Exact Results in the Aggregation and Disaggregation of Stochastic Petri Nets

W. Henderson and D. Lucic

Teletraffic Research Centre
Applied Mathematics Department
University of Adelaide

Abstract

In this paper we apply some results from the theory of insensitivity to stochastic Petri nets. In doing so exact equilibrium distributions are derived for the aggregation and disaggregation of a class of stochastic Petri nets.

1 Introduction

Norton's theorem, proved by Chandy, Herzog and Woo [5] for product form queueing networks, leads to an aggregation and disaggregation procedure which yields exact, joint queue length, equilibrium distributions. The use of Norton's theorem is invaluable when networks have large state spaces and has proved useful in the analysis of SPNs which have product form queueing network components (Balbo, Bruell and Ghanta [2]). In this paper we take a preliminary step to a Norton's theorem for SPNs.

The key to our procedure is to aggregate until the resultant "skeleton" net has the "insensitivity" property, a property which historically has been studied within the framework of generalised semi-Markov processes (GSMPs) [9],[15],[18]. The relevant results from insensitivity theory are contained in Section 2 but the basic definition is that a process is ***insensitive*** to it's generally distributed lifetimes if the equilibrium distribution of the process depends upon these general distributions only through their means. Consequently a process which is known to be insensitive can be studied, in equilibrium, by replacing the general distributions with any distribution of our choice; usually the negative exponential distribution because of its mathematically elegant memoryless property.

Insensitivity occurs often in applications. For example semi-Markov processes (SMPs) are insensitive to the distribution of time spent in each state before a state change occurs. To date the analysis of SPNs with generally distributed transition firing times, has revolved around either modelling the transitions using specifically designed SPNs [3] or restricting the SPN so that the process can be modelled as a SMP [1],[6]. SMPs however are insensitive, suggesting the possibility that the results developed in [6] may be due to the insensitivity property of the SMP and therefore may naturally extend to any insensitive system. It is this idea which we develop and generalise in this paper.

Insensitivity also appears in the theory of networks of queues. Baskett, Chandy, Muntz and Palacios [4] observed that certain queues with generally distributed service times could be placed at the nodes of a Jackson network (see [12]) and a product form equilibrium distribution obtained. At the same time Kelly [13],[14] was showing that the key to this phenomenon involved a class of queues called symmetric queues. Symmetric queues with arbitrary service time distributions can be placed at the nodes of a Jackson network without altering the simple form of the equilibrium distribution. The property being observed by both Kelly and Baskett *et al* is insensitivity within queueing networks, hinted at originally by König and Jansen [15], and later established by Schassberger [20] directly from the theory of insensitivity within generalised semi-Markov processes.

Since our objective is to create a skeleton SPN with the insensitivity property we do not propose a step by step approach to aggregation. Instead the analyst aims for a skeleton SPN which is insensitive. This may be accomplished either by using the partial balance test illustrated in Section 3 or, from experience, by recognising the SPN as having a known insensitive structure. This task may not be trivial until some expertise has been gained but it is worth noting that for each insensitive skeleton SPN there are an infinite number of

TH0386-3/91/0000/0166/$01.00 © 1991 IEEE

original SPNs for which the skeleton SPN provides exact results. Consequently, a library of insensitive SPNs, or an understanding of the properties of insensitivity, is essential for effectively utilising this approach. An earlier paper by the authors [10] may prove relevant in this respect.

Norton's theorem [5] requires a product form equilibrium distibution for the original network for the aggregation and disaggregation procedure to give exact results. In this approach we do not assume a product form solution but do rely on the skeleton SPN to be insensitive. In both cases partial balance equations are being imposed on the process, neither of which are necessarily a subset of the other. This suggests that there is a more general unifying theory containing both results where forms of partial balance on invariant measures hold iff certain aggregation techniques and results apply. We refer to our approach as a preliminary step to a Norton's theorem for SPNs because we feel the more general results will appear in the near future. In our case we require the strong condition of insensitivity for the skeleton SPN and consequently do not expect the technique to be generally applicable. However when it is appropriate much can be done. For instance, Example 3 of Section 3 includes forks, joins and generally distributed, concurrent, firing times. A strength of the approach is that, once an insensitive skeleton SPN has been achieved each aggregated subnet of the original SPN along with the relatively simple insensitive skeleton SPN can be analysed in isolation, without losing accuracy but significantly increasing computational efficiency. Exact marginal equilibrium distributions for the original SPN can be derived from the skeleton SPN and the complete equilibrium distribution can be evaluated, when necessary, using the disaggregation procedure of Section 4.

The paper is constructed as follows: In Section 2 we discuss insensitivity, generalised semi-Markov processes and stochastic Petri nets and present the ideas surrounding the proposed aggregation-disaggregation technique. Section 3 contains examples of the aggregation procedure with the disaggregation procedure discussed in Section 4.

2 Definitions and Basic Results

Definition: Stochastic Petri Net

A SPN has a finite set, $\mathcal{P}$, of places, a finite set, $\mathcal{T}$ of transitions and tokens chosen from a finite set $\mathcal{C}$ of colours. The *markings* of the SPN are $\mathbf{m} = (m(i,c),\ \ i \in \mathcal{P},\ \ c \in \mathcal{C})$ where $m(i,c)$ gives the number of tokens of colour c in place i. When transition $t \in \mathcal{T}$ fires it releases an input bag of tokens $I(t)$ and transforms them immediately into an output bag from the set $\{O_j(t)\ 1 \le j \le N(t)\}$ with associated probabilities $\{p(j,t)\ 1 \le j \le N(t),\, t \in \mathcal{T}\}$. We require that $\sum_{j=1}^{N(t)} p(j,t) = 1\ \forall\, t \in \mathcal{T}$. Transition $t \in \mathcal{T}$ is *enabled* in marking $\mathbf{m}$ if $\mathbf{m} - I(t) \ge \mathbf{0}$, in which case it can fire, creating the marking $\mathbf{m} - I(t) + O_j(t)$ with probability $p(j,t)$. The time between transition $t \in \mathcal{T}$ becoming enabled and firing is the *firing time*.

The standard definition of a SPN assumes that the transition firing times are negative exponentially distributed. In this paper, however, we retain the flexibility that some of the transitions may have arbitrary general distributions.

Probabilistic output bags can be modelled using additional places with immediate transitions. We retain the former for mathematical convenience and so that results from GSMP theory can be directly applied to SPNs.

Definition: Generalised Semi-Markov Process

A GSMP is defined on a set of states $\{\mathbf{g} : \mathbf{g} \in G\}$. Within each of these states are active elements s, from the set S which decay at the rate $c(s,\mathbf{g})$, $s \in S$. When the active element s dies, the process moves to state $\mathbf{g}' \in G$ with probability $p(\mathbf{g}, s, \mathbf{g}')$. With $S = S' \cup S^\star$, 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. In order to establish insensitivity results it is standard, and in most cases necessary, 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 active 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 [18]. This assumption shall be retained throughout the paper and its relevance discussed later.

The marking of a SPN, giving the number and colours of tokens in each place, contains, in general, insufficient information to define a Markov process corresponding to the SPN. The generally distributed transition firing times cause the process to be non Markovian. However the standard technique of appending the marking with supplementary variables retrieves the

Markov property and also realises a GSMP. The approach that we use throughout this paper is to model a SPN as a GSMP and apply the insensitivity results directly. Example 1 of Section 3 illustrates this step and also highlights the care which must be taken when including supplementary variables in the state space if insensitivity is to be achieved in the skeleton SPN.

Excellent papers by Haas and Shedlar [7],[8] also use GSMP theory to examine SPNs and include a formal definition of the underlying Markov process of a SPN as a GSMP. Although we use the same GSMP modelling framework our reasons for doing so are different. Haas and Shedlar are interested in discrete event simulation techniques whereas we seek exact equilibrium distributions.

Throughout this paper the assumption is made that any stochastic process representing the SPN is stationary.

Theorem 1

Within the framework of GSMPs Matthes [18] showed that the following two statements are equivalent:

1. The process is insensitive with respect to the active elements of $S^\star$. That is, the general distributions of the lifetimes of the elements of $S^\star$ can be replaced by any other distributions with the same mean, and yet the process still retains the same equilibrium distribution.
2. When all active elements of $S^\star$ are assumed to be negative exponentially distributed, the flux out of each state due to the death of an element of $S^\star$ is equivalent to the flux into that state due to the birth of that element.

The second statement describes the set of ***insensitivity balance equations*** for each active element and will be illustrated in Example 1. In some circumstances insensitivity balance is equivalent to the detailed balance associated with reversibility and, in product form queueing networks, when the queue is symmetric, it is equivalent to job local balance. In SMPs insensitivity balance arises naturally as 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 partial balance equations.

Let $\pi(\mathbf{g})$, $\mathbf{g} \in G$ be the equilibrium distribution for the purely Markov form of the GSMP, i.e where the distribution of each of the active elements in $S^\star$ is negative exponentially distributed.

For each $\mathbf{g}$ of the GSMP let $\mathbf{y}$ be a vector comprising either spent or residual lifetimes of each of the active elements in $S^\star$. For our purposes it does not matter which of these lifetimes are used. When the GSMP is insensitive the equilibrium distribution is the same when the state space is supplemented with either spent or residual lifetimes (see [9]).

Let $(\mathbf{g}, \mathbf{y})$ be the state of the GSMP supplemented by these lifetimes.

Theorem 2

If the GSMP is insensitive the equilibrium density of being in state $(\mathbf{g}, \mathbf{y})$ is

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

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

Theorem 2 gives the equilibrium distribution of an insensitive GSMP when the state space includes the supplementary variables necessary to make the process Markovian. In Section 4 we show that Theorem 2 is the pivotal idea in reversing the aggregation procedure to give exact results for the original SPN.

3 Examples

In this section we introduce insensitivity through Example 1, followed by the aggregation results and some special cases. Example 1 contains a number of general features but in some respects has been kept as simple as possible for ease of explanation. In particular there are no probabilistic output bags. Examples with probabilistic output bags are not difficult to find and include, for example, any skeleton SPN which takes the form of a Kelly or BCMP queueing network with the aggregated subnets corresponding to the symmetric queues.

Example 1

Consider the SPN of Figure 1.

Place p_1 contains a group of customers which are trying to access a collection of resources. The set of resources available for use are in place p_2 and the customers using resources and the resources being used are located in place p_3. The complete set of resources in the net, in places p_2 and p_3, is $\mathcal{R}$.

There are N customers in the SPN individually labelled from the set of coloured tokens $\{c_j \ 1 \le j \le N\}$. Customer c_j requires the use of the set of resources R_j with $R_j \subset \mathcal{R}$. When the resources R_j are accessed c_j holds them in place p_3 for a generally distributed time with mean $[\lambda_j]^{-1}$. After using the resources, c_j returns to p_1 and waits for a generally distributed time with mean $[\gamma_j]^{-1}$ before again accessing the resources.

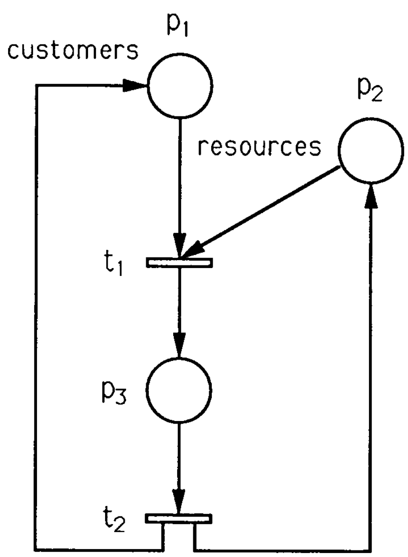

Figure 1

For reasons to be made clear soon, a preemptive resume protocol is assumed at the access stage so that, whenever a subset of the required resources is removed from p_2, the residual time to access is frozen. Transition t_1 thus represents an active element for each of the customers in p_1, some of which may have their firing times frozen. Transition t_2 similarly represents a number of concurrent lifetimes, one for each customer accessing resources. Consequently the SPN of Figure 1 is a folded coloured SPN with t_1 and t_2 each representing N original transitions.

Note that, with the numerous concurrent, generally distributed, active elements and the preemptive resume protocol, this example cannot be modelled either as a semi-Markov process or as a BCMP queueing network. We shall however show that it is insensitive to all of its active elements.

Let the marking of the SPN be the N-vector $\mathbf{m} = (\delta_j \ 1 \le j \le N, \ \delta_j = 1 \text{ or } 3)$ where for each customer c_j, $\delta_j = i$ in the marking $\mathbf{m}$ indicates that customer c_j is in place p_i.

To model the SPN of Figure 1 as a GSMP, $\mathcal{G}$ say, create the state $\mathbf{g}$ by appending to each marking $\mathbf{m}$ an N vector $\mathbf{s}$ of active elements s_{j,δ_j}, one for each customer, representing the required time to access the resources R_j (when $\delta_j = 1$) or the time using resources R_j (when $\delta_j = 3$).

An active element has been included in the state space of $\mathcal{G}$ for each interrupted time to access resources to record the residual lifetime of the interrupted firing time.

In the case $\delta_j = 3$ c_j is using resources and the residual lifetime of the active element $s_{j,3}$ is decreasing at a constant rate. Thus the speed on active element $s_{j,3}$ in state $\mathbf{g}$ is $c(s_{j,3}, \mathbf{g}) = 1$. The same is true when $\delta_j = 1$ and the resources R_j are available. However when a customer attempting to access resources has been preempted the preemptive resume protocol enforces a zero speed on the residual time to access resources. That is $c(s_{j,1}, \mathbf{g}) = 0$ whenever the resources R_j are unavailable in state $\mathbf{g}$.

To ensure insensitivity the definition of a GSMP requires that a state change cannot activate or kill two generally distributed active elements simultaneously and all other generally distributed active elements carry over their spent lifetimes to the next state. The preemptive resume protocol automatically makes this assumption. Although the disabled transition has been preempted its active element is not dead and will resume where it left off when the transition is again enabled.

To establish insensitivity for $\mathcal{G}$, via Theorem 1, consider the purely Markov version of $\mathcal{G}$, $\mathcal{M}$ say, where all active elements are negative exponentially distributed. In this case the preemptive resume and preemptive repeat different protocols are mathematically identical because of the memoryless property of the negative exponential distribution.

Let $j : \delta_j = 3$ signify the set of customers in p_3 and $\mathbf{m}(j)$ be the marking created from marking $\mathbf{m}$ by customer c_j moving from p_1 to p_3 or vice versa.

Result

The purely Markov process $\mathcal{M}$ is reversible with equilibrium distribution

$$\pi(\mathbf{m}) = K \prod_{j:\delta_j=3} \frac{\gamma_j}{\lambda_j} \qquad (2)$$

and $\mathcal{G}$ is insensitive to all of its active elements. K is a normalising constant.

Proof

We prove this result by satisfying the detailed balance equations of reversibility and noting that these equations, when appropriately summed, produce the global balance equations.

Consider any marking $\mathbf{m}$ and any j. Let $\mathbf{m}(j)$ be the marking differing from $\mathbf{m}$ by c_j moving from p_1 to p_3 when $\delta_j = 1$ in $\mathbf{m}$ or from p_3 to p_1 when $\delta_j = 3$.

When $\delta_j = 1$ and the resources R_j are available in marking $\mathbf{m}$ equation (2) is seen to satisfy

$$\pi(\mathbf{m})\gamma_j = \pi(\mathbf{m}(j))\lambda_j . \tag{3}$$

If R_j is not available for c_j in marking $\mathbf{m}$ when $\delta_j = 1$ then $\pi(\mathbf{m}(j)) = 0$, $s_{j,1}$ has zero speed and the detailed balance still holds as

$$\pi(\mathbf{m})0 = 0\lambda_j . \tag{4}$$

For those j with $\delta_j = 3$ in $\mathbf{m}$

$$\pi(\mathbf{m})\lambda_j = \pi(\mathbf{m}(j))\gamma_j \tag{5}$$

is again satisfied by equation (2). Summing equations (3), (4) and (5) gives the global balance equations for the arbitrary state $\mathbf{m}$ and the result is proved.

Using Theorem 1, the insensitivity balance equations must be satisfied for each marking to which the active element under scrutiny belongs. These are:

for $s_{j,1}$ and R_j available in $\mathbf{m}$
$$\pi(\mathbf{m})\gamma_j = \pi(\mathbf{m}(j))\lambda_j$$
for $s_{j,1}$ and R_j unavailable in $\mathbf{m}$
$$\pi(\mathbf{m})0 = 0\lambda_j$$
for $s_{j,3}$
$$\pi(\mathbf{m})\lambda_j = \pi(\mathbf{m}(j))\gamma_j$$

The insensitivity balance equations are the detailed balance equations (3), (4) and (5). Consequently the SPN of Figure 1 is insensitive to all customers access and usage times.

Example 2

Figure 2 is the unfolded SPN model for the resource access problem with two customers (the tokens in P_1 and P_3) sharing a common resource (the token in P_2). Figure 3, on the other hand, is a resource access problem analysed by Lazar and Robertazzi [16] with a resource access phase which appears to be more complex than that of Figure 2.

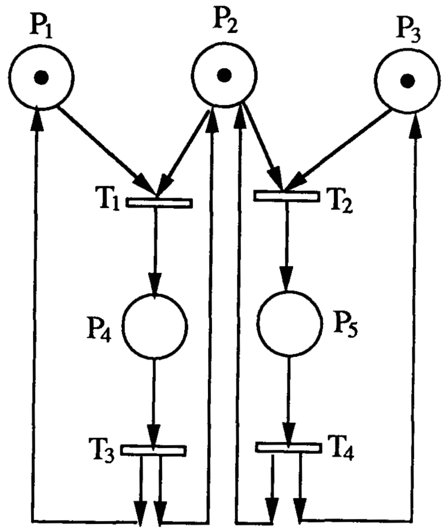

Figure 2

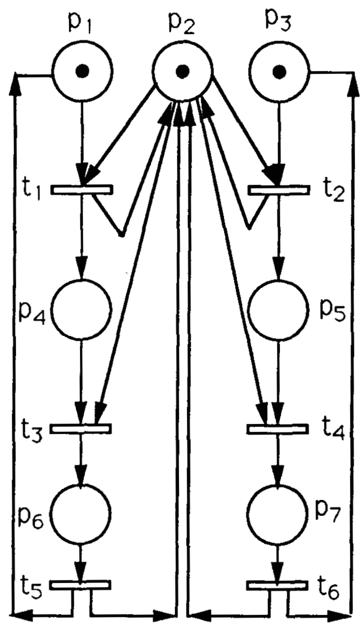

Figure 3

Lazar and Robertazzi assume that transitions t_i $1 \le i \le 6$ are negative exponentially distributed with means $[\mu_i]^{-1}$ $1 \le i \le 6$ and hence that the time to access is governed by the convolution of two negative exponential distributions. They also assume a preemptive resume protocol in the access stage since if customer 1, for example, has reached stage p_4 when customer 2 accesses the resource, customer 1 will resume from p_4 when the resource is returned to p_2 by customer 2. As each stage has a negative exponential distribution, and is therefore memoryless, the assumed protocol is equivalent to the access time being frozen until the access begins again; i.e. a preemptive resume policy.

Thus Lazar and Robertazzi find the equilibrium distribution for a resource access problem in which the time to access is the convolution of two negative exponential distributions and obeys a preemptive resume protocol. Since Example 1 established that this process is insensitive the marginal equilibrium distribution for markings $\mathbf{m} = (\delta_1, \delta_2)$ is given by equation (2) with:

$$[\gamma_1]^{-1} = [\mu_1]^{-1} + [\mu_3]^{-1} \quad [\lambda_1]^{-1} = [\mu_5]^{-1}$$
$$[\gamma_2]^{-1} = [\mu_2]^{-1} + [\mu_4]^{-1} \quad [\lambda_2]^{-1} = [\mu_6]^{-1}.$$

Note that $\mathbf{m}$ is a marking of the insensitive SPN of Figure 2 but not of the SPN of Figure 3 as it lacks information on the phase of access. In Section 4 we show how to retrieve the equilibrium distribution for the original SPN of Figure 3.

Note also that the above marginal distribution is still valid when all firing times in the Figure 3 SPN are generally distributed with means $[\mu_i]^{-1} \quad 1 \leq i \leq 6$.

Alternatively if T_1 and T_2 are given firing times which are the sum of four negative exponential random variables and T_3 and T_4 have firing times as the sum of two negative exponential random variables the extension of [16] given by Wang and Robertazzi [21] has been achieved. Furthermore the model can be extended to any number of negative exponential access and resource usage stages by the appropriate use of sums of negative exponential random variables.

Example 3

The SPN of Figure 4 is to be used in Section 4 to illustrate disaggregation. It again has a skeleton SPN as in Figure 2 and represents a problem in which resource usage involves some parallel activity represented by the places p_5 and p_6 for customer 1 and places p_9 and p_{10} for customer 2.

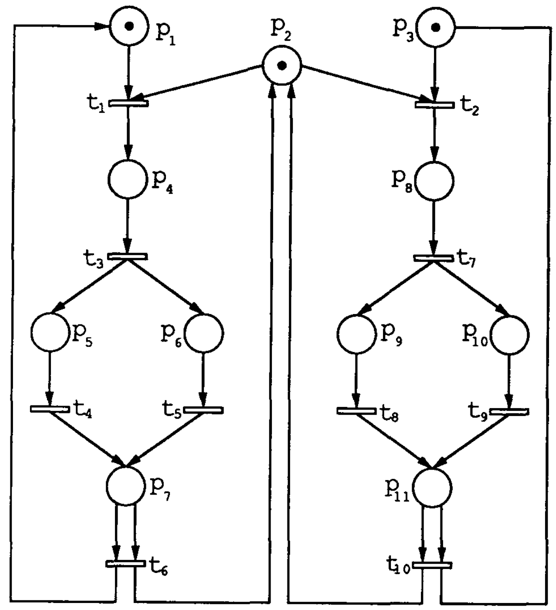

Figure 4

The above results indicate that each generally distributed active element in an insensitive SPN can be the result of aggregating a subnet. The time associated with an active element is the time taken for an unimpeded token to pass through the subnet but, as we have seen in the examples, this time can be worked off at different speeds depending on the state of the SPN. For example, a server sharing facility works the time off more slowly when more customers are present in the subnet and a preemptive repeat discipline freezes the residual time for a while. The subnets need not consist only of strings of places and transitions with negative exponentially distributed firing times as in [16] and [21]. They can equally well include forks, joins and generally distributed firing times (see Example 3). A rough rule of thumb for aggregation is that any subnet can be replaced by a single place-transition combination if tokens entering the subnet are allocated a time from a fixed general distribution for traversing the subnet with that time possibly worked off at a marking dependent speed.

It should now be clear, as stated in the introduction, that each skeleton net gives exact marginal distributions for an infinite number of original SPNs. For example the skeleton SPN of Example 1, with its $2N$ generally distributed active elements, can be the end result of up to $2N$ subnet aggregations each of which could have been any of an infinite number of subnets.

4 Disaggregation

Now let us consider the task of retrieving information apparently lost in the aggregation procedure. In Section 3 we noted that the SPN of Figure 2 is insensitive to all four of its transitions and its purely Markov version had an equilibrium distribution given by equation (2).

Label the markings of the SPN of Figure 2 as:

$$\overline{\mathbf{m}}_1 = \{1,1,1,0,0\}$$
$$\overline{\mathbf{m}}_2 = \{0,0,1,1,0\}$$
$$\overline{\mathbf{m}}_3 = \{1,0,0,0,1\}$$

representing the number of tokens in places P_1 to P_5. The $\overline{\mathbf{m}}$ notation is used to indicate that the marking is of the skeleton net. Using the notation of Example 1 the equilibrium distribution can be completely specified, in this case as:

$$\pi(\overline{\mathbf{m}}_1) = \frac{\lambda_1\lambda_2}{(\lambda_1\lambda_2+\gamma_1\lambda_2+\gamma_2\lambda_1)} \tag{6}$$

$$\pi(\overline{\mathbf{m}}_2) = \frac{\gamma_1\lambda_2}{(\lambda_1\lambda_2+\gamma_1\lambda_2+\gamma_2\lambda_1)} \tag{7}$$

$$\pi(\overline{\mathbf{m}}_3) = \frac{\gamma_2\lambda_1}{(\lambda_1\lambda_2+\gamma_1\lambda_2+\gamma_2\lambda_1)}. \tag{8}$$

Since we are now dealing with generally distributed transition firing times, a finer description of the markings of the SPN of Figure 2 can be achieved by including supplementary variables representing the spent or residual lifetimes for each active element. Thus, due to the insensitivity of the process, the supplemented equilibrium distribution of the SPN in Figure 2 is given by Theorem 2. That is,

$$\pi(\overline{\mathbf{m}}_1, y_1, y_2) = \pi(\overline{\mathbf{m}}_1)\prod_{i=1}^{2}\gamma_i(1 - G_i(y_i)) \tag{9}$$

$$\pi(\overline{\mathbf{m}}_2, y_3) = \pi(\overline{\mathbf{m}}_2)\lambda_1(1 - G_3(y_3)) \tag{10}$$

$$\pi(\overline{\mathbf{m}}_3, y_4) = \pi(\overline{\mathbf{m}}_3)\lambda_2(1 - G_4(y_4)). \tag{11}$$

where y_i is the supplementary variable corresponding to $G_i(\cdot)$, the distribution function for the firing time of transition t_i $1 \leq i \leq 4$.

Except for the P_5-T_4 aggregated subnet expand the SPN given in Figure 2 into its original Figure 4 form. The result is the SPN of Figure 5.

The markings for Figure 5 are:

$$\mathbf{m}_1 = (1,1,1,0,0,0,0,0)$$
$$\mathbf{m}_2 = (0,0,1,1,0,0,0,0)$$
$$\mathbf{m}_3 = (1,0,0,0,0,0,0,1)$$
$$\mathbf{m}_4 = (0,0,1,0,1,1,0,0)$$
$$\mathbf{m}_5 = (0,0,1,0,1,0,1,0)$$
$$\mathbf{m}_6 = (0,0,1,0,0,1,1,0)$$
$$\mathbf{m}_7 = (0,0,1,0,0,0,2,0)$$

giving the number of tokens in places p_1 to p_7 and P_5 respectively.

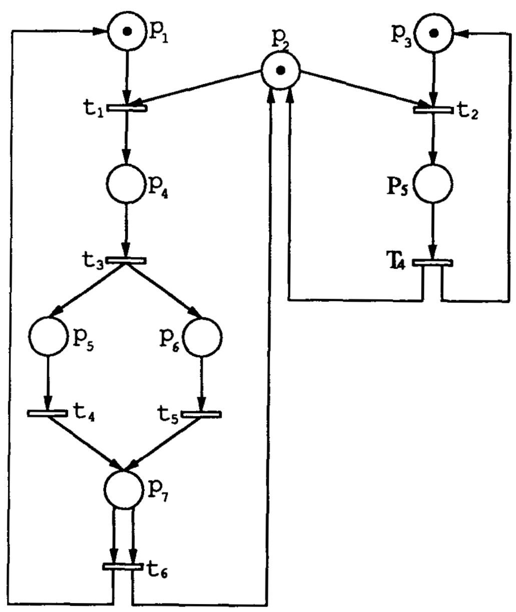

Figure 5

Now assume that the firing time distribution of t_i is $H_i(\cdot)$ with mean $[\mu_i]^{-1}$ for $3 \leq i \leq 6$

Using the state $\overline{\mathbf{m}}_2$ to indicate that a token is in one of the places p_4, p_5, p_6 or p_7 and the supplementary variable y_3 to again give the time the token has had worked off its lifetime in this subnet, basic probabilistic arguements lead to the following supplemented marking equilibrium densities.

Let $\pi(\overline{\mathbf{m}}_2, 0) = B$, given by equations (7) and (10).

$$\pi(\mathbf{m}_2, y_3) = B(1 - H_3(y_3)) \tag{12}$$

$$\pi(\mathbf{m}_4, y_3) = B\int_{u=0}^{y_3}(1-H_4(y_3-u))(1-H_5(y_3-u))dH_3(u) \tag{13}$$

$$\pi(\mathbf{m_5}, y_3) = B\int_{u=0}^{y_3} (1-H_4(y_3-u))H_5(y_3-u)dH_3(u) \tag{14}$$

$$\pi(\mathbf{m_6}, y_3) = B\int_{u=0}^{y_3} (1-H_5(y_3-u))H_4(y_3-u)dH_3(u) \tag{15}$$

$$\pi(\mathbf{m_7}, y_3) = B\Big[\int_{u=0}^{y_3} dH_3(u)\int_{v=u}^{y_3} dH_4(v-u) \int_{w=v}^{y_3} dH_5(w-u)(1-H_6(y_3-w)) + \int_{u=0}^{y_3} dH_3(u)\int_{v=u}^{y_3} dH_5(v-u) \int_{w=v}^{y_3} dH_4(w-u)(1-H_6(y_3-w))\Big] \tag{16}$$

The equilibrium densities $\pi(\overline{\mathbf{m}}_1, y_1, y_2)$ and $\pi(\overline{\mathbf{m}}_3, y_4)$ are unchanged from equations (9) and (11). In order to remove the supplementary variables we integrate equations (12) to (16) over y_3. In the negative exponential case the unsupplemented equilibrium probabilities are then given by,

$$\pi(\mathbf{m_2}) = \pi(\overline{\mathbf{m}}_2)\frac{\lambda_1}{\mu_3} \tag{17}$$

$$\pi(\mathbf{m_4}) = \pi(\overline{\mathbf{m}}_2)\frac{\lambda_1}{\mu_4+\mu_5} \tag{18}$$

$$\pi(\mathbf{m_5}) = \pi(\overline{\mathbf{m}}_2)\frac{\lambda_1\mu_4}{\mu_5(\mu_4+\mu_5)} \tag{19}$$

$$\pi(\mathbf{m_6}) = \pi(\overline{\mathbf{m}}_2)\frac{\lambda_1\mu_5}{\mu_4(\mu_4+\mu_5)} \tag{20}$$

$$\pi(\mathbf{m_7}) = \pi(\overline{\mathbf{m}}_2)\frac{\lambda_1}{\mu_6} \tag{21}$$

where $\pi(\overline{\mathbf{m}}_2)$ is given by equation (7). The right hand side of the SPN of Figure 6 can be disaggregated in a similar manner.

Although the above procedure works for this example the complexity increases when larger subnets are being aggregated. In its present form, with intricate combinations of 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. Referring back to Theorem 2 we note that an insensitive SPN has a product form solution involving the marginal distribution $\pi(\overline{\mathbf{m}})$ and a product of terms $\mu_i(1-G_i(y_i))$. Each y_i, being the supplementary variable corresponding to the time to traverse the appropriate aggregated subnet must, probabilistically, contain the information on the state of the subnet, i.e. if the equilibrium distribution of the y_i's are known then the equilibrium distribution for the state of the subnet is also known, at least in theory. However the product form of equation (1) corresponds to independence between the marginal state and the supplementary variables, and equivalently, the states of the aggregated subnets.

Consequently to find the equilibrium distribution for the original SPN we need to find:

1. The corresponding marginal equilibrium distribution of the skeleton SPN and
2. The equilibrium distribution for the aggregated subnets taken in isolation.

To find any of these distributions we can use simulation, global balance equations, product form or any other technique available. The advantage is that each of the resultant processes has a state space considerably smaller than the original and is therefore much easier to handle.

Consider, for example, the marking $\overline{\mathbf{m}}_2$ of Figure 2 in which a token is in P_3.

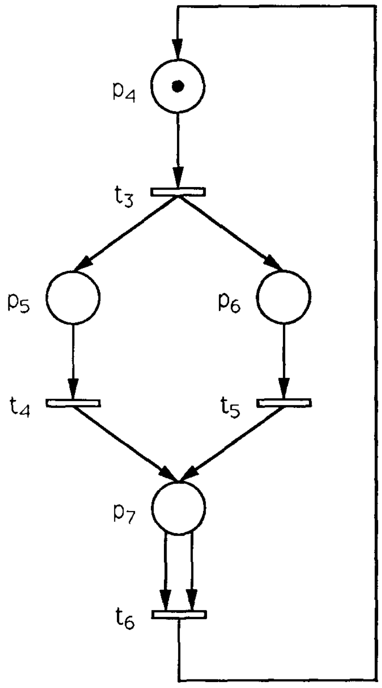

Figure 6

Isolate this subnet with its token in place p_4, as in Figure 6, and close the subnet by looping the token back to place p_4 when it fires out through t_6. Let (i, j, k, l) be a state of the subnet representing i, j, k, l tokens in places p_4, p_5, p_6 and p_7 respectively. An invariant measure for the subnet of Figure 6 is,

$$\pi(0,1,0,1) = \pi(1,0,0,0)\frac{\mu_3\mu_5}{\mu_4(\mu_4+\mu_5)} \quad (22)$$

$$\pi(0,1,1,0) = \pi(1,0,0,0)\frac{\mu_3}{(\mu_4+\mu_5)} \quad (23)$$

$$\pi(0,0,1,1) = \pi(1,0,0,0)\frac{\mu_3\mu_4}{\mu_5(\mu_4+\mu_5)} \quad (24)$$

$$\pi(0,0,0,2) = \pi(1,0,0,0)\frac{\mu_3}{\mu_6} \quad (25)$$

Using the independence property between the marginal equilibrium distribution and those of the aggregated subnets we can construct the equilibrium probabilities for the original SPN. For this example it is achieved by multiplying each of the equations (22) to (25) by $\pi(\overline{\mathbf{m_2}})$ and reconstructing the token distributions for the original SPN. Let $\pi(1,0,0,0) = C$, then,

$$\pi(\mathbf{m_2}) = C\pi(\overline{\mathbf{m_2}}) \quad (26)$$

$$\pi(\mathbf{m_4}) = C\pi(\overline{\mathbf{m_2}})\frac{\mu_3\mu_5}{\mu_4(\mu_4+\mu_5)} \quad (27)$$

$$\pi(\mathbf{m_5}) = C\pi(\overline{\mathbf{m_2}})\frac{\mu_3}{(\mu_4+\mu_5)} \quad (28)$$

$$\pi(\mathbf{m_6}) = C\pi(\overline{\mathbf{m_2}})\frac{\mu_3\mu_4}{\mu_5(\mu_4+\mu_5)} \quad (29)$$

$$\pi(\mathbf{m_7}) = C\pi(\overline{\mathbf{m_2}})\frac{\mu_3}{\mu_6} \quad (30)$$

It is clear that equations (17) to (21) and equations (26) to (30) are equivalent invariant measures with $C = \frac{\lambda_1}{\mu_3}$.

Exactly the same disaggregation procedure can be applied to the right hand side of the SPN of Figure 5.

5 Conclusion

Insensitivity is a strong property of skeleton SPNs. It allows general distributions to be replaced by negative exponential distributions and disaggregation to be performed with the exact equilibrium distribution of the original SPN as the prize. In this paper we have described the process by which these procedures yield exact results. More recent results on insensitivity ([11],[19]) allow the aggregation procedure, outlined herein, to incorporate skeleton SPNs in which the probabilistic choice of output bag when a transition fires depends upon its firing time. The concept is sometimes useful when, for example, time out transitions are contained in the subnet. This work can be found in [17] and will appear elsewhere.

The approach we describe produces generally distributed transitions. There is, however, nothing in insensitivity theory to prevent the active elements of the SPN to be times in places or times in which tokens are absorbed by transitions. The aggregation results of this paper are equally valid in these cases also.

It is possible that the true worth of this theory is as a measuring stick to the accuracy of other aggregation techniques. For example it may be possible to develop techniques in which an insensitive SPN is used to provide a bound for a performance measure of another SPN.

Further research, which should prove fruitful, is to provide a quick approach to identifying insensitive SPNs, to create a step by step aggregation procedure and to examine the insensitivity bounds alluded to above.

Acknowledgement

The authors would like to thank the Teletraffic Research Centre at the University of Adelaide and Telecom Australia for supporting the research presented in this paper and the referees for some constructive comments.

References

[1] Ajmone Marsan M., Balbo G., Bobbio A., Chiola G., Conte G., Cumani A., (1985)- *On Petri Nets with Stochastic Timing*, Proceedings from the International Workshop on Timed Petri Nets, Torino, Italy.

[2] 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**, No. 10.

[3] Bruel, S.C., Chen P-Z. and Balbo G. (1989)-*Alternative Methods for Incorporating Non-exponential Distributions into Stochastic Timed Petri Nets*, Proc. Third Int. Workshop on Petri Nets and Perf. Models, Kyoto, 187-199.

[4] 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.

[5] K.M. Chandy, U. Herzog and L. Woo, (1975) - *Parametric Analysis of Queueing Networks,* IBM. J. Res. Develop., pp 36-42.

[6] Dugan J.B., Geist R.M., Nicola V.F., Trivedi K.S. (1984) - *Extended Stochastic Petri Nets: Applications and Analysis,* Performance 84, Paris, France, pp 507-519.

[7] Haas P. and Shedlar G. S. (1986) - *Regenerative Stochastic Petri Nets,* Performance Evaluation **6**, pp 189-204.

[8] Haas P. and Shedlar G. S. (1988) - *Modelling Power of Stochastic Petri Nets for Simulation,* Prob. Eng. Informational Sci. **2** pp 435-459.

[9] Henderson W. (1983) - *Insensitivity and Reversed Markov Processes,* Adv. Appl. Prob, **15**, pp 752-768.

[10] 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.

[11] Henderson W. and Taylor P.G. (1989) - *Insensitivity of processes with Interruptions,* J.Appl. Prob., **26**, pp 242-258.

[12] Jackson J. (1957) - *Networks of Waiting Lines,* Operations Research, **5** , pp 518-521.

[13] Kelly F. (1975) - *Networks of Queues With Customers of Different Types,* J. Appl. Prob., **12**, pp 542-544.

[14] Kelly F. (1976) - *Networks of Queues,* Adv. Appl. Prob., **8**, pp 416-432.

[15] 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.

[16] Lazar A.A. and Robertazzi T.G. (1986) - *The Algebraic and Geometric Structure of Markovian Petri Network Latices,* 24th. Allerton Conference on Communication, Control and Computing, University of Illinois.

[17] Lucic D. (1990) - *On Exact Equilibrium Distributions of Stochastic Petri Nets,* Ph.D. Thesis, Adelaide University.

[18] Matthes K. (1962) - *Zur Theorie der Bedienungsprozesse,* Trans. 3rd Prague Conference Inf. Th., Stat. Dec. Fns. and Rand. Proc., pp 513-528.

[19] Rumsewicz M. and Henderson W., (1989) - *Insensitivity with Age Dependent routing,* Adv. Appl. Prob., **21**, pp 398-408.

[20] Schassberger R. (1978) - *The Insensitivity of Stationary Probabilities in Networks of Queues,* Adv. Appl. Prob., **10**, pp 906-912.

[21] Wang I.Y. and Robertazzi T.G. (1988)-*Service Stage Petri Net Protocols with Product Form Solution,* CEAS Technical Report 534, Dec. 29