability distributions, there are conflict problems which arise with deterministic delays that require special consideration.

Consequently, although there are sections of this thesis in which SPNs can have deterministic delays, we will treat them separately in the following sections.

As mentioned in Chapter 1, the firing delays may be associated with either the transitions or the places in the PN. In the following section, we separate the two types of time-extended PNs and discuss them individually.

2.4 Deterministic Delays

2.4.1 Timed Transitions

One of the first works to involve time delays in PNs can be attributed to Ramchandani [81]. The model required that the transitions had to wait a deterministic length of time between becoming enabled and firing. This is called a Timed PN (TdPN).

Another model of importance, first introduced by Merlin [68] and later used by Merlin [69] and Merlin and Farber ([70], [71]), called a Time PN (TPN) introduced a firing interval associated with each transition. The conditions for enablement remained unchanged from the classical structure, however the firing had to occur somewhere in the interval $[t_{\text{min}}, t_{\text{max}}]$, a minimum and maximum time, beginning from the moment the transition was enabled. In both the first and second model, the firing of a transition is instantaneous, that is, the tokens are removed from all of the input places and are then deposited immediately into their output places. Recently, Menasche [67] noted that a TPN can be considered a generalisation of a TdPN.

A variation is presented in Razouk [83]. In this model transitions are allocated two times, an enabling time, t_e, and a firing time, t_f. The transition waits for a time t_e after it is first enabled, before it begins firing, at which point it absorbs the tokens from its input places. The tokens remain absorbed, while the transition continues to fire for a period t_f. After the transition finishes firing, tokens are deposited into its output places. This model differs from the first two because tokens disappear for a period of time, and because two times, t_e and t_f, are required to execute the firing process.

Since groups of transitions can become enabled and fire concurrently, when using deterministic lifetimes, the following authors address the issues involved in analysing models with conflicting transitions.

Zuberek [102], assigns a probability of firing to every transition in a conflict set, with these probabilities summing to one. The firing time is restricted to be a non-negative integer and the nets are restricted to be safe and free-choice. This ensures, that among any set of conflicting transitions, either all or none are enabled, and firing one disables the firing of all other transitions in the set. No subset of a set of conflicting transitions can be enabled in any marking. The disabled transitions follow the pre-emptive repeat discipline.

In the paper described above, and then later in ([103],[104],[105],[106]), Zuberek introduces an extension called escape arcs. They are represented by an arc, from a place to a transition, with a black circle on the end connected to the transition, rather than an arrowhead. The escape arcs remove a token, from the place connected to the escape arc, but no tokens from these input places are deposited into the transitions output places. These arcs can be used to model timeout, without having to interrupt a transitions firing time, a property required in a realistic timeout.

Zuberek's use of time, assumes that when a transition is enabled it starts firing by absorbing its input tokens. The tokens remain absorbed until the end of the firing time, at which point they are released and deposited into the output places. Since the firing time is deterministic, the underlying process is not memoryless and therefore not Markovian. In order to make the equilibrium solution tractable, it is necessary to record the history of the transition firings, in other words, record the residual lifetimes of the transitions. Zuberek follows the traditional "supplementary variable" technique, by supplementing the state description with a vector of residual lifetimes, and so creates a discrete time Markov chain with an extended, but finite, state space.

To aid in evaluating the performance of a model, Zuberek then creates a reduced state space, retaining the original supplemented states only if they are connected to more than one state in the state space. The remaining states are aggregated, and the time delays taken to traverse these aggregates is evaluated by adding the time delays for each state within the aggregate. The reduced state space is then used to measure throughput and resource allocation.

Razouk [83] and Razouk and Phelps [84], introduce a different version of time-extended PNs by incorporating Merlin's $t_{\min}$ time as an enabling time. This way a transition in the net must remain enabled, for a period of time, before it can fire. The enabling time is also restricted to be a non-negative integer. The transition then follows the same firing rules as in Zuberek's model. The firing time is also restricted to be a non-negative integer. With the introduction of enabling times, they include a vector of residual enabling lifetimes as well as the vector of residual firing times, to the state description, in order to create a finite state, discrete time, Markov chain. They relax the free-choice restriction but retain the restriction that conflicting transitions be mutually disabling. They refer to each set of conflicting transitions as a conflict set, defined in the following way: Every t_i belongs to exactly one conflict set, C, such that,

$$C = \{\, t_j |\, \mathbf{I}(t_i) \bigcap \mathbf{I}(t_j) \neq 0\}.$$

With each transition t_i in the conflict set, let f_i be the firing frequency. The next marking probabilities are evaluated by working out the probability of the first transition to fire. Transition t_i fires first, and so defines the next marking, with probability given by,

$$\frac{f_i}{\sum\limits_{j|t_j \text{ fireable in C}} f_j}.$$

Holliday and Vernon [43] relax all net restrictions, except, that the state space be finite and call their model a generalised timed PN (GTPN). As opposed to Zuberek and Razouk and Phelps, who allow only non-negative integer valued firing times, this model allows non-negative real valued firing times. They introduce a generalised conflict set, which is the transitive closure of the following relation defined on $\mathcal{T}$,

$$t_i \text{ conflicts with } t_j \text{ iff } \mathbf{I}(t_i) \bigcap \mathbf{I}(t_j) \neq 0.$$

Note, that by including the transitive closure, this definition generalises the definition of the conflict sets in Razouk and Phelps. They identify a set of transitions, called a local maximal, in the generalised conflict set, which when fired, disables the remaining transitions in the set. They define the probability of such a set, s_i, firing, as

$$P(s_i) = \frac{Comb(s_i) \prod_{k|t_k \in s_i} f_k}{\sum_{s_j} Comb(s_j) \times \prod_{n|t_n \in s_j} f_n}, \tag{2.5}$$

where $Comb(s_i)$ is the number of ways tokens may be removed from the input places. The next state probability, is then simply the product of all the probabilities of local maximals, s_i, selected to fire,

$$\prod_{s_i \text{ selected}} P(s_i).$$

With this model the authors compute performance estimates. They note, that the times at which state changes occur, form an embedded, discrete time Markov chain. They then proceed, by aggregating the states, to form a reduced state space, as do Zuberek and Razouk and Phelps, and from there find performance estimates of the reduced state space.

In Woo, Phelps and Sidwell [101], the conflict resolution probabilities are determined using simple random sampling without replacement. They call the next marking the outcome, and a transition firing an event. For example, consider the conflict situation portrayed in Figure 2.3. They show how to find the probabilities for the outcomes, consisting of the events such as $\{t_1, t_2\}$ and $\{t_1, t_1\}$. They

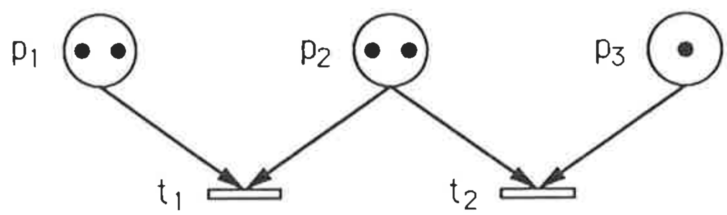

Figure 2.3: An example of conflict.

evaluate these probabilities, using simple random sampling, where the sample is considered to be an ordered sequence of transition firings. If each sample consists of one outcome, the probability model is termed simple. Strictly, simple random sampling describes a probability model that gives equal probability to all of its ordered samples. Woo, Phelps and Sidwell [101], deviate slightly by defining the probabilities on the discrete events comprising the sample space, rather than on the samples. A sample's probability is therefore determined by the product of it's discrete events' probabilities. An outcome probability is then the sum of the probabilities of the samples that comprise the outcome. Say, we wish to evaluate the probability of the sequence of transition firings $\{t_1, t_2\}$. Since each event is equally likely,

$$P(\{t_1, t_2\}) = P(\{t_1, t_2\}) + P(\{t_2, t_1\}) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times 1.$$

Also,

$$P(\{t_1, t_1\}) = \frac{1}{2} \times \frac{1}{2}.$$

2.4.2 Timed Places

Sifakis [91], first proposed a PN model whereby a delay is associated with each place. This model is called a Timed Place Transition Net (TPTN). When a token is deposited in a place, it stays there for a duration of time, after which it becomes available and is allowed to participate in the enabling of a transition. The transitions have no delays associated with firing and therefore fire instantaneously as in the classical PN model.

2.5 Stochastic Petri Nets

In this section, we will consider SPNs where the firing time of a transition is a random variable drawn from a continuous distribution. In these nets, the conflict problems which arise, take on different aspects from those of TPNs. Within a conflict set, transitions can be pre-empted by the removal of all or part of its input bag. Most authors assume that when the pre-empted transition is again enabled, the new lifetime is drawn from the same distribution (pre-emptive repeat different). Other possible choices are to restart with the same lifetime (pre-emptive repeat identical) or to restart the lifetime from where it was pre-empted (pre-emptive resume). The overall effect, is that the lifetimes themselves dictate which transitions will fire. Marsan, Balbo, Bobbio, Chiola, Conte and Cumani ([61], [62]), describe these policies and their impact on modelling. We will give an account of their papers, in more detail, in Section 2.5.1.

2.5.1 Timed Transitions

Molloy [73] initiated the study of performance evaluation in SPNs with negative exponential distributions, by noting that these SPNs are isomorphic to homogeneous Markov chains. The key factors, allowing the isomorphism to be constructed, are the countability of the markings and the memoryless property of the negative exponential distributions. That is, if a transition fires to change the marking, the distribution of time remaining on the other enabled transitions' lifetimes is the same as if they were first enabled. It is therefore possible to apply the standard techniques in Markov chains. As a precursor to Chapter 4, where we examine the global balance equations for insensitivity, the following example illustrates how to find the equilibrium distribution of the SPN given in Figure 2.4.

Example 2.1

The markings in the reachability graph are,

$$\begin{aligned}
\mathbf{m}_1 &= \{1,0,0,0,0\},\\
\mathbf{m}_2 &= \{0,1,1,0,0\},\\
\mathbf{m}_3 &= \{0,1,0,0,1\},\\
\mathbf{m}_4 &= \{0,0,1,1,0\},\\
\mathbf{m}_5 &= \{0,0,0,1,1\}.
\end{aligned}$$

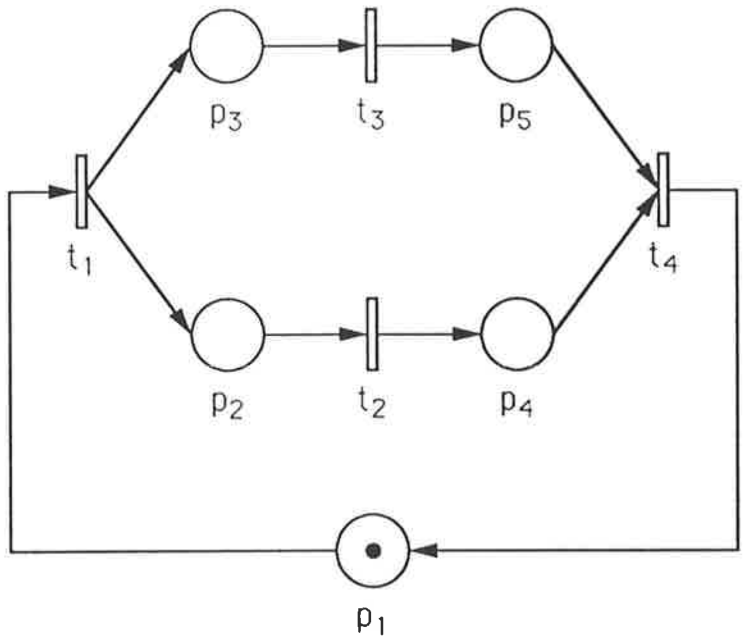

Figure 2.4: A simple SPN representing concurrent activities.

Let transitions t_i fire with rate λ_i, for $1 \leq i \leq 4$, where λ_i^{-1} is the mean of the negative exponential distribution. The equilibrium distribution, is found by solving the set of global balance equations given below, where $\pi(\mathbf{m}_i)$ is the equilibrium probability of being in state $\mathbf{m}_i$.

$$\pi(\mathbf{m}_1)\lambda_1 = \pi(\mathbf{m}_5)\lambda_4. \tag{2.6}$$

$$\pi(\mathbf{m}_2)(\lambda_2 + \lambda_3) = \pi(\mathbf{m}_1)\lambda_1. \tag{2.7}$$

$$\pi(\mathbf{m}_3)\lambda_2 = \pi(\mathbf{m}_2)\lambda_3. \tag{2.8}$$

$$\pi(\mathbf{m}_4)\lambda_3 = \pi(\mathbf{m}_2)\lambda_2. \tag{2.9}$$

$$\pi(\mathbf{m}_5)\lambda_4 = \pi(\mathbf{m}_3)\lambda_2 + \pi(\mathbf{m}_4)\lambda_3. \tag{2.10}$$

$$\sum_{i=1}^{5} \pi(\mathbf{m}_i) = 1. \tag{2.11}$$

The solution to these equations is given by,

$$\pi(\mathbf{m}_2) = \pi(\mathbf{m}_1)\frac{\lambda_1}{(\lambda_2 + \lambda_3)}, \tag{2.12}$$

$$\pi(\mathbf{m}_3) = \pi(\mathbf{m}_1)\frac{\lambda_1\lambda_3}{\lambda_2(\lambda_2 + \lambda_3)}, \tag{2.13}$$

$$\pi(\mathbf{m}_4) = \pi(\mathbf{m}_1)\frac{\lambda_1\lambda_2}{\lambda_3(\lambda_2 + \lambda_3)}, \tag{2.14}$$

$$\pi(\mathbf{m}_5) = \pi(\mathbf{m}_1)\frac{\lambda_1}{\lambda_4}, \tag{2.15}$$

subject to Equation (2.11). Using the equilibrium distribution, the throughput or average delay (using Little's result) can be evaluated.

Even though we stress the use of time extended PNs in this thesis, one must recognise the usefulness of classical PNs in the modelling of interactions between activities, which are of a logical nature and which do not consume time. The transitions which fire in zero time are called immediate transitions and therefore must have a higher priority than timed transitions. Nets which have both immediate transitions and transitions which fire according to negative exponential distributions, are called Generalised SPNs (GSPNs) and are discussed in Marsan, Conte and Balbo [65]. As with deterministically timed transitions, in Section 2.4.1, it is necessary to have some way of resolving conflict among a set of concurrently enabled immediate transitions. This is achieved, by what is called a random switch in Marsan, Conte and Balbo [65]. The random switch is defined on the set of immediate transitions and an associated probability distribution. The definition of a random switch is dependent on the marking, therefore, the reachability graph of the GSPN must be constructed before it can be assigned. Such an assignment depends on the local and already identified behaviour of the model, and so a formal definition of a random switch is not possible, but is illustrated by the use of examples in Marsan, Conte and Balbo [65]. Marsan, Balbo, Chiola and Conte [63], identify a class of GSPNs in which the random switch can be defined at the net level, implying that the reachability graph does not have to be constructed. This is achieved by using a priority structure, for the immediate transitions, by assigning weights, w_i, to each transition, t_i, in the set of simultaneously enabled immediate transitions, C. Following Razouk [83] and Razouk and Phelps [84], the probability that t_i fires in this set, C, is then given by,

$$\frac{w_i}{\sum\limits_{j|t_j \in C} w_j}.$$

Again, these weights depend on the local behaviour of the model. If a mixture of immediate and negative exponentially distributed transitions are enabled concurrently, the immediate transition, chosen by the random switch, will take priority. If a set of negative exponentially distributed transitions are enabled concurrently, the first to fire will be given by the probability,

$$\frac{\lambda_i}{\sum\limits_{k} \lambda_k},$$

where, λ_i is the firing rate of transition t_i, and the sum in the denominator is taken over all the enabled transitions.

The evaluation of equilibrium probabilities, for the markings in the GSPN, is not as straight forward as the approach given in Example 2.1. This becomes clear, if we note that the reachability set of a GSPN is a subset of the reachability set of the associated PN. The reason, is due to the firing priority introduced with the immediate transitions, resulting in not all of the states being reachable. Marsan, Conte and Balbo [65] call markings that enable negative exponentially distributed transitions, tangible markings, and markings which enable immediate transitions, vanishing markings. The approach taken, is to make use of the one to one correspondence between the GSPN markings and the states in a stochastic point process. Further, a stationary embedded Markov chain is recognised within the stochastic point process. The analysis of the embedded Markov chain allows the evaluation of the equilibrium probabilities of the GSPN markings.

An obvious extension to introduce now, is to mix different types of firing distributions in the net in order to increase modelling power.

Natkin [77] and Bertoni and Torelli [11], were the first authors to move in this direction. They restrict the SPN structure to be a semi-Markov process (to be discussed in Section 4.1). Although this approach is appropriate in some cases, it fails when parallel activities need to be modelled, because of the lack of memory necessary after every generally distributed firing time ends.

As mentioned at the start of this section, the Markovian analysis performed on SPNs requires the memoryless property. Only negative exponential and the discrete time analogue, the geometric distribution, have this property. As a new marking is entered, the geometric distribution has the property that the probability that a particular enabled transition will fire, in the next discrete time interval, is the same after any multiple of that interval. Results concerning the combination of deterministic transitions, with geometrically timed PNs, have been reported by Molloy [74] and Holliday and Vernon [43].

In Molloy [74], the deterministic firing times are restricted to be non-zero multiples of some unit time step, and all the conflicting transitions which fire with deterministic delays, must have equal probability of firing. In the paper by

Holliday and Vernon [43], discussed in Section 2.4.1, we noted that the GTPN allowed firing times to be any non-negative real value. This implies that the GTPN can represent geometric holding times, thus bridging the gap between the TPN and SPN models.

In Bobbio and Cumani [12], probabilistic branching (through the use of immediate transitions) and general distributions, are mixed in the SPN. However, the general distributions cause the resulting associated stochastic process to be no longer Markovian. They use phase type distributions (refer to Section 4.2.2 for details) to approximate these general distributions. After expanding the state space, with the appropriate SPN representation of the phase type distribution, a continuous time, discrete state, homogeneous Markov process is constructed, which can then be analysed in the usual way.

Dugan, Geist, Nicola and Trivedi [28], introduce more generality for modelling, by classifying a type of SPN which includes general distributions and negative exponential distributions. They add inhibitor arcs to the PN structure and introduce probabilistic branching. An inhibitor arc, from a place to a transition, has a small circle, rather than an arrowhead, connected to the transition. The firing rule is changed in the following way. A transition is enabled when tokens are present in all of its (normal) input places and no tokens are present in the inhibiting input places. When a transition fires, the tokens are removed from all the normal (without inhibitor arcs) input places and deposited in the output places as usual, with the number of tokens in the inhibiting input place remaining zero. A probabilistic arc, from a transition to a set of output places, deposits a token in one of the places in the set. The choice of which place receives the token is determined by the probabilities assigned to each arc. The net is called an Extended SPN (ESPN). The structure of the ESPN is restricted to be of the following form:

1. The firing time of transitions, which are enabled concurrently, must be exponentially distributed.

2. The firing time of an exclusive transition (that is, no other transition is enabled at the same time) may be generally distributed.

3. A transition in conflict (that is when the firing of a transition disables the firing of all others) may be generally distributed. It is assumed that the transition, when re-enabled, has a firing time independent of, but identically distributed to, the pre-empted firing time.

Dugan, Geist, Nicola and Trivedi [28] also consider marking amalgamation for ESPNs, where at most one generally distributed transition is enabled in any marking, and where each marking is visited only once. They suggest the amalgamation procedure, of combining all markings which have a common generally distributed transition enabled. In Chapter 4, we derive further results along these lines, using the theory of insensitivity applied to marking amalgamation.

Marsan and Chiola [64], consider an extension of the GSPN model by incorporating deterministic firing times. They refer to this model, including the combination of deterministic, exponential and immediate firing times, as a Deterministic and Stochastic PN (DSPN). The restrictions placed on the DSPN are, that exclusive transitions may have any firing time, transitions in conflict may have any firing time, and not more than one transition with deterministic firing time may be enabled concurrently, unless the set of concurrently enabled transitions satisfy the following conditions.

1. The set of concurrently enabled transitions (which may include both deterministic and exponentially distributed firing times) must be enabled at the same instant.

2. Each of these transitions can also be connected to a chain of transitions (that is a transition connected to a place connected to another transition etc.), however, the type of firing distribution must be the same for each transition in the chain.

The time points of the process are then taken to be the instant that the deterministic transition starts firing, or has just been disabled. At the end of each interval, the process will change states, probabilistically, according to whether or not an exponentially distributed transition (or a set of exponentially distributed transitions) fired within the interval.

Marsan, Balbo, Bobbio, Chiola, Conte and Cumani's ([61], [62]) definition of their model, includes a specification of the selection of the next transition to fire, and they refer to these specifications as an execution policy. They use supplementary variables to record the spent lifetimes in each marking. Note that supplementary variables are not required, if the execution of the SPN only relies on the firing of exclusive transitions. However, for combinations of competitive and concurrent transitions it is necessary to record spent lifetimes, for the case when transitions pre-empt the firing of others. This publication gives a good account of the impact of the execution policies, on the analysis of SPN models which have generally distributed firing times. We will now give a summary of these policies and the effect they have on the SPN.

The most common way of selecting which transition will fire, from a set of enabled transitions, is to choose the one whose firing time is statistically the minimum. This is what is referred to as the *race* policy model. The other policy, known as *preselection*, chooses the next transition to fire according to a probability mass function.

The supplementary variables are recorded in three different ways.

1. If a transition is pre-empted, the supplementary variable is set to zero, when once again enabled. This is termed pre-emptive repeat.

2. If a transition is pre-empted, the supplementary variable is set equal to the spent lifetime and when re-enabled, the firing will resume from where it was pre-empted. This is termed pre-emptive resume.

3. If a transition is pre-empted, the supplementary variable is set to zero, unless it remains enabled in the next marking, when it is set to the spent lifetime. From Marsan, Balbo, Bobbio, Chiola, Conte and Cumani ([61], [62]) this is called age-enabling.

Where possible, we have tried to remain consistent with the literature in other areas of operations research. In this case we follow the nomenclature of the literature on queueing networks. The concept of age-enabling, is the status quo assumption in stochastic systems and is not usually given a special title.

Marsan, Balbo, Bobbio, Chiola, Conte and Cumani ([61], [62]) consider combinations of both the race and pre-selection policy, together with the three alternative choices in recording the supplementary variable, are then presented and the corresponding resultant process is described. For both the pre-selection and race policies, with pre-emptive repeat, the process is a SMP with well known solution techniques. For the race policy, with pre-emptive resume or age-enabling, the corresponding process is no longer a SMP and the authors suggest the use of supplementary variables, to extract the equilibrium distribution.

Marsan, Balbo, Bobbio, Chiola, Conte and Cumani [62] also mention the potential for introducing marking dependent firing times. In addition, the paper describes how to approximate a generally distributed firing time, by using a phase type distribution, substituted directly into the state space rather than at the net level. The reason they do this, is to avoid enlarging the reachability tree unnecessarily (for more details see Section 4.2.2).

Henderson and Taylor [41], discuss discrete time SPNs, which can have mixtures of negative exponentially distributed and generally distributed transitions, that form a SMP. Both firing and enabling intervals are used in their model, by assuming alternating enabling and firing time points. They define a class of net, which has an extended product form solution, based on the work of Henderson, Lucic and Taylor [37], (see Chapter 3). This will be discussed in more detail in

Chapter 5.

This is a small sample of the different types of suggested models for SPNs used in the literature, more detail will follow in later chapters.

2.5.2 Timed Places

The condition/event interpretation of nets with timed places is as follows. The firing of a transition, which inputs into a place, represents the start of execution. The presence of tokens in a place, represents a condition of the process in execution, and the firing of output transitions of a place, represents the completion of a process.

Wong, Dillon and Forward [99] consider PNs with timed places (TPPN). These nets are restricted to be safe, marked graphs, in order to make the following analysis tractable. They decompose the marked graph into elementary loops, such that each place and transition has one input and one output arc. These elementary loops are found by the S-invariants. The method then describes how to find the distribution of the cycle time of the marked graph. They define a random variable, X_i, to be the loop time of the elementary loop, i. The cycle time is then the maximum of the loop times. Hence, the distribution function of the cycle time is the product of the distribution functions of the loop times. Such a result is shown to be useful in providing upper and lower bounds on the execution times of component modules.

Chapter 3

Closed Form Solutions

The major problem in finding the performance measures of SPNs, is the need to work with the equilibrium distribution based on the reachability graph. The size of the reachability graph grows exponentially, with both the number of places and the initial token distribution, making such analysis very difficult for even moderately sized SPNs. Since this creates such a major problem, we aim to devote a large proportion of this thesis to finding some form of solution technique, which eliminates the problem of having to generate the reachability graph. In this chapter, we are concerned with finding a closed form solution for SPNs, and in Chapter 5, we give a technique for aggregation and disaggregation which will hopefully shed some light onto finding equilibrium distributions, without necessarily generating the entire reachability graph.

A similar problem in queueing theory, lead researchers to look for a simplified method of finding the equilibrium distribution. The investigation produced approximation techniques, and a class of Markov processes for which there exists a product form equilibrium solution, (see Jackson ([45], [46]), Kelly [50] and Baskett, Chandy, Muntz and Palacios [8]). Along with this product form solution also came the creation of computationally efficient algorithms for their evaluation (see Buzen [14] and Reiser and Lavenberg [85]).

When there is a product form solution, the joint equilibrium distribution of the network involves a product, over the nodes of the network of the equilibrium distribution of the individual nodes, treated as though in isolation (a normalising constant is involved in the expression which normalises the invariant measure). Product form solutions are of great interest, because they avoid the direct numerical solution of the global balance equations, which is computationally expensive. While the calculation of the normalising constant may not be trivial, the evaluation of product form solutions is still computationally and analytically preferable, to solving very large systems of linear equations.

3.1 Product Form Solutions

Product form equilibrium solutions have their origin with Jackson [45] and Gordon and Newell [31]. Jackson [45], considered an open network of K, many server queues and one customer type. Let $s(i)$ be the number of servers at queue i, where $s(i)$ is an integer. Customers arrive at the network to queue i, according to a Poisson process with rate $\lambda(i)$, that is, according to the distribution function,

$$p(n \text{ arrivals in } (0,t)) = \frac{(\lambda(i)t)^n}{n!} e^{-\lambda(i)t}.$$

The service rate at queue i is negative exponentially distributed with mean $[\mu_i]^{-1}$.

After completion of service, customers leave queue i and move to queue j with probability $p(i,j)$, or leave the network with probability $q(i)$. Let $\mathbf{n} = (n(1), n(2), \ldots, n(K))$ be the state of the network. The joint equilibrium distribution for the state of the network is of the form,

$$\pi(\mathbf{n}) = B(K) \prod_{i=1}^{K} \zeta_i(n(i)), \tag{3.1}$$

where $\zeta_i(n(i))$ is the equilibrium distribution of the ith queue, treated as though in isolation and is given by,

$$\zeta_i(n(i)) = \begin{cases} \left[\dfrac{y(i)}{\mu_i}\right]^{n(i)} \dfrac{1}{n(i)!} & \text{if } 0 \le n(i) \le s(i), \\ \left[\dfrac{y(i)}{\mu_i}\right]^{n(i)} \dfrac{1}{s(i)!s(i)^{n(i)-s(i)}} & \text{if } n(i) \ge s(i), \end{cases} \tag{3.2}$$

where $\{y(i),\ 1 \leq i \leq K\}$ satisfy,

$$y(i) = \lambda(i) + \sum_{j=1}^{K} y(j)p(j,i). \tag{3.3}$$

$B(K)$ is the normalising constant, and, for a Jackson open network, is the product of the normalising constants for each queue in the network.

Jackson [46], extended this work to include service rates $\mu_i(n(i))$ where $n(i)$ is the number of customers at queue i. In this case,

$$\zeta_i(n(i)) = [y(i)]^{n(i)} \prod_{l=1}^{n(i)} [\mu_i(l)]^{-1}. \tag{3.4}$$

He also included an arrival rate which could depend upon the total number of customers in the network. Gordon and Newell [31], considered a closed network of queues of the type defined by Jackson [45]. The joint equilibrium distribution of this network was shown to be of the same form as Equation (3.1).

Whittle [97], referred to the network described by Jackson [45] as a migration process, rather than a network of queues, because no account is made of the positions of the individuals in the queues. When a service facility is ready to serve a customer, the choice of which customer to serve is random. Other later extensions, by authors such as Baskett, Chandy, Muntz and Palacios [8] and Kelly [49], include customers of different types, different service disciplines and generally distributed service times.

Jackson ([45], [46]), proved the product form result by substituting into the global balance equations. Whittle ([97], [98]), provided a valuable new view on the theory, by showing that the global balance equations could be broken down into partial balance equations, each of which were satisfied by the product form solution. Define the flux, from state $\mathbf{n}$ to state $\mathbf{m}$ to be $\pi(\mathbf{n})q(\mathbf{n}, \mathbf{m})$, where $q(\mathbf{n}, \mathbf{m})$ is the rate/intensity from state $\mathbf{n}$ to state $\mathbf{m}$. The global balance equations can then be interpreted as, the flux out of state $\mathbf{n}$ = flux into state $\mathbf{n}, \forall\ \mathbf{n}$. Whittle's partial balance equations are:

1. Flux out of $\mathbf{n}$, due to an external arrival = flux into $\mathbf{n}$, due to an external

departure. (3.5)

2. Flux out of $\mathbf{n}$, due to a departure from queue i = flux into $\mathbf{n}$, due to an arrival at queue i. (3.6)

Kelly [49], extended the theory further, by generalising the service rate at queue i to be of the form,

$$\mu_i(\mathbf{n}) = \frac{\phi(\mathbf{n} - \mathbf{e}_i)}{\phi(\mathbf{n})}, \tag{3.7}$$

where $\mathbf{e}_i$ is a vector, with a 1 in the ith position and zeros elsewhere, representing the state with a single customer at the ith queue. The joint equilibrium distribution becomes,

$$\pi(\mathbf{n}) = B(K)\phi(\mathbf{n}) \prod_{i=1}^{K} [y(i)]^{n(i)}. \tag{3.8}$$

When applied to networks of queues, product form originally referred to the product over the nodes of the network. The product form of Equation (3.8), is no longer a product over the nodes, rather it is a product of two functions. $\phi(\mathbf{n})$ is related to the service rate and the product of the $y(i)$'s is a function of the routeing of the network.

Standard assumptions in queueing theory, are that customers arrive, are served singly, and after service, are routed through the network in a manner which may depend on the state, but does not depend on the routeing of other customers. The state machine, which is analogous to a queueing network, has this property. However, these properties are rarely found in general SPNs. In general, tokens are fired in a batch, by a transition, and transformed into an output bag. Tokens are not routed independently of one another, since their movement through the net depends on the input bag to which they belong. In summary, the tokens move in batches, and the routeing of the individual tokens is correlated. Consequently, the theory on product form solutions has not been applicable to SPNs. Authors such as Walrand [95] and Henderson, Pearce, Taylor and Van Dijk [38], have found equilibrium distributions for queueing networks with batch movement and independent routeing. Henderson and Taylor [42], however, incorporated both batch movement and correlated routeing into their queueing network. In applying the same approach to SPNs, Henderson, Lucic and Taylor [37] found classes of SPNs with an extended product form solution. Generally, tokens in SPNs do not have a position in a place so, this extended product form solution is a solution to the migration process discussed earlier with, batch arrivals and batch services. The following work uses the approach developed by Henderson and Taylor [42], but differs in one significant aspect, which makes it more relevant to SPNs. The difference is illustrated in Example 3.5 in Section 3.3.

3.2 Description of the Model

Since the problem of finding a product form solution for SPNs has been approached by extending the present queueing network literature, the group of SPNs that can be analysed using this method, do not naturally fit into the standard SPN classifications. For example, the equilibrium distribution of arbitrary marked graphs can not be found. However, the theory unifies a number of aspects not collectively covered by general SPNs, such as probabilistic routeing, arbitrary initial markings, marking dependent firing rates, marking dependent blocking and coloured tokens. Although we do not classify the type of SPN for which the theory can be applied in standard terms, as we progress through this chapter, we will identify a new class which supports the extended product form solution. As a start, we note that SPNs need to be live, but not necessarily bounded, provided the underlying Markov Process is stationary.

Consider a SPN, with a finite set $\mathcal{P} = \{p_1, \ldots, p_{|\mathcal{P}|}\}$ of places, and a finite set, $\mathcal{T} = \{t_1, \ldots, t_{|\mathcal{T}|}\}$ of transitions. The Markov process representing the SPN, has markings $\mathbf{m} \in \mathbb{Z}^{|\mathcal{P}|}$, to denote the state when there are $m(i)$ tokens at place p_i. When the state is $\mathbf{m}$, transition $t \in \mathcal{T}$ has a negative exponentially distributed firing time, which is worked off at a state dependent rate $q(\mathbf{m}, t)$.

When transition t fires, tokens move from an input bag, $\mathbf{I}(t) \in \mathbb{Z}^{|\mathcal{P}|}$, to one of a set of possible output bags, $\mathbf{O}_j(t) \in \mathbb{Z}^{|\mathcal{P}|}$, with probability $p(\mathbf{I}(t), t, \mathbf{O}_j(t))$, where $\sum_j p(\mathbf{I}(t), t, \mathbf{O}_j(t)) = 1$. Thus the marking of the net changes from $\mathbf{m}$ to $\mathbf{m} - \mathbf{I}(t) + \mathbf{O}_j(t)$ with probability $p(\mathbf{I}(t), t, \mathbf{O}_j(t))$.

It is worth noting that there are a variety of SPN structures which can be transformed to give a net of the above form. Two examples follow.

1. If a set, $\mathcal{A}$, of transitions enabled together can fire simultaneously, as well as independently, we can alter the original SPN by adding a new transition to replace the firing time "clock" for the set, $\mathcal{A}$. The new transition will have input bag given by $\bigcup_{t \in \mathcal{A}} \mathbf{I}(t)$ and output bag $\bigcup_{t \in \mathcal{A}} \mathbf{O}(t)$, where $\mathbf{I}(t)$ and $\mathbf{O}(t)$ are input and output bags respectively for transition t. This new SPN is equivalent to the original. The reason for making such a technical change is to prevent transitions firing simultaneously and so keep the notation relatively simple.

2. Without loss of generality, we can assume that there is a one to one correspondence between input bags and transitions. Any SPN in which a set of transitions, $\mathcal{A} \subseteq \mathcal{T}$ have a common input bag, can be modelled by amalgamating the transitions in $\mathcal{A}$ into a single transition. Note, that whenever one of these transitions is enabled, all are enabled. For any state $\mathbf{m}$ in which the transitions in $\mathcal{A}$ are enabled, the firing time distribution of the new transition will be negative exponential, with parameter $\sum_{s \in \mathcal{A}} q(\mathbf{m}, s)$. The routeing probabilities for the next marking are weighted by the probabilities, $q(\mathbf{m}, t) \left[\sum_{s \in \mathcal{A}} q(\mathbf{m}, s) \right]^{-1}$, $t \in \mathcal{A}$, indicating which transition has fired.

Assumption 1

The above procedures can be performed on any SPN before analysis takes place. Thus, without loss of generality, we can assume henceforth, that the SPN is structured so that transitions fire at different times from one another, and no two transitions have the same input bag.

If t is not enabled in $\mathbf{m}$ then $q(\mathbf{m}, t)$ is equal to zero. Otherwise, we assume that $q(\mathbf{m}, t)$ is of the form,

$$q(\mathbf{m} + \mathbf{I}(t), t) = \frac{\varphi(\mathbf{m})}{\Phi(\mathbf{m} + \mathbf{I}(t))}\chi(\mathbf{I}(t)), \tag{3.9}$$

where $\varphi(\cdot)$, $\Phi(\cdot)$ and $\chi(\cdot)$ are arbitrary, but given non-negative functions.

It is worth pointing out the similarities of the two structures in Equation (3.7) and Equation (3.9). In Equation (3.7), $\mathbf{n}$ is the state of the process in which t is enabled, and $\mathbf{n} - \mathbf{e}_i$ is a pseudo state which is unobservable, indicating the state of the process when customer i has been served, and is in transit to its next destination. In Equation (3.9), $\mathbf{m} + \mathbf{I}(t)$ is a marking of the process and $\mathbf{m}$ is a pseudo state that is unobservable, indicating the state of the process when a batch of tokens, $\mathbf{I}(t)$, are removed by the firing of t, and have not yet been deposited into the corresponding output bag. Note also, that only one function is involved in Equation (3.7) and three different functions are involved in Equation (3.9), which therefore has a greater degree of freedom. The following examples illustrate the flexibility of this definition.

Example 3.1

Most SPN models analysed in the literature, assume that transition firing times are marking independent. In such a SPN, transition t fires with rate $q(t)$ (that is, has mean firing time $[q(t)]^{-1}$). This can be modelled using Equation (3.9), by choosing $\varphi(\cdot) = \Phi(\cdot) \equiv 1$ and $\chi(\mathbf{I}(t)) = q(t)$.

Example 3.2

The functions $\varphi(\cdot)$ and $\Phi(\cdot)$ have an important role, when the nets being modelled have marking dependent transition firing rates. For example, consider a net which has the structure in Figure 3.1, where C is an arbitrary SPN configuration. When the marking is $\mathbf{m}$, we wish to choose $\varphi(\cdot)$ and $\Phi(\cdot)$ so that the transitions in C have marking independent firing times and so that the firing rate for t^* is $q(t^*, m(1), m(2))$, that is, dependent on the number of tokens in $p(1)$ and $p(2)$.

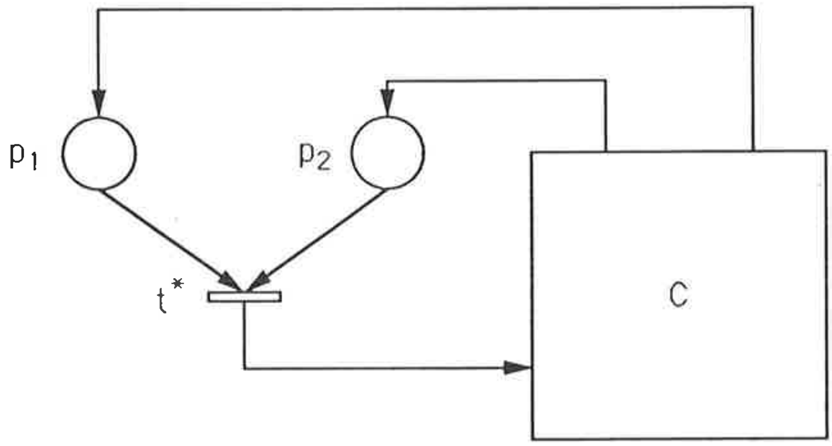

Figure 3.1: An example of a SPN with marking dependent firing rates.

Define

$$\chi(\mathbf{I}(t)) = \begin{cases} 1 & \text{if } t = t^*, \\ q(t) & \text{otherwise,} \end{cases}$$

and

$$\Phi(\mathbf{m}) = \varphi(\mathbf{m}) = \prod_{l=0}^{min(m(1),m(2))} [q(t^*, m(1) - l, m(2) - l)]^{-1}.$$

Substitution into Equation (3.9) gives,

$$q(\mathbf{m}, t) = \begin{cases} q(t^*, m(1), m(2)) & \text{if } t = t^*, \\ q(t) & \text{otherwise,} \end{cases}$$

as required.

Example 3.3

Using Figure 3.1, we can introduce more generality by letting, $\chi(\mathbf{I}(t^*)) = 1$, and $\Phi(\mathbf{m}) = q(t^*, m(1), m(2))\varphi(\mathbf{m} - \mathbf{e}_1 - \mathbf{e}_2)$. This again gives the result $q(\mathbf{m}, t^*) = q(t^*, m(1), m(2))$, without placing any restrictions on the transitions in C.

Example 3.4

Consider Figure 3.2. We will show how marking dependent firing rates, including blocking, can be incorporated into the model, using the appropriate choice of the $\varphi(\cdot)$ and $\Phi(\cdot)$ functions. Suppose that we wish to model the situation where t_1, t_2 and t_3 have firing restrictions, which halt firing when the number of tokens in places p_3, p_4 and p_5 satisfy certain criteria. To model this, let,

$$\varphi(\mathbf{m}) = \begin{cases} 0 & \text{if } \mathbf{m} \in \mathcal{D}, \\ \varphi'(\mathbf{m}) & \text{otherwise,} \end{cases}$$

for an arbitrary choice of $\mathcal{D} \in \mathcal{M}$. For example, let $\mathcal{D}$ be defined by the expression, $m(1) + m(2) < W$, for some integer W. Then $q(\mathbf{m} + \mathbf{I}(t), t) = 0$, whenever $\mathbf{m}$ has insufficient tokens in places p_1 and p_2. Equivalently, transitions t_1, t_2 and t_3 can not fire, when p_3, p_4 and p_5 have too many tokens in them.

The above is a simple example to illustrate blocking in a SPN, with the firing rates given by Equation (3.9). Now that a blocking assumption is in place, we can add to the model, by incorporating marking dependent firing rates, as follows.

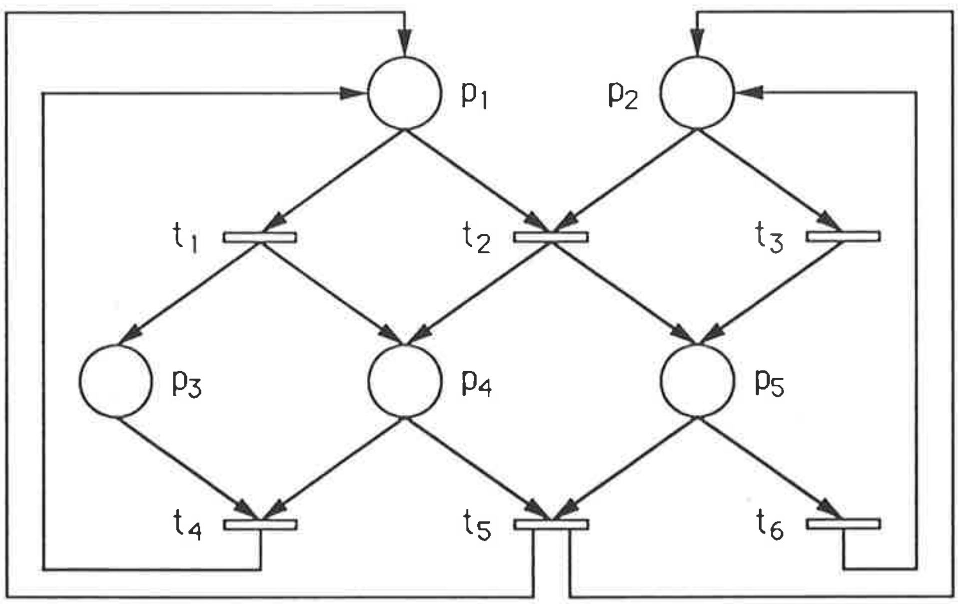

Figure 3.2: The SPN for Example 3.4.

Let,

$$\varphi'(\mathbf{m} - \mathbf{e}_1 - \mathbf{e}_2) = r_1(m(1))r_2(m(2))\eta(m(1), m(2))\Phi^*(m(1), m(2)),$$

and let,

$$\Phi^*(m(1), m(2)) = \prod_{l=1}^{m(1)} \frac{r_3(l-1)}{r_1(l)} \prod_{k=1}^{m(2)} \frac{r_4(k-1)}{r_2(k)}.$$

According to this description, transitions t_i, $1 \leq i \leq 3$, will fire at the following state dependent rate,

$$q(\mathbf{m}, t_i) = \frac{\varphi'(\mathbf{m} - \mathbf{I}(t_i))}{\Phi^*(m(1), m(2))}\chi(\mathbf{I}(t_i)). \tag{3.10}$$

Consider $q(\mathbf{m}, t_1)$ and note that $\mathbf{I}(t_1) = (1, 0, 0, 0, 0)$ so that,

$$\begin{aligned} \varphi'(\mathbf{m} - \mathbf{I}(t_1)) &= \varphi'(\mathbf{m} - \mathbf{e}_1 - \mathbf{e}_2 + \mathbf{e}_2), \\ &= r_1(m(1))r_2(m(2)+1)\eta(m(1), m(2)+1)\Phi^*(m(1), m(2)+1), \end{aligned}$$

Substitution in Equation (3.10) yields,

$$q(\mathbf{m}, t_1)$$

$$= r_1(m(1))r_2(m(2)+1)\eta(m(1), m(2)+1)\frac{\prod_{l=1}^{m(1)} \frac{r_3(l-1)}{r_1(l)} \prod_{k=1}^{m(2)+1} \frac{r_4(k-1)}{r_2(k)}}{\prod_{l=1}^{m(1)} \frac{r_3(l-1)}{r_1(l)} \prod_{k=1}^{m(2)} \frac{r_4(k-1)}{r_2(k)}}\chi(\mathbf{I}(t_1)).$$

After cancellation this gives,

$$q(\mathbf{m}, t_1) = r_1(m(1))r_4(m(2))\eta(m(1), m(2)+1)\chi(\mathbf{I}(t_1)).$$

Similarly,

$$q(\mathbf{m}, t_2) = r_1(m(1))r_2(m(2))\eta(m(1), m(2))\chi(\mathbf{I}(t_2)),$$

$$q(\mathbf{m}, t_3) = r_3(m(1))r_2(m(2))\eta(m(1)+1, m(2))\chi(\mathbf{I}(t_3)).$$

In conclusion, we have shown that the firing rates can depend on the number of tokens, not only in the input places of the transition, which is conventional, but on the token distribution in any of the places.

3.3 The Routeing Process

The central feature of our analysis is to consider the transitions of the SPN, to be themselves states in a Markov chain, which we call the routeing process. This is achieved, by first considering the input and output bags of the SPN to be the states of a Markov chain, and under suitable conditions, finding a one to one correspondence between the states of this Markov chain and the transitions of the net.

Although each transition, t, of our SPN has a unique input bag, the probabilistic routeing allows for a number of different output bags. Let $\mathcal{O}(t)$, be the set of output bags of transition t and define,

$$\mathcal{R} = \bigcup_{t\in\mathcal{T}} \mathcal{O}(t) \cup \bigcup_{t\in\mathcal{T}} \{\mathbf{I}(t)\}.$$

$\mathcal{R}$ is the set of all vectors which are either output or input bags for the SPN. We assume the net is such that $\mathcal{R}$ is finite.

Define single step transition probabilities on the set $\mathcal{R}$ as, whenever there exists a transition t, with $\mathbf{r} = \mathbf{I}(t)$,

$$\overline{p}(\mathbf{r},\mathbf{r}') = \begin{cases} p(\mathbf{I}(t), t, \mathbf{O}_j(t)) & \text{if } \exists\, j \mid \mathbf{r}' = \mathbf{O}_j(t), \\ 0 & \text{otherwise.} \end{cases}$$

If $\mathbf{r}$ is not the input bag of some transition, then $\overline{p}(\mathbf{r},\mathbf{r}') = \delta_{\mathbf{rr}'}$ where $\delta_{\mathbf{rr}'}$ is the Kronecker delta. Note, the requirement that no vector can be the input bag of two distinct transitions, means these probabilities are well defined.

Now define the set,

$$\mathcal{F} = \{f(\cdot) : f(\mathbf{r}) > 0,\ \chi(\mathbf{r})f(\mathbf{r}) = \sum_{\mathbf{r}'} \chi(\mathbf{r}')f(\mathbf{r}')\overline{p}(\mathbf{r}',\mathbf{r}),\ \forall\, \mathbf{r} \in \mathcal{R}\},$$

where $\chi(\mathbf{r})$ is the function given in Equation (3.9), when $\mathbf{r} = \mathbf{I}(t)$ for some t. $\chi(\mathbf{r}) = 1$ otherwise. $\chi(\cdot)f(\cdot)$ is therefore an invariant measure for the routeing process. In Section 3.4, we need $\mathcal{F}$ to be non-empty. The effect of this assumption on the structure of the net is examined in the following results.

Lemma 3.1

For $\mathcal{F}$ to be non-empty, all vectors $\mathbf{I}(t), t \in \mathcal{T}$, must be in positive recurrent communicating classes of the routeing process.

Proof

For t to fire in some marking, $\chi(\mathbf{I}(t)) > 0$. Assume there exists a function, $f(\cdot) \in \mathcal{F}$, implying therefore, that $\chi(\mathbf{I}(t))f(\mathbf{I}(t))$ is positive. Since the set $\mathcal{R}$ is finite, and $\chi(\cdot)f(\cdot)$ is an invariant measure of the routeing process, $\mathbf{I}(t)$ is in a positive recurrent communicating class. ■

Corollary 3.1

For $\mathcal{F}$ to be non-empty,

(a) all $\mathbf{r} \in \mathcal{R}$ must be the input bag for some transition, t,

(b) all $\mathbf{r} \in \mathcal{R}$ must be an output bag for some transition, t.

Proof

(a) Assume $\mathcal{F}$ is non-empty and that $\mathbf{r}$ is not an input bag for any transition. Since state $\mathbf{r}$ is not the input bag to some transition, $\overline{p}(\mathbf{r}, \mathbf{r}') = \delta_{\mathbf{rr}'}$, and $\mathbf{r}$ is an absorbing state in the routeing process. Hence $\mathbf{r}$ resides in its own communicating class. Also, $\mathbf{r} \in \mathcal{O}(t)$ for some transition, t, means that there exists an $\mathbf{r}' \neq \mathbf{r}$, such that $\overline{p}(\mathbf{r}', \mathbf{r}) > 0$, implying that $\mathbf{r}'$ is not in a positive recurrent communicating class. This contradicts Lemma 3.1.

(b) If $\mathbf{r}$ is not the output bag for any transition, then there exists no, $\mathbf{r}' \neq \mathbf{r}$, with $\overline{p}(\mathbf{r}', \mathbf{r}) > 0$. Thus $\mathbf{r}$ must be a transient state. This contradicts Lemma 3.1 in conjunction with Corollary 3.1(a). ■

Corollary 3.2

If $\mathcal{F}$ is non-empty, there exists a one to one correspondence between distinct elements of $\mathcal{R}$ and elements of $\mathcal{T}$.

Proof

It follows from Corollary 3.1(a), that each vector $\mathbf{r} \in \mathcal{R}$ is the input bag for some transition $t \in \mathcal{T}$. By the assumption in Section 3.2, each input bag has a unique transition. The result follows. ∎

Remark 3.1

Corollary 3.1, states that for nets with non-empty $\mathcal{F}$, each output bag of a transition is the input bag of another transition. This implies that whenever an output bag $O_j(t)$ is deposited by transition t, then $O_j(t) = I(s)$, for some transition s. Note that this matching input and output bag criterion does not imply that s is the only transition enabled by $O_j(t)$. There are many other transitions enabled by $O_j(t)$, that may even be more likely to fire. For example, in Figure 3.3, the firing of transition t_3 enables both transitions t_1 and t_2, but $\mathbf{I}(t_1) = \mathbf{O}(t_3)$, so $\overline{p}(t_3, t_1) > 0$ and $\overline{p}(t_3, t_2) = 0$.

Example 3.5

At this point we will reveal the difference between the extended product form solution for queueing networks, developed by Henderson and Taylor [42], and the adapted theory for SPNs.

Henderson and Taylor use a vector, $\mathbf{a} = (a(0), a(i))$, to represent the total number of customers released into the network. $a(i)$ represents the total number of customers released from node i and $a(0)$ represents the number of customers which enter (leave) the network from (to) the outside. Using Figure 3.3, we will show that it is not possible to achieve the matching input and output bag criterion given by Corollary 3.1, if the vector $\mathbf{a}$ is used in place of input and output bags. That is, let $\mathbf{I}'(t)$ be the normal input bag of transition t appended by an entry placed at the beginning of the vector, which represents the tokens that enter the SPN from the outside. $\mathbf{O}'(t)$ is formed in a similar way. For Figure 3.3, we give $\mathbf{I}'(t)$ and the corresponding $\mathbf{O}'(t)$ for each transition.

$$\mathbf{I}'(t_1) = (0,1,1,0,0), \quad \mathbf{O}'(t_1) = (0,0,0,1,1),$$

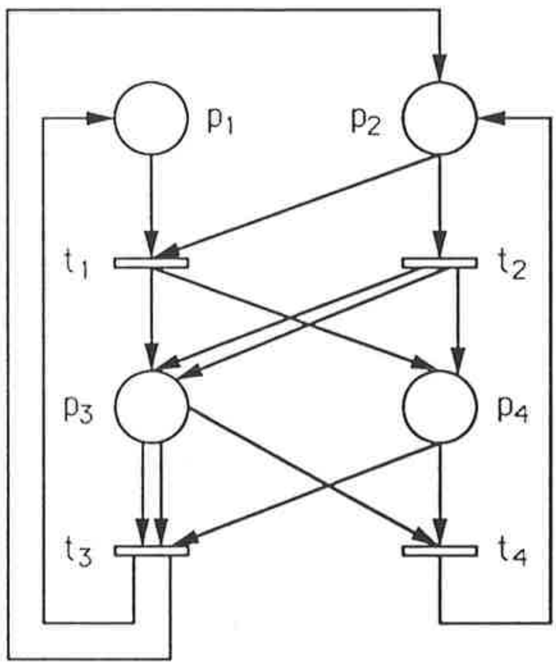

Figure 3.3: An unbounded SPN.

$$\mathbf{I}'(t_2) = (2,0,1,0,0), \quad \mathbf{O}'(t_2) = (0,0,0,2,1),$$

$$\mathbf{I}'(t_3) = (0,0,0,2,1), \quad \mathbf{O}'(t_3) = (1,1,1,0,0),$$

$$\mathbf{I}'(t_4) = (0,0,0,1,1), \quad \mathbf{O}'(t_4) = (1,0,1,0,0).$$

Corollary 3.1 is not satisfied, since there is no matching output bag for $\mathbf{I}'(t_1)$ or $\mathbf{I}'(t_2)$ and therefore the theory of Henderson and Taylor can not applied directly. However, this does not imply that tokens can not be created or destroyed, when the theory is applied to SPNs, since, the $a(0)$ term has been incorporated into the original input and output bags $\mathbf{I}(t)$ and $\mathbf{O}(t)$. Since the definition of a SPN allows the destruction and creation of tokens, $\mathbf{I}(t)$ and $\mathbf{O}(t)$ automatically absorb the $a(0)$ term. Therefore, accounting for the $a(0)$ term is superfluous and can destroy the matching input and output bag criterion.

Remark 3.2

Let the transitions corresponding to the non-zero entries of a T-invariant be rewritten as a set of transitions, $\mathcal{C}_i$. If the SPN, described in Section 3.2, does not have probabilistic routeing, $\mathcal{C}_i$ indicates which transitions belong to the ith positive recurrent communicating class. If the SPN, described in Section 3.2, has probabilistic routeing the following alteration must be made to eliminate the probabilistic routeing (recall that the T-invariant can only be found for deterministic PNs). For each transition, t, with more than one output bag, replace t with a set of transitions, $\mathcal{A}(t)$, one for each output bag. Each transition in $\mathcal{A}(t)$ will share $\mathbf{I}(t)$, and have one output bag. This is effectively the reverse operation described in item 2 in Section 3.2. The firing rates do not need to be evaluated, since we are only interested in the structure of the net. From this altered net, the positive recurrent communicating classes for the original net are evaluated in the following way. For each transition, t, and for each $\mathcal{C}_i$, such that $\mathcal{C}_i \cap \mathcal{A}(t) \neq \{\}$, rewrite $\mathcal{C}_i$ with the elements of $\mathcal{A}(t)$ replaced by t, where $\{\}$ is the empty set. Now, for each i, j, such that $\mathcal{C}_i \cap \mathcal{C}_j \neq \{\}$, let $\mathcal{C}_k = \mathcal{C}_i \cup \mathcal{C}_j$. Continue this process until all of the remaining sets, $\mathcal{C}_i$, are mutually disjoint. The remaining sets, $\mathcal{C}_i$, are the positive recurrent communicating classes for the original net. For an illustration of this process on a net which includes probabilistic routeing, see Example 3.8.

Remark 3.3

Note that Corollary 3.2 implies that each $\mathbf{r} \in \mathcal{R}$ can be uniquely identified with a transition. Consequently, we can define the routeing process to be a Markov chain on the set of transitions, $\mathcal{T}$, rather than on the set of input and output bags, $\mathcal{R}$. Without loss of generality, we will retain the same notation, by assuming that the one step transition probabilities for the routeing process on $\mathcal{T}$ are,

$$\overline{p}(t,s) = \overline{p}(\mathbf{I}(t), \mathbf{I}(s)), \quad s, \ t \in \mathcal{T}.$$

Let us stress again that $\overline{p}(t,s)$ is not the probability that s will fire after t. In fact the set $\{s : \overline{p}(t,s) > 0\}$ is a subset of the set of transitions which can fire after t fires. Transition s will be enabled by the output bag of t with probability $\overline{p}(t,s)$. However, as we have seen in Example 3.3, and in many other examples, many other transitions will be enabled after t fires, and any of these can fire after t fires. Examples 3.8 and 3.9, illustrate this for the probabilistic routeing case and for the deterministic routeing case, respectively.

$\mathcal{F}$ can now be redefined as,

$$\mathcal{F} = \{f(\cdot) : f(t) > 0, \ \chi(t)f(t) = \sum_{s \in \mathcal{T}} \chi(s)f(s)\overline{p}(s,t), \ \forall\, t \in \mathcal{T}\}, \qquad (3.11)$$

and $\chi(\cdot)f(\cdot)$ is an invariant measure for the routeing process. As the routeing process consists of only positive recurrent communicating classes, it may be assumed that $\chi(t)f(t) > 0, \ \forall\, t \in \mathcal{T}$.

3.4 The Extended Product Form Solution

Lemma 3.2 (Theorem 1.13 from Kelly [51])

Let $X(t)$ be a stationary Markov process with transition rates $q(\mathbf{m},\mathbf{n})$, $\mathbf{m}$, $\mathbf{n} \in \mathbf{S}$, where $\mathbf{S}$ is the state space. If we can find a collection of numbers,

$q^R(\mathbf{m}, \mathbf{n})$, $\mathbf{m}$, $\mathbf{n} \in \mathbf{S}$, such that,

$$\sum_{\mathbf{n}\in\mathbf{S}} q(\mathbf{m}, \mathbf{n}) = \sum_{\mathbf{n}\in\mathbf{S}} q^R(\mathbf{m}, \mathbf{n}), \quad \forall \mathbf{m} \in \mathbf{S},$$

and a collection of positive numbers $\pi(\mathbf{m})$, $\mathbf{m} \in \mathbf{S}$, summing to unity, such that,

$$\pi(\mathbf{m})q(\mathbf{m}, \mathbf{n}) = \pi(\mathbf{n})q^R(\mathbf{n}, \mathbf{m}) \quad \mathbf{m}, \ \mathbf{n} \in \mathbf{S},$$

then $q^R(\mathbf{m}, \mathbf{n})$, $\mathbf{m}$, $\mathbf{n} \in \mathbf{S}$, are the transition rates of the reversed process $X(\tau - t)$ and $\pi(\mathbf{m})$, $\mathbf{m} \in \mathbf{S}$ is the equilibrium distribution of both $X(t)$ and $X(\tau - t)$.

Lemma 3.3 (Adapted from Lemma 3.2 for SPNs)

Let $X(t)$ be the stationary Markov process of a SPN described in Section 3.2 with transition rates $q(\mathbf{m}, t)\overline{p}(t, s)$, $\mathbf{m} \in \mathcal{M}$, $s, t \in \mathcal{T}$, which is the rate of moving from marking $\mathbf{m}$, to marking $\mathbf{m} - \mathbf{I}(t) + \mathbf{I}(s)$. If we can find a collection of numbers $q^R(\mathbf{m}, \mathbf{m} + \mathbf{I}(t) - \mathbf{I}(s))$, $\mathbf{m} \in \mathcal{M}$, $s, t \in \mathcal{T}$, such that,

$$\sum_{t\in\mathcal{T}} \sum_{s\in\mathcal{T}} q(\mathbf{m}, t)\overline{p}(t, s) = \sum_{t\in\mathcal{T}} \sum_{s\in\mathcal{T}} q^R(\mathbf{m}, \mathbf{m} + \mathbf{I}(s) - \mathbf{I}(t)),$$

or equivalently,

$$\sum_{t\in\mathcal{T}} q(\mathbf{m}, t) = \sum_{t\in\mathcal{T}} \sum_{s\in\mathcal{T}} q^R(\mathbf{m}, \mathbf{m} + \mathbf{I}(s) - \mathbf{I}(t)), \tag{3.12}$$

since $\sum_{s\in\mathcal{T}} \overline{p}(t, s) = 1$, and a collection of positive numbers $\pi(\mathbf{m}+\mathbf{I}(t))$, $\mathbf{m} \in \mathcal{M}$, $t \in \mathcal{T}$, summing to unity, such that,

$$\pi(\mathbf{m} + \mathbf{I}(t))q(\mathbf{m} + \mathbf{I}(t), t)\overline{p}(t, s) = \pi(\mathbf{m} + \mathbf{I}(s))q^R(\mathbf{m} + \mathbf{I}(s), \mathbf{m} + \mathbf{I}(t)), \tag{3.13}$$

then $q^R(\mathbf{m} + I(s), \mathbf{m} + \mathbf{I}(t))$, $\mathbf{m} \in \mathcal{M}$, $s, t \in \mathcal{T}$ are the transition rates of the reversed process $X(\tau - t)$ and $\pi(\mathbf{m} + \mathbf{I}(t))$, $\mathbf{m} \in \mathcal{M}$, $t \in \mathcal{T}$, is the equilibrium distribution of both $X(t)$ and $X(\tau - t)$.

Theorem 3.1

Assume that there exists a function $f(\cdot) \in \mathcal{F}$, and a function $g(\cdot) : \mathbb{Z}^{|\mathcal{P}|} \to \mathbb{R}$, which has the property that, for all $t \in \mathcal{T}$ and $\mathbf{m} + \mathbf{I}(t)$ in the reachability grap.

$$\frac{g(\mathbf{m} + \mathbf{I}(t))}{g(\mathbf{m} + \mathbf{I}(s))} = \frac{f(t)}{f(s)}, \tag{3.14}$$

whenever $\overline{p}(t, s) > 0$. Then the equilibrium distribution of the SPN is given by,

$$\pi(\mathbf{m}) = K\Phi(\mathbf{m})g(\mathbf{m}). \tag{3.15}$$

where $\Phi(\cdot)$ is given in Equation (3.9) and K is a normalising constant.

Theorem 3.1 gives a product form solution. The product is between two distinct terms, as in Equation (3.8), which was discussed in Section 3.1. The first of these is $\Phi(\cdot)$, a function related to the transition firing rates and to marking dependent properties of the SPN. The second, $g(\cdot)$, can be evaluated by analysing the network skeleton provided by the transitions and the routeing process. The effect is to reduce the problem of finding an equilibrium distribution for the number of tokens in each place of a SPN, from that of solving balance equations for a Markov process on the markings, which grows exponentially in size with an increase in the number of tokens, to solving balance equations for a Markov chain on the transitions, the size of which does not change with an increase in the number of tokens. The examples in Section 3.5 illustrate this effect.

Proof

Let us propose, that the reversed time transition probabilities for the routeing process are,

$$\overline{p}^R(s, t) = \frac{\chi(t)f(t)\overline{p}(t, s)}{\chi(s)f(s)}, \tag{3.16}$$

and postulate an equilibrium distribution for the SPN given by (3.15) subject to (3.14).

Note that any marking change of the SPN begins with a marking $\mathbf{m} + \mathbf{I}(t)$ with $t \in \mathcal{T}$ enabled. Transition t fires, producing an output bag, which because $\mathcal{F}$ is non-empty, is the input bag of another transition, s say. The resultant marking is $\mathbf{m} + \mathbf{I}(s)$, and transition s is enabled. In reversed time, the markings and enabled transitions are unchanged but the reverse operation occurs.

Postulate that the marking change rate for the SPN in reversed time is,

$$q^R(\mathbf{m} + \mathbf{I}(s), \mathbf{m} + \mathbf{I}(t)) = q(\mathbf{m} + \mathbf{I}(s), s)\overline{p}^R(s, t). \tag{3.17}$$

From Lemma 3.3, the validity of Equations (3.12) and (3.13) is sufficient to establish that $\pi(\mathbf{m})$, as given by Equation (3.15) subject to (3.14), is the equilibrium distribution of the SPN and that the postulate of Equation (3.17) is correct.

Consider the right hand side of Equation (3.13). Substituting the form for $\pi(\mathbf{m}+I(s))$, from Equation (3.15) and $q^R(\mathbf{m}+I(s),\mathbf{m}+\mathbf{I}(t))$, from Equations (3.16) and (3.17) yields,

$$K\Phi(\mathbf{m}+I(s))g(\mathbf{m}+I(s))\frac{\varphi(\mathbf{m})\chi(s)\chi(t)f(t)\overline{p}(t,s)}{\Phi(\mathbf{m}+I(s))\chi(s)f(s)}$$

$$=Kg(\mathbf{m}+I(s))\frac{f(t)}{f(s)}\varphi(\mathbf{m})\chi(t)\overline{p}(t,s), \qquad (3.18)$$

but from Equation (3.14),

$$g(\mathbf{m}+I(s))\frac{f(t)}{f(s)}=g(\mathbf{m}+\mathbf{I}(t)).$$

Substitution into Equation (3.18) yields,

$$Kg(\mathbf{m}+\mathbf{I}(t))\varphi(\mathbf{m})\chi(t)\overline{p}(t,s),$$

$$=K\Phi(\mathbf{m}+\mathbf{I}(t))g(\mathbf{m}+\mathbf{I}(t))\frac{\varphi(\mathbf{m})\chi(t)}{\Phi(\mathbf{m}+\mathbf{I}(t))}\overline{p}(t,s),$$

$$=\pi(\mathbf{m}+\mathbf{I}(t))q(\mathbf{m}+\mathbf{I}(t),t)\overline{p}(t,s),$$

by Equations (3.15) and (3.9). This is the left hand side of Equation (3.13), which shows that Equation (3.13) is satisfied, $\forall\ \mathbf{m}\in\mathcal{M},\ s,t\in\mathcal{T}$. Consider now the right hand side of Equation (3.12) and substitute,

$$q^R(\mathbf{m},\mathbf{m}+I(s)-\mathbf{I}(t))=q(\mathbf{m},t)p^R(t,s),$$

which gives,

$$\sum_{t\in\mathcal{T}}\sum_{s\in\mathcal{T}}q(\mathbf{m},t)p^R(t,s)=\sum_{t\in\mathcal{T}}q(\mathbf{m},t),$$

since $\sum_{s\in\mathcal{T}}p^R(t,s)=1$. This shows that Equation (3.12) is satisfied and therefore proves that, $\pi(\mathbf{m})$ as given by Equation (3.15), is the equilibrium distribution of the SPN. ∎

Theorem 3.2

For the SPN of Theorem 3.1, the equilibrium distribution given by Equation (3.15) subject to Equation (3.14), satisfies the following partial balance equation.

$$\begin{aligned} &\text{flux out of } \mathbf{m}+\mathbf{I}(t), \text{ due to the firing of } t \\ &= \text{flux into } \mathbf{m}+\mathbf{I}(t), \text{ due to the bag } \mathbf{I}(t) \text{ being deposited by a transition.} \end{aligned} \tag{3.19}$$

Proof

Consider transitions s and t, in the same positive recurrent communicating class C. The flux into $\mathbf{m}+\mathbf{I}(t)$ from $\mathbf{m}+\mathbf{I}(s)$, where the output bag of s is the input bag of t, is

$$\sum_{s\in C} \pi(\mathbf{m}+\mathbf{I}(s))q(\mathbf{m}+\mathbf{I}(s), s)p(s,t).$$

By Equation (3.9),

$$= \sum_{s\in C} \pi(\mathbf{m}+\mathbf{I}(s))\frac{\varphi(\mathbf{m})}{\Phi(\mathbf{m}+\mathbf{I}(s))}\chi(s)p(s,t).$$

Using Equation (3.15),

$$= \varphi(\mathbf{m})\sum_{s\in C} Kg(\mathbf{m}+\mathbf{I}(s))\chi(s)p(s,t).$$

Substituting for $g(\mathbf{m}+\mathbf{I}(s))$ using Equation (3.14),

$$= K\varphi(\mathbf{m})\sum_{s\in C} g(\mathbf{m}+\mathbf{I}(t))\frac{f(s)}{f(t)}\chi(s)p(s,t),$$

$$= K\varphi(\mathbf{m})\frac{g(\mathbf{m}+\mathbf{I}(t))}{f(t)}\sum_{s\in C} f(s)\chi(s)p(s,t).$$

But $\sum_{s\in C} f(s)\chi(s)p(s,t) = f(t)\chi(t)$, is the global balance equation on the routeing chain. So,

$$= K\varphi(\mathbf{m})g(\mathbf{m}+\mathbf{I}(t))\frac{f(t)}{f(t)}\chi(t),$$

$$= K\varphi(\mathbf{m})g(\mathbf{m}+\mathbf{I}(t))\chi(t).$$

Once again, using Equation (3.15) and (3.9),

$$= \varphi(\mathbf{m})\frac{\pi(\mathbf{m}+\mathbf{I}(t))}{\Phi(\mathbf{m}+\mathbf{I}(t))}\chi(t),$$

$$= \pi(\mathbf{m}+\mathbf{I}(t))q(\mathbf{m}+\mathbf{I}(t),t).$$

Which is the flux out of $\mathbf{m}+\mathbf{I}(t)$ due to the firing of transition t. ■

Remark 3.4

The partial balance equations given by Equation (3.19) above, are a generalisation of Whittle's, given by Equations (3.5) and (3.6). Whittle's partial balance equations apply to open queueing networks, where there is purely single customer movement. In a SPN, this corresponds to having a single input and output arc per transition and an extra transition, t_0, to represent the "outside". Let $\mathcal{T}' = \mathcal{T} \bigcup t_0$ and let the input bag of t_i be place p_i, say. Now assume that a token will arrive at place p_i, with rate $\lambda(i)$. Therefore, let t_0 fire with rate $\sum_{i=1}^{|\mathcal{P}|}\lambda(i)$. Connect one probabilistic arc from t_0, to every place p_i in the SPN, such that t_0 will deposit a token in p_i with probability $\lambda(i)\left[\sum_{j=1}^{|\mathcal{P}|}\lambda(j)\right]^{-1}$. Also, let the tokens depart the SPN with rate $q(i)$, from place p_i. Consequently, $\mathbf{O}(t_i) = \mathbf{0}$, which represents the "outside", with probability $q(i)$, or is the next place p_j (since $p_j \in \mathcal{O}(t_i)$) with probability $\overline{p}(t_i, t_j)$. Similarly, $\mathbf{I}(t_0) = \mathbf{0}$. Note that, we have maintained the matching input and output bag criterion, in using probabilistic output bags.

Figure 3.4 is a SPN representation of a simple open queueing network. Tokens arrive to place p_1 from the "outside", with probability $\frac{\lambda(1)}{\lambda(1)+\lambda(2)}$, or internally with probability $\overline{p}(t_2, t_1)$. Tokens leave p_1, and depart externally, with probability $q(1)$, or move on to the next place with probability $\overline{p}(t_1, t_2)$.

Equation (3.19), relates the flux out of $\mathbf{m}+\mathbf{I}(t_j)$, due to a token leaving p_j with the flux into $\mathbf{m}+\mathbf{I}(t_j)$, due to a token being deposited into p_j. This is Equation (3.6), if $\mathbf{m}+\mathbf{I}(t_j) = \mathbf{n}$ and p_j represents queue j. Equation (3.19) also relates the flux out of $\mathbf{m}+\mathbf{0}$, due to a token entering the SPN from the "outside", with the flux into $\mathbf{m}+\mathbf{0}$, due to a token leaving the SPN to the "outside". This is

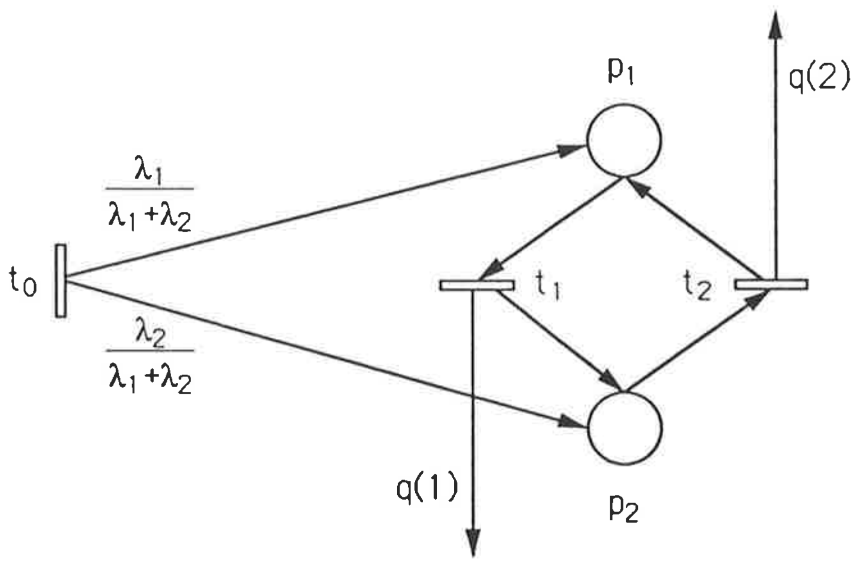

Figure 3.4: A SPN representation of a simple open Jackson network.

Equation (3.5), if $\mathbf{m} + \mathbf{0} = \mathbf{n}$.

The firing rates in our model also incorporate more general marking dependent firing rates and blocking, as in Example 3.4, which Whittle does not consider.

In Example 3.6, it is shown how the solution to the Jackson networks can be derived as a special case of the results in this section.

3.5 Examples

Example 3.6: Jackson Networks

Consider a SPN as defined in Remark 3.4. Since each place is connected to a unique transition by one arc, the number of times transition t_i must fire to remove $m(p_i)$ tokens from place p_i is given by $m(p_i)$, which in this case can be written, without loss of generality, as $m(t_i)$. Let $\chi(t_i) = 1, \forall t_i \in \mathcal{T}$, $\chi(t_0) = \sum_{i=1}^{|\mathcal{P}|} \lambda(i)$ and $\Phi(\mathbf{m}) = \varphi(\mathbf{m}) = \prod_{t_i \in \mathcal{T}} \prod_{l=1}^{m(t_i)} (q(t_i, l))^{-1}$. With these parameter values and assumptions on the net structure, $f(\cdot) \in \mathcal{F}$ satisfies,

$$f(t_i) = \sum_{t_j \in \mathcal{T}} f(t_j)\overline{p}(t_j, t_i) + f(t_0)\chi(t_0)\overline{p}(t_0, t_i), \ \forall\ t_i \in \mathcal{T}, \tag{3.20}$$

and

$$q(\mathbf{m}, t_i) = \frac{\varphi(\mathbf{m} - \mathbf{I}(t_i))}{\Phi(\mathbf{m})}$$

$$= \frac{\prod_{l=1}^{m(t_i)-1} q(t_i, l)^{-1} \prod_{t_j \in \mathcal{T}\setminus\{t_i\}} \prod_{l=1}^{m(t_j)} q(t_j, l)^{-1}}{\prod_{l=1}^{m(t_i)} q(t_i, l)^{-1} \prod_{t_j \in \mathcal{T}\setminus\{t_i\}} \prod_{l=1}^{m(t_j)} q(t_j, l)^{-1}} = q(t_i, m(t_i)).$$

Equation (3.20), is the standard traffic equation for the open Jackson network, given by Equation (3.3) in Section 3.1, as can be shown by the following. Let $f(t_0) = 1$, so, $f(t_0)\chi(t_0)\overline{p}(t_0, t_i) = \lambda(i)$. Substitution yields,

$$f(t_i) = \lambda(i) + \sum_{t_j \in \mathcal{T}} f(t_j)\overline{p}(t_j, t_i). \tag{3.21}$$

$\overline{p}(t_j, t_i)$ gives the probability that the output bag of t_j is the input bag of t_i. Since the input and output bags of the transitions are just the single places connected to the transitions, $\overline{p}(t_j, t_i)$ is equivalent to the probability that a single token moves from $\mathbf{I}(t_j)$ to $\mathbf{O}(t_i)$. This is analogous to single customers moving from one queue to another. Equation (3.21), is therefore of the same form as Equation (3.3), with $f(t_i) = y(i)$. The function $g(\cdot)$ is given below.

$$g(\mathbf{m}) = \prod_{t_i \in \mathcal{T}} f(t_i)^{m(t_i)}. \tag{3.22}$$

$g(\cdot)$ satisfies Equation (3.14), for all $\mathbf{m}$, since,

$$\frac{g(\mathbf{m} + \mathbf{I}(t_i))}{g(\mathbf{m} + \mathbf{I}(t_j))} = \frac{f(t_i) \prod_{t_k \in \mathcal{T}} f(t_k)^{m(t_k)}}{f(t_j) \prod_{t_k \in \mathcal{T}} f(t_k)^{m(t_k)}} = \frac{f(t_i)}{f(t_j)}.$$

The equations are satisfied and the equilibrium solution is therefore,

$$\pi(\mathbf{m}) = K \prod_{t_i \in \mathcal{T}} f(t_i)^{m(t_i)} \prod_{l=1}^{m(t_i)} [q(t_i, l)]^{-1}. \tag{3.23}$$

With the above choices, the SPN under consideration is a Jackson network of queues. Each place has it's own transition, that is, each queue has it's own server. Tokens are served one at a time, through transition t_i with service rate $q(t_i, m(t_i))$, when $m(t_i)$ tokens are available. These tokens are then independently routed, according to the probability distribution $\overline{p}(t_i, t_j)$, t_i, $t_j \in \mathcal{T}$. We established, that Equation (3.20) is the traffic equation for the Jackson network given in Equation (3.3), so therefore the solution in (3.23) is equivalent to the standard solution given by Equations (3.1) and (3.4).

Example 3.7

In this example, we will show how to find the function $g(\cdot)$ for a simple SPN, which will hopefully lay the foundations for the examples to follow. Figure 3.5, is a SPN which represents a simple closed Jackson network, with K customers. Let,

$$\Phi(\mathbf{m}) = \varphi(\mathbf{m}) = \prod_{t_i \in \mathcal{T}} \prod_{l=1}^{m(t_i)} (q(t_i, l))^{-1},$$

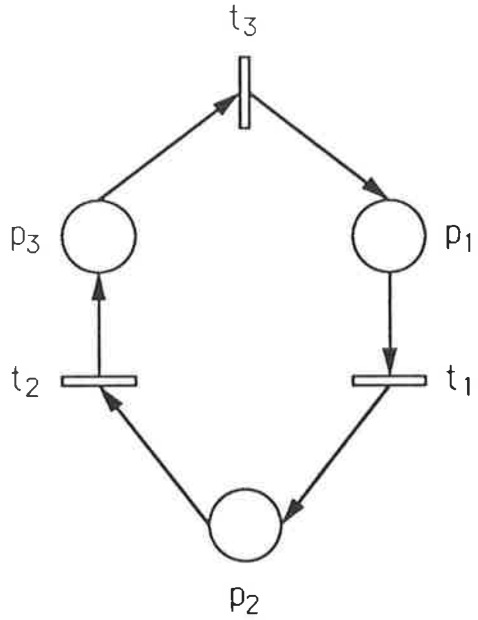

Figure 3.5: A SPN representation of a simple closed Jackson network.

as in the previous example. The procedure to find $g(\cdot)$, is to first identify one state of the reachability graph and label it as a base state, $\mathbf{b}$, say. In many ways, the initial marking is a natural choice for the base state, if we happen to know what it is. Otherwise, we chose any marking in the reachability graph, some of which seem more "natural" than others. Since each place is connected to a unique transition, by one arc, the number of tokens in a place p is the same as the number of times the transition connected to p must fire to remove all of its tokens. Assume that, in marking $\mathbf{m}$, p_i contains $m(i)$ customers. In this example, a natural base state is the marking with all tokens in p_1, that is, given a marking $\mathbf{m}$ with an arbitrary number of tokens in each place, $\mathbf{b} = (\mathbf{m} - m(2)\mathbf{e}_2 - m(3)\mathbf{e}_3 + m(2)\mathbf{e}_1 + m(3)\mathbf{e}_1)$. The procedure to find $g(\cdot)$, is then to empty every other place in the net of its tokens, until they are all in p_1. This requires t_2 to fire $m(2)$ times, to remove the tokens from p_2, and t_3 to fire $m(2)$ times, to move these same tokens from p_3 to p_1. This procedure must be repeated, until all of the tokens in the net have been deposited in p_1.

Each time a transition t_i is fired, to change the marking of the SPN, Equation (3.14) provides a relation between $g(\cdot)$ of the marking enabling t_i and $g(\cdot)$ of the new marking. For the SPN in Figure 3.5, these are,

$$\frac{g(\mathbf{m}+\mathbf{I}(t_1))}{g(\mathbf{m}+\mathbf{I}(t_2))} = \frac{f(t_1)}{f(t_2)}, \tag{3.24}$$

$$\frac{g(\mathbf{m}+\mathbf{I}(t_2))}{g(\mathbf{m}+\mathbf{I}(t_3))} = \frac{f(t_2)}{f(t_3)}, \tag{3.25}$$

$$\frac{g(\mathbf{m}+\mathbf{I}(t_3))}{g(\mathbf{m}+\mathbf{I}(t_1))} = \frac{f(t_3)}{f(t_1)}. \tag{3.26}$$

Using this procedure, we can begin with any marking $\mathbf{m}$ in the reachability graph and recursively evaluate $g(\mathbf{m})$ in terms of $g(\mathbf{b})$. In this example, let us begin by imagining that we fire t_2, $m(2)$ times. Each time we do, we transfer a token from place p_2 to p_3 and Equation (3.14) suggests the result,

$$g(\mathbf{m}) = g(\mathbf{m} + m(2)\mathbf{e}_3 - m(2)\mathbf{e}_2)\left[\frac{f(t_2)}{f(t_3)}\right]^{m(2)}. \tag{3.27}$$

Now we must transfer the $m(2)$ tokens from p_2, that have been moved to p_3 due to the previous firing sequence, to place p_1. This suggests the relation,

$$g(\mathbf{m} + m(2)\mathbf{e}_3 - m(2)\mathbf{e}_2) = g(\mathbf{m} + m(2)\mathbf{e}_1 - m(2)\mathbf{e}_2) \left[\frac{f(t_3)}{f(t_1)}\right]^{m(2)} . \qquad (3.28)$$

The last place to empty is p_3, so we must transfer the $m(3)$ tokens from p_3 to p_1, giving the result,

$$\begin{aligned} & g(\mathbf{m} + m(2)\mathbf{e}_1 - m(2)\mathbf{e}_2) \\ & = g(\mathbf{m} + m(2)\mathbf{e}_1 + m(3)\mathbf{e}_1 - m(2)\mathbf{e}_2 - m(3)\mathbf{e}_3) \left[\frac{f(t_3)}{f(t_1)}\right]^{m(3)}, \\ & = g(\mathbf{b}) \left[\frac{f(t_3)}{f(t_1)}\right]^{m(3)} . \end{aligned} \qquad (3.29)$$

Using relations (3.27), (3.28) and (3.29) recursively, gives,

$$g(\mathbf{m}) = g(\mathbf{b}) \left[\frac{f(t_2)}{f(t_3)}\right]^{m(2)} \left[\frac{f(t_3)}{f(t_1)}\right]^{m(2)+m(3)} .$$

Treating $g(\mathbf{b})$ as a constant, normally set to unity, this reduces to,

$$g(\mathbf{m}) = \frac{f(t_2)^{m(2)} f(t_3)^{m(3)}}{f(t_1)^{m(2)+m(3)}}.$$

Since the network is a closed network with K customers, $m(2)+m(3) = \mathrm{K} - m(1)$, therefore,

$$g(\mathbf{m}) = C f(t_2)^{m(2)} f(t_3)^{m(3)} f(t_1)^{m(1)},$$

where C is a constant and is absorbed into the normalising constant. Note that, this is of the same form as Equation (3.22), as required. In performing this procedure, we have found a specific $g(\cdot)$ by following a particular path to the base state $\mathbf{b}$. We must now check to see that $g(\cdot)$ satisfies Equations (3.24), (3.25) and (3.26) for an arbitrary marking, $\mathbf{m}$.

$$\begin{aligned} \frac{g(\mathbf{m} + \mathbf{I}(t_1))}{g(\mathbf{m} + \mathbf{I}(t_2))} & = \frac{f(t_2)^{m(2)} f(t_3)^{m(3)} f(t_1)^{m(1)+1}}{f(t_2)^{m(2)+1} f(t_3)^{m(3)} f(t_1)^{m(1)}}, \\ & = \frac{f(t_1)}{f(t_2)}. \end{aligned}$$

$$\frac{g(\mathbf{m}+\mathbf{I}(t_2))}{g(\mathbf{m}+\mathbf{I}(t_3))} = \frac{f(t_2)^{m(2)+1} f(t_3)^{m(3)} f(t_1)^{m(1)}}{f(t_2)^{m(2)} f(t_3)^{m(3)+1} f(t_1)^{m(1)}},$$

$$= \frac{f(t_2)}{f(t_3)}.$$

$$\frac{g(\mathbf{m}+\mathbf{I}(t_3))}{g(\mathbf{m}+\mathbf{I}(t_1))} = \frac{f(t_2)^{m(2)} f(t_3)^{m(3)+1} f(t_1)^{m(1)}}{f(t_2)^{m(2)} f(t_3)^{m(3)} f(t_1)^{m(1)+1}},$$

$$= \frac{f(t_3)}{f(t_1)}.$$

The equations are satisfied, and using the results of Example 3.6, the equilibrium distribution is given by Equation (3.23).

Example 3.8

Consider the SPN illustrated in Figure 3.6. Since the input bags of transitions t_1' and t_1'' are identical, we follow the pre-analysis procedure outlined in Section 3.2, to achieve the correct structure for the extended product form solution. This consists of amalgamating the two transitions and introducing probabilistic routeing, to create a SPN of the form given in Figure 3.7.

If the SPN of Figure 3.6 has mean transition firing times, $[\chi(t_1')]^{-1}$, $[\chi(t_1'')]^{-1}$, $[\chi(t_2)]^{-1}$, $[\chi(t_3)]^{-1}$, $[\chi(t_4)]^{-1}$, for transitions $t_1', t_1'', t_2, t_3, t_4$ respectively, then the SPN of Figure 3.7 has mean firing times $[\chi(t_1)]^{-1}, [\chi(t_2)]^{-1}, [\chi(t_3)]^{-1}, [\chi(t_4)]^{-1}$, with $\chi(t_1) = \chi(t_1') + \chi(t_1'')$. The probabilistic routeing, when t_1 fires, produces output bags (0,1,2,0,0) (solid arcs) and (0,0,1,1,0) (dashed arcs) with probabilities, $a = \frac{\chi(t_1')}{\chi(t_1)}$ and $b = \frac{\chi(t_1'')}{\chi(t_1)}$, respectively.

The routeing process of the SPN has one step probabilities $\overline{p}(t_1, t_2) = a$, $\overline{p}(t_1, t_3) = b$, $\overline{p}(t_2, t_1) = \overline{p}(t_3, t_4) = \overline{p}(t_4, t_2) = 1$. Consider the one step transition probability matrix of the routeing process below,

$$\begin{bmatrix} 0 & a & b & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Since the output bag of t_1 is the input bag of either t_2 or t_3 with probability a and b, respectively and since $\mathbf{O}(t_3) = \mathbf{I}(t_4)$, $\mathbf{O}(t_4) = \mathbf{I}(t_2)$, $\mathbf{O}(t_2) = \mathbf{I}(t_1)$ with

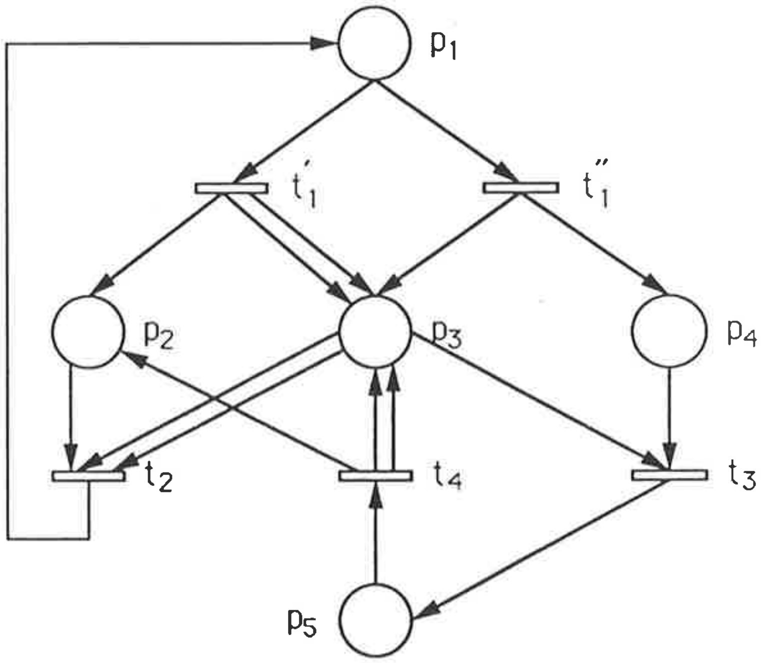

Figure 3.6: The SPN for Example 3.8.

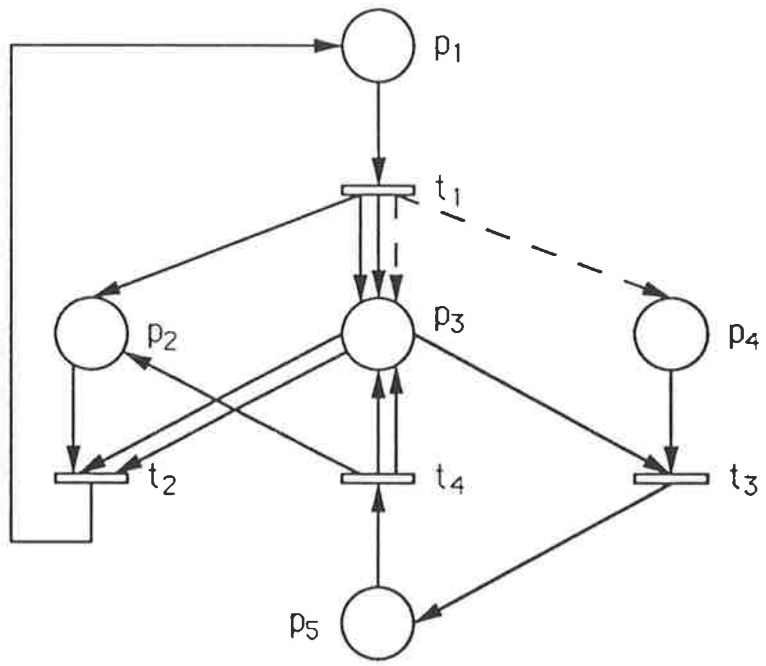

Figure 3.7: The transition merged SPN for Example 3.8.

probability one, there exists one communicating class $C_1 = \{t_1, t_2, t_3, t_4\}$. Using Remark 3.2, the same result can be achieved if we find the T-invariants for the SPN of Figure 3.6. The incidence matrix is given by,

$$\begin{bmatrix} -1 & 1 & 2 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 \\ 1 & -1 & -2 & 0 & 0 \\ 0 & 0 & -1 & -1 & 1 \\ 0 & 1 & 2 & 0 & -1 \end{bmatrix}.$$

The T-invariants are given by the firing count vectors (0,1,1,1,1) and (1,0,1,0,0). Following Remark 3.2, $\mathcal{C}_1 = \{t_1'', t_2, t_3, t_4\}$ and $\mathcal{C}_2 = \{t_1', t_3\}$. The union of these sets is given by $\{t_1, t_2, t_3, t_4\}$ by merging t_1' and t_1'' into t_1. $\mathcal{C}_1$ is therefore given by the set $\{t_1, t_2, t_3, t_4\}$, which is the positive recurrent communicating class as given above.

To find the invariant measure of the routeing process, we must solve the set of Equations (3.11), that are,

$$\chi(t_1)f(t_1) = \chi(t_2)f(t_2), \tag{3.30}$$

$$\chi(t_2)f(t_2) = a\chi(t_1)f(t_1) + \chi(t_4)f(t_4), \tag{3.31}$$

$$\chi(t_3)f(t_3) = b\chi(t_1)f(t_1), \tag{3.32}$$

$$\chi(t_4)f(t_4) = \chi(t_3)f(t_3), \tag{3.33}$$

which has solution,

$$\chi(t_1)f(t_1) = 1 = \chi(t_2)f(t_2), \tag{3.34}$$

$$\chi(t_3)f(t_3) = b = \chi(t_4)f(t_4). \tag{3.35}$$

Consequently,

$$f(t_1) = [\chi(t_1)]^{-1}, \tag{3.36}$$

$$f(t_2) = [\chi(t_2)]^{-1}, \tag{3.37}$$

$$f(t_3) = b\,[\chi(t_3)]^{-1}, \tag{3.38}$$

$$f(t_4) = b\,[\chi(t_4)]^{-1}. \tag{3.39}$$

To satisfy Equation (3.14), we need a function $g(\mathbf{m})$, such that,

$$\frac{g(\mathbf{m}+\mathbf{I}(t_1))}{g(\mathbf{m}+\mathbf{I}(t_2))} = \frac{f(t_1)}{f(t_2)} = \frac{\chi(t_2)}{\chi(t_1)}, \tag{3.40}$$

$$\frac{g(\mathbf{m}+\mathbf{I}(t_1))}{g(\mathbf{m}+\mathbf{I}(t_3))} = \frac{f(t_1)}{f(t_3)} = \frac{\chi(t_3)}{b\chi(t_1)}, \tag{3.41}$$

$$\frac{g(\mathbf{m}+\mathbf{I}(t_3))}{g(\mathbf{m}+\mathbf{I}(t_4))} = \frac{f(t_3)}{f(t_4)} = \frac{\chi(t_4)}{\chi(t_3)}, \tag{3.42}$$

$$\frac{g(\mathbf{m}+\mathbf{I}(t_4))}{g(\mathbf{m}+\mathbf{I}(t_2))} = \frac{f(t_4)}{f(t_2)} = \frac{b\chi(t_2)}{\chi(t_4)}, \tag{3.43}$$

$$\frac{g(\mathbf{m}+\mathbf{I}(t_2))}{g(\mathbf{m}+\mathbf{I}(t_1))} = \frac{f(t_2)}{f(t_1)} = \frac{\chi(t_1)}{\chi(t_2)}. \tag{3.44}$$

To find a function $g(\mathbf{m})$, we follow the same procedure described in Example 3.7. To simplify the notation, we will not give the recursive relation for each sequence of transition firings, when removing tokens from places, as in the previous example. Instead, we will give the coefficient obtained after each sequence of transition firings as this provides sufficient information in itself.

Let the base state be the marking with as many tokens as possible in p_1. As the example unfolds, it will become clear that there may be some residual tokens remaining in some of the places of the net. Assume p_5 contains $m(5)$ tokens. To empty p_5 we require t_4 to fire and thus deposit $\mathbf{I}(t_2)$, $m(5)$ times. The coefficient created in doing this is,

$$\left[\frac{f(t_4)}{f(t_2)}\right]^{m(5)}. \tag{3.45}$$

At this stage, there are $m(3)+2m(5)$ tokens in p_3, since two tokens are deposited into p_3 for each token in p_5, and there are $m(2)+m(5)$ tokens in p_2.

Now, assume that there are $m(4)$ tokens in p_4. To empty p_4 we require the number of tokens in p_3, given by $m(3)+2m(5)$, to be greater than zero. We will consider the cases where $m(3)+2m(5) = 0$ and $m(3)+2m(5) \geq 1$, separately. If $m(3)+2m(5) \geq 1$ we can fire t_3 (which deposits $\mathbf{I}(t_4)$) and then, so as not to deplete p_3 of its tokens, we can fire t_4 (which deposits $\mathbf{I}(t_2)$). This sequence is repeated a total of $m(4)$ times. The coefficient created is given by,

$$\left[\frac{f(t_3)}{f(t_4)}\frac{f(t_4)}{f(t_2)}\right]^{m(4)}. \tag{3.46}$$

At this stage there are $m(3)+2m(5)+m(4)$ tokens in p_3 and $m(2)+m(5)+m(4)$ tokens in p_2 and there are no tokens in either p_5 or p_4.

If $m(3)+2m(5)=0$, we require t_1 to fire, which deposits tokens into $\mathbf{I}(t_2)$ with probability a. Now fire t_3 (which deposits $\mathbf{I}(t_4)$) and then t_4 (which deposits $\mathbf{I}(t_2)$) a total of $m(4)$ times. Lastly, fire t_2 which redeposits the token removed from p_1 in the first step. The coefficient created is given by,

$$\frac{f(t_1)}{f(t_2)}\left[\frac{f(t_3)f(t_4)}{f(t_4)f(t_2)}\right]^{m(4)}\frac{f(t_2)}{f(t_1)},$$

$$=\left[\frac{f(t_3)}{f(t_2)}\right]^{m(4)},$$

which is equivalent to Equation (3.46). Consequently we can conclude that the same coefficient is obtained, whatever way p_4 is emptied.

Assume now, that we have a token distribution given by,

$$(m(1), m(2)+m(5)+m(4), m(3)+2m(5)+m(4), 0, 0).$$

The number of times that t_2 can fire, is given by,

$$\delta = min(m(2)+m(5)+m(4), [(m(3)+2m(5)+m(4))/2]),$$

where $m(2)+m(5)+m(4)$ is the number of times that t_2 can fire, before emptying p_2 and $[(m(3)+2m(5)+m(4))/2]$ is the number of times that t_3 can fire before emptying p_3, since two tokens are removed from p_3 at a time. $[(m(3)+2m(5)+m(4))/2]$ denotes the integer part of $(m(3)+2m(5)+m(4))/2$. The maximum number of times that t_2 can fire is given by the minimum of these two terms, which is δ. The coefficient created is given by,

$$\left[\frac{f(t_2)}{f(t_1)}\right]^{\delta}. \tag{3.47}$$

Note that if $m(2) + m(5) + m(4) \leq [(m(3) + 2m(5) + m(4))/2]$ there may be some residual tokens left in p_3. The base state in this case is given by the token distribution,

$$(m(1) + \delta, 0, m(3) + 2m(5) + m(4) - 2\delta, 0, 0)$$

If $m(2) + m(5) + m(4) \geq [(m(3) + 2m(5) + m(4))/2]$ then there may be some residual tokens left in p_2, and possibly one in p_3. The base state in this case, is given by the token distribution,

$$(m(1) + \delta, (m(2) + m(5) + m(4)) - \delta, m(3) + 2m(5) + m(4) - 2\delta, 0, 0)$$

The function $g(\mathbf{m})$, is given by the product of these coefficients, that is,

$$g(\mathbf{m}) = g(\mathbf{b}) \left(\frac{f(t_3)}{f(t_4)}\right)^{m(4)} \left(\frac{f(t_4)}{f(t_2)}\right)^{m(4)+m(5)} \left(\frac{f(t_2)}{f(t_1)}\right)^{\delta}, \tag{3.48}$$

$$= g(\mathbf{b}) \left(\frac{\chi(t_4)}{\chi(t_3)}\right)^{m(4)} \left(\frac{b\chi(t_2)}{\chi(t_4)}\right)^{m(4)+m(5)} \left(\frac{\chi(t_1)}{\chi(t_2)}\right)^{\delta}. \tag{3.49}$$

Treating $g(\mathbf{b})$ as a constant, set to unity, we must show that $g(\mathbf{m})$ satisfies Equations (3.40) to (3.44).

$$\frac{g(\mathbf{m} + \mathbf{I}(t_1))}{g(\mathbf{m} + \mathbf{I}(t_2))} = \frac{1}{\frac{f(t_2)}{f(t_1)}} = \frac{f(t_1)}{f(t_2)} = \frac{\chi(t_2)}{\chi(t_1)},$$

$$\frac{g(\mathbf{m} + \mathbf{I}(t_1))}{g(\mathbf{m} + \mathbf{I}(t_3))} = \left[\frac{f(t_3}{f(t_4)}\frac{f(t_4)}{f(t_2)}\frac{f(t_2)}{f(t_1)}\right]^{-1} = \frac{f(t_1)}{f(t_3)} = \frac{\chi(t_3)}{b\chi(t_1)},$$

$$\frac{g(\mathbf{m} + \mathbf{I}(t_3))}{g(\mathbf{m} + \mathbf{I}(t_4))} = \frac{\frac{f(t_3)}{f(t_4)}\frac{f(t_4)}{f(t_2)}\frac{f(t_2)}{f(t_1)}}{\frac{f(t_4)}{f(t_2)}\frac{f(t_2)}{f(t_1)}} = \frac{f(t_3)}{f(t_4)} = \frac{\chi(t_4)}{\chi(t_3)},$$

$$\frac{g(\mathbf{m} + \mathbf{I}(t_4))}{g(\mathbf{m} + \mathbf{I}(t_2))} = \frac{\frac{f(t_4)}{f(t_2)}\frac{f(t_2)}{f(t_1)}}{\frac{f(t_2)}{f(t_1)}} = \frac{f(t_4)}{f(t_2)} = \frac{b\chi(t_2)}{\chi(t_4)},$$

as required. The equilibrium distribution is therefore given by,

$$\pi(\mathbf{m}) = K\Phi(\mathbf{m})g(\mathbf{m}), \tag{3.50}$$

with $g(\mathbf{m})$ given in (3.49).

Note that p_3 is structurally implicit, because the vector $y = (0,2,0,1,0)$ satisfies the condition $Ay \leq A(p_3)$. However p_3 is not necessarily implicit, as condition 2 of Definition 2.10 can be violated. If $m(2) + m(3) + m(4) > [(m(3) + 2m(5) + m(4))/2]$ then there are reachable markings, such that p_2 will satisfy t_2's precondition for firing, but p_3 will not. This violates condition 2. Therefore p_3 can not in general, be removed from the net. In performing the procedure for finding $g(\cdot)$, we have made no allowances for implicit places, however, we hasten to add that the function $g(\cdot)$ we found, is valid for any arbitrary marking. In addition, we note the following points:

1. Note that neither $\varphi(\cdot)$ nor $\Phi(\cdot)$ have yet been specified, but that the solution has been found. As shown in Section 3.2, choosing $\varphi(\cdot)$ and $\Phi(\cdot)$ appropriately, can result in meaningful marking dependent firing rates.

2. The initial marking has not been used to find the solution given by Equation (3.50), therefore any choice of the initial marking would result in the SPN having the same equilibrium solution. The solution given by Equation (3.50) has normalising constant K, as it's only unknown. Once the initial marking is specified, the reachability graph comprising the set of markings, $\mathcal{M}$, may be found. By summing Equation (3.50) over the markings in $\mathcal{M}$ and noting that $\sum_{\mathbf{m}\in\mathcal{M}} \pi(\mathbf{m}) = 1$, the normalising constant can be calculated by,

$$K = \sum_{\mathbf{m}\in\mathcal{M}} [\Phi(\mathbf{m})g(\mathbf{m})]^{-1}.$$

Example 3.9

Figure 3.8 (which is equivalent to Figure 3.3), is another example of a SPN in which the input and output bags, given below, match.

$$\mathbf{I}(t_1) = (1,1,0,0), \quad \mathbf{O}(t_1) = (0,0,1,1),$$
$$\mathbf{I}(t_2) = (0,1,0,0), \quad \mathbf{O}(t_2) = (0,0,2,1),$$

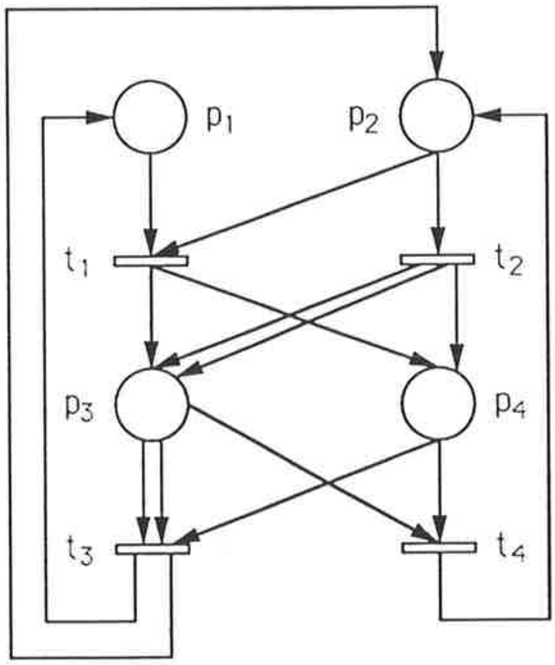

Figure 3.8: An unbounded SPN.

$$\mathbf{I}(t_3) = (0,0,2,1), \quad \mathbf{O}(t_3) = (1,1,0,0),$$
$$\mathbf{I}(t_4) = (0,0,1,1), \quad \mathbf{O}(t_4) = (0,1,0,0).$$

The one step probabilities are given by $\overline{p}(t_1,t_4)$, $\overline{p}(t_2,t_3)$, $\overline{p}(t_3,t_1)$, $\overline{p}(t_4,t_2)$ and are equal to one, so there is one positive recurrent communicating class given by $C_1 = \{t_1,t_2,t_3,t_4\}$. We can also find the recurrent classes by finding the T-invariants for the SPN of Figure 3.8. The incidence matrix is given by,

$$\begin{bmatrix} -1 & -1 & 1 & 1 \\ 0 & -1 & 2 & 1 \\ 1 & 1 & -2 & -1 \\ 0 & 1 & -1 & -1 \end{bmatrix}.$$

The T-invariant is (1,1,1,1), therefore, $\mathcal{C}_1 = \{t_1,t_2,t_3,t_4\}$ which is the positive recurrent communicating class, as given above.

Of interest in this example, is that place p_3 can accumulate tokens by firing the sequence t_2, t_4 arbitrarily often. Thus the SPN is unbounded. However, we are still able to find the equilibrium distribution for the SPN using the extended product form solution, as long as the underlying Markov process is stationary.

The invariant measure for the routeing chain is trivially,

$$f(t_i)\chi(t_i) = 1, \ \forall i.$$

Which implies that,

$$f(t_i) = [\chi(t_i)]^{-1}, \ \forall i.$$

To satisfy Equation (3.14), we need to find a function $g(\cdot)$, so that,

$$\frac{g(\mathbf{m}+\mathbf{I}(t_1))}{g(\mathbf{m}+\mathbf{I}(t_4))} = \frac{f(t_1)}{f(t_4)} = \frac{\chi(t_4)}{\chi(t_1)}, \tag{3.51}$$

$$\frac{g(\mathbf{m}+\mathbf{I}(t_2))}{g(\mathbf{m}+\mathbf{I}(t_3))} = \frac{f(t_2)}{f(t_3)} = \frac{\chi(t_3)}{\chi(t_2)}, \tag{3.52}$$

$$\frac{g(\mathbf{m}+\mathbf{I}(t_3))}{g(\mathbf{m}+\mathbf{I}(t_1))} = \frac{f(t_3)}{f(t_1)} = \frac{\chi(t_1)}{\chi(t_3)}, \tag{3.53}$$

$$\frac{g(\mathbf{m}+\mathbf{I}(t_4))}{g(\mathbf{m}+\mathbf{I}(t_2))} = \frac{f(t_4)}{f(t_2)} = \frac{\chi(t_2)}{\chi(t_4)}. \tag{3.54}$$

We use the same procedure for finding $g(\cdot)$ as in Example 3.8.

Let the base state be the marking with all of the tokens in places p_2. To empty place p_3, we need t_3 to fire, however, t_3 requires tokens in both p_3 and p_4. We consider the case that $m(4), m(1), m(2) > 0$ first. So as not to deplete p_4 of it's tokens, we fire t_1 (which deposits $\mathbf{I}(t_4)$) first, and then t_3 (which deposits $\mathbf{I}(t_1)$) next. This sequence is performed $m(3)$ times which gives the coefficient,

$$\left[\frac{f(t_1)}{f(t_4)}\frac{f(t_3)}{f(t_1)}\right]^{m(3)} = \left[\frac{f(t_3)}{f(t_4)}\right]^{m(3)}. \tag{3.55}$$

The marking of the SPN is now $(m(1), m(2), 0, m(4))$. Note that the SPN is not conservative, since the $m(3)$ tokens from p_3 have disappeared. However, this does not effect the end result. Now we take time out, to consider the boundary conditions. Since we are assuming that we are dealing with a stationary SPN, the initial markings that give deadlock can not occur.

1. If $m(1), m(2) = 0$, we must fire t_3 (which deposits $\mathbf{I}(t_1)$) first. This step removes two tokens from p_3, leaving $m(3)-2$ tokens, removes on token from p_4, leaving $m(4)-1$ tokens, and also deposits one token into places p_1 and p_2. Now fire the sequence, t_1 (which deposits $\mathbf{I}(t_4)$) and then t_3 (which deposits $\mathbf{I}(t_1)$), a total of $m(3)-2$ times. This gives the coefficient,

$$\left[\frac{f(t_3)}{f(t_1)}\right]\left[\frac{f(t_1)}{f(t_4)}\frac{f(t_3)}{f(t_1)}\right]^{m(3)-2}. \tag{3.56}$$

   The token distribution after this sequence of transition firings is now, $(1, 1, 0, m(4)-1)$.

   The next step, requires that we remove the $m(4)-1$ tokens from p_4. To do this, we fire t_2 (which deposits $\mathbf{I}(t_3)$) once, and then fire t_4 (which deposits $\mathbf{I}(t_2)$) twice, for a total of $m(4)-1$ times. This gives the coefficient,

$$\left[\frac{f(t_2)}{f(t_3)}\left(\frac{f(t_4)}{f(t_2)}\right)^2\right]^{m(4)-1}. \tag{3.57}$$

   The token distribution is now, $(1, m(4), 0, 0)$.

To reach the base state, we must remove the single token from p_1 to p_2. To do this, fire t_1 (which deposits $\mathbf{I}(t_4)$) and then fire t_4 (which deposits $\mathbf{I}(t_2)$). The coefficient created is,

$$\left[\frac{f(t_1)}{f(t_2)}\right]. \tag{3.58}$$

The token distribution is now, $(0, m(4), 0, 0)$ which is the base marking. the function $g(\cdot)$ is the product of Equations (3.56), (3.57) and (3.58), which gives,

$$g(\mathbf{m}) = g(\mathbf{b}) \left[\frac{f(t_3)}{f(t_4)}\right]^{m(3)} \left[\frac{f(t_4)^2}{f(t_3)f(t_2)}\right]^{m(4)}. \tag{3.59}$$

2. If $m(1), m(4) = 0$, we can fire t_2 (which deposits $\mathbf{I}(t_3)$) once, to deposit tokens in p_3 and p_4, and fire t_3 (which deposits $\mathbf{I}(t_1)$) which deposits a token in p_1 and p_2. Now we are able to follow the first firing sequence chosen, which fires t_1 (and deposits $\mathbf{I}(t_4)$) and then t_3 (which deposits $\mathbf{I}(t_1)$). This sequence of firings creates the coefficient,

$$\left[\frac{f(t_2)}{f(t_3)}\frac{f(t_3)}{f(t_1)}\right] \left[\frac{f(t_1)}{f(t_4)}\frac{f(t_3)}{f(t_1)}\right]^{m(3)},$$

$$= \left[\frac{f(t_2)}{f(t_1)}\right] \left[\frac{f(t_3)}{f(t_4)}\right]^{m(3)}. \tag{3.60}$$

The marking of the SPN is now (1,m(2),0,0). To get to the base state we must now remove the single token in p_1 and deposit it into p_2. The firing sequence t_1 (which deposits $\mathbf{I}(t_4)$) and then t_4 (which deposits $\mathbf{I}(t_2)$) does this and gives the coefficient,

$$\left[\frac{f(t_1)}{f(t_4)}\frac{f(t_4)}{f(t_2)}\right] = \left[\frac{f(t_1)}{f(t_2)}\right]. \tag{3.61}$$

The marking is now (0,m(2),0,0), which is the base marking since no other tokens are in any places other than p_2. The function $g(\cdot)$ is the product of the coefficients in Equations (3.60) and (3.61) which gives,

$$g(\mathbf{m}) = g(\mathbf{b}) \left[\frac{f(t_2)}{f(t_1)}\right] \left[\frac{f(t_3)}{f(t_4)}\right]^{m(3)} \left[\frac{f(t_1)}{f(t_2)}\right],$$

$$= \left[\frac{f(t_3)}{f(t_4)}\right]^{m(3)} . \tag{3.62}$$

as $g(\mathbf{b}) = 1$. Note that Equation (3.62) is the same as Equation (3.55). Consequently we obtain the same coefficient whatever way p_3 is emptied, subject to no deadlock.

Assume now that we have emptied the tokens from p_3 and are in marking $(m(1), m(2), 0, m(4))$. We now want to remove the $m(4)$ tokens from p_4. To do so, we fire t_2 (which deposits $\mathbf{I}(t_3)$) once and then fire t_4 (which deposits $\mathbf{I}(t_2)$) twice, for a total of $m(4)$ times. This creates the coefficient,

$$\left[\frac{f(t_2)}{f(t_3)}\left(\frac{f(t_4)}{f(t_2)}\right)^2\right]^{m(4)} . \tag{3.63}$$

The marking of the SPN is now given by (m(1),m(2)+m(4),0,0). We must now consider the boundary conditions.

1. Suppose $m(1), m(3) = 0$ (therefore the first step which removed tokens from p_3 was not performed). The same sequence of transition firings, given above, can be used in this situation.

Assume now that we are in marking (m(1),m(2)+m(4),0,0). To reach the base state, the last step to be performed is to remove the $m(1)$ tokens out of p_1. The sequence of transition firings given by, t_1 (which deposits $\mathbf{I}(t_4)$) and then t_4 (which deposits $\mathbf{I}(t_2)$) executed $m(1)$ times, will empty p_1. The coefficient created, is given by,

$$\left[\frac{f(t_1)}{f(t_4)}\frac{f(t_4)}{f(t_2)}\right]^{m(1)} = \left[\frac{f(t_1)}{f(t_2)}\right]^{m(1)} . \tag{3.64}$$

The marking is now given by (0,m(2)+m(4),0,0) which is the base marking.

The function $g(\cdot)$ is given by the product of Equations (3.55), (3.63) and (3.64), which gives,

$$g(\mathbf{m}) = g(\mathbf{b}) \left[\frac{f(t_3)}{f(t_4)}\right]^{m(3)} \left[\frac{f(t_4)^2}{f(t_3)f(t_2)}\right]^{m(4)} \left[\frac{f(t_1)}{f(t_2)}\right]^{m(1)} . \tag{3.65}$$

Note that, Equation (3.59) is of the same form as Equation (3.61) with $m(1) = 0$, again reinforcing the fact that, $g(\cdot)$ is found to be the same, independent of the order of transitions firings. To check that $g(\cdot)$ satisfies Equation (3.14), we let $g(\mathbf{b}) = 1$ and substitute it into Equations (3.51) to (3.54),

$$\frac{g(\mathbf{m}+\mathbf{I}(t_1))}{g(\mathbf{m}+\mathbf{I}(t_4))} = \frac{\dfrac{f(t_1)}{f(t_2)}}{\dfrac{f(t_3)}{f(t_4)}\dfrac{f(t_4)^2}{f(t_3)f(t_2)}} = \frac{f(t_1)}{f(t_4)},$$

$$\frac{g(\mathbf{m}+\mathbf{I}(t_2))}{g(\mathbf{m}+\mathbf{I}(t_3))} = \left[\left(\frac{f(t_3)}{f(t_4)}\right)^2 \frac{f(t_4)^2}{f(t_3)f(t_2)}\right]^{-1} = \frac{f(t_2)}{f(t_3)},$$

$$\frac{g(\mathbf{m}+\mathbf{I}(t_3))}{g(\mathbf{m}+\mathbf{I}(t_1))} = \frac{\left(\dfrac{f(t_3)}{f(t_4)}\right)^2 \dfrac{f(t_4)^2}{f(t_3)f(t_2)}}{\dfrac{f(t_1)}{f(t_2)}} = \frac{f(t_3)}{f(t_1)},$$

$$\frac{g(\mathbf{m}+\mathbf{I}(t_4))}{g(\mathbf{m}+\mathbf{I}(t_2))} = \frac{f(t_3)}{f(t_4)}\frac{f(t_4)^2}{f(t_3)f(t_2)} = \frac{f(t_4)}{f(t_2)}.$$

which confirms that $g(\mathbf{m})$ does satisfy (3.14). Therefore the equilibrium distribution, is given by,

$$\pi(\mathbf{m}) = K\Phi(\mathbf{m})g(\mathbf{m}), \tag{3.66}$$

with $g(\mathbf{m})$ given in Equation (3.65).

Note that there are no structurally implicit places in this SPN. Again there exist markings in which some places are implicit and other markings for which they are not, however, $g(\cdot)$ is valid for any arbitrary initial marking.

Again in this example, we have found $g(\cdot)$ and now have the freedom to choose an arbitrary $\Phi(\cdot)$ and $\varphi(\cdot)$ to determine the invariant measure, There is also no need to give the initial marking. With the appropriate choice of the functions $\Phi(\cdot)$ and $\varphi(\cdot)$, state dependent firing rates or blocking can be introduced as in Example 3.4. Again the initial marking specifies the reachability graph, and therefore influences the solution only through the normalising constant.

3.6 Other Work on Product Form solutions in SPNs

Lazar and Robertazzi [55] consider the underlying geometry of the state transition diagram of a SPN, to find some identifiable structure that has a product form equilibrium distribution. The motivation for this research, was to classify models, which contained blocking and which had a product form solution. Other papers applying the theory include, Lazar and Robertazzi ([56], [58]) and Cheng and Robertazzi [21].

They recognise that the state transition diagram of some SPNs have identifiable geometric shapes, in the form of multidimensional toroids and called them building blocks. When taken in isolation, the global balance equations for these building blocks give partial balance equations for the original process. In comparison, a queueing network with k queues, has building blocks which are formed by the set of states $\mathbf{n}$ and $\mathbf{n} - \mathbf{e}_i$, $1 \leq i \leq k$. When taken in isolation, the global balance equations for the building block are the partial balance equations for the original queueing network.

Once the geometric shapes have been identified, there is no guarantee that the global balance equations on the building blocks, are consistent over the different building blocks. To overcome this, Lazar and Robertazzi [55] construct what is called a consistency graph from the building blocks. The values associated with the movement from one state to another (indicated on the arcs of the consistency graph) are given by the ratios of the suggested equilibrium distributions at the states. For consistency validation, the product of these values around any closed path in the consistency graph must multiply to one.

The next Theorem links the routeing chain with the building blocks defined above.

Theorem 3.3

Any positive recurrent communicating class in the routeing chain determines a building block.

Proof

Consider the positive recurrent communicating class in the routeing chain, C_i, that contains transition t. From Theorem 3.2, the flux out of $\mathbf{m}+\mathbf{I}(t)$, due to the firing of t is equal to the flux into $\mathbf{m}+\mathbf{I}(t)$, due to the bag $\mathbf{I}(t)$ being deposited by another transition. Hence,

$$\pi(\mathbf{m}+\mathbf{I}(t))q(\mathbf{m}+\mathbf{I}(t),t)\sum_{v\in C_i}\overline{p}(t,v)=\sum_{s\in C_i}\pi(\mathbf{m}+\mathbf{I}(s))q(\mathbf{m}+\mathbf{I}(s),s)\overline{p}(s,t). \tag{3.67}$$

Note that $\sum_{v\in C_i}\overline{p}(t,v)=1$, but we leave it in for clarity.

This relation can be expressed equivalently, by noting the following. Let $\mathbf{m}$ be a base state. $\mathbf{m}+\mathbf{I}(t)$ is a marking of the process and $\mathbf{m}$ is a pseudo state as defined in Section 3.2. Let $q(\mathbf{m}+\mathbf{I}(t),t)\overline{p}(t,v)$, which is the rate of moving from $\mathbf{m}+\mathbf{I}(t)$ to $\mathbf{m}+\mathbf{I}(v)$ due to the firing of t and deposition of $\mathbf{I}(v)$ (where $\mathbf{O}_j(t)=\mathbf{I}(v)$ for some j), be equal to $q(\mathbf{m}+\mathbf{I}(t),\mathbf{m}+\mathbf{I}(v))$. Note that $q(\mathbf{m}+\mathbf{I}(t),t)\sum_{v\in C_i}\overline{p}(t,v)$, which is the rate out of $\mathbf{m}+\mathbf{I}(t)$, due to the firing of t and deposition of $\mathbf{O}(t)$, can also be expressed as $\sum_{v\in C_i}q(\mathbf{m}+\mathbf{I}(t),\mathbf{m}+\mathbf{I}(v))$.

Thus, the left hand side of Equation (3.67) is equivalent to

$$\pi(\mathbf{m}+\mathbf{I}(t))\sum_{v\in C_i}q(\mathbf{m}+\mathbf{I}(t),\mathbf{m}+\mathbf{I}(v)),$$

and the right hand side is equivalent to,

$$\sum_{s\in C_i}\pi(\mathbf{m}+\mathbf{I}(s))q(\mathbf{m}+\mathbf{I}(s),\mathbf{m}+\mathbf{I}(t)).$$

Let $B_i(\mathbf{m})=\{\mathbf{m}+\mathbf{I}(z)|z\in C_i\}$ be the set of states created by firing successive transitions in C_i. Hence, Equation (3.67) becomes,

$$\pi(\mathbf{m}+\mathbf{I}(t))\sum_{\mathbf{n}\in B_i(\mathbf{m})}q(\mathbf{m}+\mathbf{I}(t),\mathbf{n})=\sum_{\mathbf{n}\in B_i(\mathbf{m})}\pi(\mathbf{n})q(\mathbf{n},\mathbf{m}+\mathbf{I}(t)) \tag{3.68}$$

These equations are the partial balance equations for the original Markov process, on the set of states defined by $B_i(\mathbf{m})$. These are also the global balance equations on the set of states, $B_i(\mathbf{m})$. Thus, $B_i(\mathbf{m})$ is a building block determined by C_i. ■

Remark 3.5

The building blocks of Lazar and Robertazzi are multi-dimensional toroids. We have placed no physical restriction on the shape of the building block, as it results from the topological structure of the recurrent classes of the routeing chain.

Chapter 4

Insensitivity Theory Applied to SPNs

The type of firing distribution assigned to the transitions in a SPN are sometimes crucial in order to realistically model systems. For example, as mentioned earlier, reality would suffer if timeouts in a communication protocol were assumed to be negative exponential distributions instead of deterministic delays.

When using firing times which are all negative exponentially distributed, the underlying process of the SPN is Markovian. Consequently, the evaluation of the equilibrium distribution can theoretically be found by standard techniques such as the solution of the global balance equations. Or, as we pointed out in Chapter 3, for a certain class of SPN, an extended product form solution can be found. However with the introduction of generally distributed firing times the SPN no longer yields these useful forms of solution.

This chapter will be devoted to the theory of insensitivity as applied to SPNs. Without going into too much detail at this stage, a process is said to be "insensitive" to it's generally distributed lifetimes, if the equilibrium distribution of the process depends upon these generally distributed lifetimes only through their means. Therefore, what was once a potentially difficult problem due to general distributions can be reduced to a relatively simple problem involving only negative exponential distributions or any other convenient distribution with the same mean.

In the following section we give a survey of the work using the theory of insensitivity.

4.1 A Survey of the Theory of Insensitivity

The first example of insensitivity was discovered by Erlang [29] when he showed that the n server loss system with exponential service times of mean $[\mu]^{-1}$, has the same steady state distribution as the n server loss system with deterministic service times of mean $[\mu]^{-1}$. Kosten [54] and Fortet [30] elaborated on this find, showing that the service time distribution for an n server loss system can actually be general with mean $[\mu]^{-1}$, without affecting the form of the steady state distribution. Both authors, use supplementary variables to evaluate the equilibrium distribution.

Another queueing system which has similar properties to the Erlang loss system is the Engset loss system. This was shown to be insensitive with respect to generally distributed, but independent service times by Cohen [23] and with respect to successive service times, which come from a stationary point process by König [52].

Insensitivity, or robustness to many statisticians, has been historically studied within the framework of generalised semi-Markov processes (GSMPs) and consequently it is often difficult for the practising modeller to appreciate the value and relevance of the results. The remainder of this section is a brief summary of the work conducted on insensitivity in GSMPs and then, of more practical interest, a brief summary of the insensitivity properties that appear regularly in the theory of networks of queues.

Matthes [66] created the GSMP to study insensitivity within a wide class of stochastic processes. The GSMP is a generalisation of the semi-Markov process (SMP) as suggested by its name. The SMP resides in a particular state for a generally distributed length of time and then undergoes a transition to another state for a generally distributed length of time. In a GSMP, multiple generally distributed lifetimes can be alive simultaneously and the death of any one of them causes a transition to another state. To study the insensitivity of the GSMP, only one generally distributed lifetime is allowed to die in any transition, therefore the next state to be entered is restricted to the set of states that have the remaining generally distributed lifetimes alive.

Matthes [66] showed that a GSMP is insensitive with respect to a lifetime, if and only if, a set of balance equations are satisfied by the equilibrium distribution, when all of the lifetimes are taken to be exponential. These balance equations state that in equilibrium, the probability flux into any arbitrary state due to the creation of the lifetime s, is equal to the probability flux out of the same state due to the death of that lifetime. Using the definition of a SMP above, the balance equations for a SMP are always satisfied since there is only ever one generally distributed lifetime alive at one time. Therefore every SMP is insensitive to it's generally distributed lifetimes. A formal description of a GSMP and the equations for insensitivity will be given in Section 4.3 and Section 4.4 respectively.

A further generalisation was given in König and Jansen [53] who introduced state dependent speeds $c(s, \mathbf{g})$, at which the lifetime s is worked off when the state is $\mathbf{g}$. The practical uses of these speeds are numerous, in particular, they can be used to represent zero speeds. For example, it is possible to model the situation where a lifetime is not worked on at all in some states, and yet still retains it's spent or residual lifetime. This is useful, for example, when modelling the pre-emptive resume protocol. The instant a lifetime is pre-empted, the speed at which it is worked off is set to zero. This ensures that when a lifetime is re-enabled, it will resume from where it was pre-empted.

The existence of zero speeds introduces difficulties when proving the equivalence of insensitivity, product form and partial balance. In analysing systems with zero speeds, König and Jansen [53] assumed a property called instantaneous attention. This property requires a lifetime to be worked on at a positive speed immediately after being created. König and Jansen [53] point out that if a process does not possess instantaneous attention, it can be modified into one that does, by the addition of suitable extra states. Taylor [93] showed that it is not necessary to do this, as it is possible to treat insensitivity in processes without instantaneous attention directly and that it is the behaviour of the positive speed states that is important in determining whether a process is insensitive.

König and Jansen [53] and Schassberger [90] showed that the supplemented equilibrium distribution, when using residual lifetimes, can break down into a product of two terms. One term for the discrete part, the other for the continuous part. That is, the process behaves as if the discrete and continuous parts of the states are independent.

Schassberger ([88],[89]), proved Matthes' result using the method of phases and in [90] extended the result to incorporate speeds.

Henderson [35], considered the structure of the time reverse GSMP, and showed that a number of conditions are equivalent to insensitivity in a GSMP, P. In particular he observed that P is insensitive, if and only if, the supplemented equilibrium distribution, when the supplementary variables denote residual lifetimes, is the same as the supplemented equilibrium distribution, when the supplementary variables denote the spent lifetimes.

Henderson and Taylor [39], studied processes in which the generally distributed lifetimes could be prematurely ended by the deaths of other lifetimes, with subsequent state changes depending upon how the generally distributed lifetime died. As in the theory of insensitivity outlined above, Henderson and Taylor's [39] results depend upon the generally distributed lifetimes only through one parameter of each distribution. However, in this case, the parameter is no longer the mean, but the Laplace-Stieltjes transform of the distribution evaluated at a particular value. Rumsewicz and Henderson [87] followed, by allowing state changes to depend arbitrarily on the ages of the generally distributed lifetimes at their deaths. The equilibrium distribution for the Rumsewicz and Henderson [87] process, are derived using insensitivity concepts, but no longer depend on only a single parameter of each distribution.

A major application and source of results in insensitivity theory is in product form queueing networks. A model of a queueing network where the service time at a node is not always negative exponentially distributed, was considered by Baskett, Chandy, Muntz and Palacios [8] (BCMP). They allowed the nodes in the network to be of four different service disciplines, which are

1. First come first served: The service times are then negative exponentially distributed.
2. Processor sharing: Each customer may have a general service time distribution, given that the distribution has a rational Laplace transform.
3. Last come first served: Each customer may have a general service time distribution, given that the distribution has a rational Laplace transform.
4. Infinite server: The number of servers at a queue is greater than or equal to the maximum number of customers that can be queued at each queue. Each customer may have a general service time distribution, given that the distribution has a rational Laplace transform.

The condition of having a rational Laplace transform allows the service centre to be represented by a series of negative exponential stages. In Section 4.2.2 we summarise how some of the authors in the SPN literature have approximated generally distributed firing times by a series of negative exponential phases.

Kelly [50] showed that the key to allowing generally distributed service times in queueing networks, involves a class of queues called symmetric queues. The reason is that the equilibrium distribution of the network is insensitive to the service time distribution at a symmetric queue. Note that the service disciplines in BCMP, that allow generally distributed service times, are examples of symmetric queues. Other examples such as the queues with the WIERDP discipline of Chandy and

Martin [19] and the Last Batch Processor Sharing discipline of Noetzel [78], can also be represented as symmetric queues.

Kelly's techniques of proof were different to those of earlier authors. Instead of showing that symmetric queues satisfy partial balance equations, he used the time reversed process to show that postulated forms of the equilibrium distribution were correct. He also showed that a symmetric queue is quasi-reversible (see Kelly [51], pg 65) and hence a network involving symmetric queues with generally distributed service times has a product form given by Jackson or BCMP.

Chandy, Howard and Towsley [18], looked at what is essentially a symmetric queue in terms of local balance and station balance, rather than quasi-reversibility and derived insensitivity and product form results. Jansen and König [47], modified this work to include different classes of customers and showed that the output processes for nodes in a queueing network are Poisson, when the network possesses product form and insensitivity.

The results of Henderson and Taylor [39] and Rumsewicz and Henderson [87] on interruption processes and age dependent routeing, were successfully applied to Jackson/Kelly networks to show that a generalised product form solution still occurs, when the routeing of a departure from a symmetric queue is an arbitrary function of the length of service that the customer requested at the queue.

In the following sections we apply the theory of insensitivity to SPNs.

4.2 SPNs with Generally Distributed Firing Times

The techniques used in handling generally distributed firing times, can be classified into two categories. The first analyses the corresponding process of the SPN and the second replaces transitions which have generally distributed firing times, by a SPN which best approximates the distribution.

4.2.1 Use of the semi-Markov Process

In Section 2.5.1 we gave a summary of some of the different SPN models which use generally distributed firing times. These include deterministic firing times and arbitrarily distributed firing times. The one feature that the models such as Natkin [77] Bertoni and Torelli [11] and Dugan, Geist, Nicola and Trivedi [28] have in common, is that their underlying process is semi-Markovian.

Marsan and Chiola [64] impose the same restrictions as Dugan, Geist, Nicola and Trivedi in the DSPN model that governs which transitions can be given deterministic firing times, however, under special circumstances they introduce an extension. Recall from Section 2.5.1.

1. The set of concurrently enabled transitions (which may include exponentially distributed firing times) must be enabled at the same instant.

2. Each of these transitions can also represent a chain of transitions (that is a transition connected to a place connected to another transition etc.), however the firing distribution must be the same for each transition in the chain.

For the purposes of finding the equilibrium distribution, the time points of the process are then taken to be the instant that the deterministic transition starts firing, or has just been disabled. At the end of each interval the process will change states probabilistically, according to whether an exponentially distributed transition (or a set of exponentially distributed transitions) does, or does not fire within the interval. Under these conditions and using the suggested embedding procedure the process is a SMP.

Marsan, Balbo, Bobbio, Chiola, Conte and Cumani [61], [62] also use SMPs. As discussed in Section 2.5.1, the combination of the pre-selection and race policy with pre-emptive repeat gives a SMP. The combination of the race policy with pre-emptive resume or age-enabling gives a process that is not a SMP and is not necessarily a GSMP, so therefore the solution can often be intractable.

4.2.2 SPN Approximation of General Distributions

The second category approximates the general distributions, by using a series or parallel (or both) combination of exponential stages. The basic idea was first established by Erlang [29], who showed that a general distribution with coefficient of variation less than one, can be represented by a series of $r \geq 2$ exponential stages. This is the well known Erlang-r distribution. For a general distribution with coefficient of variation greater than one, $r \geq 2$ parallel exponential stages can be used as a representation. This is the hyperexponential-r distribution.

Cox [25], contributed further by showing that any distribution with a rational Laplace transform can be modelled by the combination of series and parallel stages.

One of the first applications of this idea to SPNs, was by Molloy [72]. Figure 4.1 is an example of a SPN representation of an Erlang-r distribution from Chen Bruell and Balbo [20], with mean service time $\frac{1}{\mu}$. In order to achieve this representation, it is necessary to have r transitions and $r + 1$ places. For the hyperexponential-r distribution, it is necessary to have $r + 2$ places and r^2 transitions.

It is clear to see, that the problem with direct substitution of these subnets is that the transitions and places, required to approximate the general distributions, enlarge the state space unnecessarily.

Chen, Bruell and Balbo [20] improve on this technique, by creating compact GSPN representations for Erlang's method of stages. In addition, the paper describes how the race policy with pre-emptive repeat, pre-emptive resume and age-enabling are incorporated into the model; this becomes necessary when there exist concurrent and conflicting transitions in the SPN. Figure 4.2 is an example of their representation of an Erlang-r distribution. This construction requires only 5 places and 3 transitions for the race policy with pre-emptive resume. The GSPN

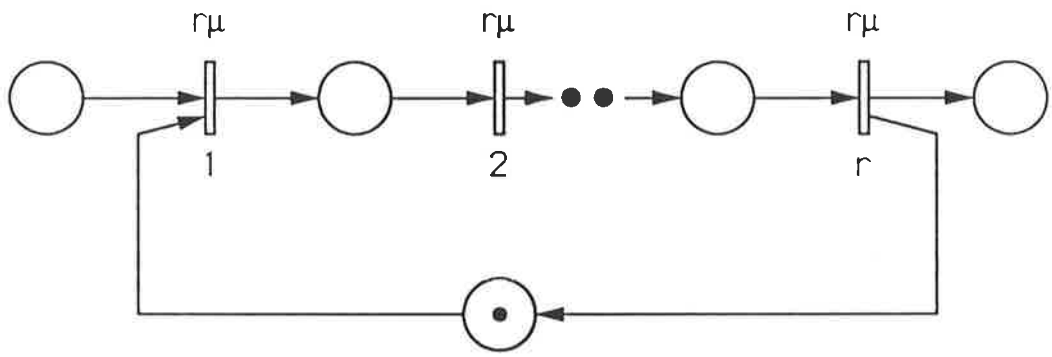

Figure 4.1: A SPN representation of the Erlang-r distribution.

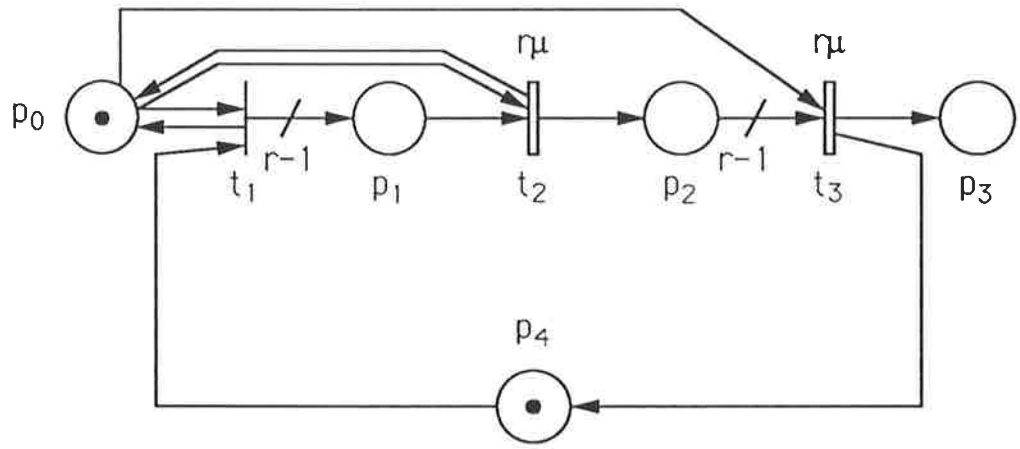

Figure 4.2: A GSPN representation of the Erlang-r distribution.

representations are automatically inserted into the SPN during the solution evaluation stage, the user is only required to specify the number of stages and the mean of the distribution. The resulting net is a GSPN, which can be solved using the techniques described in Section 2.5.1.

Another technique by Marsan, Balbo, Chiola, Conte and Cumani [62] mentioned in the last section, performs a stage expansion on the reachability graph. Each marking in the reachability graph is supplemented by a vector, which records the stage of firing for each transition. The different execution policies, race with pre-emptive repeat, pre-emptive resume or age-enabling are also incorporated into this technique. The expansion of the reachability tree is performed by an algorithm which has been automated in a software package by Cumani [27], called ESP.

Our aim in this chapter is to generalise the current work on SPNs with generally distributed firing times. This entails introducing the GSMP in detail, which we do in the following section. We then proceed by applying known insensitivity results for GSMPs to SPNs in Section 4.4 and then give the conditions for insensitivity. However, to obtain more powerful results we turn to some extensions to GSMPs, namely the age dependent routeing processes of Rumsewicz and Henderson [87] given in Section 4.5. These results give the foundations for the work on transition merging in Section 4.6 and marking amalgamation in Section 4.7. We introduce another extension to insensitivity in Sections 4.8 which is generalised in Section 4.9, where we show how to obtain equilibrium distributions for nets, which allow simultaneously enabled generally distributed firing times. Sections 4.3 to 4.7 are based on the paper by Henderson and Lucic [36].

4.3 A Generalised semi-Markov Process

A GSMP is defined on a set of states $\{\mathrm{g} : \mathrm{g} \in G\}$. Within each of these states are active elements s, from the set S which decay at the rate $c(s, \mathrm{g})$, $s \in S$. When the active element s dies, the process moves to state $\mathbf{g}' \in G$, with probability $p(\mathbf{g}, s, \mathbf{g}')$. If $s \in S'$, the element s has a negative exponentially distributed lifetime, and if $s \in S^{\star}$ it has an arbitrary general distribution, $S = S' \cup S^{\star}$. In order to establish insensitivity results it is standard to include the restriction that when the process changes from state $\mathbf{g}$ to state $\mathbf{g}'$, due to the death of s, no two elements from the set $S^{\star}$ may be activated or die simultaneously, and the remaining elements from $\mathbf{g} \cap S^{\star}$ retain their spent lifetimes.

We usually define the marking of a SPN, by the number and colours of tokens in each place. If the state of the SPN comprises this marking, along with the set of transitions enabled by this marking, then a GSMP structure is realised. State changes may be caused by either the firing of exponentially, or generally distributed transitions.

To model a SPN using a GSMP, the above definition requires that a state change cannot enable or fire, more than one generally distributed transition simultaneously and all other generally distributed transitions carry over their spent lifetimes to the next marking. Therefore, at first glance it would appear that no marking change can enable more than one transition simultaneously, nor can a transition firing disable other transitions. Fortunately we can, up to a point, relax both of these assumptions by utilising the results of Henderson and Taylor [39] and Rumsewicz and Henderson [87] and by noting that the GSMP definition allows the residual lifetimes of the active elements to decay at a state dependent speed. In particular, if transitions in conflict disable one another, we can set the decay rate of the disabled transitions to zero. The effect is to introduce a pre-emptive resume policy into the SPN, so that any transition re-enabled after it has earlier been disabled (without having fired), restarts its firing time from where it left off. We give examples in Section 4.5 which illustrate insensitivity in SPNs with conflict, when pre-emptive resume policies are adopted for disabled transitions.

4.4 Balance Equations, Insensitivity and Applications

In this section, we present the balance equations, that are necessary for insensitivity of a process with respect to a lifetime. Following this, we give examples to help reinforce the concept of insensitivity as applied to SPNs.

Within the framework of GSMPs, Matthes [66], showed that the following two statements are equivalent:

1. The process is insensitive with respect to the elements of S^*. That is, the general distributions of the lifetimes of the elements of S^*, can be replaced by any other distributions with the same mean, and yet the process still retains the same equilibrium distribution.

2. When all elements of S^* are assumed to be negative exponentially distributed, the flux out of each state due to the death of an element of S^*, is equivalent to the flux into that state due to the birth of that element.

The equivalent statements for SPNs with generally distributed transitions are:

1. The SPN is insensitive to each generally distributed transition t. That is, transition t's firing distribution can be replaced by any other distribution with the same mean, without altering the equilibrium distribution of the SPN.

2. The purely Markov process, i.e. when $S' = S$, has the property that for all transitions t, and for all the markings $\mathbf{m}_j$ that enable transition t, the flux into marking $\mathbf{m}_j$ which enables t, is balanced by the flux out of marking $\mathbf{m}_j$ due to the death of transition t.

Statement 2 is the "insensitivity balance equation" for transition t. There are some special circumstances, in which insensitivity balance is equivalent to the detail balance associated with reversibility and others when it is equivalent to the

Whittle (or local) balance (see Section 3.1), linked with product form queueing networks. In some cases insensitivity balance arises naturally in the global balance equations but, in general, it is none of the above and has a niche of its own in the fascinating relationship between properties of processes and certain kinds of balance equations.

As an illustration consider the following example.

Example 4.1

The SPN of Figure 4.3 represents a simplified model of a dual processor computer system given by Lazar and Robertazzi [57]. It consists of two processors vying for a common shared memory. The model Lazar and Robertazzi [57] use, is shown in Figure 4.4. The difference in the models, is that the access phase now has a more general distribution consisting of the additional property that a processor, interrupted in its attempt at accessing the memory, retains its access stage while the memory is being utilised by the other processor. The latter is an example of a pre-emptive resume policy. Lazar and Robertazzi [57] find the equilibrium distribution, after assuming that all transitions have negative exponentially distributed firing times. Let us return to Figure 4.3 and try to incorporate both of Lazar and Robertazzi's extensions into the simplified model, by introducing general distributions and examining the system for insensitivity. Consider the SPN of Figure 4.3 modelled as a GSMP with the initial marking as illustrated. The states of the GSMP are

$$\mathbf{m}_1 = (1,1,1,0,0,s_1,s_2),$$
$$\mathbf{m}_2 = (0,0,1,1,0,s_2,s_3),$$
$$\mathbf{m}_3 = (1,0,0,0,1,s_1,s_4).$$

where s_i is the element of S corresponding to transition t_i, $1 \le i \le 4$. Note the important point, that although states $\mathbf{m}_2$ and $\mathbf{m}_3$ contain the active transitions t_2 and t_1, respectively, they are being worked off at a zero speed, that is $c(s_2, \mathbf{m}_2) = 0 = c(s_1, \mathbf{m}_3)$. As the definition of a GSMP does not allow two lifetimes to die

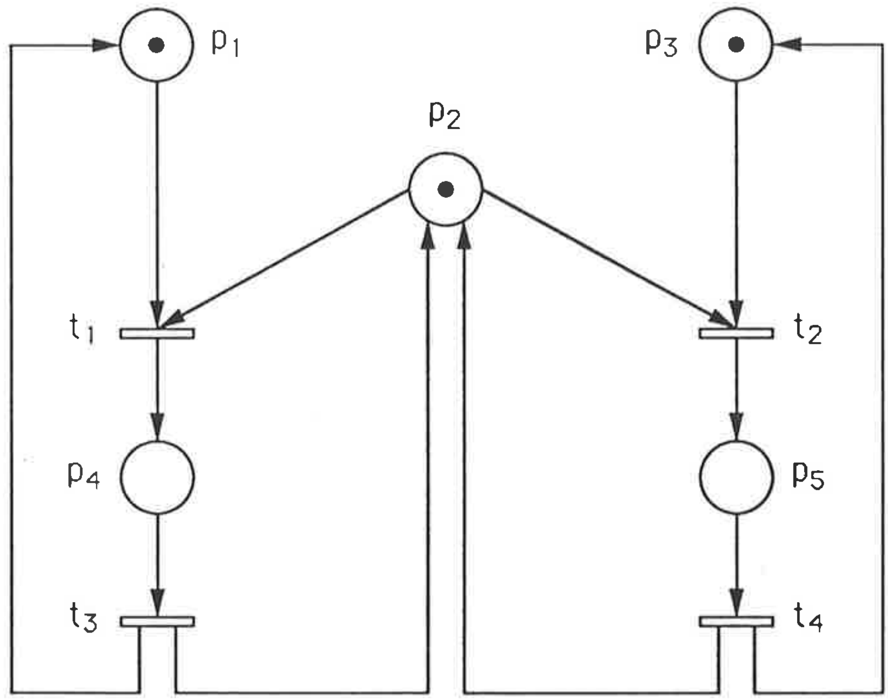

Figure 4.3: A SPN representation of a simplified dual processor model.

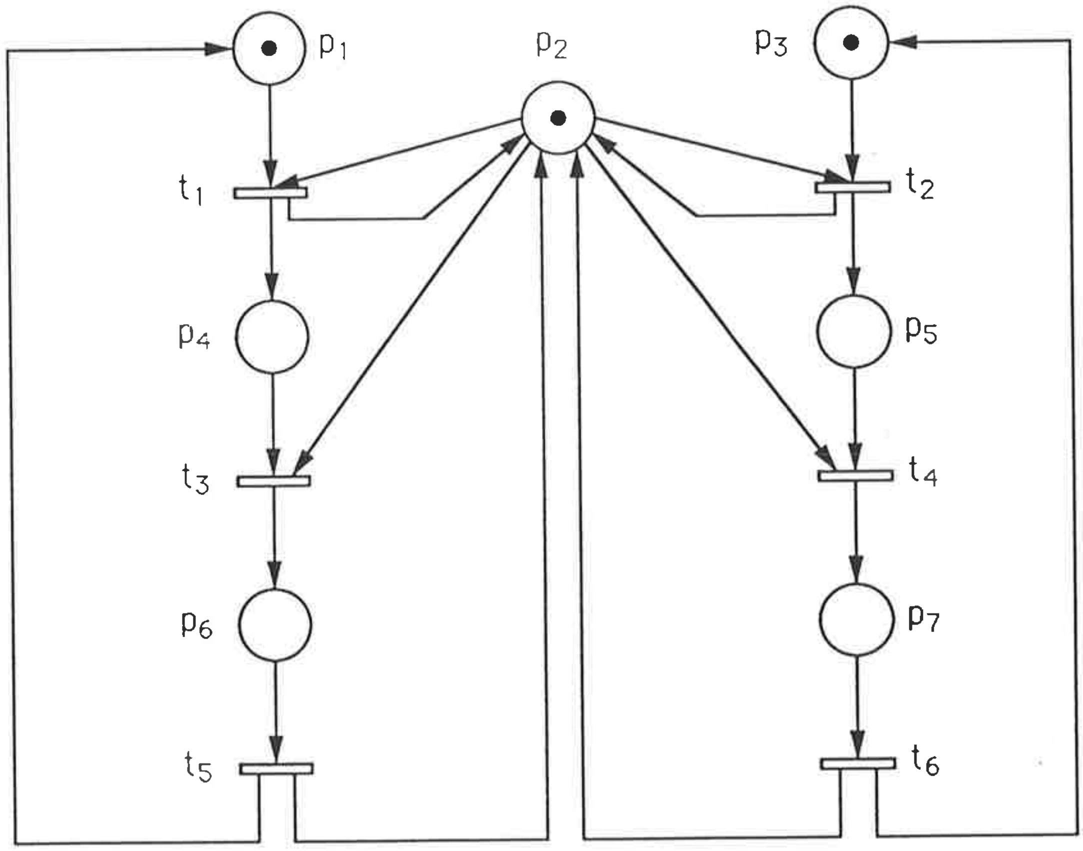

Figure 4.4: The dual processor model of Lazar and Robertazzi.

simultaneously, the firing of transition t_1 halts the firing of t_2, but it retains the spent lifetime until the token is returned to p_2. With this feature, even the simple example above is not a SMP. To test the transitions for insensitivity consider first the purely Markov version of the SPN in Figure **4.3** in which transition t_i fires with a negative exponentially distributed time with mean $(\mu_i)^{-1}$, $1 \leq i \leq 4$. We note that with negative exponential distributions, the cases of pre-emptive resume and pre-emptive repeat are identical. The global balance equations are,

$$\pi(\mathbf{m}_1)(\mu_1 + \mu_2) = \pi(\mathbf{m}_2)\mu_3 + \pi(\mathbf{m}_3)\mu_4, \tag{4.1}$$

$$\pi(\mathbf{m}_2)\mu_3 = \pi(\mathbf{m}_1)\mu_1, \tag{4.2}$$

$$\pi(\mathbf{m}_3)\mu_4 = \pi(\mathbf{m}_1)\mu_2. \tag{4.3}$$

The insensitivity balance equations must be satisfied for each state, from which the transition under scrutiny can fire. In this case, the equations are simple with only one relevant state for each transition.

$$\text{for } t_1 \qquad \pi(\mathbf{m}_1)\mu_1 = \pi(\mathbf{m}_2)\mu_3, \tag{4.4}$$

$$\text{for } t_2 \qquad \pi(\mathbf{m}_1)\mu_2 = \pi(\mathbf{m}_3)\mu_4, \tag{4.5}$$

$$\text{for } t_3 \qquad \pi(\mathbf{m}_2)\mu_3 = \pi(\mathbf{m}_1)\mu_1, \tag{4.6}$$

$$\text{for } t_4 \qquad \pi(\mathbf{m}_3)\mu_4 = \pi(\mathbf{m}_1)\mu_2. \tag{4.7}$$

It is obvious that the solution to the global balance equations, in satisfying Equations (4.2) and (4.3) automatically satisfy Equations (4.4), (4.5), (4.6) and (4.7). Consequently, the SPN is insensitive to each of the transitions t_1, t_2, t_3 and t_4.

Since the SPN is insensitive to t_1 and t_2, we are able to assign general firing time distributions to t_1 and t_2, without altering the equilibrium distribution obtained from Equations (4.1), (4.2) and (4.3). In particular, transitions t_1 and t_2 can be replaced by two consecutive negative exponentially distributed transitions separated by a place. Note that the pre-emptive resume assumption for disabled transitions is forced upon the model by the restrictions on insensitive GSMPs. As the two stages of the distribution are memoryless, we need not record the exact age of the spent lifetime, we need only ensure that the token freezes in its current place until the next transition in the sequence is re-enabled. This is Lazar and Robertazzi's model.

Wang and Robertazzi [96] generalised the SPN of Figure 4.4 by allowing the firing time distribution of each transition t_i, $1 \leq i \leq 6$ to be the sum of two negative exponential distributions. As in the previous example, this can be incorporated into the simplified model in Figure 4.3 using insensitivity. From the same argument used for Lazar and Robertazzi's model, we replace the firing time distribution of t_1 and t_2 by the sum of four negative exponential distributions and assume, as do Wang and Robertazzi [96], that the sequence of stages obey the pre-emptive resume protocol. Transitions t_3 and t_4 are also replaced by the sum of two negative exponential stages. There is no question that t_3 and t_4 obey the pre-emptive resume protocol, as they are never enabled together. The simplified model can be extended further, by allowing the firing time distributions of t_1, t_2, t_3 and t_4 to be the sum of an arbitrary number of negative exponential distributions. Again, this translates to having an arbitrary number of consecutive transitions separated by places.

Example 4.2

Now consider the same problem but in a more general setting. The structurally simple SPN of Figure 4.5, has a collection of resources available in place p_1 and a set of labelled customers in place p_2. The resources can, for example, be buses and memories as in the last example or circuits in a circuit switched network. Note that different customers require different sets of resources.

Label the customers 1,2,...,T and assume that a customer $j \in \{1, 2, \ldots, T\}$ requires the set A_j of resources to complete its task. Transition t_1 is a folded transition which serves all customers. Customer j draws a time from a general distribution, with mean $(\lambda_j)^{-1}$, and waits (or works) for that length of time before

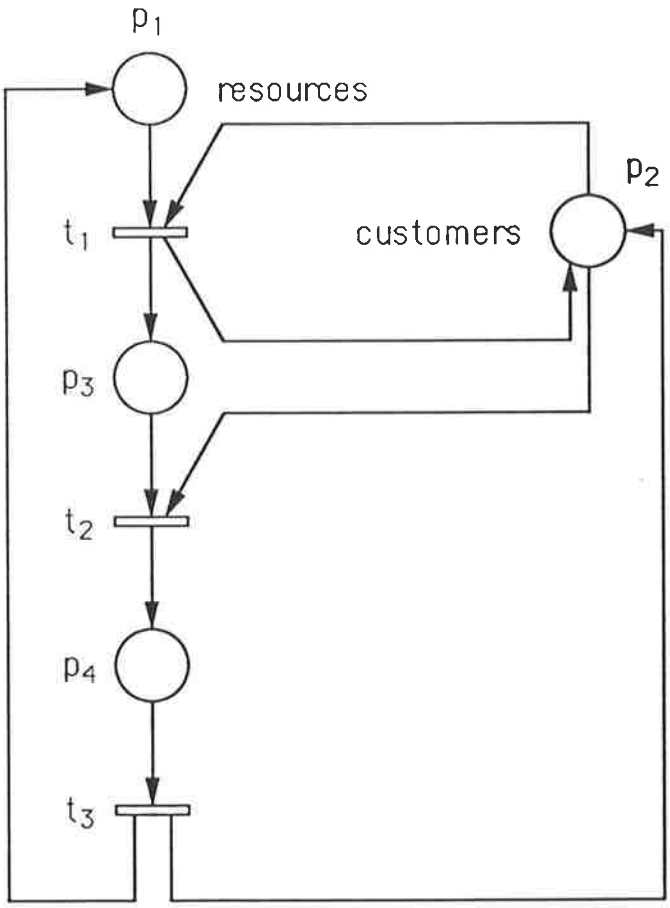

Figure 4.5: A High level PN of a circuit switched network.

accessing the resources. If the set of resources, A_j, are unavailable during that time period the countdown is temporarily halted for customer j, until the necessary resources become available in p_1. When these, "times to access service", are negative exponentially distributed, they are memoryless and many interpretations of the service procedure are possible. For example, we can assume that the countdown never stops and customers who request service but do not find the resources available, wait for another negative exponential period before making another request. Customer j accesses resources (in place p_3) for a period of time with mean, $(\mu_j)^{-1}$, before returning the resources to p_1 and themselves to p_2.

Define the state of the system as $\mathbf{n}$, a $|T|$ vector of zeros and ones, the ones indicating the customers currently utilising resources (in p_3). Consequently, $\mathbf{n}$ also indicates the active elements as s_{j0} or s_{j1}, depending upon whether customer j is waiting to access or is accessing the resources, respectively. The lifetimes s_{j0}, s_{j1} have means $(\lambda_j)^{-1}$ and $(\mu_j)^{-1}$ respectively.

If all lifetimes are negative exponentially distributed and $\mathbf{n} = (n_1, n_2, \ldots, n_T)$, $n_j = 0, 1, \forall j = 1, 2, \ldots T$, then the process is reversible with equilibrium distribution,

$$\pi(\mathbf{n}) = C \prod_{j=1}^{T} \left(\frac{\lambda_j}{\mu_j}\right)^{n_j}. \tag{4.8}$$

Where C, the normalising constant, is evaluated over all the feasible states of the process.

If $\mathbf{n}$ is a feasible state with $n_j = 1$, and $\mathbf{e}_j = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$ represents the state where only customer j is using resources, then the insensitivity balance equation for s_{j1} is found in the following way. The flux out of state $\mathbf{n}$ due to the death of s_{j1},

$$= \pi(\mathbf{n})\mu_j.$$

The flux into $\mathbf{n}$ due to the birth of s_{j1},

$$= \pi(\mathbf{n} - \mathbf{e}_j)\lambda_j.$$

Consequently, the insensitivity balance equation for s_{j1} is,

$$\pi(\mathbf{n})\mu_j = \pi(\mathbf{n} - \mathbf{e}_j)\lambda_j. \tag{4.9}$$

Equation (4.9) follows immediately from Equation (4.8) and is the insensitivity equation for s_{j0} also. For the case where customer j's resources are available in p_1, this is clearly true. Now consider the implication of Equation (4.9) on the behaviour of the Markov process when $\mathbf{n}$ is an infeasible state and $\mathbf{n} - \mathbf{e}_j$ is feasible. This occurs when customer j's resources are not available in $\mathbf{n} - \mathbf{e}_j$. For this case, $\pi(\mathbf{n}) = 0$ and the pre-emptive resume policy freezes the decay of the lifetime s_{j0} in state $\mathbf{n} - \mathbf{e}_j$, so $c(s_{j0}, \mathbf{n} - \mathbf{e}_j) = 0$, and thus we have equality in Equation (4.9). Therefore with zero speed in this situation, we have a pre-emptive resume protocol at t_1 which satisfies the necessary balance equations for all markings and is therefore insensitive to all of its active elements. Relating back to Example 4.1, we can therefore find performance measures based on the equilibrium distribution, when any number of processors are trying to access any set of memories with an arbitrary bus architecture, with any sequence of negative exponentially distributed transitions as per Wang and Robertazzi and with generally distributed memory utilisation times.

Equation (4.9) not only gives information on insensitivity, it also tells us that the SPN of Example 4.2 is reversible. We can further conclude with a little thought that the characteristic detailed balance equations of reversibility always indicate some form of insensitivity. However, as we pointed out, care must be taken when establishing the meaning of the active elements, and their behaviour when state changes occur, otherwise the insensitivity property can be lost.

Let us again stress, that there are many insensitive processes which are not reversible and many reversible purely Markov processes which, when arbitrarily given general distributions, are not insensitive.

4.5 Age Dependent Routeing

The results derived in the previous sections relied on basic insensitivity theory. In this section, we use the extensions to insensitivity from the work on GSMPs with interruptions by Henderson and Taylor [39], and GSMPs with age dependent routeing by Rumsewicz and Henderson [87], to analyse SPNs further. For example, in order to simplify the SPN structure, and underlying process, we turn to techniques such as transition merging in Section 4.6, marking amalgamation in Section 4.7 and aggregation of subnets of the SPNs in Chapter 5.

Although it is instructive to understand the work of Henderson and Taylor [39], from both an information and a modelling point of view, it is sufficient, for the purposes of this thesis, to present the definitions and results on age dependent routeing in GSMPs, due to Rumsewicz and Henderson [87].

The following results were given originally in terms of GSMPs. As we are simply translating them into an SPN structure, to obtain Theorem 4.1, we give no proof here but instead refer the interested reader to Rumsewicz and Henderson [87].

Let the process of the original SPN be denoted by P. Now alter the original process P by either performing marking amalgamation or aggregating subnets of the SPN. Assume that in performing any of these techniques we have created a process with markings $\overline{\mathrm{n}} \in \overline{\mathcal{M}}$, which has the following routeing structure. Define $p_{\overline{\mathrm{n}},\overline{\mathrm{m}}}(y,t)$ to be the probability of moving from marking $\overline{\mathrm{n}}$ to marking $\overline{\mathrm{m}}$ due to transition t firing at age y. Let $\overline{P}$ be the process which has, after the firing of a transition t with general distribution $G_t(\cdot)$, age dependent routeing probabilities given by $p_{\overline{\mathrm{n}},\overline{\mathrm{m}}}(\cdot,t)$. Assume that the age dependent routeing probabilities of $\overline{P}$ are sufficient to define P. Thus, $\overline{P}$ and P are structurally different but mathematically equivalent.

Let Q be the *averaged* process of $\overline{P}$ which has, after the firing of transition t

with general distribution $G_t(\cdot)$, age <u>*independent*</u> routeing probabilities given by,

$$p_{\overline{\mathbf{n}},\overline{\mathbf{m}}}(t) = \int_0^\infty p_{\overline{\mathbf{n}},\overline{\mathbf{m}}}(y,t)dG_t(y).$$

Define M to be the process which differs from the Q process, only in that t is now negative exponentially distributed with mean given by the mean of $G_t(\cdot)$. The next marking probability for M is given by $p_{\overline{\mathbf{n}},\overline{\mathbf{m}}}(t)$ for the merged transitions or otherwise from the original SPN. Let $\pi(\overline{\mathbf{n}}),\ \overline{\mathbf{n}} \in \overline{\mathcal{M}}$ be the equilibrium distribution of being in state $\overline{\mathbf{n}}$ of the M process.

In general, the $\overline{P}$, Q and M processes, will have different equilibrium distributions. Theorem 4.1 tells us when they all have the same equilibrium distribution, and therefore when the distribution of M is an exact marginal distribution of P.

Theorem 4.1 (Theorem 2 from Rumsewicz and Henderson [87])

If Q is insensitive to its generally distributed transitions, then the equilibrium distribution of Q is identical to the equilibrium distribution of $\overline{P}$ and is given by the equilibrium distribution of M.

Proof

See Rumsewicz and Henderson [87] ∎

Theorem 4.2 follows immediately from Theorem 4.1 and the definition of insensitivity.

Theorem 4.2

If the Q process is insensitive to its generally distributed transitions then the marginal equilibrium distribution of the original process P is $\pi(\overline{\mathbf{n}}), \overline{\mathbf{n}} \in \overline{\mathcal{M}}$.

Proof

Since we have stipulated that the age dependent routeing probabilities of $\overline{P}$ are sufficient to define P, the two processes are structurally different, but mathematically equivalent and therefore the equilibrium distribution for $\overline{P}$ is the marginal distribution for P. From Theorem 4.1, the equilibrium distribution of the Q process, which is $\pi(\overline{\mathbf{n}}), \overline{\mathbf{n}} \in \overline{\mathcal{M}}$, is the marginal distribution for the P process. ∎

Procedure:

The procedure, therefore, is to find the averaged Q process, establish the insensitivity, construct the related purely Markov process and solve for $\pi(\overline{\mathbf{n}}), \overline{\mathbf{n}} \in \overline{\mathcal{M}}$. Such a procedure is much simpler than working with phase type distributions, or supplementary variables.

In the next section we consider transition merging. The next corollary applies to transition merging and follows directly from Theorem 4.2.

Corollary 4.1

If $\overline{P}$ is formed from P by transition merging then, if Q is insensitive to its generally distributed transitions, the equilibrium distribution of P is $\pi(\mathbf{n}), \mathbf{n} \in \mathcal{M}$.

Proof

When applying this procedure the markings in $\overline{P}$ remain unchanged from the original process. Therefore $\mathcal{M} = \overline{\mathcal{M}}$ and so let $\mathbf{n} = \overline{\mathbf{n}}$ in Theorem 4.2. As the state space of Q is the same as the state space of P, the marginal distribution equilibrium distribution, in Theorem 4.2, becomes the equilibrium distribution. ∎

4.6 Transition Merging

Recall that the definition of a GSMP entails the creation and death of at most one generally distributed lifetime per marking change. In a SPN it is common for a number of transitions to be enabled simultaneously. In this section, we seek the equilibrium distribution for SPNs with sets of mutually disabling, simultaneously enabled, arbitrarily distributed transitions. The remainder of the SPN, can have any structure which is Markovian, that is, the other transitions are all negative exponentially distributed. We use the results of Section 4.5, to obtain equilibrium distributions for these SPNs.

Let the original SPN be P. Transform P into a SPN $\overline{P}$ which has age dependent routeing, by merging the mutually disabling transitions.

Consider any marking, $\mathbf{n}$, of P, in which there exists at least one set of mutually disabling, arbitrarily distributed transitions, which were enabled at the same time point. Assume that there are k transitions in this set, labelled (for simplicity) 1,2,...,k, and that transition t_i, $i \in \{1, 2, \ldots, k\}$ has firing time hazard function $h_i(\cdot)$, which satisfies,

$$h_i(u) = \frac{1}{1 - G_i(u)} \frac{dG_i(u)}{du}, \tag{4.10}$$

and consequently,

$$1 - G_i(y) = \exp\left\{-\int_0^y h_i(u)du\right\}. \tag{4.11}$$

Let marking $\mathbf{m}_i$ be the next marking entered if t_i fired first. Now combine the k conflicting transitions into one transition, T, which has a firing time distribution, $G_T(\cdot)$, given by the minimum of the k merged transition firing time distributions. The probability of moving into marking $\mathbf{m}_i$ from the original marking $\mathbf{n}$, when T fires at a time y, is given by,

$$p_{\mathbf{n},\mathbf{m}_i}(y, T) = \frac{h_i(y)}{\sum_{l=1}^{k} h_l(y)}. \tag{4.12}$$

It must be noted that a simplified notation has been adopted and that k and $h_i(y), 1 \leq i \leq n$ are functions of T. This procedure is followed for each set of mutually disabling, simultaneously enabled, arbitrarily distributed transitions.

Following the results of Section 4.5, we call the transition merged SPN, $\overline{P}$. The averaged SPN Q, based on $\overline{P}$, has transitions T, with firing time distributions $G_T(.)$ as defined earlier, and next marking probabilities given by,

$$p_{\mathbf{n},\mathbf{m}_i}(T) = \int_0^\infty \frac{h_i(y)}{\sum_{l=1}^{k} h_l(y)} dG_T(y). \tag{4.13}$$

Define the SPN, M, so that T is now negative exponentially distributed with mean given by the mean of $G_T(\cdot)$. The next marking probabilities for M, are given by Equation (4.13) when the merged transition fires. Otherwise, they are determined in the original SPN. Let $\pi(\mathbf{n}), \mathbf{n} \in \mathcal{M}$ be the equilibrium distribution of being in state $\mathbf{n}$ of the SPN, M. From Corollary 4.1, if Q is insensitive to the general distributions $G_T(\cdot)$, the equilibrium distribution for P is $\pi(\cdot)$.

The following example illustrates how to apply the procedure.

Example 4.3

Consider the SPN, P, in Figure 4.6. Let transition t_i have generally distributed firing times with distribution $G_i(\cdot)$, and hazard function $h_i(\cdot)$, for $5 \leq i \leq 7$. Let the remaining transitions t_i, $1 \leq i \leq 4$ have negative exponentially distributed firing times with means $(\mu_i)^{-1}$.

The states of this GSMP are given below. The state description gives the number of tokens in each place and implies the active elements (none of which have zero speeds).

$$\begin{aligned} \mathbf{m}_1 &= (1,1,1,0,0,0,0), & \mathbf{m}_2 &= (1,1,0,0,1,0,0), \\ \mathbf{m}_3 &= (0,1,1,1,0,0,0), & \mathbf{m}_4 &= (0,1,0,1,1,0,0), \\ \mathbf{m}_5 &= (1,0,0,0,0,0,1), & \mathbf{m}_6 &= (0,0,1,0,0,1,0), \\ \mathbf{m}_7 &= (0,0,0,1,0,0,1), & \mathbf{m}_8 &= (0,0,0,0,1,1,0). \end{aligned}$$

Applying the procedure of transition merging to the transitions t_5 and t_6 to create transition T, we are now able to construct the SPN, $\overline{P}$, given in Figure 4.7. Let the firing time distribution of T be $G_T(\cdot)$ with mean given by $(\mu)^{-1}$. $G_T(\cdot)$ is the distribution of the minimum of the firing times of t_5 and t_6.

The next marking probabilities, when T fires at time y, are given by,

$$q_i(y) = \frac{h_i(y)}{h_5(y) + h_6(y)}, \qquad i = 5, 6,$$

which route the tokens to places p_1 and p_2 with probability $q_6(y)$ (solid arcs) and places p_2 and p_4 with probability $q_5(y)$ (dashed arcs).

Define the SPN, Q, to be as in Figure 4.7, but with averaged probabilities,

$$q_i = \int_0^\infty \frac{h_i(y)}{h_5(y) + h_6(y)} dG_T(y), \qquad i = 5, 6.$$

Now define the purely Markov SPN, M, to have routeing probabilities given by q_5 and q_6 and transition firing rates given by $\mu_1, \mu_2, \mu_3, \mu_4, \mu$ and μ_7 for transitions t_1, t_2, t_3, t_4, T and t_7 respectively.

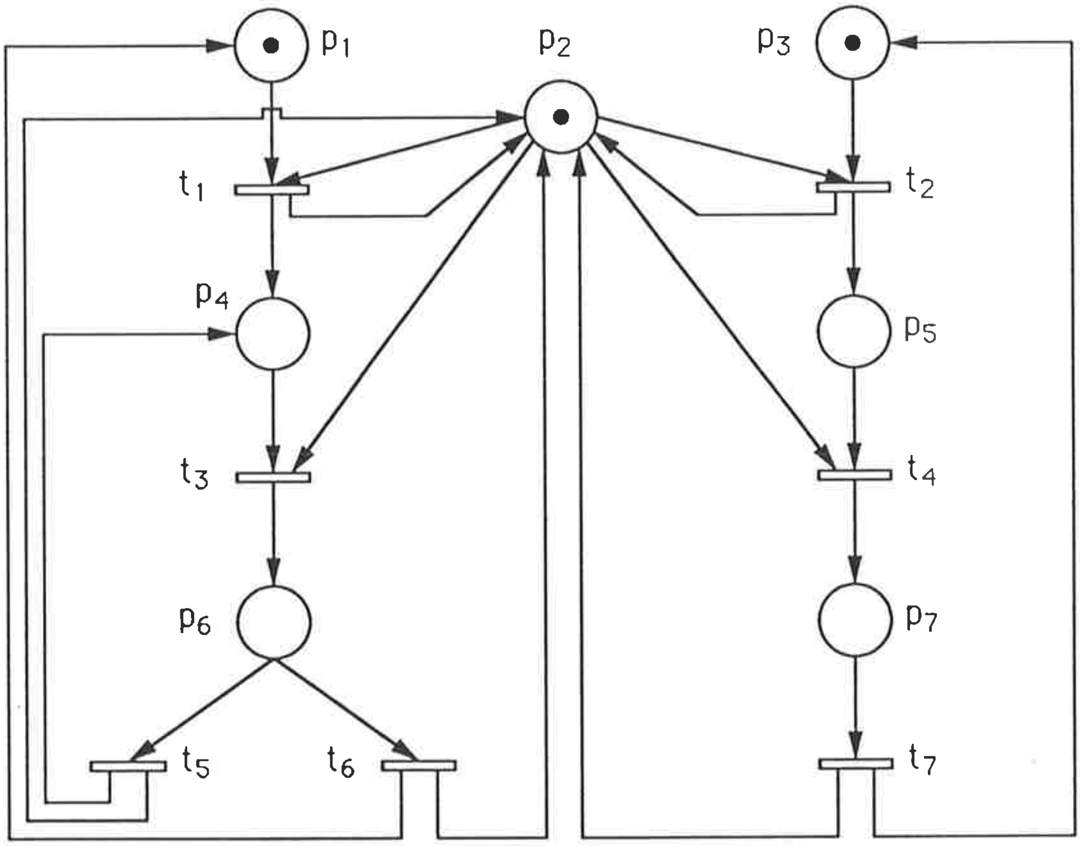

Figure 4.6: An altered dual processor model to illustrate transition merging.

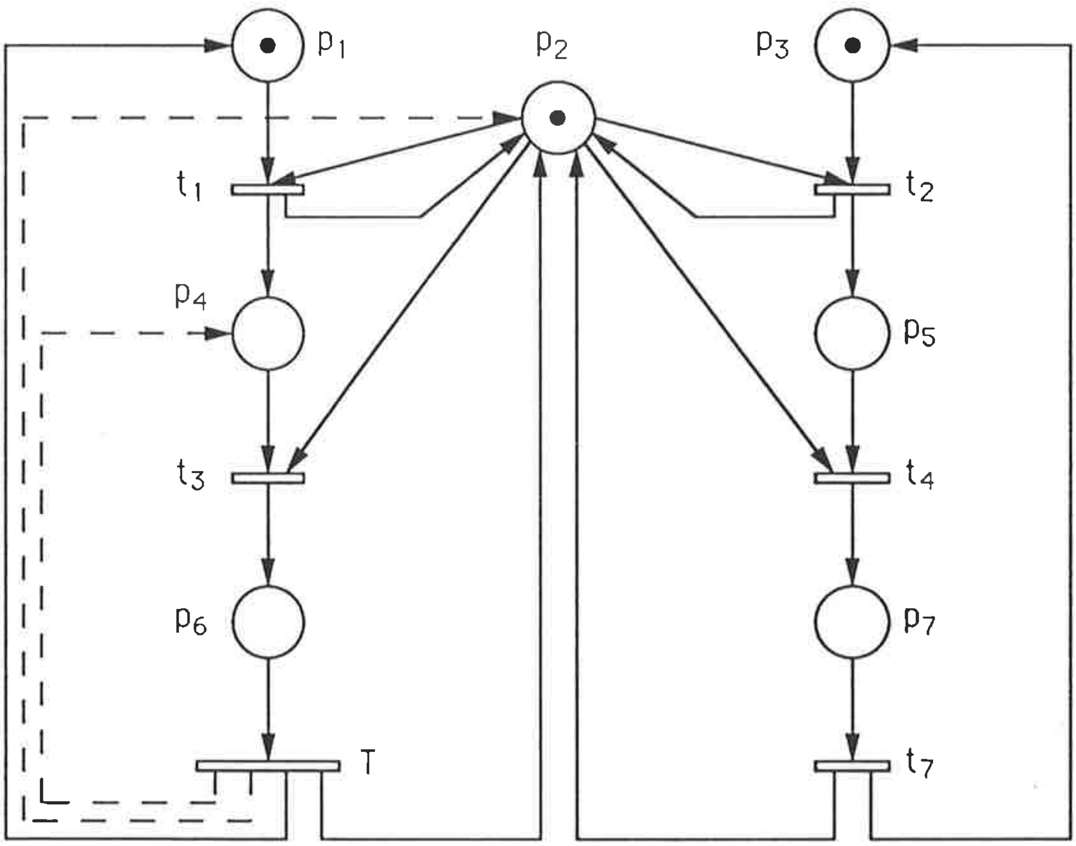

Figure 4.7: The dual processor of Figure 4.6 with a merged transition.

The global balance equations are,

$$\pi(\mathbf{m}_1)(\mu_1 + \mu_2) = \pi(\mathbf{m}_5)\mu_7 + \pi(\mathbf{m}_6)\mu q_6, \tag{4.14}$$

$$\pi(\mathbf{m}_2)(\mu_1 + \mu_4) = \pi(\mathbf{m}_1)\mu_2 + \pi(\mathbf{m}_8)\mu q_6, \tag{4.15}$$

$$\pi(\mathbf{m}_3)(\mu_2 + \mu_3) = \pi(\mathbf{m}_1)\mu_1 + \pi(\mathbf{m}_7)\mu_7 + \pi(\mathbf{m}_6)\mu q_5, \tag{4.16}$$

$$\pi(\mathbf{m}_4)(\mu_3 + \mu_4) = \pi(\mathbf{m}_2)\mu_1 + \pi(\mathbf{m}_3)\mu_2 + \pi(\mathbf{m}_8)\mu q_5, \tag{4.17}$$

$$\pi(\mathbf{m}_5)\mu_7 = \pi(\mathbf{m}_2)\mu_4, \tag{4.18}$$

$$\pi(\mathbf{m}_6)\mu = \pi(\mathbf{m}_3)\mu_3, \tag{4.19}$$

$$\pi(\mathbf{m}_7)\mu_7 = \pi(\mathbf{m}_4)\mu_4, \tag{4.20}$$

$$\pi(\mathbf{m}_8)\mu = \pi(\mathbf{m}_4)\mu_3, \tag{4.21}$$

$$\sum_{i=1}^{8} \pi(\mathbf{m}_i) = 1. \tag{4.22}$$

The insensitivity balance equations for the two generally distributed lifetimes are,

$$\text{for } T \qquad \pi(\mathbf{m}_6)\mu = \pi(\mathbf{m}_3)\mu_3, \tag{4.23}$$

$$\pi(\mathbf{m}_8)\mu = \pi(\mathbf{m}_4)\mu_3, \tag{4.24}$$

$$\text{for } t_7 \qquad \pi(\mathbf{m}_5)\mu_7 = \pi(\mathbf{m}_2)\mu_4, \tag{4.25}$$

$$\pi(\mathbf{m}_7)\mu_7 = \pi(\mathbf{m}_4)\mu_4. \tag{4.26}$$

Clearly the insensitivity balance equations are satisfied by the global balance equations, and thus the SPN of Figure 4.7 is insensitive to both T and t_7. The Q process is therefore insensitive and by Corollary 4.1, the equilibrium distribution of the SPN of Figure 4.6 is given by solving Equations (4.14) to (4.22).

This example illustrates the power of insensitivity theory in solving potentially difficult problems. In general, once the insensitivity of a SPN to some of its transitions is established, a large class of generalisations of the original SPN can be defined, all of which have the same equilibrium distribution. In the following section, we show how the same insensitivity results of Section 4.5, can be used with some kinds of marking amalgamation.

4.7 Marking Amalgamation

Many authors have had to tackle the problem of large state spaces whether their modelling tool was queueing networks, SPNs or any of the other frameworks useful for modelling systems. Aggregation is a necessary procedure in each discipline, in order to simplify the state space and thereby extract results. In queueing networks, authors such as Courtois [24], Vantilborgh [94], Balsamo and Iazeolla [7] and Chandy, Herzog and Woo [17], have used aggregation procedures on product form networks, principally so that the normalising constant could be evaluated more readily.

In this section we will suggest a procedure for marking amalgamation in SPNs, using the results of Section 4.4 and 4.5.

Dugan, Geist, Nicola and Trivedi [28], consider SPNs where at most one generally distributed transition is enabled in any marking and where each marking is visited only once. They suggest the amalgamation procedure of combining all markings which have a common generally distributed transition enabled. In the example provided in [28] the resultant amalgamated process is a SMP. In general, the amalgamation procedure described above, does not necessarily yield a SMP, since, the routeing between amalgamated markings may depend upon the past history. This problem does not occur, as was the case in the example of Dugan, Geist, Nicola and Trivedi [28], when the routeing probabilities are only functions of the marking itself and the time spent in that marking.

An additional problem with this amalgamation procedure can be seen by considering a G/M/1 queue. Combining all states with the same set of generally distributed active element, results in a simple renewal process, but provides no information on the state of the queue. In this case too much detail has been lost because of the over zealous amalgamation.

In this section, we suggest a general amalgamation procedure for use when the SPN exhibits markings, with at least one generally distributed enabled transition.

Using the results of Section 4.4 and 4.5, we derive conditions under which this procedure gives an exact marginal equilibrium distribution for the original SPN.

Procedure:

From the original SPN, P, with state space $\mathcal{M}$ amalgamate arbitrary sets of markings which have in common the same generally distributed enabled transitions.

The amalgamation will always define a simplified state space, $\overline{\mathcal{M}}$. It may not, however, naturally define a new process. For instance, there is no guarantee that, in amalgamating markings, the routeing probabilities between markings of $\overline{\mathcal{M}}$ can be evaluated. However, since the procedure involves ***arbitrary*** sets of markings, there is a great deal of flexibility and therefore many alternative choices in creating the markings in $\overline{\mathcal{M}}$. Finding the best $\overline{\mathcal{M}}$, which has well defined routeing probabilities, retains all the necessary information and is manageable, is a skill required of the analyst when using this procedure. When the routeing probabilities between amalgamated markings can be determined exactly, we label the new process $\overline{P}$.

The $\overline{P}$ process in the example of Dugan, Geist, Nicola and Trivedi [28] is insensitive and, as we have observed, insensitivity is a powerful modelling tool. Our aim in this section is to create an insensitive $\overline{P}$ process, by marking amalgamation and show that exact marginal distributions for P can be found from this $\overline{P}$ process. Theorem 4.3 is a preliminary theorem which illustrates that the insensitivity of P implies the insensitivity of any $\overline{P}$, when an amalgamation is performed in the suggested way.

Theorem 4.3

If P is insensitive to the transitions with generally distributed lifetimes, then $\overline{P}$ is also insensitive to these lifetimes.

Proof

For each marking $\mathbf{n}$, let $\mathbf{y}$ be a vector of, either the spent or the residual lifetimes of each of the generally distributed transitions enabled in $\mathbf{n}$. Let $(\mathbf{n}, \mathbf{y})$, be the marking of the SPN supplemented by these lifetimes.

As P is insensitive, the results of Matthes [66] and the subsequent extensions of König and Jansen [53] on spent lifetimes, and Henderson [35] on both spent and residual lifetimes, give the result,

$$\pi(\mathbf{n}, \mathbf{y}) = \pi(\mathbf{n}) \prod_i \mu_i(1 - G_i(y_i)). \tag{4.27}$$

$G_i(\cdot)$ is the distribution associated with the lifetime y_i and $[\mu_i]^{-1}$ is the mean of this distribution.

Summing (4.27) over all markings in the amalgamated marking $\overline{\mathbf{n}}$ and noting that each of these markings has the same set of generally distributed enabled transitions gives,

$$\pi(\overline{\mathbf{n}}, \mathbf{y}) = \pi(\overline{\mathbf{n}}) \prod_i \mu_i(1 - G_i(y_i)). \tag{4.28}$$

As the form of Equation (4.27) is necessary and sufficient for insensitivity and Equation (4.28) is valid for every marking $\overline{\mathbf{n}} \in \overline{\mathcal{M}}$, $\overline{P}$ is insensitive with equilibrium distribution given by (4.28). ■

Even though well defined routeing probabilities are not guaranteed when amalgamating markings, the consequence of Theorem 4.3, is that the equilibrium distribution on the amalgamated markings $\overline{\mathcal{M}}$ has a simple form, given by (4.28) whenever P is insensitive.

To calculate the equilibrium distribution $\pi(\overline{\mathbf{n}})$, we have two choices. We can recognise the insensitivity of the P process, replace the general distributions with negative exponential distributions, perform the amalgamation and solve the resulting system. Alternatively, we can amalgamate first, taking the general distributions into account, insert negative exponential distributions in place of the general distributions and then solve. The first approach works when P is insensitive. We proceed to show that the second alternative produces exact results in many situations even when P is not insensitive.

Let us now assume that we have a well defined $\overline{P}$ process.

The age dependent probabilities for moving from the amalgamated marking $\overline{\mathbf{n}} \in \overline{\mathcal{M}}$ to the next reachable markings in the process, are found by working out the probability of traversing a path within the amalgamated marking, which will lead to the next marking after the generally distributed transition fires.

The related age independent process Q, and the purely Markovian process M are obtained as in Section 4.5. Let M have an equilibrium distribution given by $\pi(\overline{\mathbf{n}}), \overline{\mathbf{n}} \in \overline{\mathcal{M}}$. Then from Theorem 4.2, if Q is insensitive, an exact marginal equilibrium distribution for the SPN P is given by $\pi(\overline{\mathbf{n}}), \overline{\mathbf{n}} \in \overline{\mathcal{M}}$.

The natural situations in which the conditions of Theorem 4.2 are satisfied are:

(a) The marking which first enables the generally distributed transition is always the same.

(b) The routeing probabilities in the P process from each marking in the amalgamated marking are independent of the entry point.

Note that Theorem 4.2 can accommodate the situation where the only set of enabled transitions are a set of mutually disabling transitions running concurrently with a set of negative exponentially distributed transitions. By merging the generally distributed, simultaneously enabled, mutually disabling transitions into one transition, as in Section 4.6, we obtain a Q process with a single generally distributed transition, which runs concurrently with some negative exponentially distributed transitions. Under the right conditions, Theorem 4.2 can now be applied to marking amalgamation.

The following example illustrates state amalgamation.

Example 4.4

Figure 4.8 is a high level stochastic Petri net comprising the following features. The initial marking consists of a hashed and a white token in p_1 and two neutral tokens in place p_2. Transitions t_1 and t_3 are both negative exponentially distributed with rate μ_1. Transition t_2 is a folded transition representing two

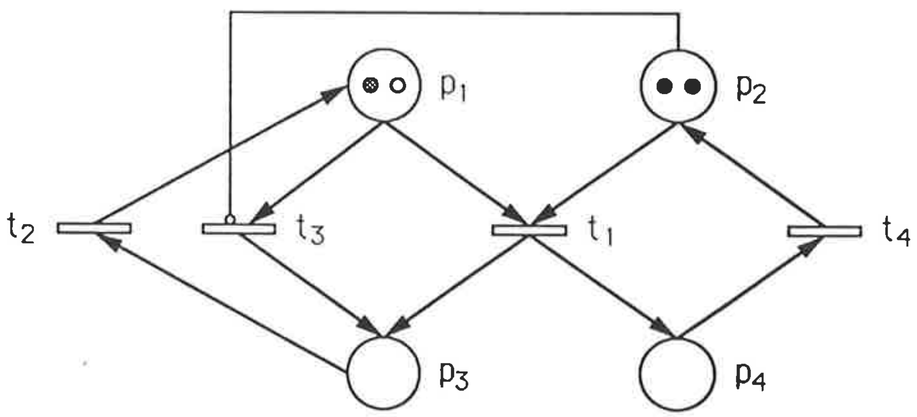

Figure 4.8: A coloured SPN to illustrate marking amalgamation.

transitions dedicated to the hashed and white tokens, respectively. The hashed (white) token remains in p_3 for a generally distributed time with mean $(\mu_h)^{-1}$ $((\mu_w)^{-1})$ before returning to p_1.

When both coloured tokens are in p_1, a coloured token is chosen randomly when t_1 or t_3 fires. The neutral tokens fire through t_1 independently and then fire through t_4, individually with rate μ_4 if both tokens are in place p_4, and rate μ_5 if only one token is in p_4. The inhibitor arc from p_2 to t_3 ensures that, even though there may be no tokens in p_2, which disables t_1, the hashed and white tokens continue being routed around the left hand side of the net.

The markings of this net can be described simply, by the colours and number of tokens in places p_1 and p_2. The marking also specifies the active lifetimes. They are given by,

$$\mathbf{m}_1 = (2,2),\ \mathbf{m}_2 = (2,1),$$
$$\mathbf{m}_3 = (2,0),\ \mathbf{m}_4 = (h,2),$$
$$\mathbf{m}_5 = (h,1),\ \mathbf{m}_6 = (h,0),$$
$$\mathbf{m}_7 = (w,2),\ \mathbf{m}_8 = (w,1),$$
$$\mathbf{m}_9 = (w,0),\ \mathbf{m}_{10} = (0,2),$$
$$\mathbf{m}_{11} = (0,1),\ \mathbf{m}_{12} = (0,0).$$

It can be shown that the process, P, comprising these markings is not insensitive to the generally distributed times that the hashed and white tokens spend in p_3. However, we will show that the following state amalgamation produces a well defined $\overline{P}$ which <u>*is*</u> insensitive to the generally distributed lifetimes. Consequently, an exact marginal distribution for P can be found by solving the global balance equations for $\overline{P}$.

Consider the following state amalgamation.

$$A = \{(2,2),(2,1),(2,0)\},$$
$$B = \{(h,2),(h,1),(h,0)\},$$

$$C = \{(w,2),(w,1),(w,0)\},$$
$$D = \{(0,2),(0,1),(0,0)\}.$$

State A represents all states which enable only negative exponentially distributed lifetimes. State B represents the set of states in which the white lifetime is enabled at t_2. State C represents the set of states in which the hashed lifetime is enabled at t_2. State D represents the set of states in which both the hashed and white lifetimes are enabled. Since both active elements are alive in state D, the amalgamated process $\overline{P}$ of Figure 4.9 is a GSMP but not a SMP.

Choosing this amalgamation has created a process with age independent routeing probabilities. Therefore, the issue of age dependent routeing does not have to be treated by averaging the routeing probabilities. Consequently the $\overline{P}$ and Q processes are equivalent and an insensitivity analysis on the $\overline{P}$ process, is all that is required.

In this case it is possible, with care, to write down the global balance equations for M directly. However, we find it more insightful to produce the M equations by summing the appropriate terms of the purely Markov version of the original process P, that is, where t_2 fires the hashed and white tokens according to negative exponential distributions with means $(\mu_h)^{-1}$ and $(\mu_w)^{-1}$ respectively. The global balance equations for this process are,

$$\pi(\mathbf{m}_1)\mu_1 = \pi(\mathbf{m}_2)\mu_5 + \pi(\mathbf{m}_4)\mu_w + \pi(\mathbf{m}_7)\mu_h, \tag{4.29}$$

$$\pi(\mathbf{m}_2)(\mu_1 + \mu_5) = \pi(\mathbf{m}_5)\mu_w + \pi(\mathbf{m}_8)\mu_h + \pi(\mathbf{m}_3)\mu_4, \tag{4.30}$$

$$\pi(\mathbf{m}_3)(\mu_1 + \mu_4) = \pi(\mathbf{m}_6)\mu_w + \pi(\mathbf{m}_9)\mu_h, \tag{4.31}$$

$$\pi(\mathbf{m}_4)(\mu_1 + \mu_w) = \pi(\mathbf{m}_{10})\mu_h + \pi(\mathbf{m}_5)\mu_5, \tag{4.32}$$

$$\pi(\mathbf{m}_5)(\mu_1 + \mu_w + \mu_5) = \pi(\mathbf{m}_1)\frac{\mu_1}{2} + \pi(\mathbf{m}_6)\mu_4 + \pi(\mathbf{m}_{11})\mu_h, \tag{4.33}$$

$$\pi(\mathbf{m}_6)(\mu_1 + \mu_w + \mu_4) = \pi(\mathbf{m}_2)\frac{\mu_2}{2} + \pi(\mathbf{m}_{12})\mu_h + \pi(\mathbf{m}_3)\frac{\mu_1}{2}, \tag{4.34}$$

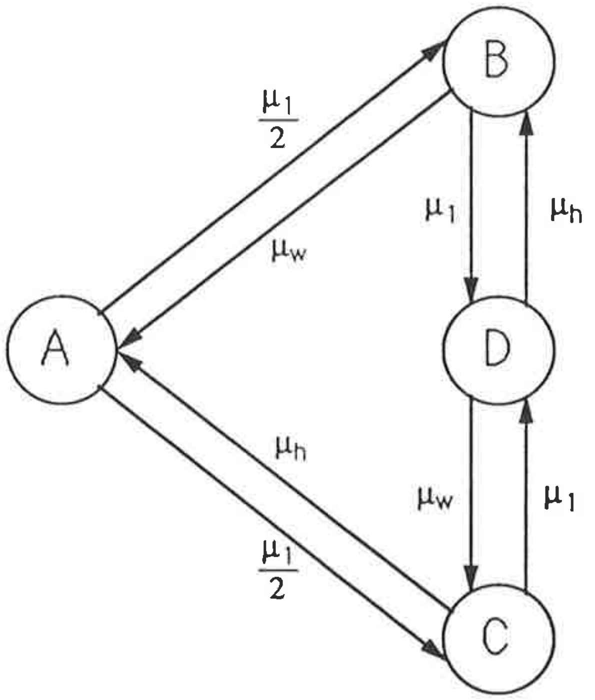

Figure 4.9: The amalgamated process of the SPN in Figure 4.8.

$$\pi(\mathbf{m}_7)(\mu_1 + \mu_h) = \pi(\mathbf{m}_{10})\mu_w + \pi(\mathbf{m}_8)\mu_5, \tag{4.35}$$

$$\pi(\mathbf{m}_8)(\mu_1 + \mu_h + \mu_5) = \pi(\mathbf{m}_1)\frac{\mu_1}{2} + \pi(\mathbf{m}_9)\mu_4 + \pi(\mathbf{m}_{11})\mu_w, \tag{4.36}$$

$$\pi(\mathbf{m}_9)(\mu_1 + \mu_h + \mu_4) = \pi(\mathbf{m}_2)\frac{\mu_1}{2} + \pi(\mathbf{m}_{12})\mu_w + \pi(\mathbf{m}_3)\frac{\mu_1}{2}, \tag{4.37}$$

$$\pi(\mathbf{m}_{10})(\mu_h + \mu_w) = \pi(\mathbf{m}_{11})\mu_5, \tag{4.38}$$

$$\pi(\mathbf{m}_{11})(\mu_h + \mu_w + \mu_5) = \pi(\mathbf{m}_{12})\mu_4 + \pi(\mathbf{m}_4)\mu_1 + \pi(\mathbf{m}_7)\mu_1, \tag{4.39}$$

$$\pi(\mathbf{m}_{12})(\mu_h + \mu_w + \mu_4) = \pi(\mathbf{m}_5)\mu_1 + \pi(\mathbf{m}_8)\mu_1 + \pi(\mathbf{m}_6)\mu_1 + \pi(\mathbf{m}_9)\mu_1, \tag{4.40}$$

$$\sum_{i=1}^{12} \pi(\mathbf{m}_i) = 1. \tag{4.41}$$

Since state A is the result of amalgamating states $\mathbf{m}_1, \mathbf{m}_2$ and $\mathbf{m}_3$, adding Equations (4.29), (4.30) and (4.31) will give the balance equation for state A. The equations for states B, C and D are found by adding the appropriate equations and are given below.

$$\pi(A)\mu_1 = \pi(B)\mu_w + \pi(C)\mu_h. \tag{4.42}$$

$$\pi(B)(\mu_1 + \mu_w) = \pi(D)\mu_h + \pi(A)\frac{\mu_1}{2}. \tag{4.43}$$

$$\pi(C)(\mu_1 + \mu_h) = \pi(D)\mu_w + \pi(A)\frac{\mu_1}{2}. \tag{4.44}$$

$$\pi(D)(\mu_h + \mu_w) = (\pi(B) + \pi(C))\mu_1. \tag{4.45}$$

$$1 = \pi(A) + \pi(B) + \pi(C) + \pi(D). \tag{4.46}$$

To show that $\overline{P}$ is insensitive to the times that the hashed and white tokens spend in p_3, we check for consistency between the global balance equations given above and the insensitivity balance equations.

The insensitivity balance equations are given by,

$$\text{for } \mathrm{t}_2 \text{ (white)}: \text{ state } B \qquad \pi(B)\mu_w = \pi(A)\frac{\mu_1}{2}, \tag{4.47}$$

$$\text{for } \mathrm{t}_2 \text{ (white)}: \text{ state } D \qquad \pi(D)\mu_w = \pi(C)\mu_1, \tag{4.48}$$

$$\text{for } \mathrm{t}_2 \text{ (hashed)}: \text{ state } C \qquad \pi(C)\mu_h = \pi(A)\frac{\mu_1}{2}, \tag{4.49}$$

$$\text{for } \mathrm{t}_2 \text{ (hashed)}: \text{ state } D \qquad \pi(D)\mu_h = \pi(B)\mu_1. \tag{4.50}$$

Equations (4.47) to (4.50) satisfy Equations (4.42) to (4.45) and therefore the $\overline{P}$ process is insensitive. Since insensitivity has been established, we are able to solve the global balance equations, (4.42) to (4.46), to obtain the marginal distribution for P as,

$$\pi(B) = \pi(A)\frac{\mu_1}{2\mu_w}, \tag{4.51}$$

$$\pi(C) = \pi(A)\frac{\mu_1}{2\mu_h}, \tag{4.52}$$

$$\pi(D) = \pi(A)\frac{\mu_1^2}{2\mu_h\mu_w}, \tag{4.53}$$

$$\pi(A) = \left[\frac{\mu_1^2}{2\mu_h\mu_w} + \frac{\mu_1}{2\mu_w} + \frac{\mu_1}{2\mu_h} + 1\right]^{-1}. \tag{4.54}$$

In this example, we have illustrated how the appropriate choice of marking amalgamation, results in an insensitive $\overline{P}$ process. There is a lot of flexibility in choosing the amalgamation, since there are many marking combinations possible. To re-iterate, the aim is to choose an amalgamation which produces well defined routeing probabilities, does not lose excessive information and which is insensitive.

4.8 Simultaneously Enabled Generally Distributed Transitions

In the previous sections, we noted that the theory of insensitivity was essentially restricted to models in which at most one lifetime could be born or die at any time instant. In this section, we present some preliminary ideas which relax the simultaneous births assumption.

Throughout this chapter, we have shown that the strength of insensitivity theory lies is in the fact that generally distributed firing times, may be replaced by negative exponentially distributed firing times with the same mean. We will aim for the same end result, when we apply our theory to fork-join sections in a SPN. In brief, so as not to pre-empt the definition, a fork-join section in a SPN may have an arbitrary number of generally distributed lifetimes enabled at the same time. As stated before, this poses a problem when trying to extract an equilibrium distribution. However, we will extend the theory of insensitivity and present a theory which allows the generally distributed firing times in the fork-join section to be replaced by state dependent negative exponentially distributed firing times with the same means. The price to be paid in removing the general distributions, are the state dependent firing rates, which are functions of the distributions themselves. We will start by presenting a simple example, to aid in the understanding of the theory, before immersing ourselves in the notation necessary for the general case.

Consider a fork-join section in a SPN. Following Baccelli, Makowski and Schwartz [5], an n-dimensional fork-join queue is represented by n parallel servers with synchronised arrival and departure streams. Customers arrive at the fork-join queue in batches less than or equal to n and then split up to be served by the n servers in parallel. This is the fork. As soon as the batch has been served, the customers are reunited. This is achieved by storing the customers, who have already been served, in an extra buffer of infinite size. They wait here until the other customers in the batch have been served. This is the join.

The analysis of fork-join queues can be applied to production systems, where a batch customer is interpreted as a customer order with several components requiring different services. Parallel processing is another area of application, where a batch customer can be viewed as a program composed of several subroutines, each to be executed by a different processor.

In the context of SPNs, an n-dimensional fork-join section is the subnet contained within an n-dimensional fork and an n-dimensional join. The subnet is executed in the same manner as described above, except, the batch is restricted to be of exactly size n. An n-dimensional fork in a SPN occurs when a transition deposits n tokens in n places simultaneously. An n-dimensional join in a SPN occurs when a transition removes all n tokens within the fork-join section at the same time. Note that the following analysis is applicable for only one batch of tokens moving through the fork-join section at a time. A new batch cannot enter the section until it is empty.

In the following example, we show how to convert the original SPN with general firing time distributions, into a structurally equivalent purely Markov SPN with state dependent firing rates that has the same equilibrium distribution.

We note, that there is a strong link between this work and that in the next chapter, which is concerned with the aggregation and disaggregation of subnets within a SPN. If we consider the sole general distribution, as the time spent in the fork-join section, we have a SMP and therefore insensitivity. It is therefore possible to use the supplementary variable to disaggregate and find detailed information about the markings in the subnet. This is the theme of the next chapter. Our aim in this section, is to find the equilibrium distribution of the detailed states of the fork-join section, using the information given about the transitions in the subnet. In doing so, we establish a different and potentially interesting result.

Example 4.5

Consider the SPN of Figure 4.10. Transition t_1 creates a 3-dimensional fork in the SPN, by depositing 3 tokens into places p_2, p_3 and p_4. The subnet of the fork-join section is formed by places p_2, p_3, p_4, p_5, p_6 and p_7 and transitions t_2, t_3

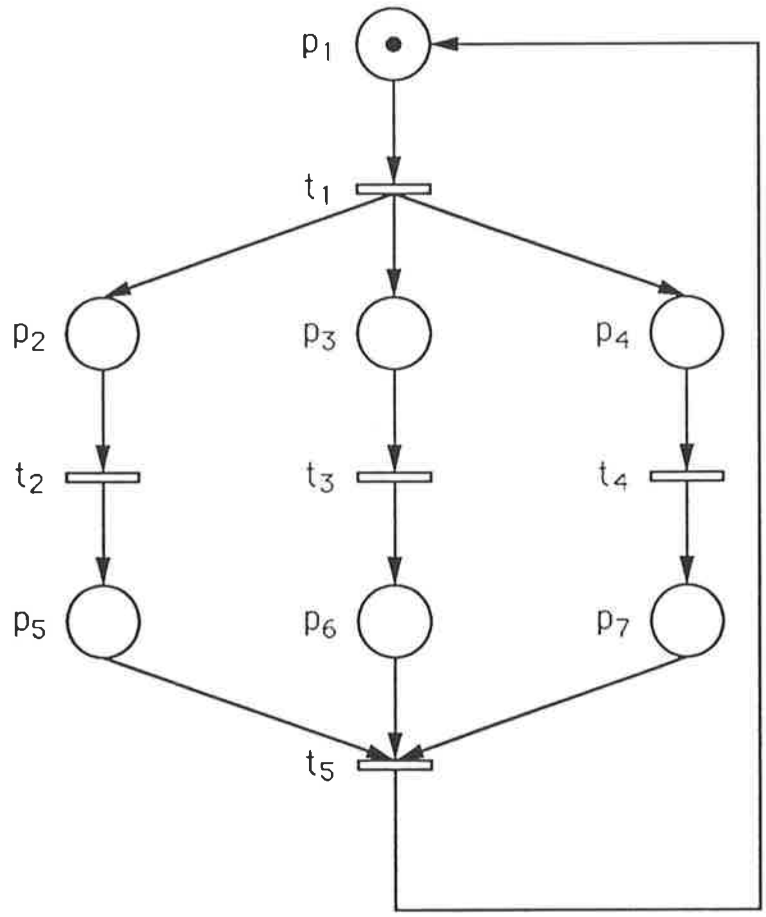

Figure 4.10: A 3-dimensional fork-join section.

and t_4. Transition t_5 constitutes a 3-dimensional join, by removing all 3 tokens from the subnet at the same time. Transitions t_1 and t_5 are negative exponentially distributed, with firing rates μ_1 and μ_5, respectively. Let transitions t_2, t_3 and t_4 have generally distributed firing times, given by the functions $G_2(\cdot), G_3(\cdot)$ and $G_4(\cdot)$, with hazard functions $h_2(\cdot), h_3(\cdot)$ and $h_4(\cdot)$, respectively. Standard insensitivity results do not apply to this example since the firing of transition t_1 enables transitions t_2, t_3 and t_4 simultaneously. We will show, that it is possible to transform the SPN into a purely Markov SPN with specific, state dependent firing rates which yields a simple solution.

The markings of the SPN are,

$$
\begin{aligned}
\mathbf{m}_1 &= (1,0,0,0,0,0,0),\\
\mathbf{m}_2 &= (0,1,1,1,0,0,0),\\
\mathbf{m}_3 &= (0,1,1,0,0,0,1),\\
\mathbf{m}_4 &= (0,1,0,0,0,1,1),\\
\mathbf{m}_5 &= (0,1,0,1,0,1,0),\\
\mathbf{m}_6 &= (0,0,1,1,1,0,0),\\
\mathbf{m}_7 &= (0,0,1,0,1,0,1),\\
\mathbf{m}_8 &= (0,0,0,1,1,1,0),\\
\mathbf{m}_9 &= (0,0,0,0,1,1,1),
\end{aligned}
$$

giving the number of tokens in each place.

For this example (and in a more general case shown later), the generally distributed transitions are enabled simultaneously, so we are able to use one supplementary variable to record the time spent in the fork-join section. We will resort to using spent time supplementary variables. In this case, the markings given above, need to be supplemented by a variable y, which records the spent time in the fork-join section. The supplemented balance equations of being in the fork-join section are given by:

$$\pi(\mathbf{m}_1)\mu_1 = \pi(\mathbf{m}_9)\mu_5.$$

$$\pi(\mathbf{m}_2, y + \Delta y) = \pi(\mathbf{m}_2, y)\left[1 - h_2(\Delta y)\Delta y - h_3(\Delta y)\Delta y - h_4(\Delta y)\Delta y\right].$$

$$\pi(\mathbf{m}_2, 0) = \pi(\mathbf{m}_1)\mu_1.$$

$$\pi(\mathbf{m}_3, y + \Delta y) = \pi(\mathbf{m}_2, y)h_4(y)\Delta y + \pi(\mathbf{m}_3, y)\left[1 - h_2(\Delta y)\Delta y - h_3(\Delta y)\Delta y\right].$$

$$\pi(\mathbf{m}_4, y + \Delta y) = \pi(\mathbf{m}_3, y)h_3(y)\Delta y + \pi(\mathbf{m}_5, y)h_4(y)\Delta y + \pi(\mathbf{m}_4, y)\left[1 - h_2(\Delta y)\Delta y\right].$$

$$\pi(\mathbf{m}_5, y + \Delta y) = \pi(\mathbf{m}_2, y)h_3(y)\Delta y + \pi(\mathbf{m}_5, y)\left[1 - h_2(\Delta y)\Delta y - h_4(\Delta y)\Delta y\right].$$

$$\pi(\mathbf{m}_6, y + \Delta y) = \pi(\mathbf{m}_2, y)h_2(y)\Delta y + \pi(\mathbf{m}_6, y)\left[1 - h_3(\Delta y)\Delta y - h_4(\Delta y)\Delta y\right].$$

$$\pi(\mathbf{m}_7, y + \Delta y) = \pi(\mathbf{m}_3, y)h_2(y)\Delta y + \pi(\mathbf{m}_6, y)h_4(y)\Delta y + \pi(\mathbf{m}_7, y)\left[1 - h_3(\Delta y)\Delta y\right].$$

$$\pi(\mathbf{m}_8, y + \Delta y) = \pi(\mathbf{m}_5, y)h_2(y)\Delta y + \pi(\mathbf{m}_6, y)h_3(y)\Delta y + \pi(\mathbf{m}_8, y)\left[1 - h_4(\Delta y)\Delta y\right].$$

$$\pi(\mathbf{m}_9) = \int_{y=0}^{\infty} \pi(\mathbf{m}_4, y)h_2(y)dy + \int_{y=0}^{\infty} \pi(\mathbf{m}_7, y)h_3(y)dy + \int_{y=0}^{\infty} \pi(\mathbf{m}_8, y)h_4(y)dy.$$

Also

$$\pi(\mathbf{m}_i, 0) = 0, \text{ for } 3 \le i \le 8,$$

since the probability of being in any of the states, other that $\mathbf{m}_2$, at $y = 0$, is zero. The supplemented differential equations are found by subtracting $\pi(\mathbf{m}_i, y)\Delta y$ from $\pi(\mathbf{m}_i, y + \Delta y)$ on the left hand side and $\pi(\mathbf{m}_i, y)\Delta y$ from the right hand side of each equation. Then, divide both sides by Δy and find the limit as $\Delta y \to 0$. The equations are then given by:

$$\frac{d\pi(\mathbf{m}_2, y)}{dy} = -\pi(\mathbf{m}_2, y)\left[h_2(y) + h_3(y) + h_4(y)\right]. \tag{4.55}$$

$$\pi(\mathbf{m}_2, 0) = \pi(\mathbf{m}_1)\mu_1. \tag{4.56}$$

$$\frac{d\pi(\mathbf{m}_3, y)}{dy} = \pi(\mathbf{m}_2, y)h_4(y) - \pi(\mathbf{m}_3, y)\left[h_2(y) + h_3(y)\right]. \tag{4.57}$$

$$\frac{d\pi(\mathbf{m}_4, y)}{dy} = \pi(\mathbf{m}_3, y)h_3(y) + \pi(\mathbf{m}_5, y)h_4(y) - \pi(\mathbf{m}_4, y)h_2(y). \tag{4.58}$$

$$\frac{d\pi(\mathbf{m}_5, y)}{dy} = \pi(\mathbf{m}_2, y)h_3(y) - \pi(\mathbf{m}_5, y)\left[h_2(y) + h_4(y)\right]. \tag{4.59}$$

$$\frac{d\pi(\mathbf{m}_6, y)}{dy} = \pi(\mathbf{m}_2, y)h_2(y) - \pi(\mathbf{m}_6, y)\left[h_3(y) + h_4(y)\right]. \tag{4.60}$$

$$\frac{d\pi(\mathbf{m}_7, y)}{dy} = \pi(\mathbf{m}_3, y)h_2(y) + \pi(\mathbf{m}_6, y)h_4(y) - \pi(\mathbf{m}_7, y)h_3(y). \tag{4.61}$$

$$\frac{d\pi(\mathbf{m}_8, y)}{dy} = \pi(\mathbf{m}_5, y)h_2(y) + \pi(\mathbf{m}_6, y)h_3(y) - \pi(\mathbf{m}_8, y)h_4(y). \tag{4.62}$$

$$\begin{aligned}\pi(\mathbf{m}_9)\mu_5 = &\int_{y=0}^{\infty} \pi(\mathbf{m}_4, y)h_2(y)dy \\ &+ \int_{y=0}^{\infty} \pi(\mathbf{m}_7, y)h_3(y)dy + \int_{y=0}^{\infty} \pi(\mathbf{m}_8, y)h_4(y)dy.\end{aligned} \tag{4.63}$$

Also

$$\pi(\mathbf{m}_i, 0) = 0, \text{ for } 3 \leq i \leq 8. \tag{4.64}$$

The solution to these supplemented differential equations, is given by the following and the proof is given in Appendix A.

$$\pi(\mathbf{m}_2) = \pi(\mathbf{m}_1)\mu_1 \int_0^{\infty} (1 - G_2(y))(1 - G_3(y))(1 - G_4(y))dy. \tag{4.65}$$

$$\pi(\mathbf{m}_3) = \pi(\mathbf{m}_1)\mu_1 \int_0^{\infty} (1 - G_2(y))(1 - G_3(y))G_4(y)dy. \tag{4.66}$$

$$\pi(\mathbf{m}_4) = \pi(\mathbf{m}_1)\mu_1 \int_0^\infty (1 - G_2(y))G_3(y)G_4(y)dy, \tag{4.67}$$

$$\pi(\mathbf{m}_5) = \pi(\mathbf{m}_1)\mu_1 \int_0^\infty (1 - G_2(y))(1 - G_4(y))G_3(y)dy, \tag{4.68}$$

$$\pi(\mathbf{m}_6) = \pi(\mathbf{m}_1)\mu_1 \int_0^\infty (1 - G_3(y))(1 - G_4(y))G_2(y)dy, \tag{4.69}$$

$$\pi(\mathbf{m}_7) = \pi(\mathbf{m}_1)\mu_1 \int_0^\infty (1 - G_3(y))G_2(y)G_4(y)dy, \tag{4.70}$$

$$\pi(\mathbf{m}_8) = \pi(\mathbf{m}_1)\mu_1 \int_0^\infty (1 - G_4(y))G_2(y)G_3(y)dy, \tag{4.71}$$

$$\pi(\mathbf{m}_9) = \pi(\mathbf{m}_1)\frac{\mu_1}{\mu_5}. \tag{4.72}$$

Note that $\pi(\mathbf{m}_1)\mu_1$ is the normalising constant.

Assume now that the SPN is purely Markov, with state dependent firing rates. If Equations (4.65) to (4.72) satisfy the global balance equations of the purely Markov SPN with state dependent firing rates, then an equivalent SPN with negative exponential state dependent firing rates can be created. Therefore, the next step is to find these state dependent firing rates from the global balance equations given below, using our knowledge of the $\pi(\cdot)$'s.

$$\pi(\mathbf{m}_1)\mu_1 = \pi(\mathbf{m}_9)\mu_5. \tag{4.73}$$

$$\pi(\mathbf{m}_2)\left[q(\mathbf{m}_2, \mathbf{m}_3) + q(\mathbf{m}_2, \mathbf{m}_5) + q(\mathbf{m}_2, \mathbf{m}_6)\right] = \pi(\mathbf{m}_1)\mu_1. \tag{4.74}$$

$$\pi(\mathbf{m}_3)\left[q(\mathbf{m}_3, \mathbf{m}_4) + q(\mathbf{m}_3, \mathbf{m}_7)\right] = \pi(\mathbf{m}_2)q(\mathbf{m}_2, \mathbf{m}_3). \tag{4.75}$$

$$\pi(\mathbf{m}_4)q(\mathbf{m}_4, \mathbf{m}_9) = \pi(\mathbf{m}_3)q(\mathbf{m}_3, \mathbf{m}_4) + \pi(\mathbf{m}_5)q(\mathbf{m}_5, \mathbf{m}_4). \tag{4.76}$$

$$\pi(\mathbf{m}_5)\left[q(\mathbf{m}_5, \mathbf{m}_8) + q(\mathbf{m}_5, \mathbf{m}_4)\right] = \pi(\mathbf{m}_2)q(\mathbf{m}_2, \mathbf{m}_5). \tag{4.77}$$

$$\pi(\mathbf{m}_6)\left[q(\mathbf{m}_6, \mathbf{m}_7) + q(\mathbf{m}_6, \mathbf{m}_8)\right] = \pi(\mathbf{m}_2)q(\mathbf{m}_2, \mathbf{m}_6). \tag{4.78}$$

$$\pi(\mathbf{m}_7)q(\mathbf{m}_7, \mathbf{m}_9) = \pi(\mathbf{m}_3)q(\mathbf{m}_3, \mathbf{m}_7) + \pi(\mathbf{m}_6)q(\mathbf{m}_6, \mathbf{m}_7). \tag{4.79}$$

$$\pi(\mathbf{m}_8)q(\mathbf{m}_8, \mathbf{m}_9) = \pi(\mathbf{m}_6)q(\mathbf{m}_6, \mathbf{m}_8) + \pi(\mathbf{m}_5)q(\mathbf{m}_5, \mathbf{m}_8). \tag{4.80}$$

$$\pi(\mathbf{m}_9)\mu_5 = \pi(\mathbf{m}_4)q(\mathbf{m}_4, \mathbf{m}_9) + \pi(\mathbf{m}_7)q(\mathbf{m}_7, \mathbf{m}_9) + \pi(\mathbf{m}_8)q(\mathbf{m}_8, \mathbf{m}_9). \quad (4.81)$$

Let us define some notation to reduce the complexity of some future expressions. First label the transitions $\{2, 3, \ldots, n+1\} = \mathcal{N}$ and let the set of labels for the transitions which have fired be given by the set $\mathcal{B} \subseteq \mathcal{N}$. Let

$$A(\mathcal{B} : i) = \int_0^\infty \prod_{j \in \mathcal{N} \setminus \{\mathcal{B} \cup \{i\}\}} (1 - G_j(y))dG_i(y), \quad (4.82)$$

where $A(\mathcal{B} : i)$ represents the probability that transition t_i fires first from the set of transitions with labels $\mathcal{N} \setminus \mathcal{B}$. Let $\{\}$ denote the empty set and so $A(\{\} : i)$ is the probability of the transition t_i firing first, when $\mathcal{B} = \{\}$. Also define $M(\mathcal{B})$ to be the expected residual time to the firing of the first transition with a label from $\mathcal{N} \setminus \mathcal{B}$, after all the transitions with labels in $\mathcal{B}$ have fired. This is written,

$$M(\mathcal{B}) = E\left[\left(\min_{k \in \mathcal{N} \setminus \mathcal{B}} X_k - \max_{l \in \mathcal{B}} X_l\right)^+\right], \quad (4.83)$$

$$= \int_0^\infty \prod_{i \in \mathcal{N} \setminus \mathcal{B}} (1 - G_i(y)) \prod_{j \in \mathcal{B}} G_j(y)dy, \quad (4.84)$$

where X_i is the random variable associated with the lifetime of t_i and $(x)^+ = \max(x, 0)$.

Let the state dependent firing rate $\mu(\mathcal{B} : j)$, be the rate of t_j firing when the transitions with labels in $\mathcal{B}$ have fired. By substituting the form for the $\pi(\cdot)$'s from Equations (4.65) to (4.72), into Equations (4.73) to (4.81), we can solve for the state dependent firing rates giving,

$$q(\mathbf{m}_2, \mathbf{m}_3) = \mu(\{\} : 4) = \frac{A(\{\} : 4)}{M(\{\})},$$

$$q(\mathbf{m}_2, \mathbf{m}_5) = \mu(\{\} : 3) = \frac{A(\{\} : 3)}{M(\{\})},$$

$$q(\mathbf{m}_2, \mathbf{m}_6) = \mu(\{\} : 2) = \frac{A(\{\} : 2)}{M(\{\})},$$

$$q(\mathbf{m}_3, \mathbf{m}_4) = \mu(\{4\} : 3) = \frac{A(\{4\} : 3) \times A(\{\} : 4)}{M(\{4\})},$$

$$q(\mathbf{m}_3, \mathbf{m}_7) = \mu(\{4\} : 2) = \frac{A(\{4\} : 2) \times A(\{\} : 4)}{M(\{4\})},$$

$$q(\mathbf{m}_5, \mathbf{m}_4) = \mu(\{3\} : 4) = \frac{A(\{3\} : 4) \times A(\{\} : 3)}{M(\{3\})},$$

$$q(\mathbf{m}_5, \mathbf{m}_8) = \mu(\{3\} : 2) = \frac{A(\{3\} : 2) \times A(\{\} : 3)}{M(\{3\})},$$

$$q(\mathbf{m}_6, \mathbf{m}_7) = \mu(\{2\} : 4) = \frac{A(\{2\} : 4) \times A(\{\} : 2)}{M(\{2\})},$$

$$q(\mathbf{m}_6, \mathbf{m}_8) = \mu(\{2\} : 3) = \frac{A(\{2\} : 3) \times A(\{\} : 2)}{M(\{2\})},$$

$$q(\mathbf{m}_4, \mathbf{m}_9) = \mu(\{3,4\} : 2),$$

$$= \frac{A(\{3,4\} : 2)\,[A(\{4\} : 3) \times A(\{\} : 4) + A(\{3\} : 4) \times A(\{\} : 3)]}{M(\{3,4\})},$$

$$q(\mathbf{m}_7, \mathbf{m}_9) = \mu(\{2,4\} : 3),$$

$$= \frac{A(\{2,4\} : 3)\,[A(\{4\} : 2) \times A(\{\} : 4) + A(\{2\} : 4) \times A(\{\} : 2)]}{M(\{2,4\}},$$

$$q(\mathbf{m}_8, \mathbf{m}_9) = \mu(\{2,3\} : 4),$$

$$= \frac{A(\{2,3\} : 4)\,[A(\{3\} : 2) \times A(\{\} : 3) + A(\{2\} : 3) \times A(\{\} : 2)]}{M(\{2,3\})}.$$

That is, we have found by using the state dependent firing rates, we obtain the same equilibrium distribution as though we used the general distributions. If we can now find a general procedure to produce these state dependent intensities, we will have side-stepped both the need to introduce supplementary variables for the general distributions and the need to solve integral equations. The concept is an extension of the standard insensitivity theory, where generally distributed firing rates could be replaced simply by their means.

In order to shed some light on how these state dependent firing rates are formed, we will discuss three representative expressions. The denominator of $q(\mathbf{m}_2, \mathbf{m}_3)$, is the expected firing time of the first transition to fire between t_2, t_3 and t_4, that is, $E[\min(X_2, X_3, X_4)]$, where X_j is the random variable associated with the lifetime of transition t_j. The numerator is the probability that transition t_4 will fire first between t_2, t_3 and t_4.

The denominator of $q(\mathbf{m}_3, \mathbf{m}_4)$, is the expected residual firing time of the first transition, out of t_2 and t_3, to fire after t_4 has fired. Mathematically, this is written as $E\left[(\min(X_2, X_3) - X_4))^+\right]$ where $(x)^+$ denotes $\max(x, 0)$. The numerator consists of the probability of t_4 firing first then t_3 firing second leaving t_2 still enabled. The denominator of $q(\mathbf{m}_4, \mathbf{m}_9)$ is again the expected residual firing time of t_2 after t_3 and t_4 have already fired. Mathematically, this is written as $E\left[(X_2 - \max(X_3, X_4))^+\right]$. The numerator consists of a sum of two probabilities, which represent two different orderings of the firing sequence. The first term represents the probability of the firing sequence $t_4 \rightarrow t_3 \rightarrow t_2$ and the second term represents the firing sequence $t_3 \rightarrow t_4 \rightarrow t_2$.

In the next section we use these observations to give state dependent intensities for the general fork-join case.

4.9 General Fork-Join Sections

Consider a SPN which enables n generally distributed transitions, within an n-dimensional fork-join section, given in Figure 4.11. Let t_1 and t_{n+2} have negative exponential firing times with mean $[\mu_1]^{-1}$ and $[\mu_{n+2}]^{-1}$, respectively. Let $t_2, t_3, \ldots, t_{n+1}$ have generally distributed firing times given by $G_i(\cdot)$, $2 \leq i \leq n+1$ with hazard functions $h_i(\cdot)$, $2 \leq i \leq n+1$. Also, let $\mathbf{m}_1$ denote the marking with one token in place p_1. Before embarking on the theory for the general case, we will give some additional definitions.

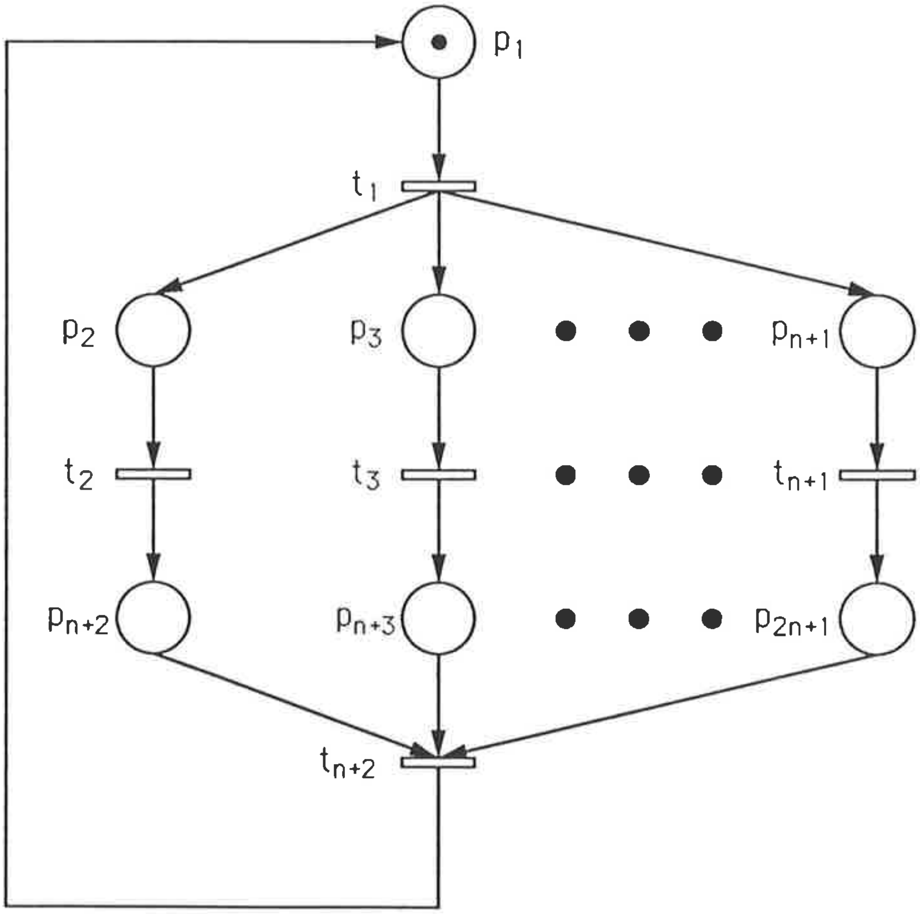

Figure 4.11: An n-dimensional fork-join section.

Definition 4.1

Recursively define, $P(\mathcal{B})$, a measure of the likelihood of being in the state where the transitions with labels in $\mathcal{B}$ have fired, as,

$$P(\mathcal{B}) = \sum_{i \in \mathcal{B}} P(\mathcal{B} \setminus \{i\}) A(\mathcal{B} \setminus \{i\} : i), \quad \mathcal{B} \neq \{\}, \tag{4.85}$$

with $P(\{\}) = 1$ and where $A(\{\} : i)$ is the probability of t_i firing first when no other transitions in $\mathcal{N}$ have fired. Note, that this can be normalised without effecting Theorem 4.5 or it's proof, however this is unnecessary.

Definition 4.2

For $j \in \mathcal{N} \setminus \mathcal{B}$,

$$\mu(\mathcal{B} : j) = \frac{P(\mathcal{B})}{M(\mathcal{B})} A(\mathcal{B} : j), \tag{4.86}$$

where $\mu(\mathcal{B} : j)$ is the firing rate of t_j when the transitions with labels in $\mathcal{B}$ have already fired, $M(\mathcal{B})$ is defined in Equations (4.83) and (4.84) and $A(\mathcal{B} : j)$ is defined in Equation (4.82).

Theorem 4.5.

The equilibrium distribution for the purely Markov n-dimensional fork-join section, with state dependent firing rates given by Equation (4.86), is given by,

$$\pi(\mathcal{B}) = KM(\mathcal{B}). \tag{4.87}$$

Proof

The global balance equations for the purely Markov n-dimensional fork-join section, are,

$$\pi(m_1)\mu_1 = \pi(\mathcal{N})\mu_{n+2}, \tag{4.88}$$

$$\pi(\{\}) \left[\sum_{j \in \mathcal{N}} \mu(\{\} : j) \right] = \pi(\mathbf{m}_1)\mu_1, \tag{4.89}$$

$$\pi(\mathcal{B}) \left[\sum_{j \in \mathcal{N} \setminus \mathcal{B}} \mu(\mathcal{B} : j) \right] = \sum_{i \in \mathcal{B}} \pi(\mathcal{B} \setminus \{i\}) \mu(\mathcal{B} \setminus \{i\} : i), \; \mathcal{B} \neq \{\}. \tag{4.90}$$

Substitute Equations (4.86) and (4.87) into the left hand side of Equation (4.89) gives,

$$KM(\{\})\left[\sum_{j\in\mathcal{N}}\frac{P(\{\})A((\{\}:j)}{M(\{\})}\right],$$

$$= K\sum_{j\in\mathcal{N}} A(\{\}:j).$$

Since $\sum_{j\in\mathcal{N}} A(\{\}:j) = 1$, this implies $\pi(\mathbf{m_1})\mu_1 = K$, as was noted in Section 4.8 with reference to the 3-dimensional fork-join section.

Substitution from Equations (4.86) and (4.87), into the right hand side of Equation (4.90) yields,

$$\sum_{i\in\mathcal{B}} KM(\mathcal{B}\setminus\{i\})\frac{P(\mathcal{B}\setminus\{i\})A(\mathcal{B}\setminus\{i\}:i)}{M(\mathcal{B}\setminus\{i\})}.$$

Using Equation (4.85) gives the right hand side as,

$$= KP(\mathcal{B}).$$

Substitution into the left hand side yields,

$$KM(\mathcal{B})\left[\sum_{j\in\mathcal{N}\setminus\mathcal{B}}\frac{P(\mathcal{B})A(\mathcal{B}:j)}{M(\mathcal{B})}\right].$$

However, $\sum_{j\in\mathcal{N}\setminus\mathcal{B}} A(\mathcal{B}:j) = 1$ so the left hand side

$$= KP(\mathcal{B}).$$

There is no need to satisfy Equation (4.88) as it is redundant. ∎

Theorem 4.6

The equilibrium distribution for the fork-join section with n generally distributed transitions is given by Equation (4.87).

Proof

Consider the supplemented equations for the n-dimensional fork-join section and conjecture that

$$\pi(\mathcal{B},y) = K\prod_{i\in\mathcal{N}\setminus\mathcal{B}}(1-G_i(y))\prod_{j\in\mathcal{B}}G_j(y)dy, \tag{4.91}$$

are the solutions to the supplemented differential equations. In general form the supplemented differential equations are given by,

$$\frac{d\pi(\mathcal{B},y)}{dy} = -\pi(\mathcal{B},y) \sum_{j\in\mathcal{N}\setminus\mathcal{B}} h_j(y) + \sum_{j\in\mathcal{B}} \pi(\mathcal{B}\setminus\{j\},y)h_j(y). \qquad (4.92)$$

Substitution from (4.91) into the left hand side of (4.92) gives,

$$\frac{d\left[\pi(\mathcal{B},y)\right]}{dy}$$

$$= -K \sum_{l\in\mathcal{N}\setminus\mathcal{B}} h_l(y) \prod_{k\in\mathcal{N}\setminus\mathcal{B}} (1-G_k(y)) \prod_{j\in\mathcal{B}} G_j(y)$$

$$+K \prod_{l\in\mathcal{N}\setminus\mathcal{B}} (1-G_l(y)) \left[\sum_{j\in\mathcal{B}} h_j(y)(1-G_j(y)) \prod_{k\in\mathcal{B}\setminus\{j\}} G_k(y)\right],$$

$$= -\pi(\mathcal{B},y)\left[\sum_{j\in\mathcal{N}\setminus\mathcal{B}} h_j(y)\right] + \sum_{j\in\mathcal{B}} h_j(y)(1-G_j(y)) \prod_{l\in\mathcal{N}\setminus\mathcal{B}} (1-G_l(y)) \prod_{k\in\mathcal{B}\setminus\{j\}} G_k(y),$$

$$= -\pi(\mathcal{B},y)\left[\sum_{j\in\mathcal{N}\setminus\mathcal{B}} h_j(y)\right] + \sum_{j\in\mathcal{B}} h_j(y) \prod_{l\in\mathcal{N}\cup\{j\}\setminus\mathcal{B}} (1-G_l(y)) \prod_{k\in\mathcal{B}\setminus\{j\}} G_k(y).$$

Since $\mathcal{N}\setminus\{\mathcal{B}\setminus\{j\}\} = \mathcal{N}\cup\{j\}\setminus\mathcal{B}$,

$$\pi(\mathcal{B}\setminus\{j\},y) = \prod_{l\in\mathcal{N}\cup\{j\}\setminus\mathcal{B}} (1-G_l(y)) \prod_{k\in\mathcal{B}\setminus\{j\}} G_k(y).$$

Hence,

$$\frac{d\left[\pi(\mathcal{B},y)\right]}{dy} = \pi(\mathcal{B},y) \sum_{j\in\mathcal{N}\setminus\mathcal{B}} h_j(y) + \sum_{j\in\mathcal{B}} \pi(\mathcal{B}\setminus\{j\},y)h_j(y),$$

which is Equation (4.92). Now integrate Equation (4.91) over y and using Equation (4.84) this gives (4.87) as required. ∎

The theory is applicable to any n-dimensional fork-join section within any SPN as long as the SPN is insensitive to the time spent in the n-dimensional fork-join section. This will be discussed in more detail in Chapter 5.

Chapter 5

Aggregation and Disaggregation in SPNs

5.1 Approximation Techniques

In Chapters 3 and 4, we provided classes of SPNs that have simple and exact solutions, using the extended product form solution and the theory of insensitivity. However, we have only just scratched the surface. Authors concerned with the performance analysis of SPNs, have responded to the need of finding a larger class of SPN with tractable solutions, by using a variety of approximation techniques. In this chapter, we give a survey of these techniques and then introduce our contribution to the area. In brief, our work is based on the aggregation and disaggregation of subnets of the SPN, but differs from previous work, by achieving exact solutions, whilst at the same time reducing the size of the reachability graph.

The approximation techniques used can be naturally divided into two categories, the first dealing with the underlying Markov process the second with the structure of the net.

Haas and Shedlar ([32], [33], [34]) find steady state estimation procedures for the discrete event simulation of SPNs, based on the underlying regenerative process structure and may therefore be included in the first category. They define the marking process of the SPN in terms of a GSMP (see Section 4.3 for a definition). The characteristic property of a regenerative stochastic process, is that there exists random time points, referred to as regeneration points, at which the process probabilistically restarts. The evolution of the process between regeneration points, is statistically identical to the process in any other cycle. Given that some mild regularity conditions are satisfied (see Crane and Iglehart [26]), a regenerative stochastic process, $\{X(t);\ t \geq 0\}$, has a limiting distribution, $\{X(t) \rightarrow X \text{ as } t \rightarrow \infty\}$, provided the expected time between regeneration points is finite. Haas and Shedlar [32] provide the conditions which ensure that the marking process is a regenerative process, in continuous time, with finite expected time between regeneration points. They extend their results to allow simultaneous transition firings in [33] and then analyse GSPNs in [34].

Ramamoorthy and Ho [82], were the first to find approximations based on the net structure. They use the property of marked graphs, that the number of tokens in a cycle, defined by the initial marking, remains the same after any firing sequence. We note that the S-invariant can be used to detect these cycles, as in Wong, Dillon and Forward [99], as outlined in Section 2.5.2. Ramamoorthy and Ho [82] compute the minimum cycle time to return to the initial marking, by finding the minimum time taken over all of the cycles. An approach for computing upper and lower bounds for cycle times, in conservative general SPNs is suggested, however, they comment that these bounds may not be tight.

Molloy [75], takes another approach, by finding fast upper bounds for the throughput in a SPN, using bottleneck analysis. Using this technique, the markings do not have to be generated. The SPN is restricted to have at most one arc between any place and any transition. By noting that average token flows in a net are conserved at steady state, a series of flow balance equations are constructed. Token flows are conserved in places, so the sum of all flows into a place is balanced by the sum of all flows out. This can be written as,

$$\sum_{t_i \in \mathcal{T}} f(t_i, p_j) = \sum_{t_i \in \mathcal{T}} f(p_j, t_i),\ \forall\, p_j \in \mathcal{P},$$

where $f(t_i, p_j)$ is the flow from transition t_i to place p_j and $f(p_j, t_i)$ is the flow from place p_j to transition t_i. Since there is at most one arc between any place and transition, the token flows, f_{t_i}, on the input and output arcs of a transition are equal. So, $\forall\ t_i \in \mathcal{T},\ \exists\ f_{t_i}$ such that,

$$f(t_i, p_j) = \begin{cases} 0 & \text{if } a_{ij}^{+} = 0, \\ f_{t_i} & \text{if } a_{ij}^{+} = 1, \end{cases}$$

$$f(p_j, t_i) = \begin{cases} 0 & \text{if } a_{ij}^{-} = 0, \\ f_{t_i} & \text{if } a_{ij}^{-} = 1, \end{cases}$$

where, a_{ij}^{+} and a_{ij}^{-} are the elements from which the incidence matrix is constructed, as defined in Section 2.1.2. The solution to these equations gives the token flows for loops in the SPN to within a constant. From the example given, the loops are chosen by the natural flows of the net, that is, by observing the cycles that tokens may flow around. Molloy [75] states that it is only possible to find an approximation for this constant, for a sufficiently large number of tokens in $\mathbf{m}_0$. As the number of tokens increases, the operation of the SPN tends toward a limit, if the average flow through a transition is less than the average firing rate. Therefore a further set of inequalities can be formed, given by,

$$f_t \le \lambda_t,\ \forall t \in \mathcal{T}.$$

These inequalities, define a feasible region for the vector of average token flows. The boundary of feasible solutions for these equations, defines the bounds of token flow through the net. That is, the boundary $f_t = \lambda_t$, constitutes a bottleneck in the system, which limits the throughput to some maximum value.

Campos, Chiola, Colom and Silva [15] create linear programs defined on the incidence matrix, and use them to compute the upper and lower bounds for the throughput of transitions. The throughput of the transition is defined as the average number of firings per unit time. The SPN is restricted to be a live and bounded marked graph, and the underlying Markov process must be ergodic. The upper and lower bounds for the throughput, are found to depend on the initial marking and the mean values of the delays, but not on the probability distributions. This approach is different from that of Molloy [75], in that only the initial marking is required, rather than having to compute the limiting behaviour for a large number of tokens. A companion paper by Campos, Chiola and Silva [16] extends these results to live and bounded nets.

Bruell and Ghanta [13], give an informal account on how to compute the upper and lower throughput bounds for GSPNs, which are bounded, live, conservative and have marking independent firing intensities. The proposed algorithm simplifies the GSPN, by removing fork-join structures such as those described in Sections 4.8 and 4.9. They find the upper bound on throughput of the simplified GSPN by using the following technique. The aim is to reduce the SPN to a simple structure, in which tokens are delayed because of the time taken for a transition to fire, rather than in the contention for the use of shared resources. The paths within the fork-join section that contain contention, are replaced by a transition whose rate is given by the maximum time taken to traverse the fork-join section. This maximum is taken over the paths without contention and the path being considered, with the contention removed. This procedure is continued until each path in the fork-join section does not contain contention. The entire fork-join section is then replaced by the slowest path that a token can follow, from the moment it forks to the moment it joins. The throughput is then evaluated on the simplified SPN in the usual way. To evaluate the lower bound, in a shared resource section, they evaluate the shortest time taken to traverse the section, by using the upper bound, which effects how often the tokens are available.

This chapter is constructed as follows: In Section 5.2 we give a survey of other work concerned with aggregation and disaggregation at the net level and also introduce our technique. In Section 5.3 we give the definitions and basic results required for the following sections. Section 5.4 illustrates the aggregation technique using an example that shows how insensitive skeleton nets provide exact marginal distributions for the original net. Two different approaches for the exact disaggregation of the skeleton net, are given in Section 5.5.

5.2 An Introduction to Aggregation and Disaggregation at the Net Level

As briefly mentioned at the start of this chapter, we are interested in extending the class of SPNs which can be analysed exactly, using aggregation and disaggregation of subnets in the SPN. Our approach, is to aggregate subnets to create "skeleton nets", which yield exact marginal distributions for the original SPN. The key to this method, is to use the results of insensitivity theory to create these "skeleton nets". Moreover, we show that whenever our approach is applicable, the strength of insensitivity theory can be used to rebuild the exact equilibrium distribution of the original net. A feature of our aggregation technique, is that we do not propose a step by step approach, but instead define an aim, which is to create a skeleton net which is insensitive to it's general distributions. It is worth noting, that for each insensitive skeleton net there are an uncountable number of original nets for which the skeleton net provides exact results. Consequently, a library of insensitive nets, or an understanding of the properties of insensitive nets, is essential for effectively utilising this approach. To fill in the entire picture, we proceed by summarising other work concerned with aggregation and disaggregation in SPNs.

Balbo, Bruell and Ghanta [6], aggregate subnets of GSPNs with their main concern being computational efficiency. Consequently, they choose subnets which have the BCMP type product form or which are GSPN structures, that can then be handled using available software. They follow the equivalent server idea of Chandy, Herzog and Woo [17], to piece together an approximation for the equilibrium distribution of the SPN.

The papers by Ammar and Islam ([2], [3]), Ammar, Huang and Liu [1], and Ammar and Liu [4], present an aggregation technique at the net level, which is used to compute bounds for SPNs. They aggregate subnets, which contain transitions with significantly faster firing rates, than those transitions which join these subnets together to form the SPN. They then replace these subnets by representative places, connected to the slow firing transitions. The number of tokens in each representative place, is the total number in the original subnet. The adjusted firing rates of the transitions with slow firing rates, are marking dependent, and are determined using their original rates and the token distribution in the connected subnet (for details see Ammar and Islam [3]). An analysis of these subnets in isolation, yields a choice of throughput rates, which are then used to give upper and lower bounds for the marking probabilities of the skeleton net.

Dugan, Geist, Nicola and Trevedi [28] and Section 4.6 of this thesis use aggregation to merge sets of transitions which are in conflict. Another form of aggregation which achieves an improvement in graphical complexity is used in the construction of high level SPNs (see Lin and Marinescu [59]). Briefly, a hierarchical representation is created which makes it possible to find a simpler representation of a SPN. A high level SPN can be considered a structurally folded version of a regular SPN, if the number of colours is finite. That is, it can be unfolded into a regular SPN by unfolding each place p into a set of places, one for each colour of token that place p may hold, and by unfolding each transition t into a set of transitions, one for each way that t may fire.

Henderson and Taylor [41], introduce flexibility into the model of Henderson, Lucic and Taylor [37], by absorbing the time spent in certain subnets into specially chosen intervals. They find the equilibrium distribution of the process, embedded at the time points at the ends of these intervals. Properties of the continuous time equilibrium distribution, can then be derived from the equilibrium distribution of the embedded process. By choosing appropriate embedding points, they show that it is possible to find marginal distributions for some SPNs, which do not fit the structure required by Henderson, Lucic and Taylor [37]. We will follow along a similar line, but do not restrict ourselves to finding an extended product form structure. There may be some value in aggregating to find an extended product form structure but it is outside the consideration of this thesis. Further, we will show that under certain conditions, we can disaggregate the subnet to find the exact equilibrium distribution for the original net.

5.3 Definitions and Basic Results

Using the notation introduced in Section 4.5, starting with the net P, the aim is to create, by aggregating subnets, an insensitive skeleton net Q for which an equilibrium distribution can be found. Recall from Section 4.5, that the process of aggregating states of the reachability graph does not necessarily achieve a new process. Likewise, aggregation at the net level, which defines a simplified net, may not naturally define a new process. For instance, there is no guarantee that, in aggregating places and transitions, the resultant routeing probabilities when transitions fire can be evaluated as functions of the new places and new transitions. For example, consider the SPN given in Figure 5.1. Imagine that we wish to aggregate t_1, t_2, p_2 and p_3 into a single transition t, followed by a single place p as in Figure 5.2. If t_1 and t_2 are negative exponentially distributed with means $(\mu_1)^{-1}$ and $(\mu_2)^{-1}$, the tokens leaving p go up with probability $\dfrac{\mu_1}{\mu_1 + \mu_2}$ and down with probability $\dfrac{\mu_2}{\mu_1 + \mu_2}$. Although in the negative exponential case these routeing probabilities are exact,this is not the case when general distributions are involved. For example, if t_1 has a deterministic firing time distribution, with mean $(\mu_1)^{-1}$, a token leaving p moves upward with probability one, if it's firing time through t is $(\mu_1)^{-1}$, and zero, if it is less than $(\mu_1)^{-1}$. With more complex distributions for the firing times of t_1 and t_2, the routeing can be explicitly defined in terms of the hazard functions of the two distributions and now becomes crucially dependent on the time at which t fires. The situation becomes even more complicated when a second transition fires tokens into, for example, p_3. The routeing out of p may then become a function of, not only how long t took to fire, but also the method by which tokens enter p.

For the purpose of explaining the aggregation technique as simply as possible,

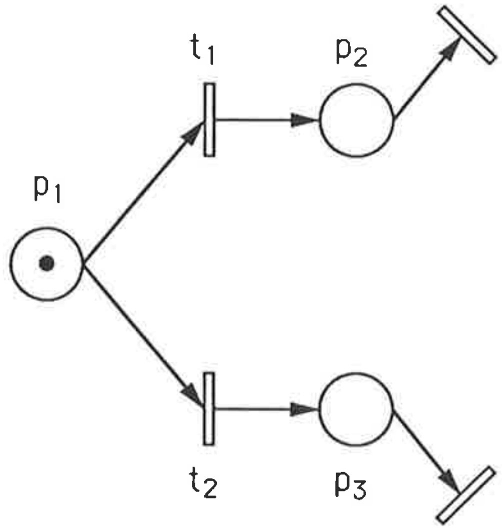

Figure 5.1: A simple SPN to illustrate aggregation.

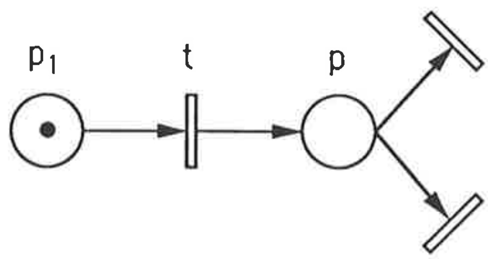

Figure 5.2: The aggregated SPN of Figure 5.1.

we shall assume that the aggregation of transitions and places, yields a net in which the future routeing through the net is only a function of the time spent in the aggregated subnet. An aggregated net has this property when, for example,

1. the aggregated subnets have only one entry point, or,

2. the routeing probabilities within the subnet are independent of the entry point. In this case, more than one entry point is allowed.

With this assumption, we define the aggregation procedure, incorporating the age dependent routeing results of Rumsewicz and Henderson [87] given in Section 4.5, as follows:

1. From the original net P, amalgamate sets of transitions and places and replace them with a box b and a transition t, to create a net $\overline{P}$. Let $G_b(\cdot)$ be the firing time distribution for t, which represents the time spent in the aggregated subnet b of the net $\overline{P}$. Assume that, after tokens have taken time y to be processed by t, the marking changes from $\overline{\mathbf{n}}$ to $\overline{\mathbf{m}}$, with age dependent probability $p_{\overline{\mathbf{n}},\overline{\mathbf{m}}}(t,y)$.

2. Create the averaged net Q from $\overline{P}$ so that, due to tokens being fired by t out of b, the marking changes from $\overline{\mathbf{n}}$ to $\overline{\mathbf{m}}$, with age independent probability $p_{\overline{\mathbf{n}},\overline{\mathbf{m}}}(t) = \int_0^\infty p_{\overline{\mathbf{n}},\overline{\mathbf{m}}}(t,y)\, dG_b(y)$.

3. Without altering the collection of routeing probabilities $p_{\overline{\mathbf{n}},\overline{\mathbf{m}}}(t)$, nor the means of the distributions $G_b(\cdot)$, create the net M, by replacing all distributions $G_b(\cdot)$ with negative exponential distributions.

As explained in the introduction, the aim is to create a net, Q, which is insensitive to its generally distributed firing times. Since the above procedure involves amalgamating *arbitrary* combinations of places and transitions, there is a great deal of flexibility and therefore many alternatives to choose from in creating the skeleton net. An aggregation which results in an insensitive net, Q, and which also has an extended product form solution is the ideal aim. Nevertheless, the advantages in succeeding to find only an insensitive Q are significant, as we will proceed to show.

Since we have stipulated that $\overline{P}$ is to be an exact amalgamated net of P, the age dependent routeing probabilities in $\overline{P}$, are sufficient to define P. Consequently, from Theorem 4.2 the equilibrium distribution of $\overline{P}$ is an exact marginal distribution of P. Unfortunately, $\overline{P}$ has general distributions as well as age dependent routeing, therefore, finding the equilibrium distribution of $\overline{P}$ is just as, if not more difficult, than finding the equilibrium distribution of P. Thus in solving one problem we have created another.

Let the state space for M be $\overline{\mathcal{M}}$. Let $\pi(\overline{\mathbf{m}})$, $\overline{\mathbf{m}} \in \overline{\mathcal{M}}$ be the equilibrium distribution for M. Note that an equilibrium distribution for M can be evaluated using simulation, direct evaluation, approximation or, occasionally, by creating M so that it has an extended product form solution. For each marking $\overline{\mathbf{m}}$ of the net, Q, let $\mathbf{y}$ be a vector of either the spent, or residual, lifetimes of each of the generally distributed transitions, enabled in $\overline{\mathbf{m}}$. For our purposes, it does not matter which of these lifetimes is used. Recall from Section 4.1, that Henderson [35] found that when Q is insensitive, the equilibrium distribution is the same when the state space is supplemented with either spent or residual lifetimes. Let $(\overline{\mathbf{m}}, \mathbf{y})$ be the marking of the net supplemented by these lifetimes. From Equation 4.27 if Q is insensitive, the equilibrium distribution of being in state $(\overline{\mathbf{m}}, \mathbf{y})$ is,

$$\pi(\overline{\mathbf{m}}, \mathbf{y}) = \pi(\overline{\mathbf{m}}) \prod_i \mu_i (1 - G_i(y_i)). \tag{5.1}$$

where $G_i(\cdot)$ is the distribution associated with the lifetime y_i and $(\mu_i)^{-1}$ is the mean of this distribution.

In Section 5.5, we show that Equation 5.1 is the pivotal idea in reversing the aggregation procedure to give exact results for the original SPN.

5.4 Aggregation

Aggregation procedures must always be questionable, unless they are supported by some measure of their error. We stress that our approach is concerned with aggregated nets, in which there is no error. The procedure provides exact equilibrium distributions, and exact marginal equilibrium distributions for the original net. That is, we achieve aggregation procedures in which no information or accuracy is lost.

As pointed out in Section 5.2, we do not propose a step by step approach to aggregation, and consequently we feel the best way to demonstrate the aggregation procedure is through an example. In the following example, we illustrate the aim in aggregating subnets of the SPN, that is to achieve exact marginal distributions for the original net. We use the simplified version of the dual processor, given in Figure 4.3, but introduce a more complex memory access stage. In reality, the access of common memory can involve more complex mechanisms, including accessing different areas of the memory and internal feedback. The SPN needed to model such a procedure, would normally be large and complex, but for simplicity, let us consider the net in Figure 5.3. Let transitions t_i, have generally distributed firing times according to the function $G_i(\cdot)$,with hazard functions $h_i(\cdot)$, and with corresponding means $(\mu_i)^{-1}$, for $1 \leq i \leq 10$. In this model, the processor is able to access two different areas of memory concurrently, represented by the places p_5 and p_6 for processor 1 and places p_9 and p_{10} for processor 2.

Even if we assume that transitions $t_3, t_4, t_5, t_6, t_7, t_8, t_9$ and t_{10} have negative exponentially distributed firing times, the equilibrium distribution of the SPN is not of a simple form. Now let us amalgamate places p_4, p_5, p_6 and p_7, and transitions t_3, t_4, t_5 and t_6, and represent the amalgamation by a box, b_1 and a transition, T_1. A symmetric procedure can be performed to create box, b_2 and transition, T_2. These amalgamations give the SPN, $\overline{P}$, of Figure 5.4. Step

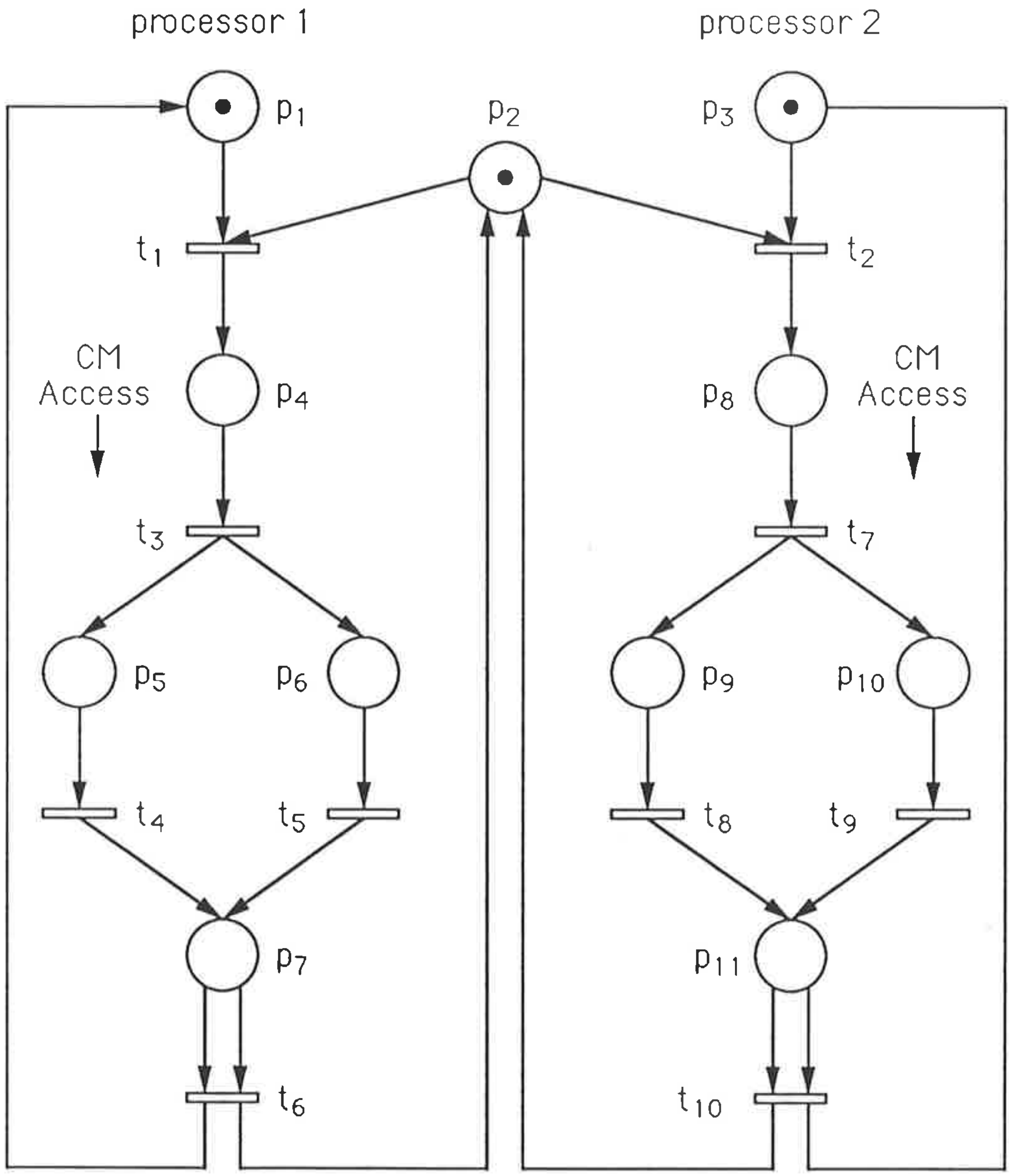

Figure 5.3: The dual processor of Figure 4.3 with complex memory access stage.

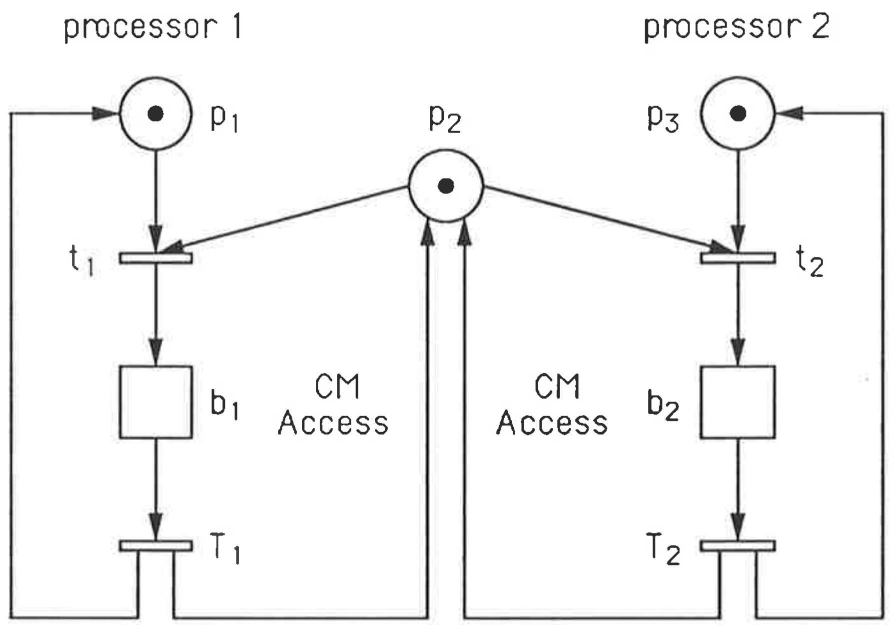

Figure 5.4: The aggregated SPN of Figure 5.3.

one of the aggregation procedure is now complete. In this case, there is no age dependent routeing, so step two of the procedure is redundant and hence $\overline{P}$ and Q are equivalent.

Because the firing times of transitions T_1 and T_2 cannot be assumed to be negative exponentially distributed, a simple exact solution can only be found if we prove insensitivity with respect to those firing times. Let $G_{b_i}(\cdot)$ $i = 1, 2$, be the firing time distributions of T_1 and T_2, with means $(\mu_{b_1})^{-1}$ and $(\mu_{b_2})^{-1}$ respectively and $h_{b_i}(\cdot)$, $i = 1, 2$ be the corresponding hazard functions. The states of the net, M, are,

$$\overline{\mathbf{m}}_1 = \{1, 1, 1, 0, 0\},$$

$$\overline{\mathbf{m}}_2 = \{0, 0, 1, 1, 0\},$$

$$\overline{\mathbf{m}}_3 = \{1, 0, 0, 0, 1\},$$

representing the number of tokens in places p_1, p_2, p_3, b_1 and b_2, respectively.

Since the aggregated SPN, $\overline{P}$, of Figure 5.4, has the same structure as the SPN of Figure 4.3, which has already been established to be insensitive to it's generally distributed lifetimes, we will not repeat the procedure given in Example 4.1 but rather quote the results of it. First, note that the means $(\mu_{b_1})^{-1}$ and $(\mu_{b_2})^{-1}$ of T_1 and T_2 in Figure 5.4 correspond to the means μ_3 and μ_4 of t_3 and t_4 in Figure 4.3, and the boxes b_1 and b_2 in Figure 5.4 correspond to places p_4 and p_5 in Figure 4.3. Therefore, from Example 4.1, the SPN is insensitive to all four transitions t_1, t_2, T_1 and T_2. Consequently, from Theorem 4.2, the marginal distribution of the P process, is the solution to the global balance equations of M, which are,

$$\pi(\overline{\mathbf{m}}_1)(\mu_1 + \mu_2) = \pi(\overline{\mathbf{m}}_2)\mu_{b_1} + \pi(\overline{\mathbf{m}}_3)\mu_{b_2}, \tag{5.2}$$

$$\pi(\overline{\mathbf{m}}_2)\mu_{b_1} = \pi(\overline{\mathbf{m}}_1)\mu_1, \tag{5.3}$$

$$\pi(\overline{\mathbf{m}}_3)\mu_{b_2} = \pi(\overline{\mathbf{m}}_1)\mu_2, \tag{5.4}$$

$$\sum_{i=1}^{3} \pi(\overline{\mathbf{m}}_i) = 1. \tag{5.5}$$

These equations have the solution,

$$\pi(\overline{\mathbf{m}}_1) = \frac{\mu_{b_1}\mu_{b_2}}{(\mu_{b_1}\mu_{b_2} + \mu_1\mu_{b_2} + \mu_2\mu_{b_1})}, \tag{5.6}$$

$$\pi(\overline{\mathbf{m}}_2) = \frac{\mu_1\mu_{b_2}}{(\mu_{b_1}\mu_{b_2} + \mu_1\mu_{b_2} + \mu_2\mu_{b_1})}, \tag{5.7}$$

$$\pi(\overline{\mathbf{m}}_3) = \frac{\mu_2\mu_{b_1}}{(\mu_{b_1}\mu_{b_2} + \mu_1\mu_{b_2} + \mu_2\mu_{b_1})}. \tag{5.8}$$

Take note, that since the SPN of Figure 5.4 is also insensitive to transitions t_1 and t_2, they too can represent the result of an aggregation procedure. However, transitions t_1 and t_2 must follow the pre-emptive resume protocol. If t_1 fires before t_2, the lifetime corresponding to t_2 will be worked off at zero speed, and vice versa. Therefore, if transitions t_1 and t_2 are to represent subnets, as do transitions T_1 and T_2, the pre-emptive protocol guarantees, that if the time taken to traverse one subnet is less than the other the transitions in the interrupted subnet must freeze for insensitivity to be applicable. At the instant the subnet is re-enabled, the transitions will resume with the same spent lifetime as before the interruption. Consequently, we are capable of further increasing the complexity of the model with regard to the transfer of the processor to main memory and still derive exact marginal equilibrium distributions.

Note that Wang and Robertazzi's [96] work, referred to in Section 4.4, is covered by this example, since we are able to use the transitions t_1 and t_2 as an aggregation of any sequence of places and transitions. This may include general distributions, feedbacks and timeout devices.

5.5 Disaggregation

In this section, we consider the task of retrieving information, apparently lost in the aggregation procedure. To describe the technique, we disaggregate the SPN of Figure 5.4. We stress again, that for each insensitive skeleton net, such as in Figure 5.4, there exists an uncountable number of original nets, for which the following approach will provide an exact result.

Since we are now dealing with generally distributed transition firing times, a finer description of the markings of the SPN is required. This is obtained by including supplementary variables, representing the spent or residual lifetimes, for each active transition. The supplemented equilibrium distribution of the SPN in Figure 5.4, is then given by Equation (5.1). Consider marking $\overline{\mathrm{m}}_1$, which enables both t_1 and t_2. For this marking, two supplementary variables y_1 and y_2, corresponding to $G_1(\cdot)$ and $G_2(\cdot)$, are required to describe the state of the SPN. If t_1 fires first, the lifetime corresponding to t_2 is worked off at zero speed, and the supplementary lifetime y_2 is set equal to the spent lifetime. When t_2 is re-enabled, it's lifetime will resume from the value y_2. The same argument holds for t_1 and it's supplementary variable y_1. The supplemented equilibrium distribution for marking $\overline{\mathrm{m}}_1$ is given by,

$$\pi(\overline{\mathrm{m}}_1, y_1, y_2) = \pi(\overline{\mathrm{m}}_1)\mu_1\mu_2(1 - G_1(y_1))(1 - G_2(y_2)), \tag{5.9}$$

where $(\mu_i)^{-1}$ is the mean of the distribution $G_i(\cdot)$, $i = 1, 2$, and $\pi(\overline{\mathrm{m}}_1)$ is given by Equation (5.6). Using a similar argument,

$$\pi(\overline{\mathrm{m}}_2, y_{b_1}) = \pi(\overline{\mathrm{m}}_2)\mu_{b_1}(1 - G_{b_1}(y_{b_1})), \tag{5.10}$$

$$\pi(\overline{\mathrm{m}}_3, y_{b_2}) = \pi(\overline{\mathrm{m}}_3)\mu_{b_2}(1 - G_{b_2}(y_{b_2})), \tag{5.11}$$

where, for $i = 1, 2$, y_{b_i} is the supplementary variable corresponding to $G_{b_i}(\cdot)$, with mean $(\mu_{b_i})^{-1}$. $\pi(\overline{\mathrm{m}}_i)$, $i = 2, 3$ are given by Equations (5.7) and (5.8).

5.5.1 The Supplemented and Unsupplemented Equilibrium Distributions for the SPN P

In this section, we will find both the supplemented and the unsupplemented equilibrium distributions for P, by using the insensitivity of the skeleton net, and an intuitive argument.

Expand the left half of the SPN, given in Figure 5.4, into its original form. The result is the SPN of Figure 5.5. Using the insensitivity of the SPN, we are able to derive from Equation (5.7), and Equations (5.9), (5.10) and (5.11), the supplemented equilibrium distribution for the SPN of Figure 5.5.

The markings are given by,

$$\begin{aligned}
\mathbf{m}_1 &= (1,1,1,0,0,0,0,0),\\
\mathbf{m}_2 &= (0,0,1,1,0,0,0,0),\\
\mathbf{m}_3 &= (1,0,0,0,0,0,0,1),\\
\mathbf{m}_4 &= (0,0,1,0,1,1,0,0),\\
\mathbf{m}_5 &= (0,0,1,0,1,0,1,0),\\
\mathbf{m}_6 &= (0,0,1,0,0,1,1,0),\\
\mathbf{m}_7 &= (0,0,1,0,0,0,2,0),
\end{aligned}$$

giving the number of tokens in places p_1 to p_7 and b_2 respectively.

We can use the information, given by the supplementary variable in the aggregated subnet, to find the equilibrium probabilities for the markings in the subnet. They are given below, and the proof is given in Appendix B.

$$\pi(\mathbf{m}_2, y_{b_1}) = \pi(\overline{\mathbf{m}}_2, 0)(1 - G_3(y_{b_1})). \tag{5.12}$$

$$\pi(\mathbf{m}_4, y_{b_1}) = \pi(\overline{\mathbf{m}}_2, 0)\int_{u=0}^{y_{b_1}}(1 - G_4(y_{b_1} - u))(1 - G_5(y_{b_1} - u))dG_3(u). \tag{5.13}$$

$$\pi(\mathbf{m}_5, y_{b_1}) = \pi(\overline{\mathbf{m}}_2, 0)\int_{u=0}^{y_{b_1}}(1 - G_4(y_{b_1} - u))\int_{w=u}^{y_{b_1}} dG_5(w - u)dG_3(u). \tag{5.14}$$

$$\pi(\mathbf{m}_6, y_{b_1}) = \pi(\overline{\mathbf{m}}_2, 0)\int_{u=0}^{y_{b_1}}(1 - G_5(y_{b_1} - u))\int_{w=u}^{y_{b_1}} dG_4(w - u)dG_3(u). \tag{5.15}$$

$$\pi(\mathbf{m}_7, y_{b_1}) = \pi(\overline{\mathbf{m}}_2, 0)\times \tag{5.16}$$

$$\left[\int_{u=0}^{y_{b_1}} dG_3(u)\int_{v=u}^{y_{b_1}} dG_4(v - u)\int_{w=v}^{y_{b_1}} dG_5(w - u)(1 - G_6(y_{b_1} - w))\right.$$

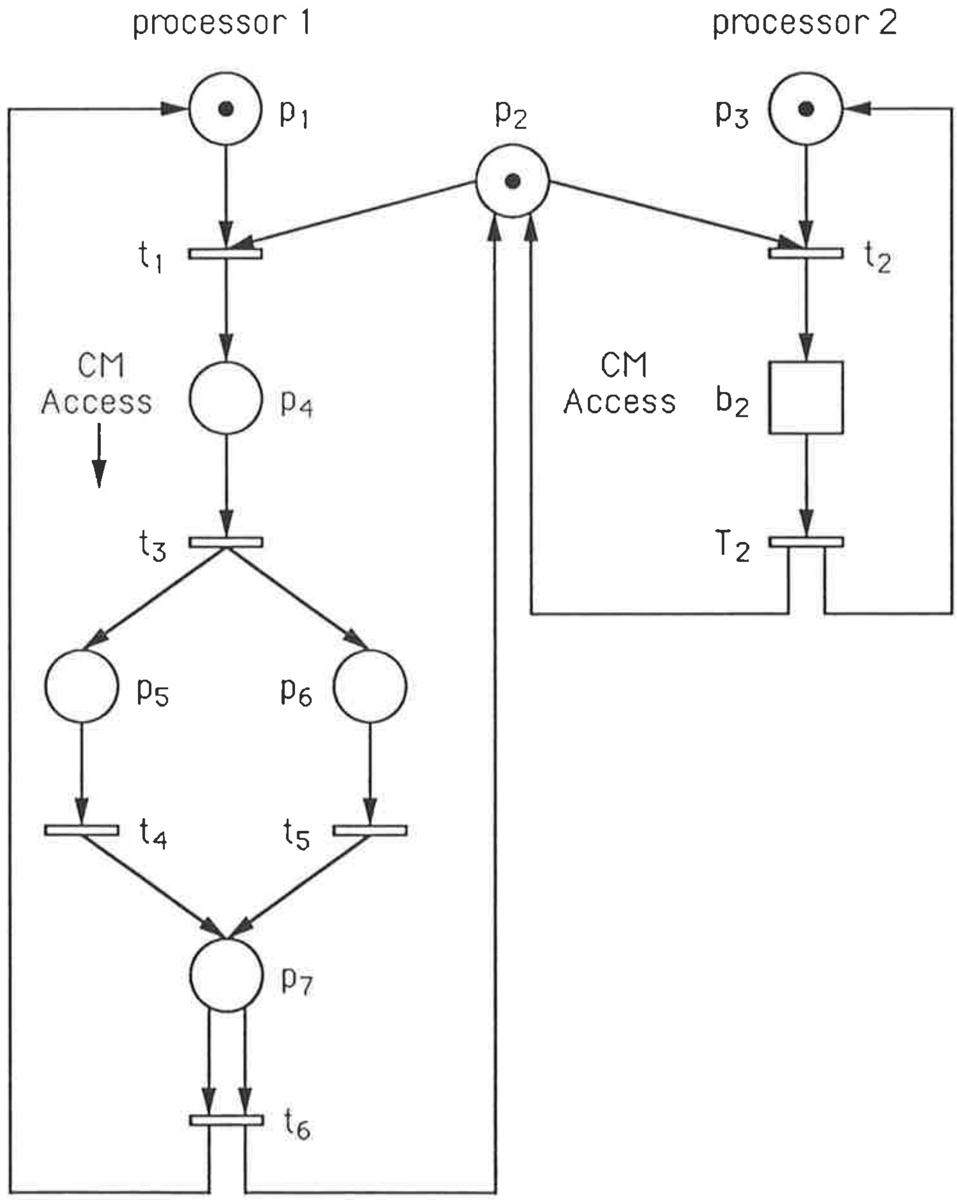

Figure 5.5: A partially expanded SPN of Figure 5.4.

$$+ \int_{u=0}^{y_{b_1}} dG_3(u) \int_{v=u}^{y_{b_1}} dG_5(v-u) \int_{w=v}^{y_{b_1}} dG_4(w-u)(1 - G_6(y_{b_1} - w))\Bigg].$$

From Equation (5.10), $\pi(\overline{\mathbf{m}}_2, 0) = \pi(\overline{\mathbf{m}}_2)\mu_{b_1}$, where $\pi(\overline{\mathbf{m}}_2)$ is the probability of being in a marking with a token in b_1 and is given by Equation (5.7) and $(\mu_{b_1})^{-1}$ is the mean time spent in b_1. The equilibrium distribution for $\pi(\overline{\mathbf{m}}_1, y_1, y_2)$ and $\pi(\overline{\mathbf{m}}_3, y_{b_2})$ remain unchanged from Equations (5.9) and (5.11).

Note that the form of $\pi(\mathbf{m}_i, y_{b_1})$, $2 \leq i \leq 7$ is intuitive. For example, consider $\pi(\mathbf{m}_2, y_{b_1})$. Marking $\mathbf{m}_2$ is defined by a token in places p_3 and p_4. The supplementary variable starts recording the time spent in b_1, as soon as a token is deposited into the subnet, that is, deposited into p_4. $\pi(\overline{\mathbf{m}}_2, 0)$ is the probability of being in a marking in which a token has just entered b_1. The term $(1 - G_3(y_{b_1}))$ is the probability that transition t_3 has not fired at time y_{b_1}. Using this information, $\pi(\mathbf{m}_2, y_{b_1})$, the probability of being in marking $\mathbf{m}_2$ at time y_{b_1}, is the probability that a token entered b_1 at $y_{b_1} = 0$ and t_3 hasn't fired at time y_{b_1}.

Now consider $\pi(\mathbf{m}_4, y_{b_1})$. $\int_{u=0}^{y_{b_1}}(1 - G_4(y_{b_1} - u))(1 - G_5(y_{b_1} - u))dG_3(u)$ can be interpreted as the probability that transition t_3 fired at time $u < y_{b_1}$, and that transitions t_4 and t_5 have not fired in the time $y_{b_1} - u$. The product of this integral and $\pi(\overline{\mathbf{m}}_2, 0)$, gives the probability of the tokens being in places p_3, p_5 and p_6, which is the probability of being in marking $\mathbf{m}_4$, at time y_{b_1}. The form for $\pi(\mathbf{m}_5, y_{b_1})$ and $\pi(\mathbf{m}_6, y_{b_1})$ can be argued in the same way.

The form of $\pi(\mathbf{m}_7, y_{b_1})$ is a little more complicated, so, we will give an explanation. The first integral in the expression can be interpreted as the probability that transition t_3 fired at time u, $u < v < w < y_{b_1}$, transition t_4 fired at time v, $v < w < y_{b_1}$ having been enabled for time $v - u$, transition t_5 fired at time w, $w < y_{b_1}$ having been enabled for time $w - u$ and transition t_6 has not fired in time $y_{b_1} - w$. The second integral can be interpreted as the probability that transition t_3 fired at time u, $u < v < w < y_{b_1}$, transition t_5 fired at time v, $v < w < y_{b_1}$ having been enabled for time $v - u$, transition t_4 fired at time w, $w < y_{b_1}$ having been enabled for time $w - u$ and transition t_6 has not fired in time $y_{b_1} - w$. Since there exist two sequences of transition firings that reach $\mathbf{m}_7$, and each integral is the probability of such a firing sequence, the addition of the two integrals, multiplied by $\pi(\mathbf{m}_2, 0)$, gives the probability of being in $\mathbf{m}_7$.

The intuitive reasoning applied to this example, can be used for any aggregated subnet, but, as can be seen with this example, the equations can become complicated. In Section 5.5.2, we will give a more general and much more useful procedure, again, making use of the insensitivity of the skeleton net.

Now we integrate out the supplementary variables from Equations (5.12)-(5.16) and Equations (5.9) and (5.11), to find the exact unsupplemented equilibrium distribution for P. The unsupplemented equilibrium equations are given by the following and the proof is given in Appendix C.

$$\pi(\mathbf{m}_2) = \pi(\overline{\mathbf{m}}_2)\frac{\mu_{b_1}}{\mu_3}, \tag{5.17}$$

$$\pi(\mathbf{m}_4) = \pi(\overline{\mathbf{m}}_2)\frac{\mu_{b_1}}{\mu_4 + \mu_5}, \tag{5.18}$$

$$\pi(\mathbf{m}_5) = \pi(\overline{\mathbf{m}}_2)\frac{\mu_{b_1}\mu_5}{\mu_4(\mu_4 + \mu_5)}, \tag{5.19}$$

$$\pi(\mathbf{m}_6) = \pi(\overline{\mathbf{m}}_2)\frac{\mu_{b_1}\mu_4}{\mu_5(\mu_4 + \mu_5)}, \tag{5.20}$$

$$\pi(\mathbf{m}_7) = \frac{\pi(\overline{\mathbf{m}}_2)\mu_{b_1}}{\mu_6}, \tag{5.21}$$

where $\pi(\overline{\mathbf{m}}_1)$, $\pi(\overline{\mathbf{m}}_2)$ and $\pi(\overline{\mathbf{m}}_3)$ are given by Equations (5.6), (5.7) and (5.8).

We note that the right hand side of the SPN of Figure 5.5 can be disaggregated in a similar manner.

5.5.2 The General Theory

Although the intuitive approach given in the previous section for disaggregation, works for the example given, the complexity increases when larger subnets are being aggregated. In its present form, with combinations of complex integrals, it is also of limited practical value. There is however an alternative approach, which naturally lends itself to a variety of simple and practical techniques and is given in Result 5.1.

Result 5.1

From Equation (5.1), we note that an insensitive net has a product form solution involving the marginal distribution $\pi(\overline{\mathrm{m}})$ and a product of terms $\mu_i(1 - G_i(y_i))$. Each y_i gives information on how long a token has been in the appropriate aggregated subnet, and hence the state of the subnet. The product form indicates independence between the marginal state and the states of the aggregated subnets. Consequently, if we wish to find the equilibrium distribution for the original SPN we need to find:

1. The corresponding marginal equilibrium distribution of the skeleton net.

2. The equilibrium distribution for the aggregated subnets considered in isolation.

To find the equilibrium distributions for the skeleton net and aggregated subnets we can use simulation, global balance equations, the extended product form solution or any other technique available. The big advantage is, that each of the resultant processes has a state space considerably smaller than the original, and therefore relatively trivial to handle.

Consider any marginal marking in which tokens are in b_1. In this example, $\overline{\mathrm{m}}_2$ is the only relevant marking, with one token somewhere inside b_1. Now isolate this subnet with one token in place p_4, as in Figure 5.6, and close the subnet by looping the token back to place p_4, when it fires out of t_6.

Let (i, j, k, l) be a state of the subnet, representing i, j, k and l tokens in places p_4, p_5, p_6 and p_7, respectively. An invariant measure for the subnet of Figure 5.6, is,

$$\pi(0,1,1,0) = \pi(1,0,0,0)\frac{\mu_3}{(\mu_4 + \mu_5)}, \tag{5.22}$$

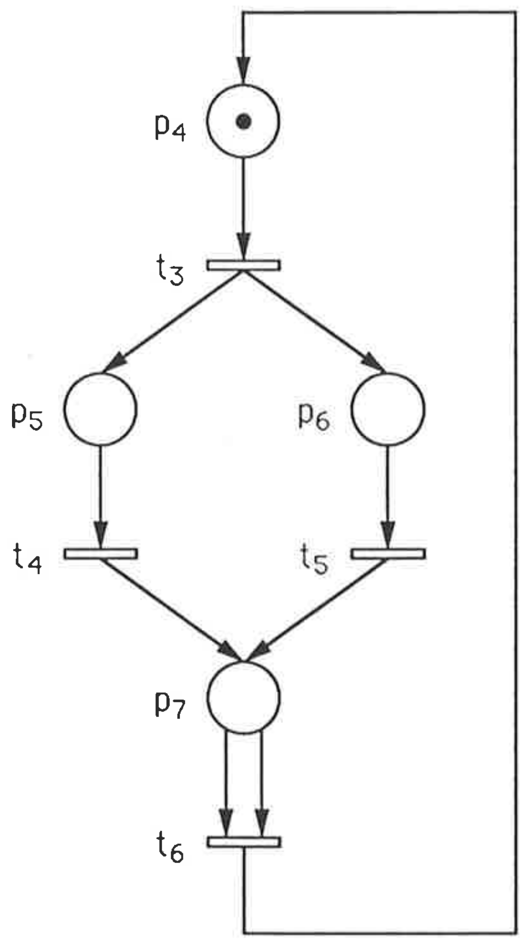

Figure 5.6: A subnet of Figure 5.5.

$$\pi(0,1,0,1) = \pi(1,0,0,0)\frac{\mu_3\mu_5}{\mu_4(\mu_4+\mu_5)}, \tag{5.23}$$

$$\pi(0,0,1,1) = \pi(1,0,0,0)\frac{\mu_3\mu_4}{\mu_5(\mu_4+\mu_5)}, \tag{5.24}$$

$$\pi(0,0,0,2) = \pi(1,0,0,0)\frac{\mu_3}{\mu_6}. \tag{5.25}$$

Using the independence established in Result 5.1, it is now possible to construct the equilibrium probabilities for the original SPN. This is achieved by multiplying each of the Equations (5.22) to (5.25) by $\pi(\overline{\mathbf{m}}_2)$ and reconstructing the token distributions for the original SPN. For example, (1,0,0,0) corresponds to marking $\mathbf{m}_2$ in the original net since the entries for p_4, p_5, p_6 and p_7 are equivalent. Let $\pi(1,0,0,0) = C$, then,

$$\pi(\mathbf{m}_2) = C\pi(\overline{\mathbf{m}}_2), \tag{5.26}$$

$$\pi(\mathbf{m}_4) = C\pi(\overline{\mathbf{m}}_2)\frac{\mu_3}{(\mu_4+\mu_5)}, \tag{5.27}$$

$$\pi(\mathbf{m}_5) = C\pi(\overline{\mathbf{m}}_2)\frac{\mu_3\mu_5}{\mu_4(\mu_4+\mu_5)}, \tag{5.28}$$

$$\pi(\mathbf{m}_6) = C\pi(\overline{\mathbf{m}}_2)\frac{\mu_3\mu_4}{\mu_5(\mu_4+\mu_5)}, \tag{5.29}$$

$$\pi(\mathbf{m}_7) = C\pi(\overline{\mathbf{m}}_2)\frac{\mu_3}{\mu_6}. \tag{5.30}$$

It is clear, that Equations (5.17) to (5.21), and Equations (5.26) to (5.30), are equivalent invariant measures, with $C = \frac{\mu_{b_1}}{\mu_3}$. μ_{b_1} which can be estimated using simulation, if necessary, but in this example it can be evaluated analytically, since the mean time to traverse the subnet is,

$$\frac{1}{\mu_{b_1}} = \frac{1}{\mu_3} + \frac{1}{\mu_6} + \frac{1}{\mu_4} + \frac{1}{\mu_5} - \frac{1}{\mu_4+\mu_5}.$$

This is an interesting result, which may be useful in evaluating the normalising constant for some SPNs. We will discuss this idea further in Chapter 6.

We are now able to link the results of Section 4.9, with the results of this section. Let an aggregated subnet be an n-dimensional fork-join section, as described in Section 4.9. Using the results of this Chapter we are able to disaggregate the subnet in the following way. Suppose, marking $\overline{\mathbf{m}}_i$ of the skeleton net, is the only marking with a token in the subnet. Let $\mathbf{m}_j$ be a marking of the original SPN, with token distribution in the subnet corresponding to the token distribution of state $\mathcal{B}$ (as defined in Sections 4.8 and 4.9). The Equilibrium distribution of the original SPN, is given by,

$$\pi(\mathbf{m}_j) = \pi(\mathcal{B})\pi(\overline{\mathbf{m}}_i),$$

where $\pi(\mathcal{B})$ is defined in Equation (4.87).

Chapter 6

Conclusions

In Chapter 3 of this thesis, we gave the theory on the extended product form solution for SPNs. Whilst finding the invariant measure for some nets is greatly simplified using this theory, the normalising constant must still be evaluated by generating the reachability graph. Note that, there are exceptions where the normalising constant can be found in a closed form, as illustrated in Example 3 in Henderson and Taylor [41]. Also, the normalising constant for the SPN of Figures 4.11 and the subnet of Figure 5.6 were found in closed form. It may be possible to use these examples as a basis for finding an approach to evaluate the normalising constant in closed form. Future research in this area should include finding a technique, similar to Buzen's algorithm in queueing networks, which evaluates the normalising constant without having to generate the reachability graph.

In Chapters 4 and 5 we illustrated that insensitivity is a strong property of SPNs. It allows general distributions to be replaced by negative exponential distributions, age dependent routeing to be replaced by age independent routeing, aggregations to yield exact marginal distributions and disaggregations to yield the exact equilibrium distribution for the original SPN. We suspect, but have not proved, that in general, insensitivity is the only property that will allow the above procedures to be performed, while yielding an exact result. Also, we would have liked to present a complete theory involving the insensitivity balance equations, reversed time structures and the necessary and sufficient conditions but, if this is at all possible, it requires much more work, which is beyond the scope of this thesis.

Further work that should prove fruitful, is to first, simplify the conversion from age dependent routeing to age independent routeing. Secondly, since the modeller applying the theory of insensitivity must rely on their experience of structures possessing this property. It would beneficial to provide an approach which easily identifies these structures, such as the symmetric queue in queueing theory.

In this thesis we have concentrated on finding the exact equilibrium distribution for SPNs. In queueing theory, networks which have a product form solution or which are insensitive (or both) have been used to find bounds for queueing networks which do not possess these useful properties. We suggest, that a similar tactic ought to be applicable for SPNs. Using a suitable choice of SPN which displays an exact solution, it may be possible to give performance bounds for nets which are otherwise difficult to solve.

Appendix A

In this appendix, we will prove that the supplemented differential equations, given by Equations (A.1) to (A.10) have solution given by Equations (A.11) to (A.18).

$$\frac{d\pi(\mathbf{m}_2, y)}{dy} = -\pi(\mathbf{m}_2, y)\left[h_2(y) + h_3(y) + h_4(y)\right]. \tag{A.1}$$

$$\pi(\mathbf{m}_2, 0) = \pi(\mathbf{m}_1)\mu_1. \tag{A.2}$$

$$\frac{d\pi(\mathbf{m}_3, y)}{dy} = \pi(\mathbf{m}_2, y)h_4(y) - \pi(\mathbf{m}_3, y)\left[h_2(y) + h_3(y)\right]. \tag{A.3}$$

$$\frac{d\pi(\mathbf{m}_4, y)}{dy} = \pi(\mathbf{m}_3, y)h_3(y) + \pi(\mathbf{m}_5, y)h_4(y) - \pi(\mathbf{m}_4, y)h_2(y). \tag{A.4}$$

$$\frac{d\pi(\mathbf{m}_5, y)}{dy} = \pi(\mathbf{m}_2, y)h_3(y) - \pi(\mathbf{m}_5, y)\left[h_2(y) + h_4(y)\right]. \tag{A.5}$$

$$\frac{d\pi(\mathbf{m}_6, y)}{dy} = \pi(\mathbf{m}_2, y)h_2(y) - \pi(\mathbf{m}_6, y)\left[h_3(y) + h_4(y)\right]. \tag{A.6}$$

$$\frac{d\pi(\mathbf{m}_7, y)}{dy} = \pi(\mathbf{m}_3, y)h_2(y) + \pi(\mathbf{m}_6, y)h_4(y) - \pi(\mathbf{m}_7, y)h_3(y). \tag{A.7}$$

$$\frac{d\pi(\mathbf{m}_8, y)}{dy} = \pi(\mathbf{m}_5, y)h_2(y) + \pi(\mathbf{m}_6, y)h_3(y) - \pi(\mathbf{m}_8, y)h_4(y). \tag{A.8}$$

$$\begin{aligned}\pi(\mathbf{m}_9)\mu_5 = &\int_{y=0}^{\infty} \pi(\mathbf{m}_4, y)h_2(y)dy \\ &+ \int_{y=0}^{\infty} \pi(\mathbf{m}_7, y)h_3(y)dy + \int_{y=0}^{\infty} \pi(\mathbf{m}_8, y)h_4(y)dy.\end{aligned} \tag{A.9}$$

Also

$$\pi(\mathbf{m}_i, 0) = 0, \text{ for } 3 \le i \le 8. \tag{A.10}$$

The solution to these supplemented differential equations is given by,

$$\pi(\mathbf{m}_2) = \pi(\mathbf{m}_1)\mu_1 \int_{y=0}^{\infty} (1 - G_2(y))(1 - G_3(y))(1 - G_4(y))dy, \qquad \text{(A.11)}$$

$$\pi(\mathbf{m}_3) = \pi(\mathbf{m}_1)\mu_1 \int_{y=0}^{\infty} (1 - G_2(y))(1 - G_3(y))G_4(y)dy, \qquad \text{(A.12)}$$

$$\pi(\mathbf{m}_4) = \pi(\mathbf{m}_1)\mu_1 \int_{y=0}^{\infty} (1 - G_2(y))G_3(y)G_4(y)dy, \qquad \text{(A.13)}$$

$$\pi(\mathbf{m}_5) = \pi(\mathbf{m}_1)\mu_1 \int_{y=0}^{\infty} (1 - G_2(y))(1 - G_4(y))G_3(y)dy, \qquad \text{(A.14)}$$

$$\pi(\mathbf{m}_6) = \pi(\mathbf{m}_1)\mu_1 \int_{y=0}^{\infty} (1 - G_3(y))(1 - G_4(y))G_2(y)dy, \qquad \text{(A.15)}$$

$$\pi(\mathbf{m}_7) = \pi(\mathbf{m}_1)\mu_1 \int_{y=0}^{\infty} (1 - G_3(y))G_2(y)G_4(y)dy, \qquad \text{(A.16)}$$

$$\pi(\mathbf{m}_8) = \pi(\mathbf{m}_1)\mu_1 \int_{y=0}^{\infty} (1 - G_4(y))G_2(y)G_3(y)dy, \qquad \text{(A.17)}$$

$$\pi(\mathbf{m}_9) = \pi(\mathbf{m}_1)\frac{\mu_1}{\mu_5}, \qquad \text{(A.18)}$$

Consider Equation (A.1). The integrating factor is the term,

$$\exp\left\{\int_{v=0}^{y} h_2(v) + h_3(v) + h_4(v)dv\right\}.$$

Now multiply both sides of Equation (A.1) by this integrating factor giving,

$$\frac{d\pi(\mathbf{m}_2, y)}{dy} \exp\left\{\int_{v=0}^{y} h_2(v) + h_3(v) + h_4(v)dv\right\}$$

$$= -\pi(\mathbf{m}_2, y)\left[h_2(y) + h_3(y) + h_4(y)\right]\exp\left\{\int_{v=0}^{y} h_2(v) + h_3(v) + h_4(v)dv\right\}.$$

This is equivalent to,

$$\frac{d}{dy}\left[\pi(\mathbf{m}_2, y)\exp\left\{\int_{v=0}^{y} h_2(v) + h_3(v) + h_4(v)dv\right\}\right] = 0,$$

or, after integration,

$$\pi(\mathbf{m}_2, y) = C_1 \exp\left\{-\int_{v=0}^{y} h_2(v) + h_3(v) + h_4(v)dv\right\},$$

where C_1 is the constant of integration. Using Equation (4.11) gives,

$$\exp\left\{-\int_{v=0}^{y} h_2(v)+h_3(v)+h_4(v)dv\right\}=(1-G_2(y))(1-G_3(y))(1-G_4(y)),$$

and hence the solution,

$$\pi(\mathbf{m}_2,y)=C_1(1-G_2(y))(1-G_3(y))(1-G_4(y)).$$

Using Equation (A.2), the relevant initial condition, we obtain,

$$\pi(\mathbf{m}_2,y)=\pi(\mathbf{m}_1)\mu_1(1-G_2(y))(1-G_3(y))(1-G_4(y)). \tag{A.19}$$

Integrating both sides with respect to y yields,

$$\pi(\mathbf{m}_2)=\pi(\mathbf{m}_1)\mu_1\int_{y=0}^{\infty}(1-G_2(y))(1-G_3(y))(1-G_4(y))dy,$$

which is Equation (A.11) as required. Now consider Equation (A.3), the integrating factor is the term,

$$\exp\left\{\int_{v=0}^{y} h_2(v)+h_3(v)dv\right\}.$$

Now multiply both sides of Equation (A.3) by this integrating factor giving,

$$\begin{aligned}\frac{d\pi(\mathbf{m}_3,y)}{dy}&\exp\left\{\int_{v=0}^{y} h_2(v)+h_3(v)dv\right\}\\ &=-\pi(\mathbf{m}_3,y)\left[h_2(y)+h_3(y)\right]\exp\left\{\int_{v=0}^{y} h_2(v)+h_3(v)dv\right\}\\ &\quad+\pi(\mathbf{m}_2,y)h_4(y)\exp\left\{\int_{v=0}^{y} h_2(v)+h_3(v)dv\right\}.\end{aligned}$$

This is equivalent to

$$\begin{aligned}\frac{d}{dy}&\left[\pi(\mathbf{m}_3,y)\exp\left\{\int_{v=0}^{y} h_2(v)+h_3(v)dv\right\}\right]\\ &=\pi(\mathbf{m}_2,y)h_4(y)\exp\left\{\int_{v=0}^{y} h_2(v)+h_3(v)dv\right\},\end{aligned}$$

or, after integration,

$$\begin{aligned}\pi(\mathbf{m}_3,y)&\exp\left\{\int_{v=0}^{y} h_2(v)+h_3(v)dv\right\}\\ &=\left[C_2+\int_{w=0}^{y}\pi(\mathbf{m}_2,w)h_4(w)\exp\left\{\int_{v=0}^{w} h_2(v)+h_3(v)dv\right\}dw\right],\end{aligned}$$

where C_2 is the constant of integration. Using equation (A.19) and (4.11) this reduces to,

$$\pi(\mathbf{m}_3, y) = (1 - G_2(y))(1 - G_3(y)) \times \left[C_2 + \right.$$

$$\left. \int_{w=0}^{y} \pi(\mathbf{m}_1)\mu_1(1-G_2(w))(1-G_3(w))(1-G_4(w))h_4(w)\frac{1}{(1-G_2(w))(1-G_3(w))}dw \right].$$

That is,

$$\pi(\mathbf{m}_3, y) = (1 - G_2(y))(1 - G_3(y)) \left[C_2 + \pi(\mathbf{m}_1)\mu_1 \int_{w=0}^{y} (1 - G_4(w))h_4(w)dw\right].$$

Using the initial condition, Equation (A.10), gives $C_2 = 0$. In addition, Equation (4.10) implies,

$$\int_{w=0}^{y} (1 - G_4(w))h_4(w)dw = G_4(y),$$

and hence,

$$\pi(\mathbf{m}_3, y) = \pi(\mathbf{m}_1)\mu_1(1 - G_2(y))(1 - G_3(y))G_4(y).$$

Integrating both sides with respect to y yields,

$$\pi(\mathbf{m}_3) = \pi(\mathbf{m}_1)\mu_1 \int_{y=0}^{\infty} (1 - G_2(y))(1 - G_3(y))G_4(y)dy,$$

which is Equation (A.12). We will not give the proof for Equations (A.13) to (A.18), since they follow in the same vein. ∎

Appendix B

In this appendix, we prove that the following equations are the supplemented equilibrium equations for the markings in b_1.

$$\pi(\mathbf{m}_2, y_{b_1}) = \pi(\overline{\mathbf{m}}_2, 0)(1 - G_3(y_{b_1})). \tag{B.1}$$

$$\pi(\mathbf{m}_4, y_{b_1}) = \pi(\overline{\mathbf{m}}_2, 0)\int_{u=0}^{y_{b_1}}(1 - G_4(y_{b_1} - u))(1 - G_5(y_{b_1} - u))dG_3(u). \tag{B.2}$$

$$\pi(\mathbf{m}_5, y_{b_1}) = \pi(\overline{\mathbf{m}}_2, 0)\int_{u=0}^{y_{b_1}}(1 - G_4(y_{b_1} - u))\int_{w=u}^{y_{b_1}} dG_5(w - u)dG_3(u). \tag{B.3}$$

$$\pi(\mathbf{m}_6, y_{b_1}) = \pi(\overline{\mathbf{m}}_2, 0)\int_{u=0}^{y_{b_1}}(1 - G_5(y_{b_1} - u))\int_{w=u}^{y_{b_1}} dG_4(w - u)dG_3(u). \tag{B.4}$$

$$\pi(\mathbf{m}_7, y_{b_1}) = \pi(\overline{\mathbf{m}}_2, 0)\times \tag{B.5}$$

$$\left[\int_{u=0}^{y_{b_1}} dG_3(u)\int_{v=u}^{y_{b_1}} dG_4(v - u)\int_{w=v}^{y_{b_1}} dG_5(w - u)(1 - G_6(y_{b_1} - w))\right.$$

$$\left. + \int_{u=0}^{y_{b_1}} dG_3(u)\int_{v=u}^{y_{b_1}} dG_5(v - u)\int_{w=v}^{y_{b_1}} dG_4(w - u)(1 - G_6(y_{b_1} - w))\right].$$

Consider box, b_1. Transition t_3 is enabled first, t_4 and t_5 are enabled concurrently, and t_6 is enabled last. As the three sets of transitions are enabled at different times we must use three supplementary variables to find the supplemented equilibrium distribution. The supplemented global balance equations for the markings in b_1 can be written as,

$$\pi(\mathbf{m}_2, u + \Delta u) = \pi(\mathbf{m}_2, u)\left[1 - h_3(u)\Delta u\right],$$

$$\pi(\mathbf{m}_4, y + \Delta y, u) = \pi(\mathbf{m}_4, y, u)\left[1 - h_4(y-u)\Delta y - h_5(y-u)\Delta y\right],$$

$$\pi(\mathbf{m}_5, y + \Delta y, u) = \pi(\mathbf{m}_5, y, u)\left[1 - h_4(y-u)\Delta y\right] + \pi(\mathbf{m}_4, y, u)h_5(y-u)\Delta y,$$

$$\pi(\mathbf{m}_6, y + \Delta y, u) = \pi(\mathbf{m}_6, y, u)\left[1 - h_5(y-u)\Delta y\right] + \pi(\mathbf{m}_4, y, u)h_4(y-u)\Delta y,$$

$$\pi(\mathbf{m}_7, z + \Delta z, y, u) = \pi(\mathbf{m}_7, z, y, u)\left[1 - h_6(z-y)\Delta z\right].$$

Rearranging, and taking the limit as $\Delta u, \Delta y, \Delta z \to 0$, we get a set of differential equations.

$$\frac{d}{du}\pi(\mathbf{m}_2, u) = -\pi(\mathbf{m}_2, u)h_3(u), \tag{B.6}$$

$$\frac{d}{dy}\pi(\mathbf{m}_4, y, u) = -\pi(\mathbf{m}_4, y, u)\left[h_4(y-u) + h_5(y-u)\right], \tag{B.7}$$

$$\frac{d}{dy}\pi(\mathbf{m}_5, y, u) = -\pi(\mathbf{m}_5, y, u)h_4(y-u) + \pi(\mathbf{m}_4, y, u)h_5(y-u), \tag{B.8}$$

$$\frac{d}{dy}\pi(\mathbf{m}_6, y, u) = -\pi(\mathbf{m}_6, y, u)h_5(y-u) + \pi(\mathbf{m}_4, y, u)h_4(y-u), \tag{B.9}$$

$$\frac{d}{dz}\pi(\mathbf{m}_7, z, y, u) = -\pi(\mathbf{m}_7, z, y, u)h_6(z-y). \tag{B.10}$$

Their boundary conditions are given by,

$$\pi(\mathbf{m}_2, 0) = \pi(\overline{\mathbf{m}}_2, 0), \tag{B.11}$$

$$\pi(\mathbf{m}_4, u, u) = \pi(\mathbf{m}_2, u)h_3(u), \tag{B.12}$$

$$\pi(\mathbf{m}_5, u, u) = 0, \tag{B.13}$$

$$\pi(\mathbf{m}_6, u, u) = 0, \tag{B.14}$$

$$\pi(\mathbf{m}_7, y, y, u) = \pi(\mathbf{m}_5, y, u)h_4(y-u) + \pi(\mathbf{m}_6, y, u)h_5(y-u). \tag{B.15}$$

Equation (B.6) is a first order differential equation and we can conclude that,

$$\pi(\mathbf{m}_2, u) = \pi(\mathbf{m}_2, 0) \exp\left\{-\int_{t=0}^{u} h_3(t)dt\right\}.$$

Using the boundary condition given by Equation (B.11) and Equation (4.11),

$$\pi(\mathbf{m}_2, u) = \pi(\overline{\mathbf{m}}_2, 0)\left[1 - G_3(u)\right].$$

This is Equation (B.1) with $u = y_{b_1}$

Equation (B.7) is also a first order differential equation and we can conclude that,

$$\pi(\mathbf{m}_4, y, u) = \pi(\mathbf{m}_4, u, u) \exp\left\{-\int_{t=0}^{y-u} h_4(t) + h_5(t)dt\right\},$$

$$= \pi(\mathbf{m}_4, u, u) \exp\left\{-\int_{t=0}^{y-u} h_4(t)dt\right\} \exp\left\{-\int_{t=0}^{y-u} h_5(t)dt\right\},$$

$$= \pi(\mathbf{m}_4, u, u)\left[1 - G_4(y-u)\right]\left[1 - G_5(y-u)\right].$$

Using the boundary condition given by Equation (B.12),

$$\pi(\mathbf{m}_4, y, u) = \pi(\mathbf{m}_2, u)h_3(u)\left[1 - G_4(y-u)\right]\left[1 - G_5(y-u)\right].$$

We can now integrate out the variable u, which ranges from 0 to y. Therefore,

$$\pi(\mathbf{m}_4, y) = \int_{u=0}^{y} \pi(\mathbf{m}_4, y, u)du,$$

$$= \int_{u=0}^{y} \pi(\mathbf{m}_2, u)h_3(u)\left[1 - G_4(y-u)\right]\left[1 - G_5(y-u)\right]du,$$

$$= \int_{u=0}^{y} \pi(\overline{\mathbf{m}}_2, 0)\left[1 - G_3(u)\right]h_3(u)\left[1 - G_4(y-u)\right]\left[1 - G_5(y-u)\right]du.$$

Using equation (4.10), yields,

$$\pi(\mathbf{m}_4, y) = \int_{u=0}^{y} \pi(\overline{\mathbf{m}}_2, 0)\left[1 - G_4(y-u)\right]\left[1 - G_5(y-u)\right]dG_3(u),$$

$$= \pi(\overline{\mathbf{m}}_2, 0)\int_{u=0}^{y}\left[1 - G_4(y-u)\right]\left[1 - G_5(y-u)\right]dG_3(u).$$

This is Equation (B.2) with $y = y_{b_1}$.

The third differential equation is slightly more complicated and requires the use of an integrating factor. A suitable choice is $\exp\left\{\int_{t=0}^{y-u} h_4(t)dt\right\}$. This gives,

$$\frac{d}{dy}\left[\pi(\mathbf{m}_5, y, u)\exp\left\{\int_{t=0}^{y-u} h_4(t)dt\right\}\right] = \pi(\mathbf{m}_4, y, u)h_5(y-u)\exp\left\{\int_0^{y-u} h_4(t)dt\right\}.$$

Hence,

$$\pi(\mathbf{m}_5, y, u)\exp\left\{\int_{t=0}^{y-u} h_4(t)dt\right\}$$

$$= C_1 + \int_{w=u}^{y} \pi(\mathbf{m}_4, w, u)h_5(w-u)\exp\left\{\int_{t=0}^{w-u} h_4(t)dt\right\}dw,$$

where C_1 is the constant of integration. The boundary condition, Equation (B.13), implies that $C_1 = 0$. So,

$$\pi(\mathbf{m}_5, y, u)\left[1 - G_4(y-u)\right]^{-1}$$

$$= \int_{w=u}^{y} \pi(\mathbf{m}_4, w, u)h_5(w-u)\left[1 - G_4(w-u)\right]^{-1}dw,$$

$$= \int_{w=u}^{y} \pi(\mathbf{m}_2, u)h_3(u)\left[1 - G_4(w-u)\right]\left[1 - G_5(w-u)\right]$$

$$\times\, h_5(w-u)\left[1 - G_4(w-u)\right]^{-1}dw,$$

$$= \pi(\overline{\mathbf{m}}_2, 0)\int_{w=u}^{y}\left[1 - G_3(u)\right]h_3(u)dG_5(w-u).$$

Then,

$$\pi(\mathbf{m}_5, y, u) = \pi(\overline{\mathbf{m}}_2, 0)\left[1 - G_3(u)\right]h_3(u)\left[1 - G_4(y-u)\right]\int_{w=u}^{y} dG_5(w-u). \quad (B.16)$$

Once again we can integrate out the dependence on u, so,

$$\pi(\mathbf{m}_5, y) = \int_{u=0}^{y} \pi(\mathbf{m}_5, y, u)du,$$

$$= \int_{u=0}^{y} \pi(\overline{\mathbf{m}}_2, 0)\left[1 - G_3(u)\right]h_3(u)\left[1 - G_4(y-u)\right]\int_{w=u}^{y} dG_5(w-u)du,$$

$$= \pi(\overline{\mathbf{m}}_2, 0)\int_{u=0}^{y}\left[1 - G_4(y-u)\right]\int_{w=u}^{y} dG_5(w-u)dG_3(u).$$

This is Equation (B.3) with $y = y_{b_1}$.

We will not give the details for the solution of Equation (B.9) as it is similar to that given above. However, for future use,

$$\pi(\mathbf{m}_6, y, u) = \pi(\overline{\mathbf{m}}_2, 0)\,[1 - G_3(u)]\,h_3(u)\,[1 - G_5(y-u)] \int_{w=u}^{y} dG_4(w-u). \quad \text{(B.17)}$$

Equation (B.10) is a first order differential equation. Therefore,

$$\begin{aligned}
\pi(\mathbf{m}_7, z, y, u) &= \pi(\mathbf{m}_7, y, y, u) \exp\left\{-\int_{t=0}^{z-y} h_6(t)dt\right\} \\
&= [\pi(\mathbf{m}_5, y, u)h_4(y-u) + \pi(\mathbf{m}_6, y, u)h_5(y-u)] \exp\left\{-\int_{t=0}^{z-y} h_6(t)dt\right\}, \\
&= [\pi(\mathbf{m}_5, y, u)h_4(y-u) + \pi(\mathbf{m}_6, y, u)h_5(y-u)]\,[1 - G_6(z-y)].
\end{aligned}$$

Substitution for $\pi(\mathbf{m}_5, y, u)$ and $\pi(\mathbf{m}_6, y, u)$ from Equations (B.16) and (B.17), gives,

$$\begin{aligned}
&\pi(\mathbf{m}_7, z, y, u) = \pi(\overline{\mathbf{m}}_2, 0)\,[1 - G_6(z-y)] \times \\
&\Bigg[[1 - G_3(u)]\,h_3(u)\,[1 - G_4(y-u)]\,h_4(y-u) \int_{v=u}^{y} dG_5(v-u) \\
&+ [1 - G_3(u)]\,h_3(u)\,[1 - G_5(y-u)]\,h_5(y-u) \int_{v=u}^{y} dG_4(v-u)\Bigg].
\end{aligned}$$

Integrating out the dependence on u gives,

$$\begin{aligned}
\pi(\mathbf{m}_7, z, y) &= \int_{u=0}^{z} \pi(\mathbf{m}_7, z, y, u)du, \\
&= \int_{u=0}^{z} \pi(\overline{\mathbf{m}}_2, 0)\,[1 - G_6(z-y)] \times \\
&\quad \Bigg[[1 - G_3(u)]\,h_3(u)\,[1 - G_4(y-u)]\,h_4(y-u) \int_{v=u}^{y} dG_5(v-u) \\
&\quad + [1 - G_3(u)]\,h_3(u)\,[1 - G_5(y-u)]\,h_5(y-u) \int_{v=u}^{y} dG_4(v-u)\Bigg] du,
\end{aligned}$$

$$= \pi(\overline{\mathbf{m}}_2, 0)\,[1 - G_6(z-y)] \times$$

$$\left[\int_{u=0}^{z} dG_3(u)\,[1 - G_4(y-u)]\,h_4(y-u)\int_{v=u}^{y} dG_5(v-u)\right.$$

$$\left. + \int_{u=0}^{z} dG_3(u)\,[1 - G_5(y-u)]\,h_5(y-u)\int_{v=u}^{y} dG_4(v-u)\right].$$

Now integrating out the dependence on y gives,

$$\pi(\mathbf{m}_7, z) = \int_{y=u}^{z} \pi(\mathbf{m}_7, z, y)dy,$$

$$= \pi(\overline{\mathbf{m}}_2, 0)\int_{u=0}^{z} dG_3(u) \times$$

$$\left[\int_{y=u}^{z} [1 - G_4(y-u)]\,h_4(y-u)\,[1 - G_6(z-y)]\int_{v=u}^{y} dG_5(v-u)dy\right.$$

$$\left. + \int_{y=u}^{z} [1 - G_5(y-u)]\,h_5(y-u)\,[1 - G_6(z-y)]\int_{v=u}^{y} dG_4(v-u)dy\right],$$

$$= \pi(\overline{\mathbf{m}}_2, 0) \times$$

$$\left[\int_{u=0}^{z} dG_3(u)\int_{y=u}^{z} dG_4(y-u)\,[1 - G_6(z-y)]\int_{v=u}^{y} dG_5(v-u)\right.$$

$$\left. + \int_{u=0}^{z} dG_3(u)\int_{y=u}^{z} dG_5(y-u)\,[1 - G_6(z-y)]\int_{v=u}^{y} dG_4(v-u)\right].$$

Changing the order of integration from $\int_{y=u}^{z}\int_{v=u}^{y}$ to $\int_{v=u}^{z}\int_{w=u}^{z}$ gives,

$$\pi(\mathbf{m}_7, z)$$

$$= \pi(\overline{\mathbf{m}}_2, 0)\int_{u=0}^{z} dG_3(u)\int_{v=u}^{z} dG_5(v-u)\int_{w=u}^{z} dG_4(w-u)\,[1 - G_6(z-w)]\,dy$$

$$+ \pi(\overline{\mathbf{m}}_2, 0)\int_{u=0}^{z} dG_3(u)\int_{v=u}^{z} dG_4(v-u)\int_{w=u}^{z} dG_5(w-u)\,[1 - G_6(z-w)]\,dy.$$

Which is Equation (B.5) with $z = y_{b_1}$. ∎

Appendix C

In this appendix, we will derive the following unsupplemented equations from the supplemented balance equations given by Equations (B.1)-(B.5) in Appendix B.

$$\pi(\mathbf{m}_2) = \pi(\overline{\mathbf{m}}_2)\frac{\mu_{b_1}}{\mu_3}, \tag{C.1}$$

$$\pi(\mathbf{m}_4) = \pi(\overline{\mathbf{m}}_2)\frac{\mu_{b_1}}{\mu_4 + \mu_5}, \tag{C.2}$$

$$\pi(\mathbf{m}_5) = \pi(\overline{\mathbf{m}}_2)\frac{\mu_{b_1}\mu_5}{\mu_4(\mu_4 + \mu_5)}, \tag{C.3}$$

$$\pi(\mathbf{m}_6) = \pi(\overline{\mathbf{m}}_2)\frac{\mu_{b_1}\mu_4}{\mu_5(\mu_4 + \mu_5)}, \tag{C.4}$$

$$\pi(\mathbf{m}_7) = \frac{\pi(\overline{\mathbf{m}}_2)\mu_{b_1}}{\mu_6}, \tag{C.5}$$

To remove the supplementary variable, we integrate Equations (B.1)-(B.5) over y_{b_1} from 0 to ∞. Equations (B.1) and (5.10) give,

$$\pi(\mathbf{m}_2) = \int_{y_{b_1}=0}^{\infty} \pi(\mathbf{m}_2, y_{b_1})dy_{b_1} = \pi(\overline{\mathbf{m}}_2)\mu_{b_1}\int_{y_{b_1}=0}^{\infty}(1 - G_3(y_{b_1}))dy_{b_1}.$$

However,

$$\frac{1}{\mu_i} = \int_{y_{b_1}=0}^{\infty}(1 - G_i(y_{b_1}))dy_{b_1}, \ \forall i. \tag{C.6}$$

Therefore,

$$\pi(\mathbf{m}_2) = \pi(\overline{\mathbf{m}}_2)\frac{\mu_{b_1}}{\mu_3}, \tag{C.7}$$

which is Equation (C.1), as required. Equation (B.2) gives,

$$\pi(\mathbf{m}_4) = \int_{y_{b_1}=0}^{\infty} \pi(\mathbf{m}_4, y_{b_1})dy_{b_1},$$

$$= \pi(\overline{\mathbf{m}}_2)\mu_{b_1} \int_{y_{b_1}=0}^{\infty} \int_{u=0}^{y_{b_1}} (1 - G_4(y_{b_1} - u))(1 - G_5(y_{b_1} - u)) dG_3(u) dy_{b_1}.$$

Swapping the order of integration, gives,

$$\pi(\mathbf{m}_4) = \pi(\overline{\mathbf{m}}_2)\mu_{b_1} \int_{u=0}^{\infty} \int_{y_{b_1}=u}^{\infty} (1 - G_4(y_{b_1} - u))(1 - G_5(y_{b_1} - u)) dG_3(u) dy_{b_1}$$

$$= \pi(\overline{\mathbf{m}}_2)\mu_{b_1} \int_{u=0}^{\infty}) \int_{y_{b_1}=u}^{\infty} (1 - G(y_{b_1} - u)) dy_{b_1} dG_3(u),$$

where $G(\cdot)$ is the distribution with hazard function $h_4(\cdot)+h_5(\cdot)$ and so, by equation (C.6), has mean $\frac{1}{\mu_4 + \mu_5}$. Hence,

$$\pi(\mathbf{m}_4) = \pi(\overline{\mathbf{m}}_2)\mu_{b_1} \int_{u=0}^{\infty} dG_3(u) \frac{1}{\mu_4 + \mu_5},$$

$$= \pi(\overline{\mathbf{m}}_2) \frac{\mu_{b_1}}{\mu_4 + \mu_5},$$

which is Equation (C.2), as required. Equation (B.3) gives,

$$\pi(\mathbf{m}_5) = \int_{y_{b_1}=0}^{\infty} \pi(\mathbf{m}_5, y_{b_1}) dy_{b_1},$$

$$= \pi(\overline{\mathbf{m}}_2)\mu_{b_1} \int_{y_{b_1}=0}^{\infty} \int_{u=0}^{y_{b_1}} (1 - G_4(y_{b_1} - u)) \int_{w=u}^{y_{b_1}} dG_5(w - u) dG_3(u) dy_{b_1},$$

$$= \pi(\overline{\mathbf{m}}_2)\mu_{b_1} \int_{y_{b_1}=0}^{\infty} \int_{u=0}^{y_{b_1}} (1 - G_4(y_{b_1} - u)) G_5(y - u) dG_3(u) dy_{b_1}.$$

Swapping the order of integration, gives,

$$\pi(\mathbf{m}_5)$$

$$= \pi(\overline{\mathbf{m}}_2)\mu_{b_1} \int_{u=0}^{\infty} \int_{y_{b_1}=u}^{\infty} (1 - G_4(y_{b_1} - u)) G_5(y - u) dG_3(u) dy_{b_1},$$

$$= \pi(\overline{\mathbf{m}}_2)\mu_{b_1} \int_{u=0}^{\infty} \int_{y_{b_1}=u}^{\infty} (1 - G_4(y_{b_1} - u)) \left[1 - (1 - G_5(y_{b_1} - u))\right] dG_3(u) dy_{b_1},$$

$$= \pi(\overline{\mathbf{m}}_2)\mu_{b_1} \int_{u=0}^{\infty} \int_{y_{b_1}=u}^{\infty} \Big[(1 - G_4(y_{b_1} - u))$$

$$- (1 - G_4(y_{b_1} - u))(1 - G_5(y_{b_1} - u))\Big] dG_3(u) dy_{b_1}.$$

Now, by Equation (C.6) and the comments in the derivation of Equation (C.2),

$$= \pi(\overline{\mathbf{m}}_2)\mu_{b_1} \int_{u=0}^{\infty} \left[\frac{1}{\mu_4} - \frac{1}{\mu_4 + \mu_5}\right] dG_3(u),$$

$$= \pi(\overline{\mathbf{m}}_2)\mu_{b_1} \left[\frac{1}{\mu_4} - \frac{1}{\mu_4 + \mu_5}\right],$$

$$= \pi(\overline{\mathbf{m}}_2)\mu_{b_1} \left[\frac{\mu_5}{\mu_4(\mu_4 + \mu_5)}\right].$$

This is Equation (C.3), as required, and a similar argument holds for obtaining Equation (C.4).

The form for $\pi(\mathbf{m}_7)$ is more difficult so we give the derivation below. Equation (B.5) gives,

$$\pi(\mathbf{m}_7) = \int_{y_{b_1}=0}^{\infty} \pi(\mathbf{m}_7, y_{b_1}) dy_{b_1},$$

$$= \pi(\overline{\mathbf{m}}_2)\mu_{b_1} \times$$

$$\left[\int_{y_{b_1}=0}^{\infty} \int_{u=0}^{y_{b_1}} \int_{v=u}^{y_{b_1}} \int_{w=v}^{y_{b_1}} dG_3(u) dG_4(v-u) dG_5(w-u) \left[1 - G_6(y_{b_1} - w)\right] dy_{b_1}\right.$$

$$\left. + \int_{y_{b_1}=0}^{\infty} \int_{u=0}^{y_{b_1}} \int_{v=u}^{y_{b_1}} \int_{w=v}^{y_{b_1}} dG_3(u) dG_5(v-u) dG_4(w-u) \left[1 - G_6(y_{b_1} - w)\right] dy_{b_1}\right].$$

Changing the order of integration, to enable us to remove the dependence on y_{b_1}, gives,

$$\pi(\mathbf{m}_7) = \pi(\overline{\mathbf{m}}_2)\mu_{b_1} \times$$

$$\left[\int_{u=0}^{\infty} \int_{v=u}^{\infty} \int_{w=v}^{\infty} \int_{y_{b_1}=w}^{\infty} \left[1 - G_6(y_{b_1} - w)\right] dy_{b_1} dG_5(w-u) dG_4(v-u) dG_3(u)\right.$$

$$\left. + \int_{u=0}^{\infty} \int_{v=u}^{\infty} \int_{w=v}^{\infty} \int_{y_{b_1}=w}^{\infty} \left[1 - G_6(y_{b_1} - w)\right] dy_{b_1} dG_4(w-u) dG_5(v-u) dG_3(u)\right].$$

Integrating out the dependence on y_{b_1} and using Equation (C.6) gives,

$$\pi(\mathbf{m}_7) = \pi(\overline{\mathbf{m}}_2)\mu_{b_1} \left[\frac{1}{\mu_6} \int_{u=0}^{\infty} \int_{v=u}^{\infty} \int_{w=v}^{\infty} dG_5(w-u) dG_4(v-u) dG_3(u)\right.$$

$$+\frac{1}{\mu_6}\int_{u=0}^{\infty}\int_{v=u}^{\infty}\int_{w=v}^{\infty} dG_4(w-u)dG_5(v-u)dG_3(u)\Bigg].$$

Now change the order of integration of the last two integrals, in the first expression only, to get,

$$\pi(\mathbf{m}_7) = \pi(\overline{\mathbf{m}}_2)\mu_{b_1}\Bigg[\frac{1}{\mu_6}\int_{u=0}^{\infty}\int_{w=u}^{\infty}\int_{v=u}^{w} dG_4(v-u)dG_5(w-u)dG_3(u)$$

$$+\frac{1}{\mu_6}\int_{u=0}^{\infty}\int_{v=u}^{\infty}\int_{w=v}^{\infty} dG_4(w-u)dG_5(v-u)dG_3(u)\Bigg].$$

Evaluating the two inner integrals,

$$\pi(\mathbf{m}_7) = \pi(\overline{\mathbf{m}}_2)\mu_{b_1}\Bigg[\frac{1}{\mu_6}\int_{u=0}^{\infty}\int_{w=u}^{\infty} G_4(w-u)dG_5(w-u)dG_3(u)$$

$$+\frac{1}{\mu_6}\int_{u=0}^{\infty}\int_{v=u}^{\infty} \left[1-G_4(v-u)\right] dG_5(v-u)dG_3(u)\Bigg].$$

Change the variable of integration, to get,

$$\pi(\mathbf{m}_7) = \pi(\overline{\mathbf{m}}_2)\mu_{b_1}\Bigg[\frac{1}{\mu_6}\int_{u=0}^{\infty}\int_{z=0}^{\infty} G_4(z)dG_5(z)dG_3(u)$$

$$+\frac{1}{\mu_6}\int_{u=0}^{\infty}\int_{z=0}^{\infty} \left[1-G_4(z)\right] dG_5(z)dG_3(u)\Bigg].$$

Combining the two integrals gives,

$$\pi(\mathbf{m}_7) = \pi(\overline{\mathbf{m}}_2)\mu_{b_1}\left[\frac{1}{\mu_6}\int_{u=0}^{\infty}\int_{z=0}^{\infty} dG_5(z)dG_3(u)\right].$$

Evaluate the two remaining integrals consecutively, to get

$$\pi(\mathbf{m}_7) = \frac{\pi(\overline{\mathbf{m}}_2)\mu_{b_1}}{\mu_6}, \tag{C.8}$$

which is Equation (C.5), as required. ∎

Bibliography

[1] Ammar H.H., Huang Y.F. and Liu R.W. (1987) - *Hierarchical Models for Systems Reliability, Maintainability and Availability*, IEEE Trans Circuits and Systems, **CAS-34**, no. 6, pp 629-638, 1987.

[2] Ammar H.H and Islam S.M.R. (1989) - *On Bounds for Token Probabilities in a Class of Generalised Stochastic Petri Nets*, Proc. 3rd Int. Conf. on Petri Nets and Performance Models, Kyoto, Japan, pp 221-227, Dec. 1989.

[3] Ammar H.H and Islam S.M.R (1989) - *Time Scale Decomposition of a Class of Generalised Stochastic Petri Nets*, IEEE Trans on Software Engineering, **15**, no. 6, pp 809-820, 1989.

[4] Ammar H.H and Liu R.W. (1985) - *Analysis of the Generalised Stochastic Petri Net by State Aggregation*, Proc. Int. Workshop on Timed Petri Nets, Torino, Italy, July 1985.

[5] Baccelli F., Makowski A.M. and Schwartz A. (1989) - *The Fork Join Queue and Related Systems with Synchronisation Constraints: Stochastic Ordering and Computable Bounds*, Adv. Appl. Prob., **21**, pp 621-660, 1989.

[6] Balbo G., Bruell S.C. and Ghanta S. (1988) - *Combining Queueing Networks and Generalized Stochastic Petri Nets for the Solution of Complex Models of System Behaviour*, IEEE Trans. on Computers, **37**, pp 1251-1268, 1988.

[7] Balsamo S. and Iazeolla G. (1984) - *Aggregation and Disaggregation in Queueing Networks: The Principle of Product-Form Synthesis*, Mathematical Computer Performance and Reliability, G. Iazeolla, P.J. Courtois and

A. Hordijk (editors), Elsevier Science Publishers, B.V. (North-Holland), pp 95-109, 1984.

[8] Baskett F., Chandy K., Muntz R., and Palacios J.(1975) - *Open, Closed and Mixed Networks of Queues with Different Classes of Customers*, J. Assoc. Comp. Mach., **22** , pp 248-260, 1975.

[9] Berthelot G. (1986) - *Checking Properties of Nets using Transformations*, Lecture Notes in Computer Science, New York: Springer-Verlag, **222**, pp 19-40, 1986.

[10] Berthelot G., Roucairol G. and Valk R. (1980) - *Reduction of Nets and Parallel Programs*, Lecture Notes in Computer Science, New York: Springer-Verlag, **84**, pp 277-290, 1980.

[11] Bertoni A. and Torelli M., (1981) - *Probabilistic Petri Nets and Semi-Markov Processes*, Proc. 2nd European Workshop on Petri Nets, Bad Honnef, West Germany, Sept. 1981.

[12] Bobbio A. and Cumani A. (1984) - *Discrete State Stochastic Systems with Phase-Type Distributed Transition Times*, Proc. AMSE Intl. Conf. on Modelling and Simulation, Athens, pp 173-192, 1984.

[13] Bruell S.C. and Ghanta S. (1985) - *Throughput Bounds for Generalised Stochastic Petri Nets*, Proc. Int. Workshop on Timed Petri Nets, Torino, Italy, July 1985.

[14] Buzen J.P. (1973)- *Computational Algorithms for Closed Queueing Networks With Exponential Servers*, Commun. ACM, **16**, pp 527-531, 1973.

[15] Campos J., Chiola G., Colom J.M. and Silva M. (1989) - *Tight Polynomial Bounds for Steady State Performance of Marked Graphs*, Proc. 3rd Int. Conf. on Petri Nets and Performance Models, Kyoto, Japan, pp 200-209, Dec. 1989.

[16] Campos J., Chiola G. and Silva M. (1989) - *Properties and Steady State Performance Bounds for Petri Nets with Unique Repetitive Firing Count Vector* Proc. 3rd Int. Conf. on Petri Nets and Performance Models, Kyoto, Japan, pp 210-220, Dec 1989.

[17] Chandy K.M., Herzog U. and Woo L. (1975) - *Parametric Analysis of Queueing Networks*, IBM J. Res. Develop., **19**, pp 36-42, 1975.

[18] Chandy K.M., Howard J. and Towsley D. (1977) - *Product Form and Local Balance in Queueing Networks*, J. Assoc. Comp. Mach., **24**, pp 250-263, 1977.

[19] Chandy K.M. and Martin J. (1983) - *A Characterization of Product Form Queueing Networks*, J.Assoc. Comp. Mach., **30**, pp 250-263, 1983.

[20] Chen P., Bruell S.C. and Balbo G. (1989) - *Alternative Methods for Incorporating Non-Exponential Distributions into Stochastic Timed Petri Nets*, Proc. 3rd Intl. Workshop on Petri Nets and Performance Models, Kyoto, Japan, pp 187-197, Dec. 1989.

[21] Cheng Y. and Robertazzi T.G. (1988) - *The Lattice Geometry of Petri Network Models of Multihop Radio Networks*, 22nd Annual Conference on Information Sciences and Systems, Princeton, N.J., March 1988.

[22] Chu T.A. (1987) - *A Method of Abstraction for Petri Nets*, Proc. Int. Workshop on Petri Nets and Performance Models, Madison, WI, pp 164-173, Aug. 1987.

[23] Cohen J. (1957) - *The Generalised Engset Formulae*, Philips Telecomm. Review, **18**, pp 158-170, 1957.

[24] Courtois P.J. (1977) - *Decomposability*, Academic Press, New York, 1977.

[25] Cox D.R. (1955) - *The Analysis of Non-Markovian Stochastic Processes by the Inclusion of Supplementary Variables*, Proc. Camb. Phil. Soc., **51**, pp 313-319, 1955.

[26] Crane M.A. and Iglehart D.L. (1975) - *Simulating Stable Stochastic Systems, III, Regenerative Processes and Discrete Event Simulation*, Operation Res, **23**, pp 33-45, 1975.

[27] Cumani A. (1985) - *ESP - A Package for the Evaluation of Stochastic Petri Nets with Phase-Type Distributed Transition Times*, Proc. Int. Workshop on Timed Petri Nets, Torino, Italy, pp 144-151, July 1985.

[28] Dugan J.B., Trivedi K.S., Geist R.M. and Nicola V.F. (1984) - *Extended Stochastic Petri Nets: Applications and Analysis*, Performance 84, Elsevier Sci. Publ. B.V. (N-Holland), pp 507-519, 1984.

[29] Erlang A. (1917) - *Solutions of some problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges*, Post Office Electrical Engineers Journal, **10**, pp 189-197, 1917.

[30] Fortet R. (1950) - *Calcul des probabilités*, Centre National de la Recherche Scientifique, Paris.

[31] Gordon W.J. and Newell G.F. (1967) - *Closed Queueing Systems with Exponential Servers*, Operations Res. **15**, pp 254-265, 1967.

[32] Haas P.J. and Shedlar G.S. (1986) - *Regenerative SPNs*, Perform. Eval. **6**, pp 189-204, 1986.

[33] Haas P.J. and Shedlar G.S, (1986) - *Stochastic Petri Nets with Simultaneous Transition Firings*, Proc. Int. Workshop on Petri Nets and Performance Models, Madison, WI: IEEE Computer Society Press, pp 24-34, Aug. 1987.

[34] Haas P.J. and Shedlar G.S. (1989) - *Stochastic Petri Net Representation of Discrete Event Simulations*, IEEE Trans on Software Engineering, **15**, no. 4, pp 381-393, 1989.

[35] Henderson W. (1983) - *Insensitivity and Reversed Markov Processes*, Adv. Appl. Prob, **15**, pp 752-768, 1983.

[36] Henderson W. and Lucic D. (1988)- *Applications of Generalised Semi Markov Processes to Stochastic Petri Nets*, Performance of Distributed and Parallel Systems, Ed. T. Hasegawa, H. Takagi, Y. Takahashi, pp. 315-328, 1988.

[37] Henderson W., Lucic D. and Taylor P.G. (1989) - *A Net Level Performance Analysis of Stochastic Petri Nets*, J. Aust. Math. Soc. Ser.B, **31**, pp 176-187, 1989.

[38] Henderson W., Pearce C.E.M., Taylor P.G. and van Dijk N.M. (1990) - *Closed Queueing Networks with Batch Services*, QUESTA, **6**, pp 57-70, 1990.

[39] Henderson W. and Taylor P.G. (1989) - *Insensitivity of Processes with interruptions*, J. Appl. Prob., **26**, pp 242-258, 1989.

[40] Henderson W. and Taylor P.G. (1989) - *Aggregation Methods in Exact Analysis of Stochastic Petri Nets*, The 3rd International Workshop on Petri Nets and Performance Models, Kyoto, Japan, pp 12-18, Dec. 1989.

[41] Henderson W. and Taylor P.G. (1990) - *Embedded Processes in Stochastic Petri Nets*, to be published in IEEE, Trans. on Software Engineering, 1991.

[42] Henderson W. and Taylor P.G. (1990) - *Product Form in Networks of Queues with Batch Arrivals and Batch Services*, QUESTA, **6**, pp 71-88, 1990

[43] Holliday M.A. and Vernon M.K. (1987) - *A Generalised Timed Petri Net Model for Performance Analysis*, IEEE Transactions on Software Engineering, **SE-13**, No. 12, pp 1297-1310, 1987.

[44] Holt A.W. and Commoner F. (1970) - *Events and Conditions*, Princeton, N.J., Applied Data Research Inc., Information System Theory Project. Also: Record Project MAC Conf. Concurrent Systems Parallel Computation, pp 3-52, 1970.

[45] Jackson J. (1957) - *Networks of Waiting Lines*, Operations Research, **5**, pp 518-521, 1957.

[46] Jackson J. (1963) - *Jobshop like Queueing Systems*, Management Science, **10**, pp 131-142, 1963.

[47] Jansen U. and König D. (1980) - *Insensitivity of Steady State Probabilities in Product Form for Queueing Networks*, Elektron, Informatsionsverarb. Kybernetik., **16**, pp 285-397, 1980.

[48] Johnsonbaugh R. and Murata T. (1981) - *Additional Methods for Reduction and Expansion of Marked Graphs*, IEEE Trans. Circuit Syst., **CAS-28**, no. 10, pp 1009-1014, Oct. 1981.

[49] Kelly F. (1975) - *Networks of Queues With Customers of Different Types*, J. Appl. Prob., **12**, pp 542-544, 1975.

[50] Kelly F. (1976) - *Networks of Queues*, Adv. Appl. Prob., **8**, pp 416-432, 1976.

[51] Kelly F. (1979) - *Reversibility and Stochastic Networks*, Wiley, London, 1979.

[52] König D. (1965) - *Verallgemeinerungen der Erlanschen Formeln*, Math. Nachr., **28**, pp 145-155, 1965.

[53] König D. and Jansen U., (1974) - *Stochastic Processes and Properties of Invariance for Queueing Systems with Speeds and Temporary Interruptions,* Trans 7th Prague Conference Inf. Th., Stat. Dec. Fns. and Rand. Proc., pp 335-343, 1974.

[54] Kosten L. (1942) - *On the Validity of the Erlang and Engset Loss Formulae,* Het P.T.T. Bedrijf, **2**, pp 42-45, 1942.

[55] Lazar A.A. and Robertazzi T.G. (1984) - *The Geometry of Lattices for Markovian Queueing Networks*, Columbia University and SU NY at Stony Brook, Departments of Electrical Engineering, 1984.

[56] Lazar A.A. and Robertazzi T.G. (1986) - *Markovian Petri Net Protocols with Product Form Solution*, Columbia University and SU NY at Stony Brook, Departments of Electrical Engineering, 1986.

[57] Lazar A.A. and Robertazzi T.G. (1986) - *The Algebraic and Geometric Structure of Markovian Petri Network Lattices*, 24th. Allerton Conference on Communication, Control and Computing, University of Illinois, 1986.

[58] Lazar A.A. and Robertazzi T.G. (1987) - *The Lattice Structure of a Bus Oriented Multiprocessor Model*, 25th. Allerton Conference on Communication, Control and Computing, Urbana-Champaigne, Ill, Oct. 1987.

[59] Lin C. and Marinescu D.C. (1988) - *Stochastic High-Level Petri Nets and Applications*, IEEE Trans. on Computers, **37**, pp 815-825, 1988.

[60] Marsan M.A. (1989) - *Stochastic Petri Nets: An elementary introduction*, Advances in Petri Nets 1989, Lecture Notes in Comp. Sci., Springer-Verlag, pp 1-29, 1989.

[61] Marsan M.A., Balbo G., Bobbio A., Chiola G., Conte G. and Cumani A. (1985) - *On Petri Nets with Stochastic Timing*, Proc. Int. Workshop on Timed Petri Nets, Torino, Italy, pp 80-87, July 1985.

[62] Marsan M.A., Balbo G., Bobbio A., Chiola G., Conte G. and Cumani A. (1989) - *The effect of Execution Policies on the Semantics and Analysis of Stochastic Petri Nets*, IEEE Trans. on Software Engineering, **15**, pp 832-845, 1989.

[63] Marsan M.A., Balbo G., Chiola G. and Conte G. (1987) - *Generalized Stochastic Petri Nets Revisited: Random Switches and Priorities*, Proc. Int. Workshop on Petri Nets and Performance Models, Madison, WI, pp 44-53, Aug. 1987.

[64] Marsan M.A. and Chiola G. (1986) - *On petri Nets with deterministic and Exponential Transition Firing Times*, Proc. 7th European Workshop on Application and Theory of Petri Nets, Oxford, UK, pp 151-165, June 1986.

[65] Marsan M.A., Conte G. and Balbo G. (1984) - *A Class of Generalised Stochastic Petri Nets for the Performance Analysis of Multiprocessor Systems*, ACM Transactions on Computers, **2**, no. 2, pp 93-122, 1984.

[66] Matthes K. (1962) - *Zur Theorie der Bedienungsprozesse,* Trans. 3rd Prague Conference Inf. Th., Stat. Dec. Fns. and Rand. Proc., pp 513-528, 1962.

[67] Menasche M. (1985) - *PAREDE: An Automated Tools for the Analysis of Time(d) Petri Nets*, International Workshop on Timed Petri Nets, Torino, Italy, pp 162-169, July 1985.

[68] Merlin P. (1974) - *A Study of the Recoverability of Computer Systems* Ph. D. Dissertation, University of California, Irvine 1974.

[69] Merlin P. (1979) - *Specification and Validation of Protocols*, IEEE Trans Comm, **COM-27**, No. 11, pp 1671-1680, Nov. 1979.

[70] Merlin P. and Farber D.J. (1976) - *A Methodology for the Design and Implementation of Communication Protocols*, IEEE Trans. on Communications, **COM-24**, no. 6, pp 614-621, June 1976.

[71] Merlin P. and Farber D.J. (1976) - *Recoverability of Communications Protocols - Implications of a Theoretical Study* IEEE Trans. on Communications, **COM-24**, no. 9, pp 1036-1042, Sept. 1976.

[72] Molloy M.K. (1981) - *On the Interpretation of Delay and Throughput Measures in Distributed Processing Models*, Ph. D. Dissertation, University of California, Los Angeles, 1981.

[73] Molloy M.K. (1982) - *Performance Analysis using Stochastic Petri Nets* IEEE Trans. on Computers, **C-31**, no. 9, pp 913-917, Sept. 1982.

[74] Molloy M.K. (1985) - *Discrete Time Stochastic Petri Nets*, IEEE Trans on Software Engineering, **SE-11**, no. 4, pp 417-423, April 1985.

[75] Molloy M.K. (1985) - *Fast Bounds for Stochastic Petri Nets*, Proc. Int. Workshop on Timed Petri Nets, Torino, Italy, pp 244-249, July 1985.

[76] Murata T. (1989) - *Petri Nets: Properties, Analysis and Applications.* Proc. of the IEEE, **77** No. 4, pp 541-580, 1989.

[77] Natkin S. (1980) - *Les Reseaux de Petri Stochastiques et leur Application a l'Evaluation des Systemes Informatiques*, These de Docteur Ingegneur, CNAM, Paris, France, June 1980.

[78] Noetzel A. (1979) - *A Generalized Queueing Discipline for Product Form Queueing Networks*, J. Assoc. Comp. Mach., **26**, pp 779-793, 1979.

[79] Peterson J.L. (1981) - *Petri Net Theory and the Modelling of Systems.* Englewood Cliffs, NJ: Prentice-Hall, Inc., 1981.

[80] Petri C.A. (1962) *Kommunikation mit Automaten*, Ph. D. dissertation, University of Bonn, West Germany, (In German); Also: M.I.T. Memorandum MAC-M-212, Project MAC, Massachusetts Institute of Technology, Cambridge, Massachusetts; Also: Clifford F. Greene Jr. (translator), *Communi-*

cation with Automata" Supplement 1 to Technical Report RAC-TR-65-377, Volume 1 Rome Air Development Center, Griffiss Air Force Base, New York, (Jan 1966), 89 pages.

[81] Ramchandani (1973) - *Analysis of Asynchronous Concurrent Systems by Petri Nets*, Ph. D. Dissertation, Dept. of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, July 1973. Also: Technical Report 120, Project MAC, Massachusetts Institute of Technology, Cambridge, Massachusetts, Feb. 1974.

[82] Ramamoorthy C.V. and Ho G.S. (1980) - *Performance Evaluation of Asynchronous Concurrency Systems using Petri Nets*, IEEE Trans. on Software Engineering **SE-6**, no. 5, pp 440-449, Sept. 1980.

[83] Razouk R.R (1984) - *The Derivation of Performance Expressions for Communications Protocols from Timed Petri Nets*, Proc. ACM SIGCOMM '84, Symposium on Communications Architecture and Protocols, Montreal, pp 210-217, June 1984.

[84] Razouk R.R. and Phelps C.V. (1984) - *Performance Analysis using Timed Petri Nets*, Proc. 1984 Intl. Conf. on Parallel Processing, pp 126-128, Aug. 1984.

[85] Reiser M. and Lavenberg S.S. (1980) - *Mean-Value Analysis of Closed Multichain Queueing Networks*, J. ACM **27**, pp 313-322, 1980.

[86] Reisig W. (1984) - *Petri Nets*, EATCS Monographs on Theoretical Computer Science, **4**. New York: Springer-Verlag, 1985.

[87] Rumsewicz M. and Henderson W., (1989)- *Insensitivity with Age Dependent routeing*, Adv. Appl. Prob., **21**, pp 398-408, 1989.

[88] Schassberger R. (1977) - *Insensitivity of Steady State Distributions of Generalised Semi-Markov Processes*, Part I, Ann. Prob. **5**, pp 87-99, 1977.

[89] Schassberger R. (1978a) - *Insensitivity of Steady State Distributions of Generalised Semi-Markov Processes*, Part II, Ann. Prob. **6**, pp 85-93, 1978.

[90] Schassberger R. (1978b) - *Insensitivity of Steady State Distributions of Generalised Semi-Markov Processes with speeds*, Adv. Appl. Prob., **10**, pp 836-851, 1978.

[91] Sifakis J. (1977) - *Use of Petri Nets for Performance Evaluation*, Measuring, Modelling and Evaluating Computer Systems, Ed. H. Beilner, E. Gelenbe, North-Holland, pp 75-93, 1977.

[92] Silva M., Colom J. M. (1988) - *On the Computation of Structural Synchronic Invariants in P/T nets*, Lecture Notes in Comp. Science, Advances in Petri Nets 1988, **340**, pp 386-417.

[93] Taylor P. (1987) - *Aspects of Insensitivity in Stochastic Processes*, Ph. D. dissertation, 1987.

[94] Vantilborgh H. (1978) - *Exact Aggregation in Exponential Queueing Networks*, J. ACM, **25**, pp 620-629, 1978.

[95] Walrand J. (1983) - *A Discrete-time Queueing Network*, J. Appl. Prob., **20**, pp 903-909, 1983.

[96] Wang I.Y. and Robertazzi T.G. (1988) - *Service Stage Petri Net Protocols with Product Form Solution,* CEAS Technical Report 534, 1988.

[97] Whittle P. (1967) - *Nonlinear Migration Processes*, Bull. Int. Inst. Statist., **42**, pp 642-647, 1967.

[98] Whittle P. (1968) - *Equilibrium Distributions for an Open Migration Process*, J. Appl. Prob., **5**, pp 567-571, 1968.

[99] Wong C.Y., Dillon T.S. and Forward K.E (1985) - *Timed Places Petri Nets with Stochastic Representation of Place Times*, Proc. Int. Workshop on Timed Petri Nets, Torino, Italy, pp 96-103, July 1985.

[100] Wong C.Y. and Henderson W. (1987) - *Time-Extended Petri Net Models - An Overview*, 2nd Australian Teletraffic Research Seminar, organised by Teletraffic Research Centre, University of Adelaide. 1987.

[101] Woo D.L., Phelps C.V. and Sidwell R.D. (1986) - *Timed Petri Nets Probability Semantics*, Proc. 7th European Workshop on Applications and Theory of Petri Nets, Oxford, UK, pp 131-149, June 1986.

[102] Zuberek W.M. (1980) - *Timed Petri Nets and Preliminary Performance Evaluation*, 7th Annual Symposium on Computer Architecture, La Baule, France, pp 88-96, May 1980.

[103] Zuberek W.M. (1985) - *M-Timed Petri Nets, Priorities, Preemption and Performance Evaluation of Systems*, Advances in Petri Nets (LNCS 222), ed. G. Rozenburg, Springer-Verlag, pp 478-496, July 1985.

[104] Zuberek W.M. (1986) - *Modified D-Timed Petri Nets, Timeouts and Modelling of Communication Protocols*, Proc. 6th Intl. Conf. on Distributed Computer Systems, pp 452-457, 1986.

[105] Zuberek W.M. (1986) - *Extended D-Timed Petri Nets, Timeouts and Analysis of Communication Protocols*, Proc. 1985 ACM Ann. Conf., Denver, pp 10-15, Oct. 1986.

[106] Zuberek W.M. (1987) - *Pre-emptive D-Timed Petri Nets, Timeouts, Modelling and Analysis of Communication Protocols*, INFOCOM 87, San Fransisco, pp 721-730, April 1987.