THROUGHPUT ANALYSIS IN TIMED COLORED PETRI NETS

W.M. Zuberek

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

Abstract

It is shown that the steady-state behavior of a class of timed colored Petri nets can be analyzed using the concept of throughput and simple rules of operational analysis. Since such analysis is based on the average values of firing times rather than firing time distribution functions, the same approach can be applied to a variety of net models. Throughput analysis is based on structural properties of nets only, so it avoids the potential complexity of reachability analysis, and it can be used for analysis of unbounded nets.

1. INTRODUCTION

Modeling and analysis of systems of events in which it is possible for some events to occur concurrently, but there are constraints on the precedence or frequency of these occurrences are typical applications of Petri nets [P81, A84, R85, M89]. Examples of such systems include multiprocessor and distributed systems, communication networks and data flow architectures. Traditional methods, developed for analysis of systems with sequential behavior, are insufficient or inadequate for accurate description of concurrent systems.

In order to study performance aspects of Petri net models, the duration of activities must also be taken into account and included into model specifications. Several types of Petri nets 'with time' have been proposed [A84, S77, Z21] by assigning firing times to the transitions of Petri nets or to places of a net. It should be noted, however, that the basic difference between these two approaches in not in association of time but in the mechanism of transition firings. In t-timed nets [Z21], transition firings are not instantaneous events; a firing occurs in 'real-time', i.e., the tokens are removed from input places at the beginning of the firing period, and they are deposited to the output places at the end of this period (sometimes this is called a 'three-phase' firing mechanism). In p-timed nets [A84], the firing time determines the period of time that tokens must 'wait' before firing may occur, while the firing is an instantaneous event, as in ordinary nets (so the 'firing times' are in fact 'enabling times'). A simple consequence of this difference is that in p-timed nets, the conflict resolution policies cannot be defined independently of timing specifications, while in t-timed nets conflict resolution and timing specification are independent aspects. Only t-timed nets are discussed in this paper.

In colored Petri nets [R87], information can be associated with individual tokens. This associated information is called a 'color' of a token. Token colors can be quite complex, for example, they can describe the contents of a message package or the contents of a database, but typically they are rather simple, representing a class of jobs or a type of a message in a communication systems, etc. Token colors can be modified by (firing) transitions and also the conditions enabling transitions can be different for different colors.

Attributes associated with tokens result in net models that contain much fewer places and transitions than would be required in 'ordinary' Petri nets. The basic idea behind colored nets is to 'fold' (similar) sections of an ordinary Petri net; the original set of places is partitioned into a set of disjoint classes, and each class is replaced by a single colored place with colors indicating which of the original places the tokens belong to. Similarly, the original set of transitions is partitioned into a set of disjoint classes, and each class is replaced by a single colored transition with the occurrence colors indicating which of the original transitions the occurrences belong to. In this sense, 'ordinary' Petri nets correspond to colored nets in which there is just one color (and then the color can be simply ignored).

Analysis of net models can be performed on the basis of the behavior of the net, or on the basis of its structure. The behavior of a net is usually represented by the space of reachable states; analysis of this state space is known as the reachability analysis. Although reachability analysis is quite general (e.g., it can easily handle state-dependent routing probabilities as well as state-dependent timing properties), it becomes very inefficient when the state space is large (for some models, the number of states grows exponentially with model parameters, which is known as the 'state explosion problem'). Moreover, reachability analysis is usually restricted to bounded nets. Therefore, other approaches are recently gaining popularity, and approaches based on structural properties of models are especially attractive because they avoid the derivation of the state space. Such structural methods of analysis cannot provide as much information as the reachability approach does, however, quite often all the detailed results of reachability analysis are not really needed, and more synthetic performance measures, which can also be provided by structural approach, are quite satisfactory.

Structural approach to analysis of net models is similar to operational methods developed for analysis of queueing systems [B76, D78]. It applies to the steady-state of the system; it uses only the first moment (the average values) of random variables, and general relationships and laws which do not depend upon probability distribution functions. One of the basic variables of operational analysis is throughput. Many other measures can be obtained from throughputs, for example, the throughput and the maximum service rate of a component determine its utilization factor, which is in turn an indicator of systems saturation. The throughput is one of the elements used in the Little formula, etc.

The paper first introduces the concept of throughput in timed colored Petri net models and some of its properties, and then applies throughput analysis to performance evaluation of a simple model of a computer systems. The paper is organized in 3 main sections. Section 2 overviews timed colored Petri nets as an extension of marked nets. The concept of throughput is introduced in section 3 and section 4 presents an example of performance analysis based on throughput. A short discussion with identification of a few possible extensions conclude the paper.
2. TIMED COLORED PETRI NETS

Colored Petri nets can be defined as an extension of marked Petri nets, as in [Z80]. However, the arc functions are restricted to 'simple' rather than 'multiple' arcs, so sets rather than multisets are used in the following definition. Also, inhibitor arcs are introduced explicitly.

A simple colored Petri net is a 7-tuple \( N = (P, T, A, C, w, h, m_0) \) where:

- \( P \) is a finite (nonempty) set of places,
- \( T \) is a finite (nonempty) set of transitions,
- \( A \) is a (nonempty) set of directed arcs which connect places with transitions and transitions with places, \( A \subseteq P \times T \cup T \times P \), such that there are no isolated places or transitions,
- \( C \) is a finite (nonempty) set of colors which are associated with places (place or token colors) and transitions (occurrence colors),
- \( w \) is an arc function which assigns to each directed arc from the set \( A \) a mapping from the set of occurrence colors \( C \) into the powerset of place colors \( 2^P \), \( w: A \rightarrow C 

\rightarrow 2^P \),
- \( h \) is an inhibitor function which assigns to each directed arc from the set \( A \) a mapping from the set of occurrence colors \( C \) into the powerset of (place) colors \( 2^P \), \( h: A \rightarrow C 

\rightarrow 2^P \), such that:

\[ \forall (p, t) \in A \) \forall (o \in C \) \forall (p, t)(o) \cap h(p, t)(o) = \Phi \]

where \( \Phi \) denotes the empty set,
- \( m_0 \) is the initial marking function which assigns multisets of colors to places of a net, \( m_0: P \rightarrow C \rightarrow N \).

It can be observed that marked nets correspond to such colored nets in which the set of colors \( C \) contains just one color.

For all occurrence colors \( o \in C \) of all transitions \( t \in T \), the sets of input \( \text{Inp} \), inhibitor \( \text{Inh} \) and output \( \text{Out} \) place colors are defined as:

\[ \text{Inp}(t, o) = \{(p, c) \in P \times C \mid (p, t) \in A \land c \in w(p, t)(o)\} \]
\[ \text{Inh}(t, o) = \{(p, c) \in P \times C \mid (p, t) \in A \land c \in h(p, t)(o)\} \]
\[ \text{Out}(t, o) = \{(p, c) \in P \times C \mid (t, p) \in A \land c \in w(t, p)(o)\} \]

where \( w(p, t)(o) \) and \( h(p, t)(o) \) denote applications of the occurrence color \( o \) to the arc functions \( w \) and \( h \) of the arc \( (p, t) \).

Similarly, for all place colors \( (p, c) \in P \times C \) :

\[ \text{Inp}(p, c) = \{(t, o) \in T \times C \mid (t, p) \in A \land c \in w(t, p)(o)\} \]
\[ \text{Out}(p, c) = \{(t, o) \in T \times C \mid (t, p) \in A \land c \in w(t, p)(o)\} \]

Let any function \( m \) which maps \( P \) into multisets of (place) colors, \( m: P \rightarrow C \rightarrow N \), be called a marking in a net \( N \).

An occurrence color \( o \in C \) of a transition \( t \in T \) is enabled at the marking \( m \) if and only if

\[ \forall (p, c) \in \text{Inp}(t, o) \) \forall (p, c) \in \text{Inh}(t, o) \) \forall (p, c) \in \text{Out}(t, o) \) \forall (p, c) \in \text{Inp}(t, o) \cap \text{Inh}(t, o) \cap \text{Out}(t, o) \]

If an occurrence \( o \) of the transition \( t \) is enabled at the marking \( m \), it can fire. The firing of the occurrence \( (t, o) \) transforms \( m \) into another marking \( m' \), such that for each \( (p, c) \in P \times C \) :

\[ m'(p, c) = \begin{cases} m(p, c) - 1, & \text{if (p, c)} \in \text{Inp}(t, o) \land \text{Out}(t, o) \land \text{Inh}(t, o) \\ m(p, c) + 1, & \text{if (p, c)} \in \text{Out}(t, o) \land \text{Inh}(t, o) \land \text{Inp}(t, o) \\ m(p, c), & \text{otherwise} \end{cases} \]

A place color \( (p, c) \) of a net \( N \) is shared iff there are exist occurrences \( (t_1, o_{1}) \) and \( (t_2, o_{2}) \) such that:

\[ (p, c) \in \text{Inp}(t_1, o_1) \cap \text{Inp}(t_2, o_2). \]

A shared place color \( (p, c) \) is guarded if for all pairs of occurrences \( (t_1, o_1) \) and \( (t_2, o_2) \) sharing \( (p, c) \), there exists another place \( p_0 \) such that:

\[ (p_0, c) \in \text{Inp}(t_1, o_1) \lor (p_0, c) \in \text{Inh}(t_2, o_2) \lor (p_0, c) \in \text{Out}(t_2, o_2). \]

A shared place color \( (p, c) \) is free-choice if all occurrences sharing it have the same input and inhibitor sets, i.e., iff:

\[ \forall (t_1, o_1), (t_2, o_2) \in \text{Out}(p, c) \) \text{Inp}(t_1, o_1) = \text{Inp}(t_2, o_2) \land \text{Inh}(t_1, o_1) = \text{Inh}(t_2, o_2) \]

A net \( N \) is free-choice iff all its shared place colors are either guarded or free-choice. Only free-choice nets are considered in this paper because almost all practical net models are either free-choice or can easily be converted into equivalent free-choice nets.

The relation of sharing a free-choice place is an equivalence relation in the set \( T \times C \), so it implies a partition of this set into a number of free-choice equivalence classes. It can be observed that at any marking \( m \), all occurrences within any free-choice class are simultaneously enabled or they are simultaneously disabled. It is assumed that selections of firing occurrences within free-choice classes of occurrences are random processes which can be described by probabilities associated with occurrences.

In timed colored nets, transition occurrences fire in 'real-time', which means that there is a 'firing time' associated with each occurrence of each transition. This firing time may be deterministic, as in D-timed nets [Z91], or it can be a random variable with some distribution function, for example, negative exponential distribution, as in M-timed nets [Z91]. In the second case, the average value of the firing time is used to characterize the timed behavior.

Any enabled occurrence starts its firing in the same time instant in which it becomes enabled. Each firing can be considered as a three-phase event; first, the (colored) tokens are removed from the input places as indicated by the arc functions of the firing occurrence, the second phase is the firing time period, and when it is finished, (colored) tokens are deposited to output places as indicated by the arc functions of the firing occurrence. If a transition occurrence becomes enabled while it is firing, a new independent firing cycle begins.

A timed colored net is a triple, \( T = (N, u, f) \) where

\[ N \] is a colored net, \( N = (P, T, A, C, w, h, m_0), \]
\[ u \] is a choice function which, for each marking \( m \) of \( N \), assigns the 'choice' probability to each occurrence color of each transition, \( u: T \times C \rightarrow \mathbb{R}^0 \), such that for each free-choice class of occurrences the sum of probabilities is equal to 1,
\[ f \] is a firing-time function which assigns the average firing time to each occurrence color of each transition of the net, \( f: T \times C \rightarrow \mathbb{R}^+ \), where \( \mathbb{R}^+ \) denotes the set of nonnegative real numbers.

The concept of states and derivations of state space for timed colored nets is given in [Z80].

3. THROUGHPUT IN TIMED COLORED NETS

Intuitively, throughput of a place color \( (p, c) \) in a timed net \( T \) is equal to the average number of (colored) tokens entering (or leaving) the place \( p \) in a unit time; in the steady-state of the net, the average numbers of (colored) tokens entering and leaving each place must be equal since no "accumulation" of tokens can occur. Similarly, throughput of an occurrence color \( (t, o) \) is equal to the average...
number of occurrence's firings in a unit time. It should be noted that the throughput of an occurrence does not depend upon the number of incoming or outgoing arcs; in the steady-state, the (average) numbers of tokens removed from each of the input places and deposited to each of the output places in a unit time are the same.

More formally, the throughput of a timed net $T$ is defined as a function $\theta : (P \cup T) \to C \to \mathbb{R}^+$ which assigns a non-negative number to each place color and each occurrence color of the net:

$$\theta(a \in P \cup T) \forall (c \in C) \theta(a, c) = \lim_{n \to \infty} r_n(a,c)$$

where $r_n(a,c)$ denotes the time instant at which the $n$-th occurrence token of color $c$ enters (or leaves) the place $a$.

The "non-accumulation" property of the steady-state implies the following properties of throughput:

Property 1: For each place color $(p,c)$, its throughput is equal to the sum of throughputs of its input occurrence colors as well as the sum of throughputs of its output occurrence colors:

$$\forall (p,c) \in P \times C \exists (t,o) \theta(p,c) = \sum_{t \in (p,c)} \theta(t,o) = \sum_{t \in (p,c)} \theta(t,o)$$

Property 2: For each non-shared place color $(p,c)$, the throughput of $(p,c)$'s output transitions is equal to the throughput of $(p,c)$:

$$\forall (p,c) \in P \times C \exists (t,o) \theta(p,c) = \theta(t,o)$$

Property 3: For each free-choice place color $(p,c)$, throughputs of $(p,c)$'s output transitions are determined by the choice function $u$:

$$\forall (t,o) \in \text{Out}(p,c) \exists (t,o) = u(t,o) \theta(p,c)$$

Property 4: For each place color $(p,c)$ that is shared by two occurrences $(t_1,o_1)$ and $(t_2,o_2)$, and which is guarded by a place $p_3$:

$$p_3 \in \text{In}(t_1,o_1) \implies \theta(t_1,o_1) = \theta(p_3)$$

For each place color $(p,c)$, its throughput is equal to the sum of throughputs of its input occurrence colors as well as the sum of throughputs of its output occurrence colors:

$$\forall (p,c) \in P \times C \exists (t,o) \theta(p,c) = \sum_{t \in (p,c)} \theta(t,o) = \sum_{t \in (p,c)} \theta(t,o)$$

4. EXAMPLE

In Petri net models of service systems (in general sense), places usually represent system queues, transitions servers, directed arcs model the flow of activities in the model as well as synchronisation constraints for concurrent activities; colors are used to represent different classes of activities, and arc functions describe priorities of simultaneous events, queueing disciplines, etc.

The timed net shown in Fig.1 is a typical model of a memory-constrained multiprogramming system [L84] in which colors are used to represent different classes of jobs (or customers) with different priorities of acquiring resources of the system. Only two classes of jobs are considered in this example but any generalization is quite straightforward; these two classes of jobs are denoted by colors 'H' and 'L', for 'high-priority' and 'low-priority' jobs, respectively.

Moreover, another color, 'A' is used for representing the numbers of available resource units.
\[ \theta(t_1, H) = \theta(p_1, H) = m_0(p_1, H) / f(t_1, H), \]
\[ \theta(t_1, L) = \theta(p_1, L) = m_0(p_1, L) / f(t_1, L), \]
and then the remaining (symbolic) throughputs are:
\[ \theta(t_2, H) = \theta(p_2, H) = m_0(p_2, H) / f(t_2, H), \]
\[ \theta(t_2, L) = \theta(p_2, L) = m_0(p_2, L) / f(t_2, L), \]
\[ \theta(t_3, H) = \theta(p_3, H) = m_0(p_3, H) / f(t_3, H), \]
\[ \theta(t_3, L) = \theta(p_3, L) = m_0(p_3, L) / f(t_3, L), \]
\[ \theta(t_4, H) = \theta(t_4, H) = \theta(p_4, H) = \frac{1 - q_1}{f(t_4, H)}, \]
\[ \theta(t_4, L) = \theta(t_4, L) = \theta(p_4, L) = \frac{1 - q_1 + m_0(p_4, H)}{f(t_4, L)}. \]

This solution is valid only for the steady-state of the net. The validity of this solution can be verified by checking the utilization factors of occurrence colors. Generally, for each \((t, o) \in T \times C\), the utilization factor \(u(t, o)\) is equal to the ratio of the throughput \(\theta(t, o)\) to the maximum firing rate, and the maximum firing rates, if they are limited, can be determined from the finite resources of the modeled system. For transitions \(t_5\) and \(t_6\) these maximum firing rates correspond to the maximal utilization of processors, so:
\[ u(t_5) = \frac{\theta(t_5, H) + \theta(t_5, L)}{f(t_5, H) + f(t_5, L)}, \]
\[ u(t_6) = \frac{\theta(t_6, H) + \theta(t_6, L)}{f(t_6, H) + f(t_6, L)}. \]

while the remaining transitions have unlimited maximum firing rate. The values of utilization factors beyond the interval \([0,1]\) indicate that the requirement of steady-state behavior is not satisfied, so the solution is not valid. For example, for \(m_0(p_1, H) = m_0(p_1, L) = m_0(p_2) = 1, f(t_1, H) = 2, f(t_1, L) = 1, f(t_3, H) = f(t_3, L) = 0.1\), and \(q_1 = q_2 = 0.01, w(t_5) = 15\) which clearly indicates that the net is not ergodic; actually, it indicates that the processor represented by \(t_5\) should be at least 15 times faster to handle the service demand.

Many other results can be derived in a similar way.

5. CONCLUDING REMARKS

It has been shown that, for a class of timed color Petri nets, steady-state throughput analysis can be performed on the basis of structural properties of the net rather than the state space. Moreover, since throughput analysis uses the average values of firing times only, the exact firing time distribution function is not important and they may even be unknown. Also, the approach can be applied to nets which have different distribution functions associated with different transitions of the same net; for example, some transitions may have deterministic firing times while other transitions may fire in a stochastic manner.

The timed net used in the example corresponds to an open network of the queueing theory [K90]. It is believed that the proposed approach can be extended to (at least) some classes of closed network models.

The proposed approach can be used for analysis of unbounded nets, then however the verification of the ergodicity condition must be performed. As shown in the example, quite often the solution can be ‘verified’ by checking some additional requirements (e.g., utilization factors or other known performance bounds).

An attractive aspect of throughput analysis is the possibility of obtaining the solutions in symbolic form rather than a numerical value. Different combinations of parameter values, sensitivity of the results with respect to different parameters or functional dependencies can be investigated very conveniently when symbolic solutions are available.

Finally, it should be noticed that the proposed throughput analysis can be automated quite easily and can be integrated with more general tools for analysis Petri net models as one of alternative analysis methods.

Acknowledgement

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

References