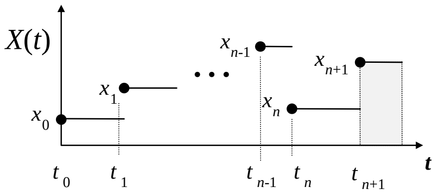
STOCHASTIC PETRI NET MODELS

Foundations

Probabilistic performance models use stochastic processes to represent the behaviour of complex systems by avoiding a detailed deterministic description. A Stochastic Petri Net (SPN) model allows the synchronization and concurrency aspects of the system model to be captured by the Petri Net (PN) and be combined with the stochastic description.

The key attribute of SPNs is that an exponentially distributed random variable is associated with each of the PN transitions. This random variable (say $X(t)$) characterising the stochastic process represents the (discrete) state in which the system resides for some *sojourn time* (i.e. the time before transition firing):



A continuous-time, discrete state-space stochastic process can be described as a *Markov process* if:

$$P[X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \ldots, X(t_0) = x_0] = P[X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n,]$$

with $t_{n+1} > t_n > t_{n-1} \cdots > t_0$ where $x_k \in S$, the state-space of $X(t)$

i.e. the stochastic process is *memoryless* in the sense that the future evolution of the process only depends on the current state and not on the previous behaviour of the process.

Let $p_{ij}(t, u) = P[X(u) = j \mid X(t) = i]$ be the probability that the process is in state j at time u given that it was in state i at time t.

Where the process is also *homogeneous* the transition probabilities only depend on the time difference $\tau = u - t$ so that:

$$p_{ij}(\tau) = p_{ij}(t, t+\tau) = P[X(t+\tau) = j \mid X(t) = i]$$

Let $\pi_i(t) = P[X(t) = i]$ be the probability that the process is in state i at time t. Then using the *Chapman-Kolomogrov* equation it can be shown that:

$$\frac{d\pi_i(t)}{dt} = \sum_{j\in S} q_{ji}\,\pi_j(t)$$

i.e. the equation describing the probability that a process is in state i is governed by a first order differential equation driven by the sum of the products of transition rates from all other states and the probability of being in those states.

This can be put more concisely in matrix form, i.e:

$\mathbf{Q} = [q_{ij}]$, $\boldsymbol{\pi}(t) = \{\pi_1(t), \pi_2(t), \ldots\}$ and the differential equation

becomes: $\dfrac{d\pi(t)}{dt} = \boldsymbol{\pi}(t)\mathbf{Q}$

where $\mathbf{Q}$ is referred to as the *infinitesimal generator* (or *transition probability matrix*).

These equations admit a general exponential solution:

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)e^{\mathbf{Q}t}$$

Thus a homogeneous Markov chain with a memoryless property admits exponential solutions for the probability of being in a state and it also admits exponential solutions for the sojourn times in that state

If the Markov chain has an equilibrium or steady-state condition, i.e. it is *ergodic*, we have:

$$\lim_{t\to\infty} \frac{d\pi(t)}{dt} = 0 = \pi\mathbf{Q}$$

Definition of the Stochastic Petri Net (SPN)

The formal definition for the SPN is:

SPN = (P, T, A, M_0, λ)

where $P = \{p_1, p_2, \ldots, p_n\}$ is the set of places
$T = \{t_1, t_2, \ldots, t_m\}$ is the set of transitions
$A = \{P \times T\} \cup \{T \times P\}$ is the set of input and output arc pairs
$M_0 = \{\mu_1, \mu_2, \ldots, \mu_n\}$ is the set of initial place markings
$\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_m\}$ is the set of transition firing rates

Because of the memoryless property of the exponential distribution of the firing delays, SPNs are *isomorphic* to continuous-time Markov chains $\rightarrow$ there is a mapping between the two state transition systems in both state spaces and transition probabilities.

If the transition firing times can be represented as a Markov Chain (MC) associated with a given SPN, then the MC can be obtained from the following rules:

1. The MC state space S corresponds to the reachability set $R(M_0)$ of the SPN.

2. The transition rate from state i (with marking M_i) to state j (with marking M_j) is given by: $q_{ij} = \sum_{k \in H_{ij}} \lambda_k$

   where H_{ij} is the set of transitions enabled by the marking M_i whose firing generates the marking M_j.

Where the generated MC is *ergodic* (i.e. all states are aperiodic, recurrent and non-null), the probability distribution of the MC converges towards a steady-state distribution which is independent of the initial state distribution.

From the above, the following matrix equation can be solved:

$$\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$$

with an additional constraint from the *law of total probability*, i.e. $\sum_i \pi_i = 1$

where **Q** is the infinitesimal generator matrix with elements given by rule 2 above, and $\boldsymbol{\pi}$ is the vector of steady state probabilities.

From the steady state distribution $\boldsymbol{\pi}$ some SPN behavioural estimates can be obtained:

1. The *probability of some condition holding* in the SPN. The condition is represented by a subset A of $R(M_0)$, is given by:

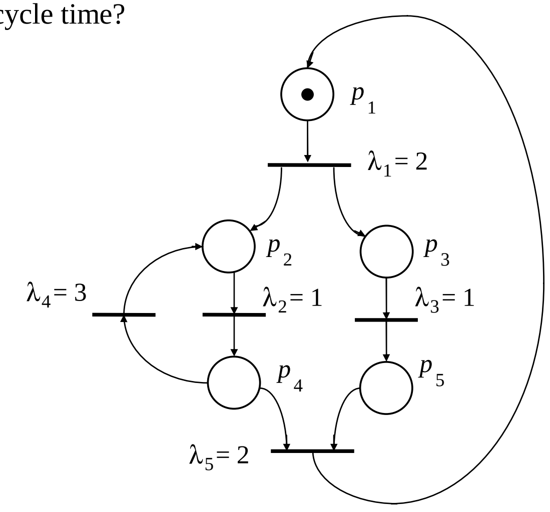
$$\mathrm{P}[A] = \sum_{i \in A} \pi_i$$

2. The *expected value of the number of tokens in a given place* of the SPN. If $A(i, x)$ is the subset of $R(M_0)$ for which the number of tokens in place p_i is x (and the place is k-bounded), then the expected value of the number of tokens in p_i is given by:

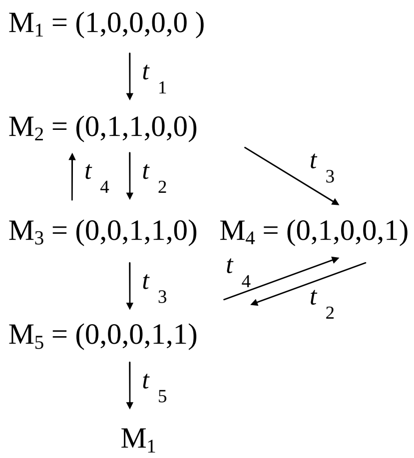
$$\mathrm{E}[\mu_i] = \sum_{n=1}^{k} nP[A(i,n)]$$

Example

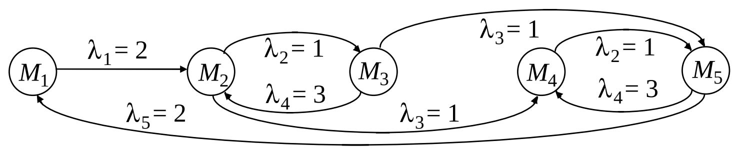
What is the cycle time?



The reachability tree is:



Which by rule 1, can be represented as a Markov chain:



Rule 2 is used to generate the infinitesimal generator matrix **Q** with elements $q_{ij} = \sum_{k \in H_{ij}} \lambda_k$ where H_{ij} is the set of transitions enabled by the marking M_i whose firing generates the marking M_j:

$$Q = \begin{bmatrix} -2 & 2 & 0 & 0 & 0 \\ 0 & -2 & 1 & 1 & 0 \\ 0 & 3 & -4 & 0 & 1 \\ 0 & 0 & 0 & -1 & 1 \\ 2 & 0 & 0 & 3 & -5 \end{bmatrix}$$

Using SPN behavioural estimate condition 1, the probability of being in state M_i is given by:

$$P[M_i] = \pi_i$$

Note also that the matrix equation $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$ can then be expressed in the form:

$$2P[M_1] = 2P[M_5] \quad (1)$$
$$2P[M_2] = 2P[M_1] + 3P[M_3] \quad (2)$$
$$4P[M_3] = P[M_2] \quad (3)$$
$$P[M_4] = 3P[M_5] + P[M_2] \quad (4)$$
$$5P[M_5] = P[M_4] + P[M_3] \quad (5)$$

These equations can be derived directly from a 'flow balance' being applied to the product of the transition rate and the steady state probabilities at each MC state. For example, in the last equation above the sum of the transition rates on the outgoing arcs of state M_5 is 5 and this is multiplied by the steady-state probability of state M_5. This must balance the sum of the transition rates on all incoming arcs multiplied by their corresponding steady-state probabilities (in this case, $\lambda_2 = 1$ from M_4 and $\lambda_3 = 1$ from M_3).

Finally, the additional constraint condition: $\sum_i \pi_i = 1$ is applied so that:

$$P[M_1] + P[M_2] + P[M_3] + P[M_4] + P[M_5] = 1 \quad (6)$$

Using (3) above we have: $8P[M_3] = 2P[M_2]$ and in (2) we have: $8P[M_3] = 2P[M_1] + 3P[M_3] \rightarrow 5P[M_3] = 2P[M_1]$

and from this and (3) we have: $5P[M_2] = 20P[M_3] = 8P[M_1]$

From (4), (1) and the above we have:

$$P[M_4] = 3P[M_5] + P[M_2] = 3P[M_1] + 8/5P[M_1] = 23/5P[M_1]$$

Thus using (6) we have:

$$P[M_1] + P[M_2] + P[M_3] + P[M_4] + P[M_5] = 1$$
$$\rightarrow\ P[M_1] + 8/5P[M_1] + 2/5P[M_1] + 23/5P[M_1] + P[M_1] = 1$$
$$\rightarrow\ 43/5P[M_1] = 1\ \rightarrow P[M_1] = 5/43 = 0.1163$$

hence $P[M_2] = 8/43 = 0.1860$, $P[M_3] = 2/43 = 0.0465$,
$P[M_4] = 23/43 = 0.5349$, and $P[M_5] = 5/43 = 0.1163$

Using SPN behavioural estimate condition 2 (the expected value of the number of tokens in a given place of the SPN), we note that no place in the reachability tree exceeds a 1-bounded condition $\rightarrow$ the expected value reduces to:

$$E[\mu_i] = \sum_{i \in A(i,1)} \pi_i$$

i.e. the expected value of the number of tokens in place p_i is the sum of the steady-state probabilities in a subset of the reachability set of the SPN (i.e. also the states of the MC) for which the number of tokens in place p_i is 1.

In this example, the expected values of token occupancy in each SPN place is given by:

$$E[\mu_1 = 1] = \pi_1 = P[M_1] = 0.1163$$
$$E[\mu_2 = 1] = \pi_2 + \pi_4 = P[M_2] + P[M_4] = 0.7209$$
$$E[\mu_3 = 1] = \pi_2 + \pi_3 = P[M_2] + P[M_3] = 0.2325$$
$$E[\mu_4 = 1] = \pi_3 + \pi_5 = P[M_3] + P[M_5] = 0.1628$$
$$E[\mu_5 = 1] = \pi_4 + \pi_5 = P[M_4] + P[M_5] = 0.6512$$

Average Performance Analysis

Little's law can be applied, i.e. the average number of customers in a queuing system ($\overline{N}$) is equal to the product of customer arrival rate (λ) and the average time spent in the system ($\overline{T}$), i.e:

$$\overline{N} = \lambda \overline{T}$$

In this example, the number of tokens entering and leaving the subsystem (made up of places p_2, p_3, p_4, p_5, and transitions t_2, t_3, t_4, t_5) is conserved. The entry transition to the subsystem (t_1) is only enabled when place p_1 has a token $\rightarrow$ the utilization of t_1 is $E[\mu_1=1] = 0.1163$.

As the average transition rate for t_1 is $\lambda_1 = 2$, the average token flow from p_1 is $0.1163 \times 2 = 0.2326$ token/unit time.

Due to the fork transition at t_1, the average token flow through the subsystem per unit time is twice the flow through $p_1 \rightarrow \lambda = 0.4652$ token/unit time.

The average number of tokens in the subsystem is the sum of the average number of tokens in all places in the subsystem, i.e:

$$\begin{aligned} \overline{N} &= \overline{\mu}_2 + \overline{\mu}_3 + \overline{\mu}_4 + \overline{\mu}_5 \\ &= E[\mu_2 = 1] + E[\mu_3 = 1] + E[\mu_4 = 1] + E[\mu_5 = 1] \\ &= 0.7209 + 0.2325 + 0.1628 + 0.6512 = 1.7674 \end{aligned}$$

Finally, from Little's law:

$$\overline{T} = \overline{N}/\lambda = 1.7674/0.4652 = 3.8 \text{ time units}$$

Thus the average time a token takes to return to p_1 is 3.8 time units.

Limitations

Note that the state probabilities are a function of the reachability set, and hence are a function of the initial marking. For example, if p_1 had a marking of 2 tokens in the previous example, the reachability set would have 14 states (and for $\mu_1 = 3$ it would be 30 states). Thus the complexity of the MC grows rapidly.

More general examples can have infinite reachability sets:

- can truncate the MC to obtain approximate results
- or reduce the SPN to a finite reachability set but retain the main characteristics of the system.

The most significant limitation is that only average performance issues can be addressed and not the dynamic temporal properties of the system.