

PERFORMANCE EVALUATION OF CONCURRENT SYSTEMS USING TIMED PETRI NETS

W.M. Zuberek

Department of Computer Science, Memorial University of Newfoundland
St. John's, NL, Canada A1C-5S7

Abstract

It is shown that the behavior of conflict-free Petri nets with exponentially distributed firing times can be represented by labeled directed "state" graphs in which labels describe the probabilities of transitions between vertices of the graph. For bounded Petri nets the corresponding state graphs are finite, and are isomorphic to finite state Markov chains, stationary descriptions can thus be obtained by standard techniques. An immediate application of such a model is performance analysis of concurrent systems, and in particular queueing systems with exponentially distributed arrival and service times. A simple example of an interactive computer system model is used as an illustration of performance evaluation.

1. INTRODUCTION

A Petri net [1,6,7,8] is known as an abstract, formal model of communication between asynchronous components of a system. The properties, concepts, and techniques of Petri nets are being developed in a search of natural and simple methods for describing and analyzing systems that may exhibit asynchronous and concurrent activities. The major use of Petri nets has been the modeling of systems of events in which it is possible for some events to occur concurrently but there are constraints on the concurrence, precedence, or frequency of these occurrences. Such a model, however, is not complete enough for the study of systems performance since no assumption is made on the duration of systems activities. The timed Petri nets have been introduced by Ramchandani [8] by assigning firing times to the transitions of Petri nets. Sifakis [9] introduced another definition of a timed Petri net by assigning time to places. Merlin and Farber [4] discussed timed Petri nets where a time threshold and maximum delay were assigned to a transition to allow the incorporation of timeouts into protocol models. Recently Molloy [5] introduced stochastic Petri nets in which transition firing times are exponentially distributed random variables, and the corresponding firing rates are assigned to transitions of a net. Stochastic Petri nets are isomorphic to homogeneous Markov processes, their analysis, however, is based on reachability sets which are generated without timing considerations. Therefore, to obtain performance measures, an additional token flow analysis is performed.

The method described in the paper is an extension of the approach originated by Ramchandani and used to model the performance of digital systems at the register transfer level [10,11]. Firing rates of exponentially distributed firing times are assigned to transitions of a Petri net, and a new continuous-time discrete-state description is introduced which is isomorphic to continuous-time discrete-state Markov chains. This directly provides such performance

measures as utilization of systems components, average queue lengths, average waiting times and turnaround times or average throughput rates, and at the same time preserves the simplicity of model specification and offers automatic generation of the state space.

The paper is organized in 3 main sections. Section 2 contains definitions of basic concepts for marked Petri nets. Timed Petri nets are introduced in section 3. Application of timed Petri nets to performance evaluation is discussed in section 4.

2. MARKED PETRI NETS

A Petri net \mathbf{N} is a triple $\mathbf{N} = (P, T, A)$ where:

P is a finite, nonempty set of places,

T is a finite, nonempty set of transitions,

A is a set of directed arcs which connect places with transitions and transitions with places, and:

$$\forall t \in T \exists p_i, p_j \in P : (p_i, t) \in A \wedge (t, p_j) \in A.$$

A place p is an input (or an output) place of a transition t iff there exists an arc (p, t) (or (t, p) , respectively) in the set A . The sets of all input and output places of a transition t are denoted by $Inp(t)$ and $Out(t)$, respectively. Similarly, the sets of input and output transitions of a place p are denoted by $Inp(p)$ and $Out(p)$.

A place p is shared iff it is an input place for more than one transition. A net is conflict-free iff it does not contain shared places. Only conflict-free Petri nets are considered in this paper. Some other classes of Petri nets are discussed elsewhere.

A marked Petri net \mathbf{M} is a pair $\mathbf{M} = (\mathbf{N}, \bar{m})$ where:

\mathbf{N} is a Petri net, $\mathbf{N} = (P, T, A)$,

\bar{m} is an initial marking function which assigns a nonnegative integer number of so called tokens to each place of the net, $\bar{m} : P \rightarrow \{0, 1, \dots\}$.

Let any function $m : P \rightarrow \{0, 1, \dots\}$ be called a marking of a net $\mathbf{N} = (P, T, A)$.

A transition t is enabled by a marking m iff every input place of this transition contains at least one token. The set of all transitions enabled by a marking m is denoted by $E(m)$.

Every transition enabled by a marking m can fire. When a transition fires, a token is removed from each of its input places and a token is added to each of its output places. This determines a new marking in a net, a new set of enabled transitions, and so on.

A marking m_j is directly reachable from a marking m_i in a net \mathbf{N} , $m_j \leftarrow m_i$, iff there exists a transition t enabled by the marking m_i , $t \in E(m_i)$, such that

$$\forall p \in P : m_j(p) = \begin{cases} m_i(p) - 1, & \text{if } p \in \text{Inp}(t) - \text{Out}(t), \\ m_i(p) + 1, & \text{if } p \in \text{Out}(t) - \text{Inp}(t), \\ m_i(p), & \text{otherwise.} \end{cases}$$

A marking m_j is (generally) reachable from a marking m_i in a net \mathbf{N} , $m_j \xleftarrow{*} m_i$, if there exists a sequence of markings $(m_{i_0} m_{i_1} m_{i_2} \dots m_{i_k})$ such that $m_{i_0} = m_i$, $m_{i_k} = m_j$, and each marking m_{i_ℓ} is directly reachable from the marking $m_{i_{\ell-1}}$ for $\ell = 1, \dots, k$.

A set $M(\mathbf{M})$ of reachable markings of a marked Petri net $\mathbf{M} = (\mathbf{N}, \bar{m})$ is the set of all markings which are reachable from the initial marking \bar{m} (including \bar{m}), $M(\mathbf{M}) = \{m \mid m \xleftarrow{*} \bar{m}\}$.

A marked net \mathbf{M} is bounded if there exists a positive integer K such that each marking in the set $M(\mathbf{M})$ assigns at most K tokens to each place of the net

$$\exists K > 0 \forall m \in M(\mathbf{M}) \forall p \in P : m(p) \leq K.$$

If a marked net \mathbf{M} is bounded, its reachability set $M(\mathbf{M})$ is finite. Only bounded Petri nets are considered in this paper.

An enable function of a marking m in a net \mathbf{N} is any function $e_m : T \rightarrow \{0, 1, \dots\}$ such that

$$\forall t \in T : \min_{p \in \text{Inp}(t)} (m(p) * e_m(t)) = 0,$$

i.e., any function which indicates (by nonzero values) all those transitions which can fire simultaneously (and some of the transitions may fire "several times"). In general case, when a net contains conflicts, there may be several different enable functions for the same marking m . For conflict-free nets, however, for each marking m there exists exactly one enable function which is determined by

$$\forall t \in T : e_m(t) = \min_{p \in \text{Inp}(t)} (m(p)).$$

3. TIMED PETRI NETS

In a timed Petri net, each transition $tt@$ takes a positive time to fire. When a transition $tt@$ is enabled, a firing can be initiated by removing a token from each of t 's input places. This token remains in the transition t for the "firing time", and then the firing terminates by adding a token to each of t 's output places (sometimes this is called a three-step firing as opposed to one-step instantaneous firing of marked nets). Each of the firings is initiated in the same instant of time in which it is enabled. If a transition is enabled while it fires, a new, independent firing can be initiated. For conflict-free nets all enabled transitions immediately initiate their firings since each marking of a net uniquely determines the enable function. If a net contains conflicts, and there are several different enable functions for the same marking, the selection of an actual enable function is usually a random process which can be described by corresponding probabilities.

The operation of a timed Petri net can thus be considered as taking place in "real time", and it is assumed that it starts at the time $\tau = 0$. At this moment the firings indicated by the enable function e_0 are initiated and the tokens are removed from the corresponding input places. Then, after the time determined by the smallest "firing time" of the transitions which initiated firings, the tokens are deposited

in appropriate output places creating a new marking, a new set of enabled transitions, and so on.

The firing times of transitions can be described in several ways. In D-timed Petri nets [3,8,10,11] they are deterministic (or constant), i.e., there is a positive (rational) number assigned to each transition of a net. In M-timed Petri nets (or stochastic Petri nets [5]), the firing times are exponentially distributed random variables, and the corresponding firing rates are assigned to transitions of a net. The memoryless property of exponential distributions is the key factor in analysis of M-timed Petri nets.

An M-timed Petri net \mathbf{T} is a pair $\mathbf{T} = (\mathbf{M}, r)$ where:

\mathbf{M} is a marked Petri net, $\mathbf{M} = (\mathbf{N}, m_0)$, $\mathbf{N} = (P, T, A)$,

r is a firing rate function which assigns a positive real number $r(t)$ to each transition t of the net, $r : T \rightarrow \mathbf{R}^+$, and \mathbf{R}^+ denotes the set of positive real numbers; the firing time of a transition t is a random variable $v(t)$ with the distribution function

$$\text{Prob}(v(t) > x) = e^{-r(t)*x}, x > 0.$$

The memoryless property of exponential distributions means that if the duration v of a certain activity (e.g., the firing time) is distributed exponentially with parameter r , and if that activity is observed at time y after its beginning, then the remaining duration of the activity is independent of y and is also distributed exponentially with parameter r :

$$\text{Prob}(v > y + x \mid v > y) = \text{Prob}(v > x) = e^{-r*x}.$$

The exponential distribution is the only continuous distribution with that property.

Also, if v and w are the durations of two independent simultaneous activities a and b , distributed exponentially with parameters q and r , respectively, then the time interval u until the first completion of an activity (a or b) is distributed exponentially with parameter $(q + r)$, and the probability that the activity a will complete first is equal to $q/(q + r)$, while the same probability for the activity b is equal to $r/(q + r)$. These results can be generalized in an obvious way to any number of activities.

A state s of an M-timed Petri net \mathbf{T} is a pair of functions $s = (m, f)$ where:

m is a marking function, $m : P \rightarrow \{0, 1, \dots\}$,

f is a firing function which indicates (for each transition of the net) the number of active firings, i.e., the number of firings which have been initiated but are not yet terminated, $f : T \rightarrow \{0, 1, \dots\}$.

The initial state s_0 of a conflict-free net \mathbf{T} is a pair $s_0 = (m_0, f_0)$ where f_0 is the enable function e_0 , and the marking m_0 is defined by

$$\forall p \in P : m_0(p) = \bar{m}(p) - \sum_{t \in \text{Out}(p)} f_0(t).$$

A state $s_j = (m_j, f_j)$ is directly t_k -reachable from the state $s_i = (m_i, f_i)$, $s_j \leftarrow s_i$, iff the following conditions are satisfied:

1. $f_i(t_k) > 0$,
2. $\forall p \in P : m_{ij}(p) = m_i(p) + \begin{cases} 1, & \text{if } p \in \text{Out}(t_k), \\ 0, & \text{otherwise,} \end{cases}$
3. $\forall p \in P : m_j(p) = m_{ij}(p)x - \sum_{t \in \text{Out}(p)} e_{ij}(t)$,

$$4. \forall t \in T : f_j(t) = f_i(t) + e_{ij}(t) - \begin{cases} 1, & \text{if } t = t_k, \\ 0, & \text{otherwise.} \end{cases}$$

The state s_j which is directly t_k -reachable from the state s_i , is thus obtained by the termination of the t_k firing (1), updating the marking of a net (2), and then initiating new firings (if any) which correspond to the enable function e_{ij} (3 and 4).

Similarly as for marked nets, a state s_j is (generally) reachable from a state s_i if there is a sequence of directly reachable states from the state s_i to the state s_j . Also, a set $S(\mathbf{T})$ of reachable states is defined as the set of all states of a net \mathbf{T} which are reachable from the initial state s_0 (including s_0). For bounded nets the sets of reachable states are finite.

A state graph \mathbf{G} of an M-timed Petri net \mathbf{T} is a labeled directed graph $\mathbf{G}(\mathbf{T}) = (V, D, b)$ where:

V is a set of vertices which is equal to the set of reachable states of the net \mathbf{T} , $V = S(\mathbf{T})$,

D is a set of directed arcs, $D \subset V \times V$, such that (s_i, s_j) is in D iff s_j is directly reachable from s_i ,

b is a labeling function which assigns the probability of transition from s_i to s_j to each arc (s_i, s_j) in the set D , $b : D \rightarrow [0, 1]$, in such a way that if s_j is directly t_k -reachable from s_i , then

$$b(s_i, s_j) = r(t_k) * f_i(t_k) / \sum_{t \in T} r(t) * f_i(t).$$

It can be observed that state graphs of bounded M-timed Petri nets are isomorphic to finite-state Markov chains and the states of Markov chains directly correspond to the states of M-timed Petri nets.

Example. For a very simple Petri net shown in Fig.1a (as usual, places are represented by circles, transitions by bars, the initial marking by dots inside circles, and the firing rate function is given as an additional description of transitions), the state graph is shown in Fig.1b, and the derivation of the set $S(\mathbf{T})$ of reachable states is shown in Tab.1.

s_i	m_i			f_i		t_k	m_{ij}			$e_{m_{ij}}$		s_j
	1	2	3	1	2		1	2	3	1	2	
0	0	0	0	1	1	1	1	0	1	0	1	1
						2	0	1	0	0	0	2
1	0	0	1	0	2	2	0	1	1	1	0	0
2	0	1	0	1	0	1	1	1	1	1	1	0

Tab.1. The set of reachable states for $\bar{m} = [1, 1, 1]$.

For the same net, and for initial marking $\bar{m} = [1, 4, 1]$, the state graph is shown in Fig.1c, and the set $S(\mathbf{T})$ of reachable states is derived in Tab.2.

s_i	m_i			f_i		t_k	m_{ij}			$e_{m_{ij}}$		s_j
	1	2	3	1	2		1	2	3	1	2	
0	0	3	0	1	1	1	1	3	1	1	1	1
						2	0	4	0	0	0	2
1	0	2	0	1	2	1	1	2	1	1	1	3
						2	0	3	0	0	0	0
2	0	4	0	1	0	1	1	4	1	1	1	0
3	0	1	0	1	3	1	1	1	1	1	1	4
						2	0	2	0	0	0	1
4	0	0	0	1	4	1	1	0	1	0	1	5
						2	0	1	0	0	0	3
5	0	0	1	0	5	2	0	1	1	1	0	4

Tab.2. The set of reachable states for $\bar{m} = [1, 4, 1]$.

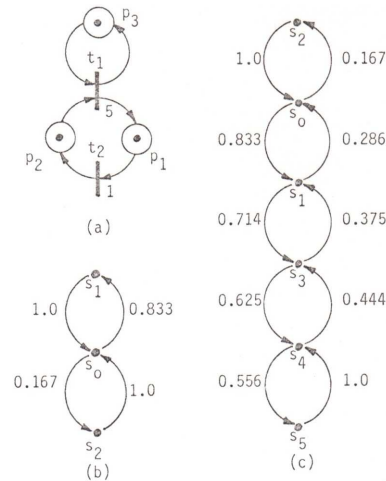


Fig.1. Timed Petri net and its state graphs.

The steady-state probabilities $x(s)$ of the states $s \in S(\mathbf{T})$ are obtained from the state-transition probabilities by solving a system of simultaneous linear equations

$$\begin{cases} \sum_{(s_j, s-i) \in D} b(s_j, s_i) * x(s_j) = x(s_i); & i = 1, \dots, K; \\ \sum_{0 \leq i \leq k} x(s_i) = 1 \end{cases}$$

where $K + 1$ is the number of states in the set $S(\mathbf{T})$.

Moreover, the normalized probabilities $y(s)$ which describe the probability that at any moment of time the net \mathbf{T} is in the state s , $s \in S(\mathbf{T})$, can be obtained from the steady-state probabilities $x(s)$:

$$y(s) = x(s) * h(s) / \sum_{u \in S(\mathbf{T})} x(u) * h(u),$$

where $h(s)$ is the average time spent in the state $s = (m, f)$:

$$h(s) = 1 / \sum_{t \in T} r(t) * f(t).$$

For the two initial markings introduced in the previous example, the steady-state probabilities $x(s)$ and the normalized probabilities $y(s)$ are given in Tab.3 and Tab.4, respectively.

s_i	$x(s_i)$	$h(s_i)$	$y(s_i)$
0	0.500	0.167	0.271
1	0.417	0.500	0.675
2	0.083	0.200	0.054

Tab.3. Steady-state probabilities for $\bar{m} = [1, 1, 1]$.

s_i	$x(s_i)$	$h(s_i)$	$y(s_i)$
0	0.046	0.167	0.055
1	0.134	0.143	0.137
2	0.008	0.200	0.011
3	0.255	0.125	0.228
4	0.359	0.111	0.285
5	0.199	0.200	0.285

Tab.4. Steady-state probabilities for $\bar{m} = [1, 4, 1]$.

4. PERFORMANCE EVALUATION

The simplest closed-network model of an interactive system is shown in Fig.2a. It contains one central server with a queue of waiting jobs, and n terminals. To simplify the solution, it is assumed that all jobs are statistically identical, that the jobs are served by the First-Come-First-Served discipline, and that the service times and the terminal times (or "thinking" times) are exponentially distributed. Under these assumptions the number of jobs in the system (i.e., in the waiting queue and in the server) is a finite continuous-time Markov chain with $n+1$ states, and its state-transition-rate diagram is shown in Fig.2b. Using well known formulas [2] and the average service time equal to 0.2 time unit, the average terminal time equal to 1 time unit (or the service rate equal to 5 jobs/time unit and the terminal rate equal to 1 job/time unit), and n equal to 2, the numerical values of some performance measures are as follows:

the equilibrium probability that the system is idle	0.675
the utilization of the system	0.325
the average throughput rate	1.625
the average turnaround time	1.231
the average time spent in the system	0.231

On the other hand, the same system can be modeled by the timed Petri net shown in Fig.1a. The transition @t sub 1@ corresponds to the central server (with the service rate or the firing rate equal to 5), the place p_2 models the waiting queue, the transition @t sub 2@ corresponds to the terminals (with the terminal rate or the firing rate equal to 1), and the initial number of tokens in the places p_1 and p_2 represent the number of terminals in the system, n . The remaining place p_3 and its initial number of tokens model the number of servers, in this case 1. For the initial marking $\bar{m} = [1, 1, 1]$ there are 3 states of the net (Tab.1) and 3 states of the Markov chain since $n = 2$ (Fig.2b). The state s_0 corresponds to the Markov state 1 (no waiting jobs and busy server), the state s_1 to the Markov state 0 (idle server), and the state s_2 to the Markov state 2 (one waiting job and busy server). Since the server is idle in the state s_1 only, the stationary probability that the system is idle is equal to the normalized probability $y(s_1) = 0.675$ (Tab.3). Then the utilization of the system is immediately $1 - y(s_1) = 0.325$. The average throughput rate can be obtained from the server's load. Since the average service time is equal to 0.2 time units, and the server utilization (for 2 terminals) is equal to 0.325, then the average throughput rate is equal to $0.325/0.2=1.625$ jobs per time unit. Since there are 2 terminals, the average turnaround time is equal to $2/1.625=1.231$ time unit. Finally, since the average terminal time is equal to 1 time unit, the average time spent in the system is equal to $1.231-1=0.231$ time unit which means that the average waiting time is only $0.231-0.2=0.031$ time unit.

Similarly, for $n = 5$ terminals, the same performance measures, obtained from Tab.2 and Tab.4, are as follows:

the equilibrium probability that the system is idle	0.285
the utilization of the system	0.715
the average throughput rate	3.575
the average turnaround time	1.399
the average time spent in the system	0.399
the average waiting time	0.199

Some other performance indices can be obtained in a very similar way.

5. CONCLUDING REMARKS

Even the very simple example of an interactive system modeling illustrates the characteristic features of timed Petri nets. Models are usually quite simple, and their parameters correspond in a very natural way to components or activities of the modeled systems (e.g., the number of users, the number of processors). The state space can easily be generated from model specifications, and since the states of the modeling net directly correspond to the "states" of the modeled system, a verification step is provided which is not available in analytical modeling. The class of timed Petri nets discussed in the paper is restricted in several ways (conflict-free, bounded nets), some of the restrictions, however, can be removed easily by appropriate extensions of the formalism. In fact, free-choice nets, and even nets with conflicts, can be handled in a very similar way provided the probabilities of conflicting firings are known and included in the state description. Also, some extensions of the (basic) Petri nets [7] require only minor modifications of the formalism given in sections 2 and 3, but can significantly increase the "modeling power" of extended timed nets.

REFERENCES

1. T. Agerwala, "Putting Petri nets to work"; IEEE Computer Magazine 1979(12)12, pp.85-94.
2. D. Ferrari, "Computer systems performance evaluation"; Prentice-Hall 1978.
3. J. Magott, "Performance evaluation of concurrent systems using Petri nets"; Information Processing Letters 1984(18)1, pp.7-13.
4. P.M. Merlin, D.J. Farber, "Recoverability of communication protocols - implications of a theoretical study"; IEEE Trans. on Communications 1976(24)9, pp.1036-1049.
5. M.K. Molloy, "Performance analysis using stochastic Petri nets"; IEEE Trans. on Computers 1982(31)9, pp.913-917.
6. J.L. Peterson, "Petri nets"; ACM Computing Surveys 1977(9)3, pp.223-252.
7. J.L. Peterson, "Petri net theory and the modeling of systems", Prentice-Hall 1981.
8. C. Ramchandani, "Analysis of asynchronous concurrent systems by timed Petri nets"; Project MAC Technical Report MAC-TR-120, Massachusetts Institute of Technology, Cambridge MA 1974.
9. J. Sifakis, "Use of Petri nets for performance evaluation"; in: Measuring, modelling and evaluating computer systems, North-Holland 1977, pp.75-93.
10. W.M. Zuberek, "Timed Petri nets and preliminary performance evaluation"; Proc. IEEE 7-th Annual Symp. on Computer Architecture, La Baule, France 1980, pp.89-96.
11. W.M. Zuberek, "Application of timed Petri nets to analysis of multiprocessor realizations of digital filters"; Proc. 25 Midwest Symp. on Circuits and Systems, Houghton MI 1982.