IMPORTANCE SAMPLING FOR THE SIMULATION OF STOCHASTIC PETRI NETS AND FLUID STOCHASTIC PETRI NETS*

Bruno Tuffin
IRISA-INRIA
Campus universitaire de Beaulieu
35042 Rennes cedex, France
btuffin@irisa.fr

Kishor S. Trivedi
CACC
Dept. of Electrical and Computer Engineering
Duke University, Durham, NC 27708-0291, U.S.A.
kst@ee.duke.edu

KEYWORDS: Performance Analysis, Stochastic Petri Nets, Hybrid Simulation, Monte Carlo, Variance Reduction.

Abstract

In this paper, we investigate the use of importance sampling techniques for the discrete event simulation of fluid or discrete stochastic Petri nets (SPNs) with general distributions of firing times. Existing research literature on importance sampling does not deal with fluid events and with policies such as PRI (Preemptive Repeat Identical) or PRS (Preemptive ReSume). Literature also does not consider the case of non-absolutely continuous distributions. Importance sampling has already been applied to a formalism similar to SPNs (the stochastic activity networks (SANs)) but no special attention was paid to the PRI and PRS policies as well as to the fluid models represented by FSPNs. The methods described here are implemented in Stochastic Petri Net Package (SPNP) which is a software package whose goal is to compute performance, availability or performability measures from Stochastic Petri Nets (SPN) and Fluid Stochastic Petri nets (FSPN).

1 Introduction

Performance, dependability and performability evaluation has become an important task in recent years. When the model is Markovian and its size "not too big", efficient analytic-numeric methods are available. Moreover, in some non Markovian cases (for example for Markov regenerative processes (Choi *et al.*, 1993)), analytic-numerical methods can be used. Unfortunately, the conditions to apply these methods are often too restrictive and the state space size of practical systems can make their use intractable. In these situations, simulation is the only available solution method.

A major limitation of the standard Monte Carlo simulation is that it is inefficient in estimating the probabilities of rare events. For example, when trying to estimate the probability of an event whose actual value is 10^{-9}, we must use on the average 10^9 replications to obtain the event just one time. Hence it will take a very long time to obtain an estimate with a high accuracy. *Importance sampling* (Fishman, 1997; Glynn & Iglehart, 1989; Goyal *et al.*, 1992) technique is often used to improve the efficiency of such simulations. Basically, it modifies the stochastic behavior of the system by making some events occur more often. The model is then biased, but the bias is compensated by the introduction of a function called the *likelihood ratio*. The key problem in importance sampling is to find an efficient change of sampling measure and to compute the likelihood ratio. Vast literature exists on importance sampling, especially when applied to highly reliable Markovian systems (see for instance (Goyal *et al.*, 1992; Heidelberger, 1995; Nakayama, 1996; Shahabuddin, 1994; Tuffin, 1999) and the references included). The case of non-Markovian systems, though less studied, has also been investigated (Glynn & Iglehart, 1989; Nicola *et al.*, 1993; Obal II & Sanders, 1994a).

We consider the simulation of stochastic Petri nets (SPNs) (Ajmone Marsan *et al.*, 1995; Choi *et al.*, 1993) with general firing times as well as their fluid extension, the fluid stochastic Petri nets (FSPNs) (Ciardo *et al.*, 1999; Horton *et al.*, 1998; Trivedi & Kulkarni, 1993). The various ilk of SPNs are found to be powerful in modeling performance and dependability measures of computer and communication systems. The simulation of non-fluid models, using importance sampling, has been studied in (Obal II & Sanders, 1994b; Obal II & Sanders, 1994a) where the model is a variant of SPNs, called *stochastic activity nets* (SANs). On the other hand, the simulation of FSPNs has been studied in (Ciardo

*This work was supported in part by the Air Force Office of Scientific Research under a MURI Grant No. F49620-00-1-0327.

et al., 1999), but without using importance sampling.

In this paper we discuss the importance sampling simulation of SPNs and FSPNs. We particularly show how to deal with different policies such as PRI (preemptive repeat identical) or PRS (preemptive resume). These cases are more complicated for generally distributed timed transitions to deal with than the usual PRD (preemptive repeat different) policy and must be carefully described because the updating of the likelihood ratio is affected. Another interesting point is to explain the complications occuring when both discrete and continuous distributions are included in the same model; case also not studied in the literature. As a matter of fact, a basic assumption of the work in (Nicola *et al.*, 1993; Obal II & Sanders, 1994a) is that the distributions are absolutely continuous. Thus, with probability one, two (non-immediate) events can not occur at the same time. We investigate here what happens if this restriction is relaxed.

The methods described here are implemented in SPNP (for Stochastic Petri Net Package) (Ciardo *et al.*, 1989; Hirel *et al.*, 2000), a software package dedicated to the analysis of different forms of SPNs.

The layout of this paper is as follows. In Section 2 we briefly introduce the SPN and FSPN model. Section 3 describes the importance sampling technique, as done in (Nicola *et al.*, 1993; Obal II & Sanders, 1994a), but generalized to fluid models. Section 4 is devoted to a further generalization, the special case of various sampling policies. We give a numerical illustration in Section 5 and we describe in Section 6 the complication when discrete and continuous distributions are used in the same model. Finally, we conclude in Section 7.

2 Stochastic Petri nets

For the sake of generality, we consider here FSPNs (Ciardo *et al.*, 1999; Horton *et al.*, 1998; Trivedi & Kulkarni, 1993), hybrid extension of SPNs, because the usual SPN case is a particular case with no fluid places.

The class of stochastic Petri nets we consider is given by an 12-tuple (Ciardo *et al.*, 1999) $(P, b, \mathcal{T}, D, F, g, m^0, >, d, r, H, \omega)$ where

- $P = P_c \cup P_d$ is the set of places partitioned into discrete places (P_d, represented by circles) and fluid places (P_c, represented by two concentric circles). Each discrete place may contain a (natural) number of *tokens* and each fluid place content is given by a non-negative real number. The *marking* (m, x) of the SPN is then defined as the number of tokens in each discrete place (vector m) and the fluid level of each fluid place (vector x).
- b gives the (discrete marking dependent) bound on each fluid place.
- $\mathcal{T}$ is the set of transitions (represented by bars).
- g is the (marking-dependent) guard function for each transition. This guard function can depend on the fluid levels in fluid places. In SPNP, this fluid dependence is on "rectangular" functions. It means that we define for each fluid place some thresholds and check the guard function value each time the thresholds are reached.
- D is the set of input arcs (from a place to a transition) and output arcs (from a transition to a place). Each arc is given by its marking dependent cardinality (an integer) when dealing with discrete places or marking dependent fluid impulse (a real number) for fluid places and is drawn as a thin arc with an arrowhead on its destination. A transition is said to be *enabled* if each of its input (discrete) places contains as many tokens as the multiplicity of the corresponding arc and if the guard function is verified. The transition which *fires* is then the one with the smallest firing time. Upon firing, the transition removes a number of tokens from each of its input places equal to the multiplicity of the corresponding input arc and it deposits in each of its output place a number of tokens equal to the multiplicity of the corresponding output arc, leading then to a (possibly) new marking. Moreover, from each input fluid place, a volume of fluid equal to the minimum of its fluid level and fluid impulse of the firing transition is removed. In each output fluid place, we add to the actual level a quantity equal to the fluid impulse to the place (if the sum is greater than the bound given by b, then b is the new level).
- F describes the marking dependent fluid rates of the input and output arcs connecting timed transitions and fluid places. These arcs are drawn as thick lines with an arrowhead on their destinations. The fluid flows through the arc at rate given by F (except if the fluid level of the input place is 0 or if the fluid level of the output place is b). In SPNP, only linear FSPNs (Ciardo *et al.*, 1999) (i.e., FSPNs for which the fluid flow in place p is of form $dx_p(t)/dt = f(m)x_p(t) + g(m)$ when the current marking is (m, x) at time t) are implemented.
- (m_0, x_0) is the initial marking.
- $>$ is the explicit priority assigned to transitions. A transition is enabled in a marking only if no higher-priority transition is enabled in that marking.

- d defines the firing time distribution for each transition. It can be marking dependent.

- r is the static resampling policy for each transition when it becomes disabled. Three policies are considered: PRI (preemptive repeat identical), PRD (preemptive repeat different) and PRS (preemptive resume).

- H defines the resampling policy for each transition when another transition fires but the considered transition remains enabled. The same three policies as above are available.

- ω is the weight function to choose between several transitions when they are all enabled and have the same firing time.

This model is used in SPNP, a versatile modeling tool for performance, dependability and performability analysis of complex systems (Hirel *et al.*, 2000; Ciardo *et al.*, 1989) including advanced constructs such as marking-dependency or the ability to define its own reward functions. The Petri nets and the solution methods selected for use are described in CSPL, an ANSI C library.

3 Importance sampling for simulating SPNs and FSPNs

The basic idea of importance sampling is the following (Fishman, 1997; Glynn & Iglehart, 1989). Assume that we wish to compute $E_P(X)$, the expectation of random variable X defined on the probability space $(\Omega, \mathcal{F}, P)$. For instance, X might be the proportion of time that a place pl is empty on the interval of time $[0, T]$, $X = (\int_0^T \mathbf{1}_{(\#pl(t)=0)} dt)/T$, the average number of tokens in this place $X = (\int_0^T \#pl(t) dt)/T$ or the number of times that a transition fires during that interval or the proportion of time that a transition is enabled. More generally, X can be the average reward during the interval $X = (\int_0^T r_{(m(t),x(t))} dt)/T$, where the reward function r depends on the marking $(m(t), x(t))$ at time t.

If P' is another probability measure defined on $(\Omega, \mathcal{F})$ such that $\forall A \in \mathcal{F}$, $P(A) \neq 0$ implies $P'(A) \neq 0$, then

$$E_P(X) = \int_\Omega X dP = \int_\Omega X \frac{dP}{dP'} dP' = E_{P'}\left(X \frac{dP}{dP'}\right).$$

$L = dP/dP'$, quotient of probability measures, is called the *likelihood ratio*. Then to simulate the random variable X under probability measure P', we have to compensate the bias by actually simulating XL. An estimator of $E_P(X)$ is then

$$\frac{1}{N} \sum_{n=1}^{N} X_n L_n$$

using N independent random variables X_n $(1 \leq n \leq N)$ following distribution P'. If P' is properly chosen, then we can have $\sigma^2_{P'}(XL) \ll \sigma^2_P(X)$, so that the size of confidence interval

$$\frac{1}{N} \sum_{n=1}^{N} X_n L_n \pm c_\alpha \frac{\sigma_{P'}(XL)}{\sqrt{N}}$$

(where c_α is the $1 - \alpha/2$ quantile of the standard normal distribution) at risk α can be reduced.

Consider now the discrete event simulation of the model defined in previous section. We are adapting the work in (Nicola *et al.*, 1993; Obal II & Sanders, 1994a) to FSPNs, and the case of special policies will be addressed in the next section. Assume that transitions are either immediate or that their firing times have (for now) absolutely continuous distributions (with probability density function $f(t, \tau)$ and complementary distribution function $\bar{F}(t, \tau)$ for a transition τ). The model is represented by a generalized semi-Markov process (GSMP) (Glynn, 1989). As a matter of fact, the fluid levels are deterministically determined at each time t knowing the state at each discrete transition (because the fluid flows are deterministic) so that events can be decomposed into two categories:

1. Firing of transitions (random events).

2. Fluid events: a fluid place becomes empty or full or a threshold is reached (to check a guard function or to modify the distribution of the firing time of a transition).

In the SPNP simulator core, the next possible fluid event time in each fluid place is computed (which can easily be done explicitly for the linear FSPNs described in previous section). The smallest of these is considered as the next event if no transition fires earlier. Assume that the marking is (m_i, x_i) after the i^{th} event. Let t_i be the time of the i^{th} event and $T_i = t_{i+1} - t_i$ (with $t_0 = 0$). Let $E_i \subset \mathcal{T}$ be the set of enabled transitions in marking (m_i, x_i), let N_i be the set of newly enabled transitions and let $O_i = E_i - N_i$ be the set of transitions which were previously enabled. We define $c_i(E_i)$ as the clock of each transition $\tau \in E_i$. Denote by $\boldsymbol{X}_i = (m_i, x_i, c_i(E_i))$ the internal state of the system upon the i^{th} event. Note that $\boldsymbol{X}$ is not constant between two event times as the fluid levels may vary, but, as said before, it can be deterministically determined. Let $\boldsymbol{X}_{0,n} = (\boldsymbol{X}_0, \cdots, \boldsymbol{X}_n)$ denote the sequence of markings at each event times. Let $\tau_i^* = \text{argmin}_{\tau \in E_i}\{c(\tau)\}$ be the transition which triggers the $(i+1)^{st}$ event, i.e., such that $T_i = c(\tau_i)$, **if it is not a fluid event**. If the next fluid event occurs before $\min_{\tau \in E_i}\{c(\tau)\}$, t_{i+1} is the time of next fluid event. In both cases, the new marking is called (m_{i+1}, x_{i+1}), E_{i+1} is updated from E_i, as well as the clocks (depending on the whole path).

Denote by $f_i(t,\tau)$ (resp. $\bar{F}_i(t,\tau)$) the probability density function (resp. the complementary distribution function) of the (conditional) remaining firing time of timed transition $\tau \in E_i$ at the i^{th} event (depending on $\boldsymbol{X}_{0,i}$ by the subscript i). Upon the next event, if timed transition τ has not fired and is still enabled, i.e. if it is in O_{i+1}, we have

$$f_{i+1}(t,\tau) = \frac{f_i(t+T_i,\tau)}{\bar{F}_i(T_i,\tau)} \tag{1}$$

and

$$\bar{F}_{i+1}(t,\tau) = \frac{\bar{F}_i(t+T_i,\tau)}{\bar{F}_i(T_i,\tau)}. \tag{2}$$

For $\tau \in N_{i+1}$, then we have
$f_{i+1}(t,\tau) = f(t,\tau)$ and $\bar{F}_{i+1}(t,\tau) = \bar{F}(t,\tau)$.

The probability of the whole sample path is

$$P(\boldsymbol{X}_{0,n}) = \prod_{i=0}^{n-1} P(\boldsymbol{X}_{i,i+1}) \tag{3}$$

where $P(\boldsymbol{X}_{i,i+1})$ has one of the following three forms, depending on the fact that the event is a fluid event, the firing of a timed transition or the firing of an immediate transition.

1. If i^{th} event is the firing of timed transition τ_i^*,

$$P(\boldsymbol{X}_{i,i+1}) = f_i(T_i,\tau_i^*) \prod_{\tau \in E_i - \{\tau_i^*\}} \bar{F}_i(T_i,\tau).$$

   With respect to (Nicola *et al.*, 1993), the equation does not involve function p_{i+1} as the new marking and set of enabled transition are deterministically determined when a transition fires.

2. Assume now that transition τ_i^* which is firing is an immediate transition. Let $\mathcal{T}^I \subset \mathcal{T}$ be the set of immediate transitions. τ_i^* is chosen with probability

$$P(\boldsymbol{X}_{i,i+1}) = \frac{\omega(\tau_i^*)}{\sum_{\tau \in \mathcal{T}^I \cap E_i} \omega(\tau)}.$$

   As remarked in (Obal II & Sanders, 1994a), the initial method in (Nicola *et al.*, 1993) did not include the case of immediate events. In SAN simulation (Obal II & Sanders, 1994a), the immediate events are chosen by the case activity probabilities (equivalent to the probabilities of immediate transitions in SPN termonology).

3. If i^{th} event is a fluid event, then we have

$$P(\boldsymbol{X}_{i,i+1}) = \prod_{\tau \in E_i} \bar{F}_i(T_i,\tau).$$

In the case of FSPNs, the fluid flows and impulses are deterministic, so they do not intervene in the computation.

To improve the efficiency of the simulation, we can use new probability distribution functions of firing times F' instead of F and new weights ω' instead of ω, leading to a new probability measure P' on the model. The likelihood ratio is then given by

$$(\boldsymbol{X}_{0,n}) = \frac{P(\boldsymbol{X}_{0,n})}{P'(\boldsymbol{X}_{0,n})} = \prod_{i=0}^{n-1} L(\boldsymbol{X}_{i,i+1}) \tag{4}$$

where

1. If i^{th} event is a firing of a timed transition

$$L(\boldsymbol{X}_{i,i+1}) = \frac{f_i(T_i,\tau_i^*)}{f_i'(T_i,\tau_i^*)} \prod_{\tau \in E_i - \{\tau_i^*\}} \frac{\bar{F}_i(T_i,\tau)}{\bar{F}_i'(T_i,\tau)}.$$

2. If i^{th} event be a firing of an immediate transition,

$$L(\boldsymbol{X}_{i,i+1}) = \frac{\omega(\tau_i^*)}{\sum_{\tau \in \mathcal{T}^I \cap E_i} \omega(\tau)} \frac{\sum_{\tau \in \mathcal{T}^I \cap E_i} \omega'(\tau)}{\omega'(\tau_i^*)}.$$

3. If i^{th} event is a fluid event

$$L(\boldsymbol{X}_{i,i+1}) = \prod_{\tau \in E_i} \frac{\bar{F}_i(T_i,\tau)}{\bar{F}_i'(T_i,\tau)}.$$

An important remark from a programming standpoint is that the updating of f_i and $\bar{F}_i$ in Equations (1) and (2) allows us to obtain the product form in Equations (3) and (4). This is very convenient because we can then easily update the likelihood ratio after each event. With respect to the implementation of crude simulation, we just add a variable representing the likelihood ratio, with initial value 1 and updated after each event using the product form (4). This variable is then used to compute the estimation.

In our simulations, we use *dynamic importance sampling* as in (Nicola *et al.*, 1993; Obal II & Sanders, 1994a). Hence, we can change the firing times of any active (new or old) transition according to some chosen new distribution at any time.

4 Importance sampling for various policies

Different policies can be applied on the firing times of transitions. Consider an enabled transition that is disabled by the firing of a concurrent transition. Then there are several possibilities (integrated in SPNP) when this transition is enabled again.

- Either a new firing time is sampled independently of the previous uncompleted one (PRD policy), which means that all the work done previously is lost and the firing time changes. It can model the case where, when returning to service, a different server is taking care of you. In SPNP, the default policy is PRD.
- Or exactly the same sampled time is used (PRI policy), which means that all the work done previously is lost but the new firing time is exactly the same as the original firing time. It is the case where the same server is serving you again, but forgot to store the work done before leaving.
- Or the transition resumes its job (PRS policy), which means that the work previously done is not lost and can be completed.

Another case where these three policies can be applied in SPNP is when the firing time of a transition is affected by the firing of another transition, even if it remains enabled. There also, PRD, PRI ans PRS policies can be applied. The default "affected" policy is PRS. We denote by $A_i(\tau_i^*) \subset O_i$ the set of transitions affected (with PRI or PRD policy) at time t_i when transition τ_i^* fires.

4.1 Preemptive repeat different (PRD) policy

If transition τ is newly enabled at time t_i (i.e., $\tau \in N_i$) and has a PRD policy or if its firing time is affected (i.e., $\tau \in A_i(\tau_i^*)$) and if the affecting policy is PRD, then we have

$$f_i(t,\tau) = f(t,\tau) \text{ and } \bar{F}_i(t,\tau) = \bar{F}(t,\tau). \qquad (5)$$

Next, the f_i and $\bar{F}_i$ are updated as described in equations (1) and (2) if $\tau \in O_i - A_i(\tau_i^*)$.

The same algorithm is applied to $f_i'(t,\tau)$ and $\bar{F}_i'(t,\tau)$, with the difference that we can change the importance sampling distribution at any discrete event time (it is what we called dynamic importance sampling). Then Equation (5) can be applied to $f_i'(t,\tau)$ and $\bar{F}_i'(t,\tau)$ even if it has been enabled since a while (Nicola *et al.*, 1993).

4.2 Preemptive repeat identical (PRI) policy

This policy is much more complicated for updating the likelihood function. Assume for example that transition τ has been enabled for a time t_0 and has been disabled by the firing of another transition. Then when it becomes enabled again, Equation (5) does not apply anymore as we already know that the firing time is larger than t. It means that the likelihood must be updated following Equations (1) and (2) only when time t is passed.

Formally, let τ be a transition with PRI policy and becoming disabled by the firing of another transition (or if it is affected by the firing of another transition and its affecting policy is PRI) at time t_i. Let cl_τ be a variable giving the maximum of the clock of τ at the times it is becoming disabled or affected before completion of its service, and set to 0 when the service is completed. Assume that τ is enabled again just after the k^{th} event ($k = i$ in the affected case). Then

$$f_k(t,\tau) = \begin{cases} 0 & \text{if } t < cl_\tau \\ f(t,\tau)/\bar{F}(cl_\tau,\tau) & \text{if } t \geq cl_\tau, \end{cases}$$

$$\bar{F}_k(t,\tau) = \begin{cases} 1 & \text{if } t < cl_\tau \\ \bar{F}(t,\tau)/\bar{F}(cl_\tau,\tau) & \text{if } t \geq cl_\tau, \end{cases}$$

and equations (1) and (2) are again used for the next step.

4.3 Preemptive resume(PRS) policy

PRS being the default and natural affecting policy, we study here only the case where it is the policy when the transition has become disabled by the firing of another transition and becomes enabled again. Then Equations (1) and (2) are used from the values of f_i and $\bar{F}_i$ (as well as f_i' and $\bar{F}_i'$) when the transition became disabled (i.e., they are not initialized using Equation (5) or as in the PRI case).

Formally, assume that transition τ becomes disabled at step $i+1$ and is enabled again at step $k > i+1$. We then have

$$f_{k+1}(t,\tau) = \frac{f_i(t+T_i,\tau)}{\bar{F}_i(T_i,\tau)}$$

and

$$\bar{F}_{k+1}(t,\tau) = \frac{\bar{F}_i(t+T_i,\tau)}{\bar{F}_i(T_i,\tau)}.$$

It means that we proceed as in Equations (1) and (2), but at stage k instead of at stage $i+1$.

5 Numerical example

We consider a variant of the FSPN example of (Ciardo *et al.*, 1999; Tuffin & Trivedi, 2000) representing a dual tank system as described in Figure 1. It contains a main tank "One" and an additional tank "Two" (of respective maximal capacity b_1 and b_2). Fluid flows with rate r_{in} from an external

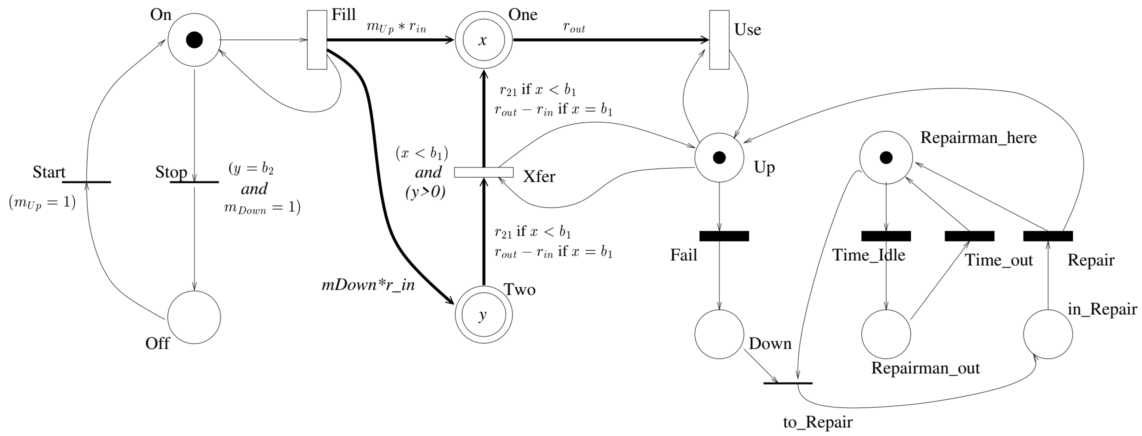

Figure 1: FSPN model of a dual tank

source into tank "One", which sends the liquid to a processing station with rate $r_{out} > r_{in}$. However, the processing station is subject to breakdown and repair. When the processing station is down, it can be repaired, depending on the presence of the repairman. During a breakdown, the station is not fueled and the liquid from the external source is immediately redirected to the additional tank "Two". The external source is shut down only when tank "Two" is full. When the processing station is repaired, the external flow is immediately switched to tank "One" which resumes its work. In addition, the liquid in tank "Two" is pumped into tank "One" with rate r_{21}. If tank "One" is full, the flow from tank "Two" to tank "One" is slowed (to rate $r_{out} - r_{in}$). If the repairman is present and the processing station is up, he can leave the system for a while and come again later. In our model, transition "Time_Idle", refering to the time the repairman is spending idle in the system before leaving for a while and coming back, may have different policies, PRI, PRS or PRD.

For our computations, the parameter values are $b_1 = 1.0$, $b_2 = 2.0$, $r_{in} = 0.30$, $r_{out} = 1.0$ and $r_{21} = 0.75$. The initial marking is such that the processing station is up and the level in both the tanks "One" and "Two" is 0.0. Transitions "Fail" and "Time_idle" are assumed to both have a Weibull distribution with scale parameter 1.0 and shape parameter 2.0. Transitions "Repair" and "Time_out" are assumed to have a constant firing time equal to 1.0. We estimate the cumulative probability that the plant is down during the interval of time $[0, T]$ with $T = 10$.

The choice of good importance sampling measures and parameters is still an open issue. Indeed, in the (discrete) SPN case some general importance sampling schemes are known to give good results (Nicola *et al.*, 1993; Obal II & Sanders, 1994a), but the fluid case is more complicated because rare events can be introduced by parameters on deterministic fluid rates which can not be changed using importance sampling. Thus, a lot of work needs to be done about how to automatically and efficiently implement importance sampling in FSPNs. For our model, rare events may be introduced by the (deterministic) fluid rates (it takes too long to fill place "Two") by taking $r_{in} = \varepsilon$ and $r_{out} = 1.05 - \varepsilon$ as $\varepsilon \to 0$. With our numerical values, $\varepsilon = 0.3$ induces small cumulative Off probabilities. Our way to increase the shut down of the source is by making the failure of the processing station occur more often (another possibility is to make the repair of the processing station to take more time, but it is not possible here as the repair does not follow a stochastic distribution). We change the distribution of transition "Fail" to an exponential law with parameter 8.0. The 95%-confidence intervals (CI) obtained for the cumulative probability that the external source is down when the resampling policy of transition "Time_Idle" is PRD, PRS or PRI are respectively in Tables 1, 2 and 3. We can observe that in each case, even after several hours of simulation, the standard simulation (i.e., without importance sampling) is still not able to obtain the event "the external source is off". Using importance sam-

Meth.	Off probability CI	Time
Std	[0.00e+00,0.00e+00]	15355 sec (10^7 replic.)
IS	[3.378e-12, 9.717e-12]	1974.8 sec (10^6 replic.)

Table 1: Results obtained for standard and importance sampling simulation methods on the dual tank example and PRD policy.

Meth.	Off probability CI	Time
Std	[0.00e+00,0.00e+00]	15441 sec 10^7 replic.)
IS	[2.410e-11,6.030e-11]	1944.2 sec (10^6 replic.)

Table 2: Results obtained for standard and importance sampling simulation methods on the dual tank example and PRS policy.

Meth.	Off probability CI	Time
Std	[0.00e+00,0.00e+00]	15917 sec (10^7 replic.)
IS	[1.567e-12, 4.845e-12]	2050 sec (10^6 replic.)

Table 3: Results obtained for standard and importance sampling simulation methods on the dual tank example and PRI policy.

pling, we are able to obtain this event and a confidence interval.

6 Complications when merging discrete and continuous distributions

In Section 3, the distributions of timed transitions are assumed to be absolutely continuous. We show here that when some distributions are discrete (or a mixture of discrete and continuous distributions), updating the likelihood ratio may become very complex. Assume that several firing of transitions happen at the same time (which can not occur with probability one to transitions with absolutely continuous distributions).

As $\bar{F}_i(T_i, \tau)$ means the probability that, knowing $\boldsymbol{X}_{0,i}$, the firing time of transition τ is *strictly* greater than T_i, let $\bar{F}_i^{(1)}(T_i, \tau)$ be the probability that the firing time of transition τ is *greater than or equal* to T_i (these values were equal for absolutely continuous variables). Let $\mathcal{T}^D$ be the subset of $\mathcal{T}$ composed of the discrete transitions. If one (of N enabled) discrete transition fires, call it τ_i^*, we need to multiply the likelihood by $P_i(T_i, \tau_i^*)/\bar{F}_i^{(1)}(T_i, \tau_i^*)$, which means that τ_i^* fires at time t_i, where P_i is defined as f_i in Equation (1) but represents the (conditional) probability that (discrete) transition τ_i^* fires at t_i. But then we must "specify" that the $N-1$ other competing discrete transitions fire

- either after time T_i;
- or at time T_i but are not chosen as the one firing, according to the weights assigned to each discrete time transition.

Taking into account all the combinations of these two cases for all the $N-1$ transitions (i.e., considering all the subsets G_i of $\mathcal{T}^D \cap E_i - \{\tau_i^*\}$, $P(\boldsymbol{X}_{i,i+1})$ is then multiplied by

$$\sum_{G_i \subset \mathcal{T}^D \cap E_i - \{\tau_i^*\}} \left(\frac{\omega(\tau_i^*)}{\sum_{e \in G_i \cup \{\tau_i^*\}} \omega(e)} \prod_{\tau \in G_i} \frac{P_i(T_i, \tau)}{\bar{F}_i^{(1)}(T_i, \tau)} \right) \prod_{\tau \notin G_i} \frac{\bar{F}_i(T_i, \tau)}{\bar{F}_i^{(1)}(T_i, \tau)} \quad (6)$$

and the same can be done for $P'(\boldsymbol{X}_{i,i+1})$. The problem is that the increments between following events when updating the likelihood ratio are not anymore "independent" in the sense that then we need to remove something from the likelihood before adding a part. As a matter of fact, if for example the next discrete transition (τ_{i+1}^*) is not affected by the firing of τ_i^* and fires immediately after τ_i^* at time t_i, we must remove in Equation (6) the cases (i.e., subsets F_i of $\mathcal{T}^D \cap E_i - \{\tau_i^*\}$) such that the firing of τ_{i+1}^* is after t_i. The same should be done when it is actually known that the transition fires later. It is rather tricky to solve this problem and it has not been implemented in SPNP yet.

7 Conclusion

In this paper we have described and implemented the importance sampling simulation of stochastic Petri nets and fluid stochastic Petri nets. We extend the literature on importance sampling, by incorporating various resampling policies (preemptive resume, preemptive repeat different and preemptive repeat identical). Furthermore, importance sampling had never been applied before to fluid stochastic Petri nets. A fluid example involving resampling policies illustrates the power of the method.

Two major points need to be dealt with to improve our work.

- First, to find a general scheme for importance sampling parameters (as done in (Nicola *et al.*, 1993; Obal II & Sanders, 1994a) for discrete models) when simulating a fluid model with rare events which are induced by fluid (and deterministic) flow rates (see Section 5).
- Next, to find an easy way to compute the likelihood ratio when discrete and continuous distributions are enabled at the same time (as described in Section 6).

References

Ajmone Marsan, M., Balbo, G., Conte, G., Donatelli, S., & Franceschinis, G. 1995. *Modelling with Generalized Stochastic Petri Nets*. John Wiley & Sons.

Choi, H., Kulkarni, V.G., & Trivedi, K.S. 1993. Markov regenerative stochastic petri nets. *Performance evaluation*, **20**(1-3), 337–357.

Ciardo, G., Muppala, J.K., & Trivedi, K.S. 1989. SPNP: Stochastic Petri Net Package. *Pages 142–151 of: Proc. third international workshop on petri nets and performance models, pnpm'89.*

Ciardo, G., Nicol, D.M., & Trivedi, K.S. 1999. Discrete-Event Simulation of Fluid Stochastic Petri-Nets. *IEEE transactions on software engineering*, **25**(2), 207–217.

Fishman, G.S. 1997. *Monte carlo: Concepts, algorithms and applications*. Springer-Verlag.

Glynn, P. W., & Iglehart, D. L. 1989. Importance Sampling for Stochastic Simulations. *Management science*, **35**(11), 1367–1392.

Glynn, P.W. 1989. A GSMP Formalism for Discrete Event Systems. *Pages 14–23 of: Proceedings of the IEEE*, vol. 77.

Goyal, A., Shahabuddin, P., Heidelberger, P., Nicola, V.F., & Glynn, P.W. 1992. A Unified Framework for Simulating Markovian Models of Highly Dependable Systems. *IEEE transactions on computers*, **41**(1), 36–51.

Heidelberger, P. 1995. Fast Simulation of Rare Events in Queueing and Reliability Models. *ACM transactions on modeling and computer simulation*, **5**(1), 43–85.

Hirel, C., Tuffin, B., & Trivedi, K.S. 2000. SPNP Version 6.0. *Pages 354–357 of:* LNCS, vol. 1786. Springer Verlag.

Horton, G., Kulkarni, V., Nicol, D., & Trivedi, K.S. 1998. Fluid Stochastic Petri Nets: Theory, Application and Solution. *European journal of operational research*, **105**, 184–201.

Nakayama, M. K. 1996. General Conditions for Bounded Relative Error in Simulations of Highly Reliable Markovian Systems. *Advances in applied probability*, **28**, 687–727.

Nicola, V. F., Nakayama, M. K., Heidelberger, P., & Goyal, A. 1993. Fast Simulation of Highly Dependable Systems with General Failure and Repair Processes. *IEEE transactions on computers*, **42**(12), 1440–1452.

Obal II, W.D., & Sanders, W.H. 1994a. An Environment for Importance Sampling Based on Stochastic Activity Networks. *Pages 64–73 of: Proceedings of the 13th symposium on reliable distributed systems*.

Obal II, W.D., & Sanders, W.H. 1994b. Importance Sampling Simulation in *ultrasan*. *Simulation*, **62**(2), 98–111.

Shahabuddin, P. 1994. Importance Sampling for the Simulation of Highly Reliable Markovian Systems. *Management science*, **40**(3), 333–352.

Trivedi, K. S., & Kulkarni, V. G. 1993. FSPNs: Fluid Stochastic Petri Nets. *Pages 24–31 of: 14th international conference on applications and theory of petri nets*.

Tuffin, B. 1999. Bounded Normal Approximation in Simulations of Highly Reliable Markovian Systems. *Journal of applied probability*, **36**(4), 974–986.

Tuffin, B., & Trivedi, K.S. 2000. Implementation of Importance Splitting Techniques in Stochastic Petri Net Package. *Pages 216–229 of:* LNCS, vol. 1786. Springer Verlag.

Biographies

Bruno Tuffin is an INRIA-researcher at IRISA, Rennes, France. He received his Ph.D. in applied Mathematics from Rennes 1 University in 1997. His areas of interest include simulation methods (Monte Carlo, quasi-Monte Carlo...) and their applications in telecommunication. In 1999, He visited Pr. Trivedi's group for 8 months to work on Stochastic Petri Nets, Fluid Stochastic Petri Nets and their analysis using simulation.

Kishor S. Trivedi received the B.Tech. degree from the Indian Institute of Technology (Bombay), and M.S. and Ph.D. degrees in computer science from the University of Illinois, Urbana-Champaign. He holds the Hudson Chair in the Department of Electrical and Computer Engineering at Duke University, Durham, NC. He also holds a joint appointment in the Department of Computer Science at Duke. He is the Duke-Site Director of an NSF Industry-University Cooperative Research Center between NC State University and Duke University for carrying out applied research in computing and communications. He has been on the Duke faculty since 1975. He has served as a Principal Investigator on various AFOSR, ARO, Burroughs, DARPA, Draper Lab, IBM, DEC, Alcatel, Telcordia, Motorola, NASA, NIH, ONR, NSWC, Boeng, Union Switch and Signals, NSF, and SPC funded projects and as a consultant to industry and research laboratories. He was an Editor of the *IEEE Transactions on Computers* from 1983-1987. He is a co-designer of HARP, SAVE, SHARPE, SPNP, and SREPT modeling packages. These packages have been widely circulated. He is the author of a well known text entitled, *Probability and Statistics with Reliability, Queuing and Computer Science Applications*, published by Prentice-Hall. He has recently published two books entitled, *Performance and Reliability Analysis of Computer Systems,* published by Kluwer Academic Publishers and *Queueing Networks and Markov Chains,* John Wiley. His research interests are in reliability and performance assessment of computer and communication systems. He has published over 250 articles and lectured extensively on these topics. He has supervised 33 Ph.D. dissertations. He is a Fellow of the Institute of Electrical and Electronics Engineers. He is a Gloden Core Member of IEEE Computer Society.