3-rd Int. Workshop on Petri Nets and Performance Models (PNPM'89), Kyoto, Japan, December 11-13, 1989, pp.180-186.

PERFORMANCE EVALUATION USING UNBOUNDED TIMED PETRI NETS

W.M. Zuberek

Department of Computer Science
Memorial University of Newfoundland
St.John's, Canada A1B 3X5

A b s t r a c t

Unbounded timed Petri nets discussed in this paper are place-unbounded free-choice place/transition nets with exponentially distributed firing times associated with transitions of a net. In such nets the infinite state space is generated by a finite set of linear equations. The regularity of this linear description can be used for a "projection" (or "folding") of the infinite state space into an equivalent finite representation that can be described by a finite set of non-linear equilibrium equations. The solution of these equations determines the stationary probabilities of the states. Many performance measures can be obtained directly from this stationary solution. Furthermore, such unbounded nets can eliminate the "state explosion" problem of some models by using unbounded but simple approximations to bounded but complex models.

1. INTRODUCTION

There are three basic methods for evaluating the performance of a complex system [Ferr]: (1) measuring the system during its operation, (2) generating the results by a simulation program that "imitates" the behaviour of the system, and (3) solving a mathematical model that captures the essential features of the system. The last of these methods is in many ways the most satisfying one since it provides many insights into dependencies and relationships between system parameters and performance characteristics. However, mathematical modelling is also the most difficult approach as it quite often uses sophisticated mathematical theories to obtain the required solutions. Therefore a number of techniques have been proposed to simplify and "automate" mathematical modelling. Stochastic Petri nets and timed Petri nets are two examples of such techniques.

Petri nets [Ager,Pete,Reis] are abstract models of systems with interacting, concurrent or parallel components. Multiprocessor systems, distributed databases, communication and computer networks are just a few examples of such systems. However, the original Petri net models do not take into account the duration of modelled activities; the changes of "states" (or markings) of a net are instantaneous events which can represent the "causality" of events, but which neglect any temporal considerations. Consequently, such models (called "ordinary Petri nets") are not complete enough for analysis and evaluation of systems performance. Several different concepts of "timed" Petri nets [HoVe,RaPh,Sifa] and "stochastic" Petri nets [AMCB,DBCT,Moll] have been proposed by assigning (deterministic or stochastic) firing or enabling times to places or transitions of Petri nets.

Timed Petri nets considered in this paper are place-unbounded free-choice Petri nets with stochastic firing times associated with transitions of a net. In timed Petri nets, the firing of an enabled transition is composed of three "conceptual" steps; the first (instantaneous) removes tokens from the input places, the second (temporal) "holds" the removed tokens within the transition for the duration of the firing time, and the third step (instantaneous) moves tokens to all transition's output places [Zub1]. The firing time of each transition is a random variable with the (negative) exponential distribution. Consequently, the holding time of a "state" (which is described by a distribution of tokens in places as well as firing transitions of a net) is also an exponentially distributed random variable. The state graph of a timed net (or the graph of reachable states [Zub1]) is thus a discrete-state continuous-time Markov process. For an unbounded net, this implied Markov process has infinitely many states. Therefore, in order to effectively find the stationary properties of unbounded nets, this infinite space must be somehow reduced (or "folded") into an equivalent finite representation. The "regularity" of state space can be used for such reduction, and for formulation of "reduced" equilibrium equations from which the stationary probabilities of states can be derived. Many performance measures can easily be obtained from these stationary probabilities.

The paper is organized in seven main sections. Section 2 surveys basic concepts and properties of ordinary Petri nets. Sections 3 and 4 cover place and transition invariants and show their use in boundedness checking. Timed nets are introduced in section 5, while section 6 deals with unbounded timed nets. Performance evaluation using unbounded timed nets is discussed in section 7, and section 8 presents and illustrative example. A few concluding remarks are given in section 9.

2. BASIC PROPERTIES

An (ordinary) place/transition 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, and A is a set of directed arcs, $A \subset P \times T \cup T \times P$. For each element of P (and T), the input and output sets denote all those elements of T (and P) which are connected by directed arcs to and from this element, respectively, i.e., $Inp(p) = \{t \in T | (t, p) \in A\}$, etc.

A place p is shared if its output set contains more than one transition. A shared place p is free-choice (or extended free-choice [Brau]) iff all transitions sharing it have identical input sets, i.e., iff:

$$\forall (t_i, t_j \in Out(p))\ Inp(t_i) = Inp(t_j)$$

A net is free-choice iff all shared places are free-choice. Only free-choice net are considered in this paper.

A marked net $\mathbf{M}$ is a pair $\mathbf{M} = (\mathbf{N}, m_0)$, where $\mathbf{N}$ is a free-choice place/transition net and m_0 is the initial marking function which assigns a nonnegative number of so called "tokens" to each place of a net, $m_0 : P \to \{0, 1, 2, ...\}$.

In a marked net a transitions t is enabled if all its input places contain at least one token. Let the set of all transitions enabled by a marking function m be denoted by $E(m)$.

Furthermore, a marking function $m_j : P \to \{0, 1, ...\}$ dominates a marking function $m_i \to \{0, 1, 2, ...\}$, $m_j \trianglerighteq m_i$, iff $m_j \neq m_i$ and for all $p \in P$, $m_j(p) \geq m_i(p)$; furthermore, m_j strongly dominates m_i, $m_j \triangleright m_i$ iff $m_j \trianglerighteq m_i$ and for all $p \in P$, if $m_j(p) > m_i(p)$ then $m_i(p) > 0$.

Property 1: In a marked net $\mathbf{M}$, for all marking functions m_i and m_j, if $m_j \trianglerighteq m_i$ then $E(m_j) \supseteq E(m_i)$ and if $m_j \triangleright m_i$ then $E(m_j) = E(m_i)$.

The property follows from the definition of $E(m)$, the set of transitions enabled by a marking function m.

Each enabled transition can fire. When a transition t fires, a token is removed from each of its input places (simultaneously), and a single token is added to each of t's output places. This determines a new marking function of a net, a new set of enabled transitions, etc.

A marking function m_j is (directly) reachable from a marking function m_i by firing an enabled transition t_k, $t_k \in E(m_i)$, iff

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

A marking function m_j is (generally) reachable from a marking function m_i in a net $\mathbf{M}$, $m_i \rightsquigarrow m_j$, iff there exists a sequence of marking functions m_{i_0}, m_{i_1}, ..., m_{i_k} such that $m_i = m_{i_0}$, $m_j = m_{i_k}$, and for all $0 \leq l < k$, $m_{i_{l+1}}$ is directly reachable from m_{i_l}. The set of all marking functions reachable from the initial marking function m_0 in a net $\mathbf{M}$ is denoted by $M(\mathbf{M})$.

A place p of a marked net $\mathbf{M}$ is bounded iff there exists a bound on the number of tokens that any marking function of the set $M(\mathbf{M})$ can assign to p:

$$\exists (k > 0) \forall (m \in M(\mathbf{M}))\ m(p) < k.$$

A marked net is bounded iff all its places are bounded. Obviously, the set $M(\mathbf{M})$ of a bounded net is finite.

Property 2: In a marked net $\mathbf{M}$, if a marking function $m_j \in M(\mathbf{M})$ is reachable from a marking function $m_i \in M(\mathbf{M})$, $m_i \rightsquigarrow m_j$, and m_j dominates m_i, $m_j \trianglerighteq m_i$, then $\mathbf{M}$ is unbounded.

To show the property, let m_j be reachable from m_i by a firing sequence $\sigma = t_{i_1} t_{i_2} ... t_{i_k}$. Since $m_j \trianglerighteq m_i$, then $E(m_j) \supseteq E(m_i)$ (by property 1), and, obviously, $t_{i_1} \in E(m_i)$ and $t_{i_1} \in E(m_j)$. Firing t_{i_1} in both m_j and m_i creates marking functions m_{j_1} and m_{i_1}, respectively, such that $m_{j_1} \trianglerighteq m_{i_1}$. Repeating the same argument for consecutive transitions in σ leads to a marking m_{j_k} which is reachable from m_j by the firing sequence σ and which dominates m_j, so the argument can be repeated for m_j and m_{j_k}, etc.

Two marking functions $m_i : P \to \{0, 1, ...\}$ and $m_j : P \to \{0, 1, ...\}$ are unordered [Reis] iff there exist elements p_k and p_l in the set P such that $m_i(p_k) < m_j(p_k)$ and $m_i(p_l) > m_j(p_l)$.

Property 3: For each (bounded or unbounded) net $\mathbf{M}$, the set $M(\mathbf{M})$ contains only a finite number of marking functions that are pairwise unordered.

The property is proven in [Reis].

A net $\mathbf{N} = (P, T, A)$ is regular iff for each transition $t \in T$ the number of input and output places is the same. It should be obvious that each regular net is "conservative" in the sense of preserving the token count, i.e., the total number of tokens in the net; for each marking function m in the set $M(\mathbf{M})$ of a regular net $\mathbf{M}$ the token count is the same.

A net $\mathbf{N_i} = (P_i, T_i, A_i)$ is a P_i-implied subnet of a net $\mathbf{N} = (P, T, A)$ iff:

(1) $T_i \ = \ \{t \in T \mid \exists (p \in P_i)\ (p, t) \in A \vee (t, p) \in A\}$,

(2) $A_i \ = \ A \cap (P_i \times T \cup T \times P_i)$,

Furthermore, if there is a family of subsets P_i of the P, $i = 1, ..., k$, that covers the set P, i.e., such that $\bigcup_{1 \leq i \leq k} P_i = P$, and if all P_i-implied subnets are regular, then the net is bounded as the number of tokens in each subnet cannot change (although some subnets can "overlap" which means that the token count of the net can increase and decrease as the transitions fire but all such changes are within fixed bounds). On the other hand, if a net is unbounded, it must contain at least one unbounded place which cannot belong to any regular subnet of this net. This is further formalized by an elegant concept of net invariants.

3. PLACE INVARIANTS

Each place/transition net $\mathbf{N} = (P, T, A)$ can be conveniently represented by a connectivity (or incidence) matrix $\mathbf{C} : P \times T \to \{-1, 0, +1\}$ in which places correspond to rows, transitions to columns, and the entries are defined as:

$$\forall (p \in P) \forall (t \in T)\ \mathbf{C}[p, t] = \begin{cases} -1, & \text{if } (p, t) \in A \wedge (t, p) \notin A, \\ +1, & \text{if } (t, p) \in A \wedge (p, t) \notin A, \\ 0, & \text{otherwise.} \end{cases}$$

It can be verified that if a marking function m_j is obtained from another marking function m_i by firing a transition t_k then (in vector notation) $m_j = m_i + \mathbf{C}[*, k]$, where $\mathbf{C}[*, k]$ denotes the k-th column of $\mathbf{C}$, i.e., the column corresponding to t_k.

Connectivity matrices disregard "selfloops", that is pairs of arcs (p, t) and (t, p); any firing of a transition t cannot change the marking of p in such a selfloop, so selfloops are neutral with respect to token count of a net. A pure net is defined as a net without selfloops [Reis].

A P-invariant (place-invariant) of a net $\mathbf{N}$ is any positive solution I of the matrix equation

$$tr(\mathbf{C}) \times I = 0,$$

where $tr(\mathbf{C})$ denotes the transpose matrix of $\mathbf{C}$. It follows immediately from this definition that if I_1 and I_2 are P-invariants of $\mathbf{N}$, then also any linear (positive) combination of I_1 and I_2 is a P-invariant of $\mathbf{N}$.

A characteristic P-invariant of a net is defined as a P-invariant which does not have simpler invariants. All characteristic P-invariants I are binary vectors [Reis], $I : P \to \{0, 1\}$.

It should be observed that in a pure net $\mathbf{N}$, each P-invariant I of a net $\mathbf{N}$ determines a P_I-implied subnet of $\mathbf{N}$, where $P_I = \{p \in P \mid I(p) > 0\}$; all nonzero elements of I select the rows of $\mathbf{C}$, and each row corresponds to a place with all input (+1) and output (–1) arcs associated with it. Furthermore, for a characteristic P-invariant, the P_I-implied net is regular since for each transition in the implied subnet the number of incoming arcs (–1) must be equal to the number of outgoing arcs (+1). Consequently, if there is a set of characteristic P-invariants that covers the net (i.e., each place belongs to (at least one) characteristic invariant), then the net is bounded. If a net is unbounded, its all unbounded places cannot belong to any characteristic P-invariants, and since all other invariants are linear combinations of characteristic invariants, unbounded places cannot belong to any P-invariants of a net.

Finding characteristic invariants is a "classical" problem of linear algebra, and there are known algorithms to solve this problem [KrJa].

4. TRANSITION INVARIANTS

A T-invariant (transition-invariant) of a net $\mathbf{N}$ with its connectivity matrix $\mathbf{C}$ is any positive solution J of the matrix equation

$$\mathbf{C} \times J = 0.$$

Similarly to P-invariants, it follows immediately from this definition that if J_1 and J_2 are T-invariants of $\mathbf{N}$, then also any linear (positive) combination of J_1 and J_2 is a T-invariant of $\mathbf{N}$.

A characteristic T-invariant of a net is defined as a T-invariant which does not have simpler invariants. Characteristic T-invariants may not be binary vectors.

Since each column k of the connectivity matrix $\mathbf{C}$ describes the change of the marking function resulting from firing the transition t_k, any T-invariant indicates (by its nonzero elements) the numbers of transition firings that are needed to derive the same marking function as the initial one (if such a sequence of transition firings is feasible in a net, which usually depends upon the function m_0). Characteristic T-invariants simply describe the "shortest" such sequences, and they may contain multiple firings of some transitions.

The T-invariants describe firing sequences which "reproduce" the marking functions; for unbounded nets more interesting are sequences which increase the number of tokens in unbounded places (identified by P-invariant analysis).

A T-(k;n)-vector (i.e., a vector of transition firings which increase the marking of place p_k by n tokens) of a net $\mathbf{N}$ with the connectivity matrix $\mathbf{C}$ is any positive solution J of the matrix equation

$$\mathbf{C} \times J = \mathbf{1}_{k;n},$$

where $\mathbf{1}_{k;n}$ denotes a vector with a single nonzero k-th element equal to n. Again, it should be observed that if J is a T-(k;n)-vector of a net, then $2J$ is a T-(k;2n)-vector for this net, $3J$ is a T-(k,3n)-vector, etc. The case $n = 1$ is of special interest as it implies the other T-(k,x)-vectors. For nets with more than one unbounded place, either there exists an "independent" T-(k;n)-vector for each unbounded place p_k, or subsets of unbounded places must be analyzed simultaneously.

It can be shown that each unbounded net has a finite number of linearly independent T-(k;n)-vectors. More specifically, it follows from the property 3 that the set of all those marking function in the set $M(\mathbf{M})$ which are pairwise unordered, is finite, and all other (reachable) marking functions must dominate one of the pairwise unordered functions. Moreover, it follows from the property 2 that if a marking function m_j is reachable from a marking function m_i and $m_j \trianglerighteq m_i$, then also $m_j + \Delta$, $m_j + 2\Delta$, etc. are reachable from m_i, where $\Delta = m_j - m_i$; this means that all marking functions $m_i + k * \Delta, k = 1, 2, ...,$ belong to the set $M(\mathbf{M})$. But since the set $M(\mathbf{M})$ can only contain a finite number of pairwise unordered marking functions, so $M(\mathbf{M})$ is composed of a finite number of subsets $\{m_i + j * \Delta_i \mid j = 0, 1, ...;\ i = 1, ..., k\}$ (not necessarily disjoint), where each Δ_i is a T-(x;y)-vector.

5. TIMED NETS

In marked nets, the firings of transitions are instantaneous events, and analysis of such nets usually assumes that the firings are performed "one at a time". Modeling of "real" systems must also take into account the duration of systems activities represented by transition firings. Therefore in timed nets [Zub1] each transition takes a "real time" to fire, and at any instant of time the tokens are distributed in places as well as (firing) transitions of a net. The "state" description is similar to marking functions of marked nets, but it contains the second component that describes the firing transitions. Moreover, to provide unambiguous modelling capability, in timed nets each firing starts at the same instant of time at which the transition becomes enabled (but some enabled transitions are disabled without firing, e.g., conflicting transitions in free-choice classes).

A timed net $\mathbf{T}$ is a triple, $\mathbf{T} = (\mathbf{M}, c, f)$, where $\mathbf{M}$ is a marked place/transition net, c is a choice function, $c : T \to [0, 1]$, such that for each free-choice place p, $\sum_{t \in Out(p)} c(t) = 1$, and for all transitions t that do not belong to free-choice classes, $c(t) = 1$, and f is a firing-rate function, $f : T \to \mathbf{R}^+$, which assigns a (positive) firing rate to each transition of the net. It is assumed that the choices within free-choice classes of enabled transitions are independent random variables with discrete distributions described by corresponding probabilities $c(t)$. Furthermore, the firing times of transitions are also independent random variables with (negative) exponential distributions described by the rates $f(t)$.

In timed nets several transitions may initiate their firings at the same instant of time. To describe all such initiations, it is convenient to introduce a "selection function" that indicates all those transitions that can initiate their firings simultaneously. For a marking function m in a net $\mathbf{N}$, a selection function e is any function $T \rightarrow \{0,1,2,...\}$ such that the set of transitions enabled by the marking m'

$$\forall (p \in P)\ m'(p) = \sum_{t \in Out(p)} e(t)$$

is empty [Zub1]. The set of all selection functions of a marking function m is denoted by $Sel(m)$.

A state s of a timed Petri net $\mathbf{T}$ is a pair $s = (m, n)$, where m is a marking function, $m : P \rightarrow \{0,1,2,...\}$, and n is a firing function, $n : T \rightarrow \{0,1,2,...\}$, which indicates all those transitions that initiated their firings but not finished them yet.

An initial state s_i of a net $\mathbf{T}$ is a pair $s_i{=}(m_i, n_i)$ where n_i is a selection function from the set $Sel(m_0)$, $n_i \in Sel(m_0)$, and the marking m_i is determined by

$$\forall (p \in P)\ m_i(p) = m_0(p) - \sum_{t \in Out(p)} n_i(t).$$

A free-choice timed net $\mathbf{T}$ may have several different initial states.

A state $s_j = (m_j, n_j)$ is directly (t_k, e_l)-reachable from a state $s_i = (m_i, n_i)$, $s_i \mapsto s_j$, iff the following conditions are satisfied:

(1) $n_i(t_k) > 0$,

(2) $\forall (p \in P)\ m'_i(p) = m_i(p) + \begin{cases} 1, & \text{if } p \in Out(t_k), \\ 0, & \text{otherwise,} \end{cases}$

(3) $e_l \in Sel(m'_i)$,

(4) $\forall (p \in P)\ m_j(p) = m'_i(p) - \sum_{t \in Out(p)} e_l(t)$,

(5) $\forall (t \in T)\ n_j(t) = n_i(t) + e_l(t) - \begin{cases} 1, & \text{if } t = t_k, \\ 0, & \text{otherwise.} \end{cases}$

The state s_j which is directly (t_k, e_l)-reachable from the state s_i is thus obtained by the termination of a t_k firing (1), updating the marking of a net (2), and then initiating new firings (if any) which are determined by a selection function e_l from the set $Sel(m'_i)$ (3, 4 and 5).

Also, a state s_j is generally reachable from a state s_i in a net $\mathbf{T}$, $s_i \rightsquigarrow s_j$, iff there exists a sequence of states s_{i_0}, s_{i_1}, ..., s_{i_k} such that $s_i = s_{i_0}$, $s_j = s_{i_k}$, and for all $0 \leq l < k$, $s_{i_{l+1}}$ is directly reachable from s_{i_l}. The sequence of transitions $t_{i_1} t_{i_2}, ... t_{i_k}$ that terminate their firings during this sequence of states changes, is called a firing sequence of $s_i \rightsquigarrow s_j$. The set of all states reachable from the initial states of net $\mathbf{T}$ is denoted by $S(\mathbf{T})$.

A timed net $\mathbf{T}$ is bounded iff

$$\exists (k > 0)\ \forall (s = (m, n) \in S(\mathbf{T}) \\ (\forall (p \in P)\ \ m(p) < k)\ \wedge\ (\forall (t \in T)\ n(t) < k).$$

A timed net $\mathbf{T}$ is place-unbounded (or p-unbounded) if there is no bound on the marking component m of states $s = (m, n) \in S(\mathbf{T})$, and it is transition-unbounded (or t-unbounded) if there is no bound on the firing component n of states.

6. UNBOUNDED TIMED NETS

It can easily be shown that any firing sequence of a timed net $\mathbf{T} = (\mathbf{M}, c, f)$ is also a firing sequence in the marked net $\mathbf{M}$ (but the opposite is not true [Zub1]); this is due to the fact that in timed nets some tokens are associated with firing transitions and cannot enable any transitions during the firing periods. Therefore, the state space of a timed net can be mapped onto a subset of the marking space of its marked net. Consequently, a bounded timed net can have unbounded marked net, but an unbounded timed net can never have a bounded marked net.

Property 4: In a timed net $\mathbf{T}$, if a state $s_j = (m_j, n_j)$ is reachable from a state $s_i = (m_i, n_i)$, $s_i \rightsquigarrow s_j$, and m_j strongly dominates m_i, $m_j \rhd m_i$, while n_j dominates n_i, $n_j \trianglerighteq n_i$, then there exist states $s_h = (m_h, n_h)$ and $s_g = (m_g, n_g)$ which are directly (t_k, e_l)-reachable from s_j and s_i, respectively, which means that $m_h - m_g = m_j - m_i$ and $n_h - n_g = n_j - n_i$.

To show this property, it should be observed that both $E(m_j)$ and $E(m_i)$ must be empty sets, so

$$\forall (t \in T)\ \min_{p \in Inp(t)}(m_i(p)) = \min_{p \in Inp(t)}(m_j(p)) = 0.$$

Since $n_j \trianglerighteq n_i$, t_k is any transition for which $n_i(t_k) > 0$, and then (from the definition of state reachability) $m'_j \rhd m'_j$, and (according to property 1) $E(m'_j) = E(m'_i)$. Moreover, $Sel(m'_i) = Sel(m'_j)$ since

$$\forall (t \in T)\ \min_{p \in Inp(t)}(m'_i(p)) = \min_{p \in Inp(t)}(m'_j(p)),$$

e_l is thus any selection function from the Sel sets. The remaining part follows from the definition of state reachability.

Property 5: In a timed net $\mathbf{T}$, if a state $s_j = (m_j, n_j)$ is reachable from a state $s_i = (m_i, n_i)$, $s_i \rightsquigarrow s_j$, and n_j dominates n_i, $n_j \trianglerighteq n_i$, while $m_j = m_i$ or $m_j \rhd m_i$, then the timed net $\mathbf{T}$ is t-unbounded.

To show the property, let s_j be reachable from s_i by a firing sequence $\sigma = t_{i_1} t_{i_2} ... t_{i_k}$. Since $n_j \trianglerighteq n_i$, there must exist states $s_{i_1} = (m_{i_1}, n_{i_1})$ and $s_{j_1} = (m_{j_1}, n_{j_1})$ which are (t_{i_1}, e_{i_1})-reachable from s_i and s_j, respectively, and then it follows immediately from the reachability of states that $m_{j_1} - m_{i_1} = m_j - m_i$ and $n_{j_1} \trianglerighteq n_{i_1}$ since $n_{j_1} - n_{i_1} = n_j - n_i$. The same argument can thus be applied to consecutive transitions in σ, resulting in a state $s_{j_k} = (m_{j_k}, n_{j_k})$ reachable from s_j such that $m_{j_k} - m_j = m_j - m_i$ and $n_{j_k} \trianglerighteq n_j$. Then the same property holds for s_j and s_{j_k}, etc.

Property 6: In a timed net $\mathbf{T}$, if a state $s_j = (m_j, n_j)$ is reachable from a state $s_i = (m_i, n_i)$, $s_i \rightsquigarrow s_j$, and m_j strongly dominates m_i, $m_j \rhd m_i$, while $n_j = n_i$ or $n_j \trianglerighteq n_i$, then the timed net $\mathbf{T}$ is p-unbounded.

To show the property, let s_j be reachable from s_i by a firing sequence $\sigma = t_{i_1} t_{i_2} ... t_{i_k}$. According to property 4, there must exist states $s_{i_1} = (m_{i_1}, n_{i_1})$ and $s_{j_1} = (m_{j_1}, n_{j_1})$ which are (t_{i_1}, e_{i_1})-reachable from s_i and s_j, respectively. Moreover, $m_{j_1} - m_{i_1} = m_j - m_i$ and $n_{j_1} - n_{i_1} = n_j - n_i$. The same argument can thus be applied to consecutive transitions in σ, resulting in a state $s_{j_k} = (m_j, n_{j_k})$ reachable from s_j such that $m_{j_k} \rhd m_j$ and $n_{j_k} - n_j = n_j - n_i$. Then the same property holds for s_j and s_{j_k}, etc.

Moreover, since in timed nets the changes of token distributions are due to transition firings in exactly the same way as in marked nets, the T-invariants and T-vectors describe the same properties of (timed) states as (ordinary) marking functions. If σ is a sequence of transition firings that corresponds to a T-invariant, and if σ is a feasible firing sequence in a timed net $\mathbf{T}$, then there exists a state $s_i \in S(\mathbf{T})$ such that σ reproduces s_i, i.e. σ transforms s_i into s_i. Similarly, if μ is a feasible firing sequence in a timed net $\mathbf{T}$ and it corresponds to a T-(k;h)-vector, then there exists a state $s_i = (m_i, n_i) \in S(\mathbf{T})$ that is transformed by μ into a state $s_j = (m_j, n_j)$ which differs at $m_j(p_k)$ element by h, i.e., $m_j(p_k) - m_i(p_k) = h$. Then, according to properties 5 and 6, the net is unbounded, and its state space is composed of a finite number of infinite classes of states $S_i = \{s_{i,j}, j = 0, 1, ...\}, i = 1, ..., k$. These infinite classes are used in "reduction" of state space.

These reduction classes of states can be identified during systematic generation of state space [Zub2] as shown in the example that follows.

7. PERFORMANCE EVALUATION

The infinite state space S of a place-unbounded timed net $\mathbf{T}$ can be subdivided into one finite (possibly empty) class S_0 and a finite number of infinite disjoint classes of states $S_i = \{s_{i,j} | j = 1, 2, ...\}, i = 1, ..., k$, such that the marking component of states in the same class is linearly increasing, i.e., $s_{i,j} = (m_i + j * \Delta, n_i), j = 1, 2, ...$. Consequently:

(1) $S = \bigcup_{0 \le i \le k} S_i$, and

(2) $\forall (1 \le i < j \le k)\ s_{i,1} \mapsto s_{j,1} \ \Rightarrow\ \forall (l = 2, 3, ...)\ s_{j,l} \mapsto s_{i,l}$

For the solution of stationary state probabilities, this infinite space can be reduced (or "folded") to a finite subset of states composed of three groups:

(1) S_0, the original finite class of states,

(2) $S_x = \{s_{i,1}, i = 1, ..., k\}$, the "bottom" layer of "folding",

(3) $S_y = \{s_{i,2}, i = 1, ..., k\}$, the "second" layer of "folding".

Stationary probabilities $x(s_i)$ of the states s_i, $i = 1, 2, ...$, are obtained by solving the following system of (nonlinear) equilibrium equations:

$$\begin{cases} \sum\limits_{s_j \in S_0 \cup S_x} q(s_j, s_i) * x(s_j)/h(s_j) = x(s_i)/h(s_i);\ s_i \in S_0 \\ \sum\limits_{s_j \in S_0 \cup S_x} q(s_j, s_i) * x(s_j)/h(s_j) + \\ \qquad \rho \sum\limits_{s_j \in S_y} q(s'_j, s_i) * x(s'_j)/h(s'_j) = x(s_i)/h(s_i);\ s_i \in S_x \\ (1-\rho) \sum\limits_{s_i \in S_0} x(s_i) + \sum\limits_{s_j \in S_x} x(s_j) = 1 - \rho \end{cases}$$

where $q(s_i, s_j)$ are transition probabilities between the states s_i and s_j, $h(s_i)$ is the average holding time in the state s_i, for each state s_j in S_y, s'_j denotes its corresponding state in S_x (i.e., its "image" in S_x), and ρ is an additional unknown which is the rate of geometric distributions within the infinite classes of states.

Many performance measures can be derived from stationary probabilities of states, as shown in the following example.

For nets with multiple unbounded places, the implied Markov chains are multidimensional (but still very regular), so the "folding" process must be performed for each unbounded place. The resulting system of reduced equations will contain an additional unknown (ρ_j) for each unbounded place.

8. EXAMPLE

The unbounded timed net $\mathbf{T_1}$ shown in Fig.1 (as usual, the firing rate f and the choice c functions are shown as additional descriptions of transitions) is a simple open network model in which a "source" with exponentially distributed interarrival times is represented by p_1 and t_1 (the arrival rate $f(t_1)$ is equal to 1 arrival per time unit), and the remaining part of the net models a single channel server; the number of channels is determined by the initial marking of the place p_3, in this case equal to 1. The place p_2 represents a queue of "jobs" waiting for the server. The server is composed of two consecutive stages. The first stage (t_2) provides exponentially distributed service with the rate equal to 4. The second stage (t_3 and t_4) has service times distributed hyperexponentially; the service rate is equal to 5 with probability 0.75 (t_3) and 2 with probability 0.25 (t_4); the place p_4 is a free-choice place. The total service time is thus hypoexponentially distributed with corresponding parameters.

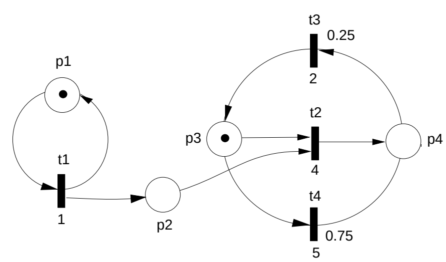

Fig.1. Unbounded timed Petri net $\mathbf{T}$.

The connectivity matrix $\mathbf{C}$ of the net from Fig.1 is as follows:

$\mathbf{C}$	t_1	t_2	t_3	t_4
p_1	0	0	0	0
p_2	+1	–1	0	0
p_3	0	–1	+1	+1
p_4	0	+1	–1	–1

There is one trivial P-invariant [1,0,0,0] which corresponds to the selfloop composed of p_1 and t_1, and one nontrivial characteristic P-invariant $[0, 0, 1, 1]$ which determines the subnet composed of p_3, p_4, t_2, t_3 and t_4, as shown in Fig.2. It should be observed that the place p_2 cannot belong to any P-invariant as it introduces a unique nonzero element in the column t_1. p_2 is an unbounded place, and its increasing token count corresponds to consecutive firings of t_1.

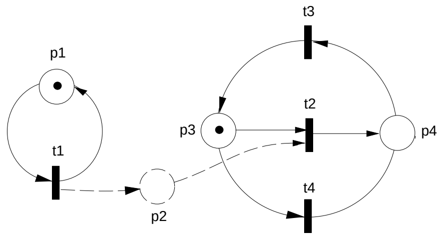

Fig.2. P-invariants of the net **T**.

There are two characteristic T-invariants for the net from Fig.1. One is [1,1,1,0] with a firing sequence (t_1, t_2, t_3); the second is [1,1,0,1] and it corresponds to a firing sequence (t_1, t_2, t_4). It can easily be checked that both these firing sequences reproduce the initial marking function m_0. Furthermore, there is one T-(2,1)-vector [1,0,0,0]; as noted earlier, each firing of t_1 increases by 1 the number of tokens assigned to p_2.

The state space of this net is infinite because of the unbounded place p_2; since the firing times of t_2, t_3 and t_4 are exponentially distributed random variables, the probability that there will be k firings of t_1 before a termination of t_2 (or t_3 or t_4) firing is nonzero, and this is true for any value of k. Therefore, the state space $S(\mathbf{T})$ contains groups of states that differ only in the marking of p_2. All such states can be "reduced" to a single state with a stationary probability that is the sum of a corresponding infinite geometric series.

The derivation of the (reduced) state space for **T** is shown in Tab.1, in which s_i and s_j are the present and the next states, respectively, m_i and n_i describe the distributions of tokens in places (m_i) and firing transitions (n_i), $h(s_i)$ if the average holding time in the state s_i, and $q(s_i, s_j)$ is the probability of transition from the state s_i to the state s_j; all details of this derivation are given in [Zub1] and [Zub2].

The initial part of the (infinite) state graph for **T** is shown in Fig.3. It can be observed that this graph has a very regular structure, and that a three-state section (6-5-7, 9-8-10, ...) is repeatedly added to a four-state "basis" (1-3-2-4). (The numbering of states is quite irrelevant; actually it is generated by a net analyzing program, which uses a rather complicated scheme for assigning consecutive state numbers.) The infinite classes of "similar" states are {6,9,12,...}, {5,8,11,...} and {7,10,13,...}.

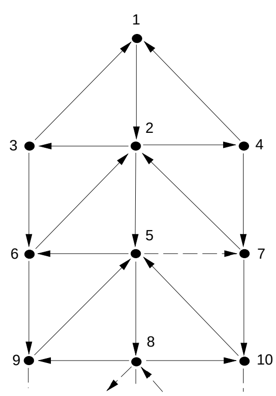

Fig.3. Initial part of the state graph for **T**.

The infinite set of states $S(\mathbf{T})$ can be subdivided into one finite class $S_0 = \{s_1, s_3, s_2, s_4\}$, and three infinite classes of "similar" states, $S_1 = \{s_6, s_9, s_{12}, ...\}$, $S_2 = \{s_5, s_8, s_{11}, ...\}$

Tab.1. The set of reachable states for **T**.

s_i	m_i 1 2 3 4	n_i 1 2 3 4	$h(s_i)$	t_k	s_j	$q(s_i, s_j)$
1	0 0 1 0	1 0 0 0	1.000	1	2	1.000
2	0 0 0 0	1 1 0 0	0.200	2	3	0.600
					4	0.200
				1	5	0.200
3	0 0 0 0	1 0 0 1	0.167	3	1	0.833
				1	6	0.167
4	0 0 0 0	1 0 1 0	0.333	5	1	0.667
				1	7	0.333
5	0 1 0 0	1 1 0 0	0.200	2	6	0.600
					7	0.200
				1	*8	0.200
6	0 1 0 0	1 0 0 1	0.167	4	2	0.833
				1	*9	0.167
7	0 1 0 0	1 0 1 0	0.333	3	2	0.667
				1	*10	0.333
*8(5)	0 2 0 0	1 1 0 0	0.200	2	*9	0.600
					*10	0.200
				1	**11	0.200
*9(6)	0 2 0 0	1 0 0 1	0.167	4	5	0.833
				1	**12	0.167
*10(7)	0 2 0 0	1 0 1 0	0.333	3	5	0.667
				1	**13	0.333

and $S_3 = \{s_7, s_{10}, s_{13}, ...\}$. Then $S_x = \{s_6, s_5, s_7\}$, $S_y = \{s_9, s_8, s_{10}\}$, and the reduced (finite) state graph is shown in Fig.4, in which broken lines indicate the arcs introduced by during reduction; the transition probabilities associated with these new arcs are "reduced" ρ times.

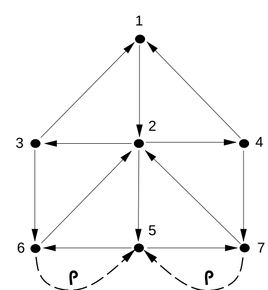

Fig.4. Reduced state graph for **T**.

The set of 8 equilibrium equations (with $x(s_i)$ simplified to x_i) is:

$$
\left\{
\begin{array}{ll}
s_1: & 0.833 * x_3/0.167 + 0.667 * x_4/0.333 = x_1/1.0 \\
s_2: & 1.0 * x_1/1.0 + 0.833 * x_6/0.167 + 0.667 * x_7/0.333 = \\
 & \hfill x_2/0.2 \\
s_3: & 0.6 * x_2/0.2 = x_3/0.167 \\
s_4: & 0.2 * x_2/0.2 = x_4/0.333 \\
s_5: & 0.2 * x_2/0.2 + 0.833 * \rho * x_6/0.167 + \\
 & \hfill 0.667 * \rho * x_7/0.333 = x_5/0.2 \\
s_6: & 0.167 * x_3/0.167 + 0.6 * x_5/0.2 = x_6/0.167 \\
s_7: & 0.333 * x_4/0.333 + 0.2 * x_5/0.2 = x_7/0.333 \\
\text{and}: & (1-\rho)(x_1 + x_2 + x_3 + x_4) + (x_5 + x_6 + x_7) = 1 - \rho
\end{array}
\right.
$$

The solution is $\rho = 0.483$, $x_1 = 0.4717$, $x_2 = 0.1489$, $x_3 = 0.0744$, $x_4 = 0.0496$, $x_5 = 0.0562$, $x_6 = 0.1319$ and $x_7 = 0.0353$. The remaining probabilities can be obtained from recursive formulas, e.g., $x_8 = \rho * x_5$, etc.

Since the server is idle only in the state s_1 ($m_1(p_3) \neq 0$ in Tab.1), the utilization of the server is equal to 1-x_1=0.5283. The probability that there is at least one job waiting (in p_2) for service is equal to $1 - x_1 - x_2 - x_3 - x_4 = 0.2554$ since only $m_1(p_2) = m_2(p_2) = m_3(p_2) = m_5(p_2) = 0$ (see Tab.1), an so on.

Many other performance measures can be derived in a very similar way.

9. CONCLUDING REMARKS

It has been shown that a class of unbounded timed nets can be used in modelling and performance evaluation of systems with parallel and concurrent activities. This opens a new direction in applications of timed Petri nets as all existing approaches [AMCB,DBCT,HoVe,RaPh] assume that the modelling nets are bounded, and their state spaces are finite.

Unbounded net models can also be used in approximate analysis of bounded models which very large state spaces, eliminating the so called “state explosion” problem. It is known that quite often the number of reachable states increases rapidly with the number of tokens introduced by the initial marking function. For example, introducing a bound on the capacity of the waiting queue in the model from the previous section may result in a net shown in Fig.5, in which “immediate” transitions t_5 and t_6 (immediate transitions fire instantaneously, i.e., in zero time [Zub1,Zub2]) are used for queue operations, and the capacity of the queue is represented by parameter K in place p_6. It should be obvious that the model is finite (it is covered by three regular subnets implied by three characteristic p-invariants) but the number of reachable states grows almost exponentially with the value of K. Therefore, for large values of K, it may be more practical to derive approximate performance indices from an unbounded model (Fig.1) restricting the analysis to a subset of reachable states, for example, the states $s_i = (m_i, n_i)$ with the $m_i(p_2)$ value (the number of waiting jobs) not grater than K.

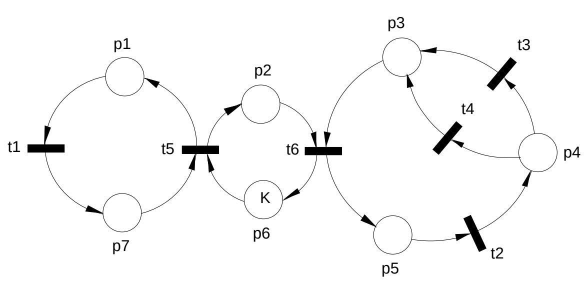

Fig.5. Bounded timed Petri net $\mathbf{T}'$.

The presented approach may seem rather complicated and thus not very practical, it should be noted, however, that the approach can easily be implemented as a computer program, and then all the detailed state descriptions and state transitions can be “invisible” for users, as is presently the case for bounded net models.

Acknowledgement

The Natural Sciences and Engineering Research Council of Canada partially supported this research through Operating Grant A8222.

References

[Ager] T. Agerwala, “Putting Petri nets to work”; IEEE Computer Magazine, vol.12, no.12, pp.85-94, 1979.

[AMCB] M. Ajmone Marsan, G. Conte, G. Balbo, “A class of generalized stochastic Petri nets for the performance evaluation of multiprocessor systems”; ACM Trans. on Computer Systems, vol.2, no.2, pp.93-122, 1984.

[Brau] W. Brauer (ed.), “Net theory and applications”; Proc. of the Advanced Course on General Net Theory of Processes and Systems, Hamburg 1979 (Lecture Notes in Computer Science 84), Springer Verlag 1980.

[DBCT] J.B. Dugan, A. Bobbio, G. Ciardo, K. Trivedi, “The design of a unified package for the solution of stochastic Petri net models”; Proc. Int. Workshop on Timed Petri Nets, Torino, Italy, pp.6-13, 1985.

[Ferr] D. Ferrari, “Computer systems performance evaluation”; Prentice-Hall 1978.

[HoVe] M.A. Holliday, M.K. Vernon, “A generalized timed Petri net model for performance evaluation”; Proc. Int. Workshop on Timed Petri Nets, Torino, Italy, pp.181-190, 1985.

[KrJa] F. Krueckeberg, M. Jaxy, “Mathematical methods for calculating invariants in Petri nets”; in: “Advances in Petri Nets 1987” (Lecture Notes in Computer Science 266), G. Rozenberg (ed.), pp.104-131, Springer Verlag 1987.

[Moll] M.K. Molloy, “Performance analysis using stochastic Petri nets”; IEEE Trans. on Computers, vol.31, no.9, pp.913-917, 1982.

[Pete] J.L. Peterson, “Petri net theory and the modeling of systems”; Prentice-Hall 1981.

[RaPh] R.R. Razouk, C.V. Phelphs, “Performance analysis using timed Petri nets”; in: “Protocol Specification, Testing, and Verification IV” (Proc. of the IFIP WG 6.1 Fourth Int. Workshop, Skytop Lodge PA, June 11-14, 1984); Y. Yemini, R. Strom, S. Yemini (eds.), pp.561-576, North-Holland 1985.

[Reis] W. Reisig, “Petri nets - an introduction” (EATCS Monographs on Theoretical Computer Science 4); Springer Verlag 1985.

[Sifa] J. Sifakis, “Use of Petri nets for performance evaluation”; in: “Measuring, modelling and evaluating computer systems”, pp.75-93, North-Holland 1977.

[Zub1] W.M. Zuberek, “M-timed Petri nets, priorities, preemptions, and performance evaluation of systems”; in: “Advances in Petri Nets 1985” (Lecture Notes in Computer Science 222), G. Rozenberg (ed.), pp.478-498, Springer Verlag 1986.

[Zub2] W.M. Zuberek, “On generation of state space for timed Petri nets”; Proc. ACM Annual Computer Science Conf., Atlanta GA, pp.239-248, 1988.