T-79.231 Parallel and Distributed Digital Systems

Stochastic Analysis

Marko Mäkelä

August 1, 2003

Why Stochastic Analysis?

In reachability analysis and model checking, the timing of the system is seldom modelled in detail. The focus is on the ordering of events: can they occur in such an order where the system malfunctions.

In practice, it takes time to perform events. For instance, if a reactive control system does not operate at expected speed, the consequences can be fatal.

Any system can be assigned some performance requirements, such that there average number of pending jobs or clients is reasonable, or the waiting times are not too long.

Stochastic analysis can be thought to be based on finite automata or reachability graphs, whose events are associated with durations or selection probabilities.

Stochastic Systems and Markov Chains

Stochastic process is a set of random states $\{\theta^{(t)} \in S | t \in T\}$ for some set of indices T (which represent moments of time).

Markov chain is a stochastic process where the next state $\theta^{(n+1)}$ only depends on the preceding state $\theta^{(n)}$.

The rest of this lecture covers *stationary* and *continuous-time* Markov chains whose transition probabilities are independent of time and which can switch states at arbitary moments of time. They can be described with a *transition rate matrix* that defines the probabilities of state transitions per unit of time.

A Markov chain can be thought as a path in some reachability graph whose transitions are associated with a probability of occurrence or a randomly distributed duration.

Stochastic Petri Nets (1/2)

Stochastic Petri Nets (SPN) are an extension of place/transition systems where the transitions are associated with time.

Enabled transitions *cannot* arbitrarily fire or refrain from firing. The delay between a transition becoming enabled to the firing of the transition follows a negative exponential distribution, that is, the transition fires in the time interval $0 \leq X \leq x$ with the probability $P\{X \leq x\} = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$, where λ is the average firing rate.

Since the exponential distribution is *memoryless*, that is, $P\{X > t + h | X > t\} = P\{X > h\}$, possible conflicts in the system can be safely ignored.

Ordinary stochastic Petri nets correspond to continuous-time Markov chains.

Stochastic Petri Nets (2/2)

Generalised Stochastic Petri Nets (GSPN) also include untimed or *immediately firing* transitions, which are associated with a weight.

If an untimed transition is enabled in a marking, it fires immediately, producing a new marking. If several untimed transitions are enabled, their firing probabilities are determined by the weights. A timed transition may only fire if no immediate transitions are enabled. GSPNs correspond to semi-Markov processes. They have the same stationary distribution as corresponding Markov processes.

Deterministic and Stochastic Petri Nets (DSPN) also contain transitions that fire in a constant time. If multiple such transitions can be enabled simultaneously, the analytical methods are challenged, and one has to resort to simulations.

Analysing Stochastic Systems

There are computer tools for modelling and analysing stochastic systems, such as GreatSPN, DSPNExpress, TimeNET and SMART. Unfortunately we are not aware of any free tools.

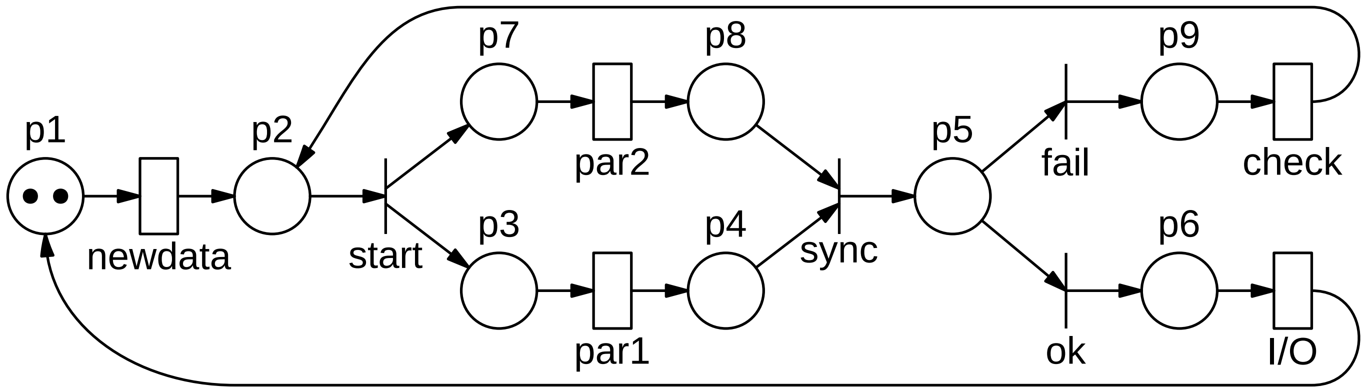
Simple systems can be analysed with a reachability analyser and a matrix calculator, such as Matlab or GNU Octave. Three steps are needed:

1. The reachability graph of the system is generated. If there are immediate transitions, the events corresponding to them are fused with timed events.

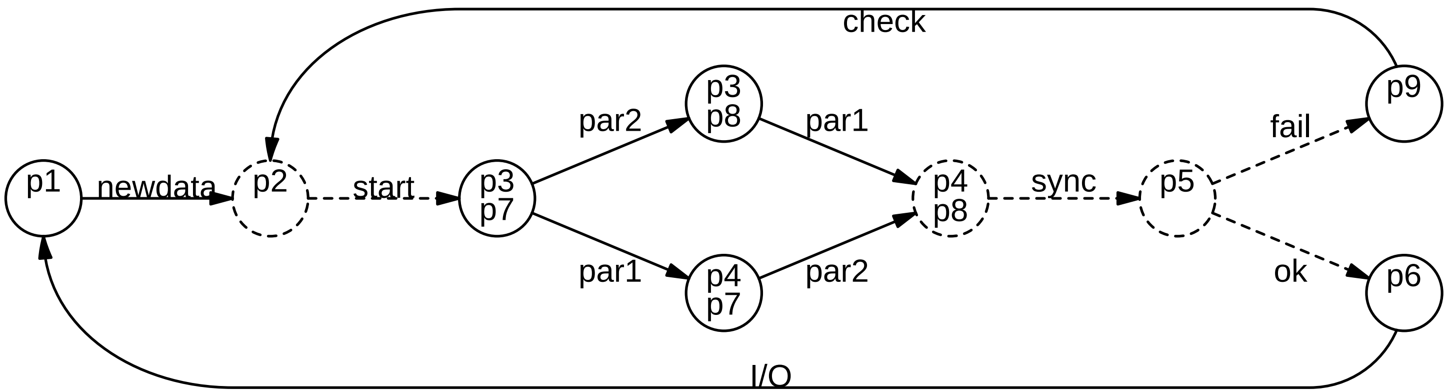
2. The events of the reachability graph are labelled with the firing rates of the transitions, and the graph is transformed into a transition rate matrix Q

3. The steady state equation is solved and the performance indices of the process are computed.

An Example System and its Reachability Graph

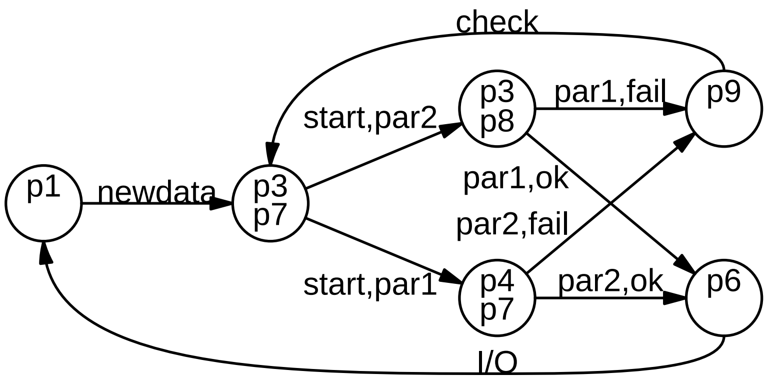


There are 38 reachable states and 62 events. When $M_0(\mathrm{p1}) = 1$, the state space shrinks:



Eliminating the Immediate Transitions

The reachability graph can be reduced by fusing the occurrences of immediate transitions with the preceding occurrences of timed transitions. The result is a Markov chain:



We have not yet associated the transitions with weights or durations. Let us assume that the firing ratio of the transitions “fail” and “ok” is $5 : 95$, that is, assign them the weights 5 and 95. The weights of the transitions “sync” and “start” are insignificant, since these transitions are not part of any conflict.

Let us assign the rates $\lambda_{\text{newdata}} = 37, \lambda_{\text{par1}} = 40, \lambda_{\text{par2}} = 60, \lambda_{\text{I/O}} = 15, \lambda_{\text{check}} = 2$. Clearly, $\lambda_{\text{start,par1}} = \lambda_{\text{par1}}$ and $\lambda_{\text{start,par2}} = \lambda_{\text{par2}}$, as “start” is enabled alone. Similarly, we can ignore “sync”.

Attention must be paid on the conflicting immediate transitions “fail” and “ok”: $\lambda_{\text{par1,ok}} = \frac{95}{95+5}\lambda_{\text{par1}} = 38$, $\lambda_{\text{par1,fail}} = \frac{5}{95+5}\lambda_{\text{par1}} = 2$, $\lambda_{\text{par2,ok}} = 57$ ja $\lambda_{\text{par2,fail}} = 3$.

Transition Rate Matrix

Let us present the reachability graph as a transition rate matrix Q. The elements of this matrix, $q_{i,j}, i \neq j$, correspond to probabilities of firing a transition leading from the state i to the state j, the transition rate.

$$\mathbf{Q} = \begin{array}{cc} \begin{array}{cccccc} \{p_1\} & \{p_3,p_7\} & \{p_4,p_7\} & \{p_3,p_8\} & \{p_6\} & \{p_9\} \end{array} & \\ \begin{pmatrix} -37 & 37 & 0 & 0 & 0 & 0 \\ 0 & -100 & 40 & 60 & 0 & 0 \\ 0 & 0 & -60 & 0 & 57 & 3 \\ 0 & 0 & 0 & -40 & 38 & 2 \\ 15 & 0 & 0 & 0 & -15 & 0 \\ 0 & 2 & 0 & 0 & 0 & -2 \end{pmatrix} & \begin{array}{c} \{p_1\} \\ \{p_3,p_7\} \\ \{p_4,p_7\} \\ \{p_3,p_8\} \\ \{p_6\} \\ \{p_9\} \end{array} \end{array}$$

If no transition is possible to a state, the transition rate is 0. The diagonal elements are assigned so that the row sums become zero: $q_{i,i} = -\sum_{j \neq i} q_{i,j}$.

Solving the Steady State Equation (1/2)

It is possible to define a probability vector $\pi(t)$ that yields the probability of each state as a function of time. The limit $\pi = \lim_{t \to \infty} \pi(t)$ is called the steady state solution, because it fulfils the equation $\pi \cdot \mathbf{Q} = \mathbf{0}$. Furthermore, the sum of all probabilities is 1, that is, $\pi \cdot \mathbf{e}^{\mathrm{T}} = 1$ or $\pi_1 + \pi_2 + \cdots + \pi_n = 1$. The condition can be written as a matrix by writing the same equation multiple times: $\pi \cdot \mathbf{E} = \mathbf{e}^{\mathrm{T}}$. (Here, $\mathbf{e}$ is a row vector consisting of ones and $\mathbf{E}$ is a square matrix consisting of ones.)

By adding the resulting matrix equations

$$\begin{aligned} \pi \cdot \mathbf{Q} &= 0 \\ \pi \cdot \mathbf{E} &= \mathbf{e} \end{aligned}$$

together we obtain $\pi \cdot (\mathbf{Q} + \mathbf{E}) = \mathbf{e}$, where we can solve $\pi = \mathbf{e} \cdot (\mathbf{Q} + \mathbf{E})^{-1}$.

Solving the Steady State Equation (2/2)

Let us compute the state probabilities of the reduced reachability graph with GNU Octave:

```
octave:1> Q=[-37,   37,   0,   0,   0,  0;
               0, -100,  40,  60,   0,  0;
               0,    0, -60,   0,  57,  3;
               0,    0,   0, -40,  38,  2;
              15,    0,   0,   0, -15,  0;
               0,    2,   0,   0,   0, -2]

octave:2> pi=ones(1,6)*inverse(Q+ones(6))*
pi =

0.176252 0.068646 0.045764 0.102968 0.434756 0.171614
```

An analytic solution would be more accurate, but in practice, approximate solutions suffice for plotting graphs (such as the probability of a given state as a function of $\lambda_{newdata}$).

*It would be a little more efficient to write `pi=((Q'+ones(6))\ones(6,1))'`.

Applications of the Steady State Distribution (1/2)

Among other things, the steady state solution indicates that the system spends $43.4\,\%$ of its time in the state $\{p_6\}$. In other words, the transition "I/O" that leaves from the state is a possible bottleneck.

In Stochastic Petri nets, it is possible to observe the average number of tokens in the places. In our example system, place p_1 contains one token in the state $\{p_1\}$, and it is empty in other reachable states. Thus, the average marking of p_1 equals the probability of the state $\{p_1\}$, 0.176252.

Generally, the average marking of a place is a weighted sum of the marking of the place in each reachable state, weighted with the state probabilities. For instance, the marking of the place p_3 in the reachable states can be written as the column vector $\mathbf{m}_{p_3} = (0\ 1\ 0\ 1\ 0\ 0)^{\mathrm{T}}$. The average marking of p_3 is `pi*[0,1,0,1,0,0]'` or 0.17161 according to GNU Octave.

Applications of the Steady State Distribution (2/2)

The probabilities of more complex conditions can be calculated in a similar way. Let us define a column vector $\mathbf{r}$, whose elements are 1 in those states where the desired property holds, and 0 elsewhere. For instance, “p_7 or p_8 is marked”, that is, $\mathbf{r} = (0\ 1\ 1\ 1\ 0\ 0)^{\mathrm{T}}$ holds at an arbitrary moment of time with the probability 0.21738.

The expected firing density of a transition can be observed as well. In this case, the *reward function* $\mathbf{r}$ is defined so that it maps those states where the transition is enabled to the rate of the transition, and other states to 0. The average firing rate is $\pi \cdot \mathbf{r}$.

The average firing density of “par1” in our example is `pi*[0,40,0,40,0,0]'` or 6.8646. Because the only pre-place of this transition is p_3 which can contain at most one token, we would have obtained the same result by multiplying the probability that p_3 is marked with the rate, $0.17161 \cdot 40$.

Summary

Performance analysis is based on state spaces (reachability graphs). If the system comprises n reachable states, analysing its performance requires operations on $n \times n$ square matrices. Because it can be computationally very expensive, the model often has to be simplified in order to reduce the number of states. We presented one way, removing immediate transitions. Other reduction methods require more detailed knowledge of the system being analysed.

This lecture only scratches the surface of the topic. We did not discuss different queuing or serving policies (how many clients can be served simultaneously, are the waiting times resampled when transitions become disabled, etc.) nor other than exponentially distributed transition durations.

Performance indices can be obtained by simulating the system long enough, for instance with the MCMC method (Monte Carlo Markov Chain). Often simulation is the only way, if the transition durations follow different kinds of distributions, or the system has a large number of reachable states.

Additional Information

1. Gianfranco Balbo: *An Introduction to Generalized Stochastic Petri Nets*, pages 217–266 in *Petri Nets 2000: Introductory Tutorial* (`http://www.daimi.au.dk/PetriNets/introductions/pn2000_introtut.pdf`)

2. Mat-2.111 Stochastic Processes (3 ov)

3. S-38.143 Queuing Theory (3 cr) L