On Markov Reward Modelling with FSPNs

Katinka Wolter

Technische Universität Berlin,
Institut für Technische Informatik,
Franklinstr. 28/29
10587 Berlin, Germany.
katinka@cs.tu-berlin.de

Andrea Zisowsky

Technische Universität Berlin,
Fachbereich Mathematik,
Strasse des 17. Juni 136,
10623 Berlin, Germany.
zisowsky@math.tu-berlin.de

September 15, 1999

Abstract

In this paper fluid stochastic Petri nets (FSPNs) will be used for modelling reward in a performability model. Two variations of a known performability model are presented. In the first one two fluid places are used, one of which represents the profit (reward) obtained by operating the system and the other one the buffer, that is approximated continuously. The continuous approximation yields randomly varying flow rates, whereas the accumulation of reward is modelled at a constant flow rate. Hence, the model is a second order FSPN with respect to one fluid place and a first order FSPN with respect to the second fluid place. In the second model only one fluid place is used, which represents the costs (negative reward) arising due to repair. The costs increase continuously at deterministic rate while the system is in state of repair (which is a rate reward in the model). Additional costs incur each time the buffer fails (which is an impulse reward in the model). With a numerical solution algorithm the distribution of the reward and its mean are computed.

1 Introduction

Fluid Petri net models have gained much attention over the last years and a number of different classes of hybrid Petri nets have been proposed. For the specification and qualitative analysis the autonomous non-timed and timed Petri nets (ACPNs) [DA93] have been used. Performance and dependability evaluation has been carried out with fluid stochastic Petri nets (FSPNs) [TK93, HKNT98]. The discrete part of an FSPN taken on its own is a generalized stochastic Petri net (GSPN) [ABCD95] and the continuous part alone is equivalent to a fluid queueing system [New71, Kob74]. In FSPNs the continuous model part is represented by fluid places and fluid arcs. The level in the fluid places changes at deterministic rate. The stochastic behaviour of the model is determined by the discrete part and the dependencies between discrete and continuous submodel. In this paper a specific class of FSPNs is considered, the second order FSPNs. In second order FSPNs the fluid places are filled and emptied at normally distributed rate. Accordingly, the continuous model part has probabilistic behaviour too. The underlying equations are second order partial differential equations, hence the prefix *second order*. Second order FSPNs can, however, in some discrete states reduce to first order FSPNs, if in these states the variances of the fluid flow are zero. Then the second order derivative vanishes from the differential equation and a first order partial differential equation (PDE) corresponding to an ordinary (first order) FSPN is left. If in all discrete states the variances of the fluid flow are zero, the whole model reduces to a first order FSPN.

In [Hor96] FSPNs have been used for modelling accumulated rate reward. Reward structures are added to Markov models, giving Markov reward models. The measures computed from a Markov reward model are the accumulated reward and the completion time. Both are random variables following an unknown probability distribution. In this paper only accumulated reward is regarded. Two kinds of rewards can be distinguished, rate reward and impulse reward. Rate reward is the reward obtained while the model is in a certain state. Impulse reward is achieved with the completion of an action, that is with the transition from one model state to the next. Both can be instantaneous transient or steady state rewards, which is the value of the reward at some point in time, or accumulated reward, the reward that has accumulated over a period of time [SM91].

The mean accumulated reward is computed through integration over the state probabilities of the Markov model multiplied with some reward rate for each state. In states where no reward is obtained the reward rate equals zero. In the computation of the mean impulse reward the state transition rate, that gives the impulse is a factor in the reward rate. The moments of the distribution of the accumulated reward are obtained by solving a differential equation using Laplace transforms. But obtaining higher moments of the accumulated reward distribution is difficult [TR99]. It has been shown in [Hor96] that the distribution of accumulated reward is computed by solving a differential equation like those underlying an FSPN.

Accumulated reward can be modelled with an FSPN in a very natural way. A fluid place holds the accumulated reward. In each state where rate reward is accumulated, the fluid place is filled at the reward rate. The main advantage of modelling reward with a fluid model is, that the full probability density is obtained automatically and from it the moments can be computed.

In this paper not only rate reward but also also impulse reward is modelled with an FSPN, which has not been done before. Impulse reward is accumulated if with a discrete state transition a non-zero amount of reward is added to the reward in the fluid place. In the FSPN a special kind of arcs, the jump arcs are introduced to model the instantaneous increase or decrease of the level in a fluid place. The jump in the fluid level happens with the firing of the transition at the arcs origin or destination, the height of the jump is determined by a probability distribution. The Dirac impulse gives a fixed height for each jump.

Jump arcs are similar to the flush-out arcs in [BGG+99, GSB99]. But flush-out arcs always empty a fluid place, whereas jump arcs remove or deposit a variable amount of fluid and hence allow for more modelling flexibility. In [BGG+99, GSB99] multidimensional fluid models are solved with an algorithm especially implemented for this specific problem. The numerical algorithms used for the computations in this paper are more general in that they can solve any FSPN with one or two fluid places. Numerical solution of second order FSPNs with two fluid places has for the first time been presented in an earlier version of this paper [Wol99].

Two variations of the performability model taken from [Mey82] will be studied. The system that is modelled is a buffered multi-processor system, where both the buffer and the processors can fail. The first model has two fluid places, one representing the buffer content and the other one holding the accumulated reward. The second model has only one fluid place representing the costs arising from the repair of system components (negative reward). As long as repair is going on inflow to the fluid place takes place. Each failure of the buffer causes additional costs once. So this is modelled with a jump arc along which a fixed amount of fluid is deposited. An earlier version of this paper is [WZ99].

In the next section the second order FSPN formalism is introduced. In section 3, the main part of the paper, the two models will be discussed.. For both models the underlying differential equations are given. In the algorithm used for their solution the equations are generated automatically from a (textual) description of the Petri net. The transient evolution of the probability density is given as a result for both examples. With section 4 the paper concludes.

2 The FSPN Formalism

For the definition of second order FSPNs the common notations for GSPNs [ABCD95] and for FSPNs [HKNT98] are taken over and extended as follows.

A second order FSPN is formally defined as a 9-tuple FSPN $= (\mathcal{P}, \mathcal{T}, \mathcal{A}, m_0, g, \lambda, r, b, w)$ where the set of places ($\mathcal{P} = \mathcal{P}_c \cup \mathcal{P}_d$) is divided into the *fluid (continuous)* and the *discrete* places. Discrete places (the elements of $\mathcal{P}_d$) are drawn as single lined circles and hold an integer number of *tokens*, whereas the continuous, or fluid, places (the elements of $\mathcal{P}_d$) are drawn as two concentrical circles and they hold a real-valued amount of *fluid*. The set of transitions $\mathcal{T} = \mathcal{T}_E \cup \mathcal{T}_I$ is composed of the exponentially distributed, and the immediate transitions, respectively. Exponentially distributed transitions are drawn as empty rectangles and the immediate transitions as thin bars.

The set of arcs $\mathcal{A} = \mathcal{A}_d \cup \mathcal{A}_c \cup \mathcal{A}_j$ is divided into three subsets, the discrete arcs (the elements of $\mathcal{A}_d$), the continuous, or fluid, arcs ($\mathcal{A}_c$) and the jump arcs ($\mathcal{A}_j$). The discrete and the jumps arcs are drawn as single lined arrows, whereas the fluid double-lined arcs are drawn like pipes. The jump arcs and the continuous arcs only connect transitions and fluid places, and vice versa. Immediate transitions can be connected to a fluid place by a jump arc only if the immediate transition is also connected to a discrete place. In other words, the fluid jumps caused by the firing of an immediate transition can not occur without a discrete model part connected to that same immediate transition. This restriction is imposed because otherwise the model cannot be mapped on a Markov process.

The *marking* $\mathbf{m} = (s, \mathbf{z}) \in \mathcal{M}$ consists of a discrete part $s = (\#p_i, i \in \mathcal{P}_d)$, where $\#p_i$ denotes the number of discrete tokens in the i-th discrete place p_i, and a continuous part, a vector representing the fluid level in each fluid place, $\mathbf{z} = (z_k, k \in \mathcal{P}_c) \in \mathcal{Z}$. The initial marking is $\mathbf{m}_0 = (s_0, \mathbf{z}_0)$. The total number of discrete states is S, the set of all discrete states is denoted $\mathcal{S}$.

$\mathrm{r}_{t,p}$, $\mathrm{r}_{p,t}$: $\mathcal{A}_c \times \mathcal{M} \rightarrow \mathbb{R}^2$ is the flow rate function along the fluid arc connecting the timed transition t and the fluid place p or vice versa. The flow rate $r = (\mu, \sigma^2)$ is a normally distributed random variable specified by expectation μ and variance σ^2. Let Z_k be the fluid level in place $k \in \mathcal{P}_c$. The change rate of the fluid level is again a normally distributed random variable with

$$\frac{dZ_k}{dt} = \sum_{t \text{ enabled in } \mathbf{m}} r_{t,p}(\mathbf{m}) - \sum_{t \text{ enabled in } \mathbf{m}} r_{p,t}(\mathbf{m}).$$

Reflecting barriers assure, that the fluid level in each place stays within its range $[z^{\min}, z^{\max}]$.

The jump height function $\mathrm{b}_{t,p}$, $\mathrm{b}_{p,t}$: $\mathcal{A}_j \times \mathcal{M} \rightarrow \mathcal{F}$ is a probability density function associated to a jump arc. The size of a fluid jump is a random variable drawn from that distribution.

g_t : $\mathcal{M} \rightarrow \mathbb{B}$ is the guard of transition t that can be a function of the discrete and the continuous state. λ_t : $\mathcal{M} \rightarrow \mathbb{R} \cup \{\infty\}$ is a function of both the continuous and the discrete marking. Immediate transitions have firing rate $\lambda_t = \infty$.

The weight function w_t : $\mathcal{S} \rightarrow \mathbb{R}$ is defined for immediate transitions. The firing probability of each of the enabled immediate transitions in a vanishing state is $w_t(s) / \sum_{t_i \text{enabled in } \mathbf{s}} w_{t_i}(s)$.

The enabling rules for the transitions are the same as in discrete Petri net models. The fluid and the jump arcs act only on the fluid places and do not influence the enabling conditions of the transitions, other than through a guard function.

2.1 The Model Parameters

The dynamics of the CTMC underlying the discrete model part are defined through the generator matrix $\mathbf{Q}$, whereas the continuous dynamics are defined by the fluid change rate $\frac{dZ_k}{dt}$ that describes the change in the fluid level over time. The change in the fluid level is determined by the difference between the inflow and the outflow, each of which is assumed to be normally distributed or deterministic, if the variance of the normal distribution equals zero. The parameters of the normal distribution are either specified by the user or calculated as the diffusion approximation to the discrete transition at the origin of the inflow arcs and at the destination of the outflow arcs [Wol97].

Let μ_a and σ_a^2 be the expectation and variance of an arbitrary inter arrival time distribution of the discrete input transition to the fluid place P_k. Then the infinitesimal mean and variance of the fluid inflow rate are

$$\mu_\alpha = \frac{1}{\mu_a}, \qquad \sigma_\alpha^2 = \frac{\sigma_a^2}{(\mu_a)^3}, \tag{1}$$

and for the fluid outflow rate μ_δ and σ_δ^2 are analogous. (See [Kle76, New71] for a derivation of this formula).

All the incoming and outgoing streams in one marking add and for the flow rate $R_k(s, \mathbf{z})$ of the fluid place k in every marking $s \in \mathcal{S}$ the following parameters are valid

$$\mu_{k,s} = \sum_{\substack{\text{inflows} \\ \text{to k in s}}} \mu_\alpha - \sum_{\substack{\text{outflows} \\ \text{of k in s}}} \mu_\delta \quad \forall k \in \mathcal{P}_c, s \in \mathcal{S}, \tag{2}$$

$$\sigma_{k,s}^2 = \sum_{\substack{\text{inflows} \\ \text{to k in s}}} \sigma_\alpha^2 + \sum_{\substack{\text{outflows} \\ \text{of k in s}}} \sigma_\delta^2 \quad \forall k \in \mathcal{P}_c, s \in \mathcal{S}, \tag{3}$$

because variances must add. Note, that this simple linear combination only holds, because all the single fluid streams are normally distributed. The expectations and variances for each fluid place are collected into diagonal matrices (if there is none in a discrete state, the entry is equal to zero):

$$\boldsymbol{M}_k = \mathbf{diag}(\mu_{k,1}, \ldots, \mu_{k,S}), \tag{4}$$

$$\boldsymbol{\Sigma}_k^2 = \mathbf{diag}(\sigma_{k,1}^2, \ldots, \sigma_{k,S}^2). \tag{5}$$

If a hybrid system is modelled $\mu_\alpha, \sigma_\alpha^2, \mu_\delta$ and σ_δ^2 are given by the user and (1) is not needed.

The jump height function $b(.)$ can be an arbitrary probability density. For all discrete states the functions are collected into a diagonal matrix $\mathbf{B} = \mathbf{diag}(b_1(.), \ldots, b_S(.))$.

3 The Models

Two models shall be studied in this paper. Both are based on the performability model of a buffered multiprocessor system presented in [Mey82]. The first one uses two fluid places, one for the buffer content and the other for accumulating reward, that flows in as long as servers and buffer are operational. The second model has only one fluid place, that represents the costs arising from repair of processors and buffer. As in the original reference in the first model processors and buffer may fail but no repair is considered. Opposed to this in the second model both processors and buffer can be repaired.

Reward increases at constant speed in each discrete state. Therefore the inflow rate to the fluid places representing the accumulated reward has zero variance. This fluid place is equivalent to a fluid place in a first order FSPN. In the first model the buffer is represented by a fluid place which is emptied and filled at randomly varying rate. So, the first model is a second order FSPN in all states that have an in- or outflow of the buffer, whereas the second model purely is a first order FSPN with fluid jumps.

3.1 Two-dimensional Fluid Rate Reward Model

At first the regarded model will be described in detail. Then the corresponding differential equations and their solutions are explained. The regarded performability model has three processors that may fail independently at rate λ_p. The buffer has L stages that can fail independently at rate λ_b. If one of the stages fails, it causes a crash of the complete system. The failure rate of the whole buffer is therefore $L\lambda_b$. Jobs arrive to the system at

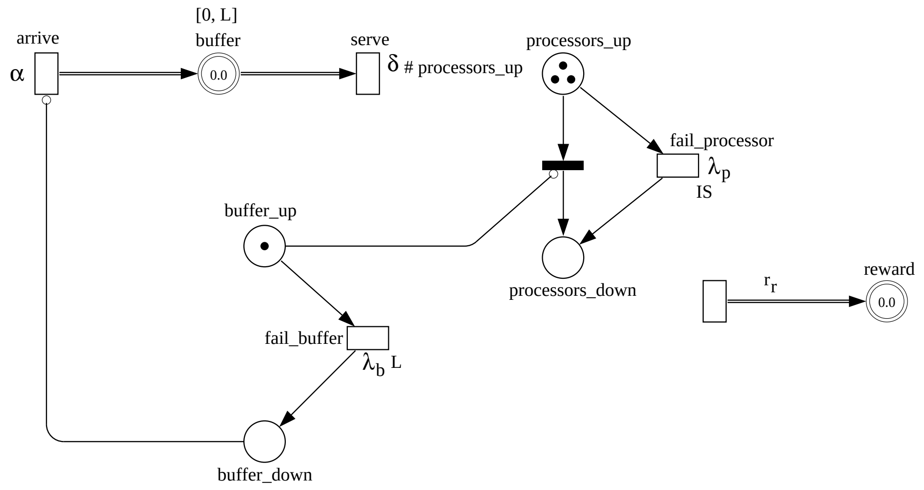

Figure 1: FSPN model of the performability model

rate α and they are served at rate δ per processor. So the buffer is being filled as long as it is operational even if all servers are down and there is an outflow of the buffer as long as at least one processor is functioning. Reward is accumulated at rate r per operational processor. The reward inflow r_r is a function of the global state of the system and is specified for each discrete state of the model.

In this model the buffer is considered as being filled with a large number of items, that are measured in some unit. So the arrival and service processes can be seen as continuous streams entering and leaving the buffer. The inflow and outflow rates of the fluid place representing the buffer are approximations computed from the corresponding discrete streams.

Figure 1 shows an FSPN model of the system. The buffer is represented by the one fluid place. Its capacity is $L = 4$. When the buffer is down all processors are put into the set of failed processors and the system terminates because none of the transitions is enabled. Complete failure of the system is an absorbing state. The second fluid place represents the accumulated reward. The rate function r_r is defined in such a way, that there is an inflow of 0.1 units per operational processor and per time unit only if the buffer is not empty. Hence the fluid flow rate accumulating reward depends on the content of the buffer.

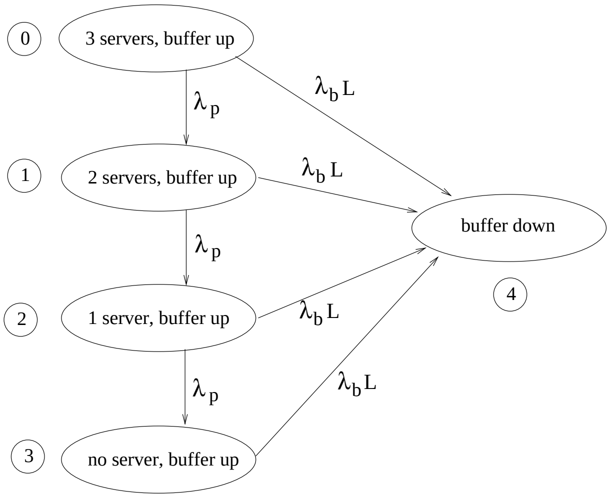

Figure 2: Reachability graph of the FSPN model

The reachability graph is shown in figure 2. Only the discrete states are labelled, the components for the two fluid levels are omitted. The numbers in the small circles denote the numbering of the discrete states as it is used for the parameter vectors and the results. The failure rate of the buffer has been set to $\lambda_b = 10^{-4}$ per hour and the processor failure rate is $\lambda_p = 10^{-3}$ per hour. The mission time is 80 hours.

For the computation of the fluid change rates (1), (2) and (3) have been used. In all states s where the buffer is operational the expectation μ_b of the fluid change rate of the buffer is computed following (2):

$$\mu_b = \mu_\alpha - \#\text{servers\_up} \cdot \mu_\delta$$

and the variance σ_b^2 equals according to (3)

$$\sigma_b^2 = \sigma_\alpha^2 + \#\text{servers\_up} \cdot \sigma_\delta^2.$$

For the exponential distribution (1) evaluates to $\mu_\alpha = \alpha$ and $\sigma_\alpha^2 = \alpha$. The same holds for the exponentially distributed service with rate δ. With $\alpha = 0.75$ and $\delta = 0.5$ the fluid flow parameters evaluate to

$$\boldsymbol{M}_b = \mathbf{diag}(-0.75, -0.25, 0.25, 0.75, 0.0) \tag{6}$$

and

$$\boldsymbol{\Sigma}_b^2 = \mathbf{diag}(2.25, 1.75, 1.25, 0.75, 0.1). \tag{7}$$

Reward is accumulated at constant rate of 0.1 per operational processor. Hence, the fluid flow parameters for the second fluid place are

$$\boldsymbol{M}_r = \mathbf{diag}(0.3 \cdot r_{rew}, 0.2 \cdot r_{rew}, 0.1 \cdot r_{rew}, 0.0, 0.0) \tag{8}$$

with

$$r_{rew} = \begin{cases} 1.0 & \text{if } z_1 \geq 0.2 \\ 0.0 & \text{else} \end{cases}$$

and $\mathbf{\Sigma}_b^2$ is the matrix with all entries equal to zero.

The reachability graph gives a graphical representation of the stochastic process underlying this model. Augmenting it with two continuous variables, one for each fluid place, gives a description of the full stochastic process which is formally defined as

$$\{(S(t), Z_1(t), Z_2(t)), t \in \mathrm{IR}_0^+\},$$

where $S(t)$ is the discrete marking at time t, and $Z_k(t)$ is a random variable, representing the fluid level in fluid place k at time t.

At time t the transient probability of the model being in discrete state i with fluid level in an infinitesimal environment around (z_1, z_2), is called the *fluid density* and is denoted by $\pi_i(t, z_1, z_2) = \frac{\partial}{\partial z_1} \frac{\partial}{\partial z_1} P(S(t) = i, Z_1(t) \leq z_1, Z_2(t) \leq z_2)$. This density is the component wise solution of the following differential equation in vector form

$$\frac{\partial \boldsymbol{\pi}(t, z_1, z_2)}{\partial t} + \frac{\partial}{\partial z_1}\Big(\boldsymbol{\pi}(t, z_1, z_2)\mathbf{M}_b(z_1)\Big) + \frac{\partial}{\partial z_2}\Big(\boldsymbol{\pi}(t, z_1, z_2)\mathbf{M}_r(z_2)\Big) - \frac{1}{2}\frac{\partial^2}{\partial z_1^2}\Big(\boldsymbol{\pi}(t, z_1, z_2)\mathbf{\Sigma}_b^2(z_1)\Big)$$
$$- \boldsymbol{\pi}(t, z_1, z_2)\mathbf{Q}(z) = 0 \forall t, \forall z \in \mathring{\Omega} = (z_1^{\min}, z_1^{\max}) \times (z_2^{\min}, z_2^{\max}). \tag{9}$$

Note that there is only one matrix of variances while there are two matrices of means, one for each continuous direction. Eqn. (9) is a system of diffusion-convection equations, also called Fokker-Planck equations. The system of equations is coupled through the generator matrix of the discrete Markov process $\mathbf{Q}$. The expectations $\mathbf{M}$ define the drift and the variances $\mathbf{\Sigma}^2$ define the diffusion.

An initial as well as a boundary condition on either end of the regarded domain must be imposed. The initial condition is defined by the model's configuration at time zero, whereas the boundary conditions are determined by the fact that the integral over a density always evaluates to one.

Let $\boldsymbol{\pi}_0$ be the vector of initial discrete state probabilities and let $\mathbf{z}_0$ be the vector of initial fluid levels, then the inital conditions are

$$\boldsymbol{\pi}(0, z_1, z_2) = \boldsymbol{\pi}_0 \cdot \delta(z_1 - z_{1,0})\delta(z_2 - z_{2,0}), \tag{10}$$

where $\delta(z_k - z_{k,0})$ is the Dirac impulse at $z_k = z_{k,0}$.

The boundary condition is equivalent to a reflecting barrier at the origin and at the upper bound [CM65]

$$\sum_{k=1}^{2}\left[\frac{1}{2}\frac{\partial}{\partial z_k}\boldsymbol{\pi}(t,z_1,z_2)\boldsymbol{\Sigma}_k^2(z_1,z_2)-\boldsymbol{\pi}(t,z_1,z_2)\mathbf{M}_k(z_1,z_2)\right]_{z=\{z_1^{\min},z_1^{\max}\}\times\{z_2^{\min},z_2^{\max}\}}=0. \tag{11}$$

Equation (9) subject to (10) and (11) has been discretized and solved numerically [Zis98]. A theta method has been applied, that is a flexible discretization having the fully implicit, explicit and the Crank-Nicolson scheme as special cases. For the convection term an upwind scheme has been used, the coupling term is always discretized explicitly. Both models presented here have been solved with the fully implicit discretization.

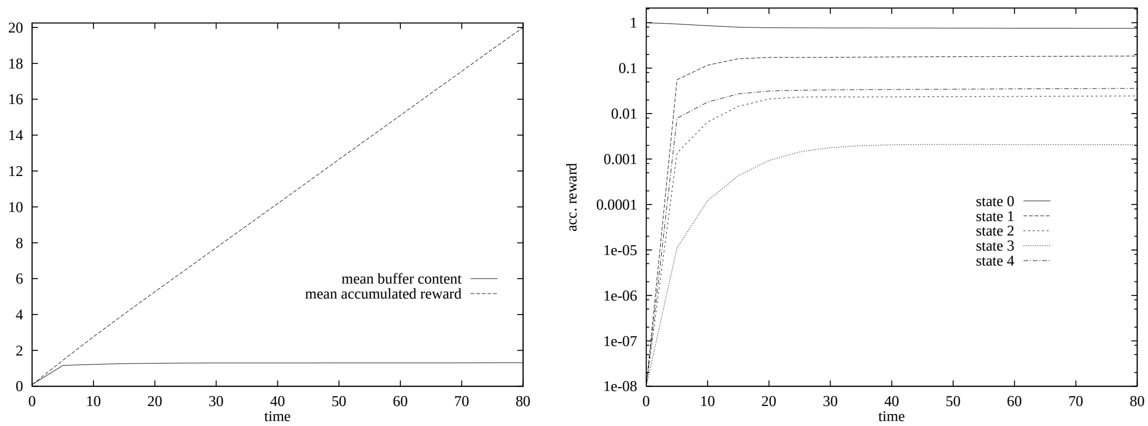

Figure 3: Means over both fluid places and discrete state probabilities

The result after applying the numerical algorithm is in each each discrete state a two-dimensional density, that evolves over time.[1] The sum over all discrete states is a two-dimensional probability density at each time point. The discrete state probabilities are obtained by integrating over both continuous dimensions for one discrete state at each point in time. The state probabilities for all discrete states are shown in figure 3 on the right in logarithmic scale. On the left in figure 3 the means for both fluid places are plotted. The accumulated reward always increases, whereas the buffer content quickly reaches steady state.

The change in the two-dimensional probability densities can be observed like a film. Figure 4 shows the probability densities in different discrete states at selected time points. Each row holds the plots for one discrete state, whereas each column is for one time point. It can be seen that in state 2 (one server, buffer up) the accumulated reward increases, as

[1] Strictly speaking this is a density only after normalization. Nonetheless, we will call the fractions of a density also density.

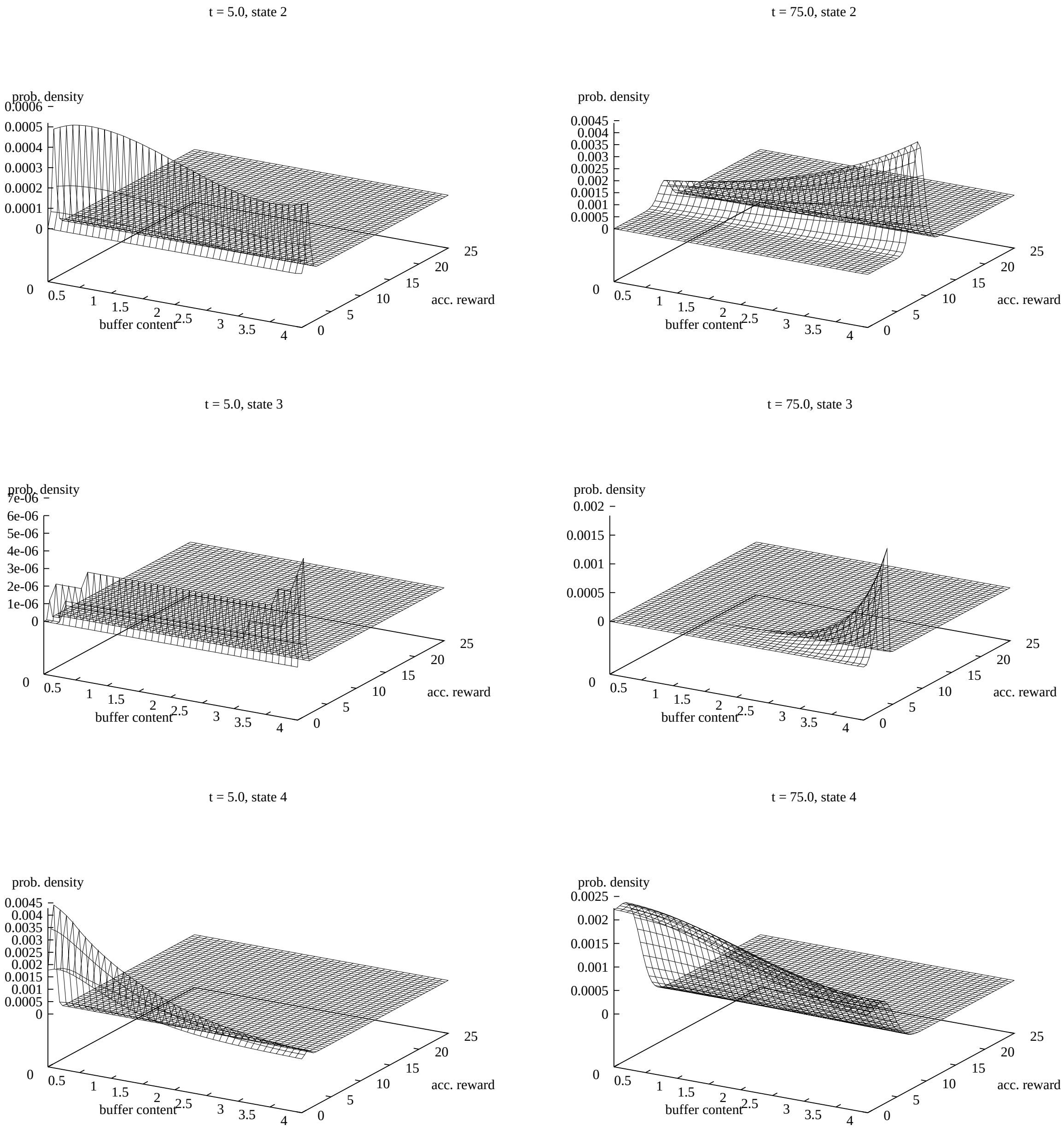

Figure 4: Transient probability density for selected discrete states

does the probability of the buffer being full. In both states 3 (no processor, buffer up) and 4 (all processors failed, buffer down) not much reward is accumulated and as long as the buffer is up, it runs full, whereas when it is down it tends to be rather empty. One does not see exactly the behaviour the system would show, was it in that state for a long time, because failures can happen at any time with reward having been already accumulated and the buffer being already filled. Therefore even in states where buffer and processors are down one will have a (though small) probability of an accumulation of reward and of the buffer not being empty.

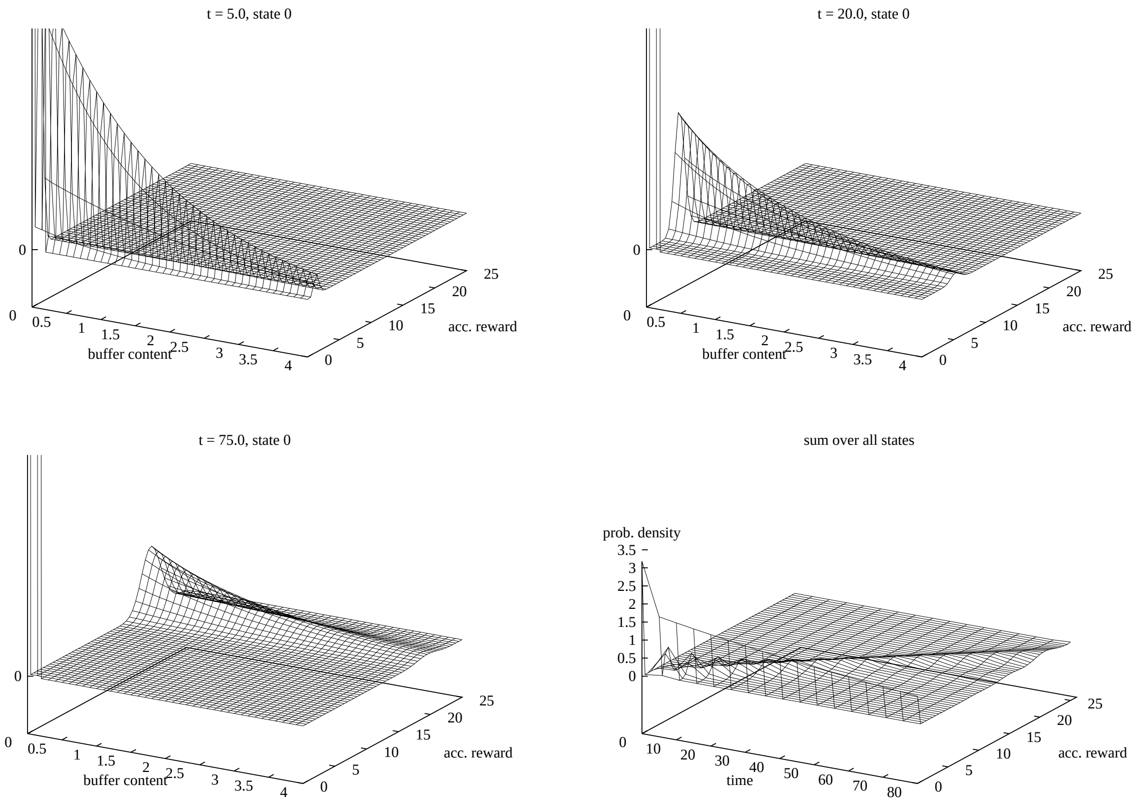

Figure 5: Transient probability density state zero and marginal density for $z2$

Figure 5 shows the evolution of probability mass in state 0 and in the lower right plot the marginal density for the accumulated reward, summed over all discrete states. In the probability density in state 0 there is a peak at zero for buffer content and accumulated reward. This peak results from the definition of the reward rate. The reward rate is defined in such a way, that if the buffer is empty no reward is accumulating. In addition, the initial state is state 0 and the initial density is concentrated in the origin. The reward rate makes sure that there always remains a probability of empty buffer and no reward. This probability mass is not transferred to the other discrete states in recognizable quantity because the fluid flow rates are extremely large compared with the discrete transition rates. The peak is still clearly visible in the marginal density. But it constantly decreases with time. Note that the z-axis is scaled differently in all plots. In those plots where the z-axis is not labelled, its range is $[0.0, 0.2]$.

The density of the accumulated reward is obtained by integrating at each time point over the buffer content, the second continuous dimension. The result has one dimension less, so it can be plotted over time. For early time points one can see oscillations in the solution. But the oscillations disappear with time. This shows a property of the diffusion equation, which is, that its solution becomes smoother over time. The oscillations happen because for early time points the stepsizes should be chosen smaller. Choosing smaller stepsizes for the whole computations would increase the computational effort and runtime. Therefore an automatic stepsize adaptation would be required to avoid such problems.

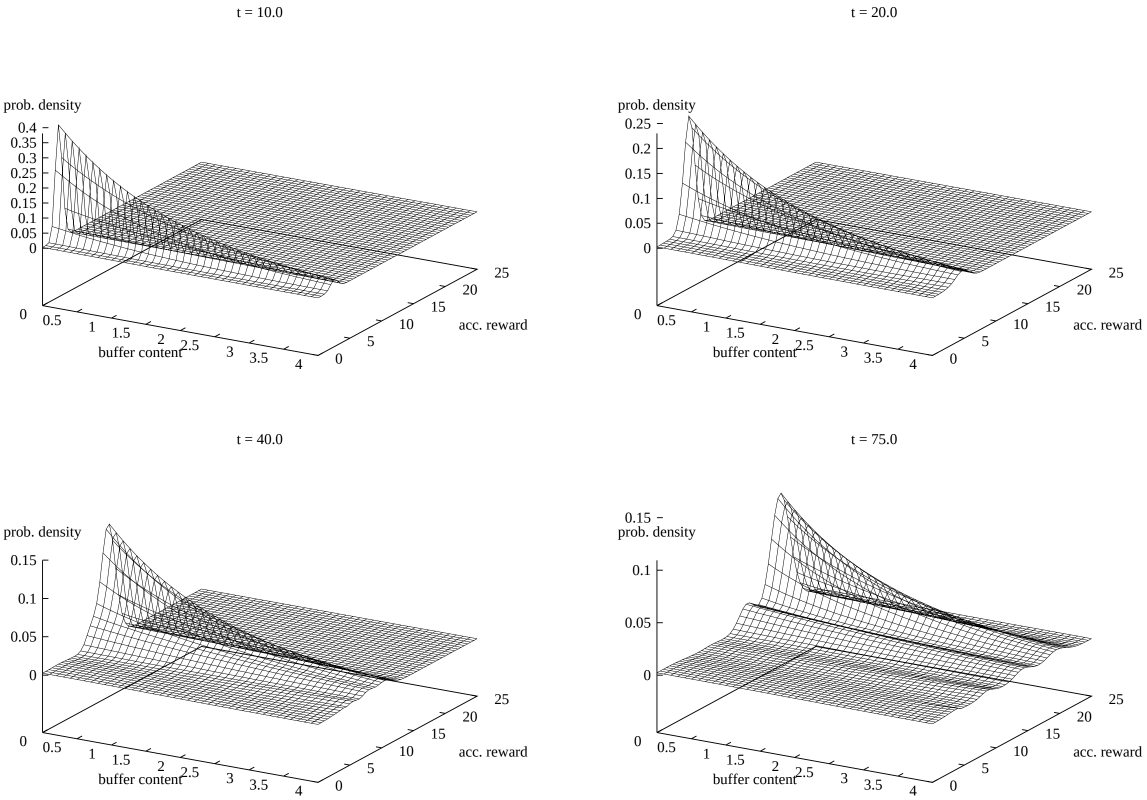

Figure 6: Transient probability density over all discrete states

The plots in figure 6 show the evolution of the overall probability density with slightly manipulated data. The peak at the origin has been removed because it dominated the plots so that the rest of the surface seemed to be completely flat. After 75 time units one can clearly see the summation of all the single waves resulting from the densities in the different discrete states. The highest one is due to the probability mass in state 0, which holds still 75 % of the total probability mass.

So, from this model the probability distribution for the accumulated reward in all discrete states and for the sum of all discrete states as well as the transient marginal densities, the discrete state probabilities and the means over the continuous dimensions have been obtained.

3.2 Rate and Impulse Reward Model

The second example regarded in this paper is another variation of the performability model. The buffer is modelled with a discrete place so there is only one fluid place. This fluid place represents the costs that arise from repair of the processors as well as the buffer. Each failure of the buffer costs an additional price once, whereas the repair costs per unit of time. The failure rate of the processors is again $\lambda_p = 0.001$ per processor and the repair rate is set to $c_p = 0.25$. The failure rate of the buffer has been increased by two orders of magnitude and is now set to $\lambda_b = 0.04$. This has been done because otherwise the effect of the fluid jumps is not visible in the solutions at all. For the same reason the repair rate of the processors has been chosen rather low. For realistic failure and repair rates there is hardly any accumulation of costs, because the probability of the system being in a failed state is extremely low. Repair of the buffer takes on the average 6 hours ($r_b = 1/6$).

The arrival rate to the buffer is $\alpha = 0.75$ and the service rate is $\delta = 0.5$ per operational processor, as in the first model. The costs that are incurred by the repair of processors and buffer are deterministic at rate 0.1 per failed processor and rate 0.5 for the repair of the buffer and maintenance of the whole system. Each failure of the buffer costs an additional deterministic amount of 0.5 units. Hence, the inflow to the fluid place representing the costs has only a mean, but no variance.

The buffer has only $L = 2$ stages. The failure rate of the whole buffer is again $L\lambda_b$ and the repair rate of the whole system after a buffer failure is r_b. Upon failure of the buffer data is saved, so that the system can continue processing after repair. With the repair of the buffer the whole system is checked, so that subsequently all processors are up again.

Figure 7 shows an FSPN model of the system. The costs are represented by the fluid place. A jump arc connects transition *fail_buffer* with the place *costs*. When the buffer is down all the failed processors are put back into the set of operational processors. As long as repair of the buffer takes all the processors are in maintenance and repair, and the transition representing the processor failure is disabled as is the arrival of jobs. The main difference between this model and the first one is that the fluid place is not only filled through the fluid arcs but also through a jump arc. The reachability graph of this model is shown in figure 8. It can be divided in three blocks, each of which represents the system with a different number of tokens in the buffer. Transitions between the blocks are due to arrivals to the buffer and the service of tokens from the buffer. To make the figure a little clearer these transitions are drawn with dashed arcs and have double arrows to represent two arcs. Again the arcs are only labelled with a description of the discrete

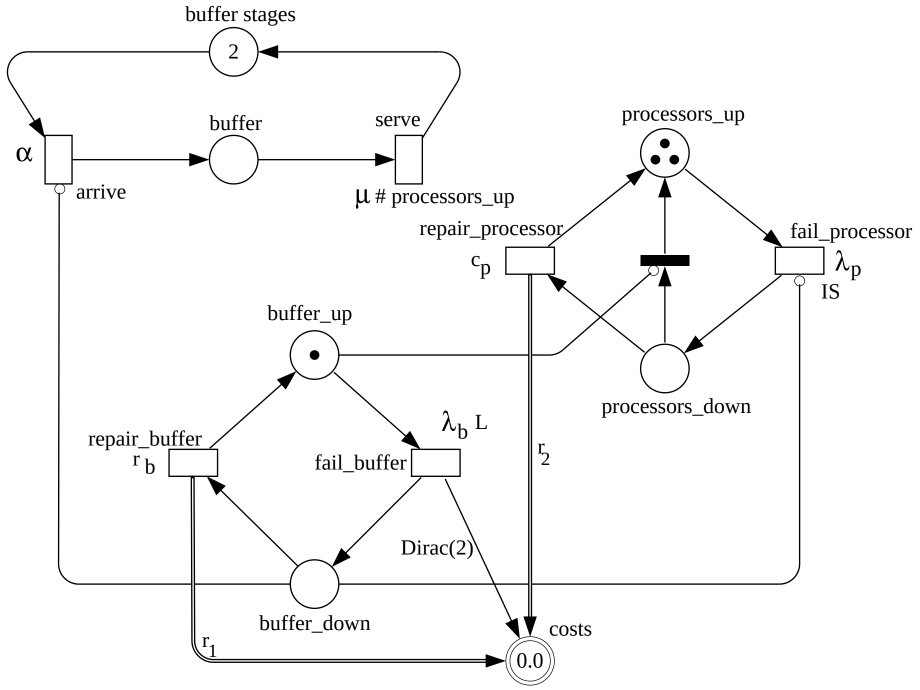

Figure 7: FSPN model of the second performability model

state and the continuous variable is omitted. The numbers in the small circles are the numbering used for the results. The model is solved transiently until time 170.0.

The fluid flow parameters are the same in all three blocks. There only is an inflow to the fluid place, so (2) evaluates to

$$\mathbf{M}_b = \mathbf{diag}(\mu_{r_1} + \mu_{r_2}) = \mathbf{diag}(\boldsymbol{\mu}, \boldsymbol{\mu}, \boldsymbol{\mu}) \quad \text{with} \quad \boldsymbol{\mu} = (0.0, 0.1, 0.2, 0.3, 0.5) \tag{12}$$

and ${\mathbf{\Sigma}^2}_b = \mathbf{0}$. The jump height matrix is

$$\mathbf{B} = \mathbf{diag}(\mathbf{b}, \mathbf{b}, \mathbf{b}) \quad \text{where} \quad \mathbf{b} = (det(0.5), det(0.5), det(0.5), det(0.5), 0.0) \tag{13}$$

and finally the jump rates are

$$\mathbf{\Lambda} = \mathbf{diag}(\boldsymbol{\lambda}, \boldsymbol{\lambda}, \boldsymbol{\lambda}) \quad \text{where} \quad \boldsymbol{\lambda} = (0.04, 0.04, 0.04, 0.04, 0.0). \tag{14}$$

The jump height distribution $det(x)$ denotes the deterministic jump height, implemented by the Dirac impulse at x.

The stochastic process under consideration is a Markov process similar to the one in the first example with only one continuous component. The fluid density in state i is $\pi_i(t, z) = d/dz\ P(S(t) = i, Z(t) \leq z)$ and the equation describing the models dynamics is

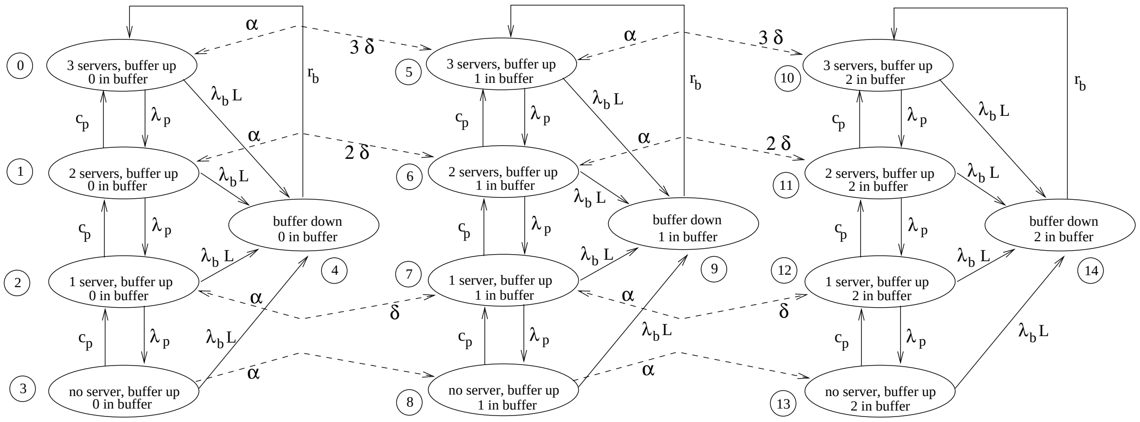

Figure 8: reachability graph of the second performability model

in vector-matrix notation

$$\frac{\partial \boldsymbol{\pi}(t,z)}{\partial t} + \frac{\partial}{\partial z}\Big(\boldsymbol{\pi}(t,z)\mathbf{M}(z)\Big) - \boldsymbol{\pi}(t,z)\Big(\mathbf{Q}(z) - \boldsymbol{\Lambda}(z)\Big) - \int_{z_{\min}}^{z} \boldsymbol{\pi}(t,y)\mathbf{B}(z-y)\,dy\,\boldsymbol{\Lambda}(z) = 0 \qquad \forall t, \forall z \in \mathring{\Omega}. \quad (15)$$

The derivatives are those of the first example reduced to one dimension. Note that there is no second order derivative and no the variances. The integral is similar to the integral in the Takács integro-differential equation [Tak55], it is the convolution of the jump height distribution with the fluid density function, multiplied with the rate λ a jump occurs at. These rates have to be subtracted from the diagonal elements of the generator matrix. This means that some mass is removed from the state probability vector and added again through the convolution in a different place on the continuous axis. The convolution gives the density of the sum of two random variables. Here these two random variables are the state vector Π and the random variable B of the jump height.

For the solution of the differential equation part of (15) the equivalent algorithm to the two-dimensional one is used. The integrals are solved explicitly on the known time step with a simple rectangular summation.

The solutions are plotted for two selected discrete states. The upper left plot in figure 9 shows the density in state 4 (buffer down and empty). This is a state where repair at high costs takes place. As opposed to the first model in this model the discrete transition rates are comparable with the fluid flow rates. Therefore the initial peak is transferred partly to the neighboring states and dominates the curves in those states as well. States 3, 9 and 13 have zero probability still after 170 time units. Only in state 14 (buffer down and holding 2 tokens), shown in the upper right plot of figure 9 the impact of the jumps can be seen clearly. The data for this plot has been slightly manipulated in that all solutions after time 120.0 have been dropped, in order to show the irregular shape of the curve more clearly.

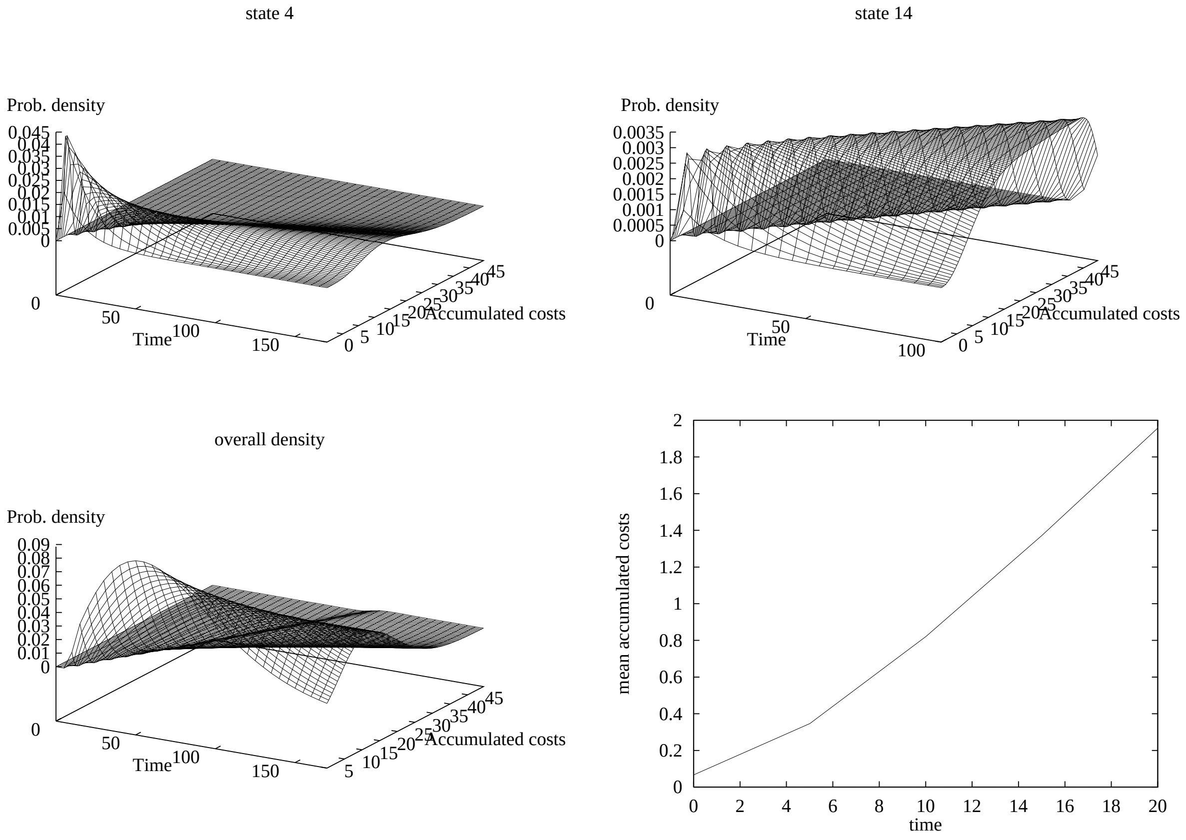

Figure 9: Transient probability density and mean accumulated costs

The lower left plot in figure 9 shows the fluid density summed over all discrete states. Here in addition all data for the accumulated reward less than 5 units has been dropped. This is done to remove the peak near the origin so that the plot scales differently and the rest of the shape becomes visible. The contribution of state 14 can be faintly seen.

The mean costs are only shown in the lower right plot in figure 9 only for the first 20 time units. They increase at slightly increasing pace as long as the discrete state probabilities are not yet in steady state.

4 Conclusions

Two models have been presented in order to demonstrate the modelling power of FSPNs in Markov reward modelling. Both are modifications of a well-known performability model. It has been shown that not only accumulated rate reward can be modelled with an FSPN, but also accumulated impulse reward.

Rate reward is accumulated in an FSPN through the inflow along a fluid arc to a fluid place while a transition is enabled, whereas impulse reward is added upon the firing of a transition as indicated by a jump arc. Jump arcs are labelled with a probability distribution. An amount sampled from that distribution is added to the content of the fluid place at the destination of the jump arc upon the firing of the transition at its origin. The example in this paper uses the Dirac impulse as probability distribution to model deterministic jump height.

The underlying equations are differential equations that also include convolution integrals. A numerical discretization method has been used to solve the equations. The numerical algorithm is also able to handle models with two fluid places. So it was possible in the first example to use one fluid place for the accumulated reward and the other one for a continuously approximated buffer. The results after applying the numerical algorithm to the differential equation is a probability density at each time instant. For models with one fluid place this density is supported on one continuous dimension and for models with two fluid places on a plane of two continuous dimensions. The means over the continuous dimensions have been regarded, the state probabilities and for the two-dimensional models the marginal densities. In the two-dimensional model there has been a peak in the density, which would be more appropriately represented with an impulse, a probability mass function. Including impulses is left for future work as are the development of good numerical algorithms that are more run-time and memory efficient, so that larger models can be analyzed.

References

[ABCD95] M. Ajmone Marsan, G. Balbo, G. Conte, and S. Donatelli. *Modelling with Generalized Stochastic Petri Nets.* Series in Parallel Computing. John Wiley & Sons, 1995.

[BGG+99] A. Bobbio, S. Garg, M. Gribaudo, A. Horváth, M. Sereno, and M. Telek. Modeling Software Systems with Rejuvenation, Restoration and Checkpointing through Fluid Stochastic Petri Nets. In *Proc. Eighth International Workshop on Petri Nets and Performance Models - PNPM'99*, Zaragoza, Spain, 1999. to appear.

[CM65] D.R. Cox and H.D. Miller. *The Theory of Stochastic Processes.* Chapman and Hall, 1965.

[DA93] R. David and H. Alla. Autonomous and Timed Continuous Petri nets. In G. Rozenberg, editor, *Advances in Petri Nets 1993*, volume 674 of *Lecture Notes in Computer Science*, pages 71 – 90. Springer-Verlag, 1993.

[GSB99] M. Gribaudo, M. Sereno, and A. Bobbio. Fluid Stochastic Petri Nets: An Extended Formalism to Include non-Markovian Models. In *Proc. Eighth International Workshop on Petri Nets and Performance Models - PNPM'99*, Zaragoza, Spain, 1999. to appear.

[HKNT98] G. Horton, V. G. Kulkarni, D. M. Nicol, and K. S. Trivedi. Fluid Stochastic Petri Nets: Theory, Applications and Solution. *European Journal of Operations Research*, 105(1):184–201, February 1998. Also published as ICASE report no. 96-5.

[Hor96] G. Horton. Computation of the Distribution of Accumulated Reward with Fluid Stochastic Petri nets. In *Proc. 2nd IEEE Int. Computer Performance and Dependability Symposium (IPDS)*, pages 90–95, Urbana-Champaign, Illinois, USA, 4-6 September 1996. IEEE-CS Press.

[Kle76] L. Kleinrock. *Queueing Systems Vol II: Computer Applications.* John Wiley, 1976.

[Kob74] H. Kobayashi. Application of the Diffusion Approximation to Queueing Networks II: Nonequilibrium Distributions and Applications to Computer Modeling. *Journal of the ACM*, 21(3):459–469, July 1974.

[Mey82] J. F. Meyer. Closed-Form Solutions of Performability. *IEEE Transactions on Computers*, C-31(7):648–657, July 1982.

[New71] G. F. Newell. *Applications of Queueing Theory.* Chapman and Hall, Ltd. (London), 1971.

[SM91] W. H. Sanders and J. F. Meyer. A Unified Approach for Specifying Measures of Performance, Dependability, and Performability. In A. Avizienis and J. Laprie, editors, *Dependable Computing for Critical Applications*, volume 4 of *Dependable Computing and Fault-Tolerant Systems*, pages 215–237. Springer, 1991.

[Tak55] L. Takács. Investigation of waiting time problems by reduction to Markov processes. *Acta Math. Acad. Sci. Hungar.*, 6:101–129, 1955.

[TK93] K. S. Trivedi and V. G. Kulkarni. FSPNs: Fluid Stochastic Petri Nets. In *Proc. 14th Int. Conf. on the Application and Theory of Petri Nets*, pages 24–31, Chicago, 1993.

[TR99] M. Telek and S. Rácz. Numerical analysis of large markov reward models. *Performace Evaluation*, 00, September 1999.

[Wol97] K. Wolter. Second Order Fluid Stochastic Petri Nets: an Extension of GSPNs for Approximate and Continuous modelling. In *Proc. 1st World Congress on Systems Simulation (WCSS'97)*, pages 328–332, Singapore, Sept. 1–3 1997.

[Wol99] K. Wolter. Jump Transitions in Second Order FSPNs. 1999. accepted for publication at MASCOTS'99.

[WZ99] K. Wolter and A. Zisowsky. On Markov Reward Modelling with Second Order FSPNs. In J. T. Bradley and N. J. Davies, editors, *Proc. 15th annual UK Performance Engineering Workshop*, pages 167–177, Bristol, UK, July, 22-23 1999. TR CSTR-99-007.

[Zis98] A. Zisowsky. Entwurf und Implementierung eines Verfahrens für die transiente Analyse fluider stochastischer Petri-Netze. Master's thesis, TU Berlin, August 1998.