PERFORMANCE EVALUATION USING TIMED COLOURED PETRI NETS

W.M. Zuberek

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

Abstract

Coloured Petri nets are Petri nets in which attributes are associated with individual tokens. These attributes are called "colours". The set of colours is finite. Colors can be modified during transition firings, and the same transition can perform different transformations for tokens of different colours. Colours can thus distinguish tokens, and this allows to "fold" similar subnets of a net into a single subnet, reducing the model complexity. In timed coloured nets, the transitions fire in "real-time", i.e., there is a firing—time associated with each colour and each transition of a net. A state description of timed nets is proposed which represents the behaviour of a timed coloured net by a probabilistic state graph. Performance analysis of timed coloured nets is based on stationary probabilities of states.

1. INTRODUCTION

Coloured Petri nets belong to the class of high-level Petri nets [GL81,J83] in which information can be associated with individual tokens. In coloured nets, this associated information is called a "colour" of a token. Token colours can be quite complex, for example, they can describe the contents of a message package or the contents of a database. Token colours can be modified by (firing) transitions and also the conditions enabling transitions can be different for different colours. The attributes attached to tokens result in net models that contain much fewer places and transitions than would be required in "ordinary" Petri nets [J87].

The basic idea of coloured nets is to "fold" 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 coloured place with token colours 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 coloured transition with the occurrence colours indicating which of the original transitions the occurrences belong to.

Any partition of places and transitions will result in a coloured net. One of the extreme partitions will put all original places into one coloured place, and all original transitions into one coloured transition; this will create a very simple net (one place and one transition only) but with quite complicated arc expressions. The other extreme partition will create one-element classes of places and transitions, so the coloured net will be isomorphic to the original net, and since there is only one colour, it contains no information at all. To be useful in practice, it is important that the coloured nets constitute a reasonable balance between the two extreme cases mentioned earlier; places and transitions should be combined if they are similar in some sense.

A coloured net can be represented in two different forms [J87], as a bipartite graph with inscriptions attached to nodes and arcs, or by defining an n-tuple containing sets and functions. The first form is normally used for the initial description of a system and for informal explanation of it, while the second form is used for formal analysis of the system. These two

forms are equivalent in the sense that the formal translations between them exist [J87]. The second representation is used in this paper.

Coloured nets are closely related to other kinds of high–level Petri nets, such as predicate/transition nets [G87], relation nets [R86], and numerical Petri nets [B82,S80].

The paper is organized in 3 main sections. Section 2 introduces coloured Petri nets as an extension of marked nets. Timed coloured nets, and characterization of their behaviour by states and state transitions is given in Section 3. Section 4 discusses performance evaluation using a simple model of a computer system as an example.

2. COLOURED PETRI NETS

Coloured Petri nets can be regarded as an extension of marked Petri nets; to show this extension, the following definition of marked nets can be used.

A marked Petri net is a quadruple $\mathbf{M} = (P, T, A, m_0)$ 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, i.e., for each place $p \in P$ there exist transitions $t_i, t_j \in T$ such that $(t_i, p) \in A$ and $(p, t_j) \in A$, and also for each $t \in T$ there exist places $p_i, p_j \in P$ such that $(p_i, t) \in A$ and $(t, p_j) \in A$,

 m_0 is the initial marking function which assigns tokens to places of a net, $m_0: P \to \mathcal{N}$, where N is the set of nonnegative integer numbers, $\mathcal{N} = \{0, 1, ...\}$.

The definitions that follow frequently use a convenient concept of "multisets". A multiset (or a bag) X over a (nonempty) set A is any function $A \to \mathcal{N}$. Intuitively, a multiset is a "set" which can contain multiple occurrences of the same elements; if X is a multiset over A, then for each $a \in A$, X(a) denotes the number of occurrences of a in X.

A coloured Petri net is a 7–tuple $\mathbf{N}=(P,T,A,C,z,w,m_0)$ 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 transition; moreover, for each $t \in T$, Inp(t) denotes the set of places which are connected by arcs directed to t, and Out(t) the set of places which are connected by arcs directed from t; Inp(p) and Out(p) are defined similarly,

C is a finite (nonempty) set of colours,

z is the colour function which assigns the set of possible token colours to each place of a net, and the set of occurrence colours to each transitions of a net, $z: P \cup T \rightarrow 2^C$,

w is the arc function which associates a linear function from a multiset of occurrence colours into a multiset of token colours with each arc of a net, i.e., for each arc $(p,t) \in A$, $w(p,t) \in \mathcal{L}[(z(t) \to \mathcal{N}) \to (z(p) \to \mathcal{N})]$, and for each arc $(t,p) \in A$, $w(t,p) \in \mathcal{L}[(z(t) \to \mathcal{N}) \to (z(p) \to \mathcal{N})]$, where $\mathcal{L}[X \to Y]$ is the set of all linear functions from X to Y,

 m_0 is the initial marking function which assigns multisets of token colours (or coloured tokens) to places of a net, m_0 : $P \to C \to \mathcal{N}$ such that for all $p \in P$ and all $c \in C$, $m_0(p)(c) = 0$ if $c \notin z(p)$.

This definition is a slightly modified version of a coloured Petri net matrix [J87]; the modification is made in order to emphasize the relation between coloured nets and marked nets. It should be observed that marked nets correspond to such coloured nets in which: (i) the set of colours C contains just one colour, (ii) the colour function z assigns this single colour to all places and all transitions, and (iii) the arc function assigns the identity function to all arcs of the net.

Let any function m that maps P into multisets of token colours, $m: P \to C \to \mathcal{N}$, such that for all $p \in P$ and all $c \in C$, m(p)(c) = 0 if $c \notin z(p)$, be called a marking of the net \mathbf{N} .

An occurrence o of the transition $t \in T$, $o(t) : z(t) \to \mathcal{N}$ is enabled at the marking m if and only if

$$\forall p \in Inp(t) : w(p,t)(o(t)) \le m(p),$$

where w(p,t)(o(t)) denotes the application of the arc function w of the arc (p,t) to the multiset of occurrence colours o(t), and \leq denotes element—wise comparison of multisets.

If an occurrence o(t) of the transition t is enabled at the marking m_i , t can fire; firing of t transforms m_i into another marking m_j which is directly o(t)-reachable (i.e., reachable in "one step") from m_i

$$\forall p \in P : m_j(p) = m_i(p) - \sum_{t \in Out(p)} w(p, t)(o(t)) + \sum_{t \in Inp(p)} w(t, p)(o(t))$$

where \sum is used for element—wise addition of multisets. During t's firing, token colours are removed from t's input places in numbers corresponding to (input) arc functions applied to the multiset of occurrence colours, and token colours are added to t's output places in numbers corresponding to (output) arc functions applied to the same multiset of occurrence colours. In other words, a transition's firing can be seen as a two—phase event, the first phase "transforms" token colours (from input places) into occurrence colours (of the transition), while in the second phase, these occurrence colours are transformed into token colours (of the output places) using functions associated with t's output arcs.

An important concept for analysis of timed nets is the "selection function" which describes all transition occurrences that can fire simultaneously (the selection function is a kind of "maximum step" [J87]).

A selection function g of a marking m is a mapping of T into multisets of occurrence colours, $g: T \to C \to \mathcal{N}$, such that:

- $\forall t \in T \ \forall c \in C : c \notin z(t) \Rightarrow g(t)(c) = 0$, and
- $\forall p \in P : m(p) \ge \sum_{t \in Out(p)} w(p,t)(g(t))$, and
- for all transitions $t \in T$, the only transition occurrence enabled at m' is the null occurrence (i.e., $C \to \{0\}$), where:

$$\forall p \in P : m'(p) = m(p) - \sum_{t \in Out(p)} w(p, t)(g(t)).$$

There may be several different selection functions for a given marking m; the set of all selection functions of a marking m is denoted by Sel(m).

3. TIMED COLOURED PETRI NETS

In timed coloured nets, the transitions fire in "real–time", which means that there is a "firing time" associated with each occurrence colour of each transition. This firing time may be deterministic, as in D–timed nets [Z87], or it can be a random variable with some distribution function, for example, negative exponential distribution, as in M–timed nets [Z86]. Only exponentially distributed random firing times are considered in this paper; the memoryless property of the negative exponential distribution simplifies behavioural description of such timed nets.

In timed coloured nets, the firing of a transition t can be considered as a three-phase event; first, the token colours are removed from t's input places (in numbers corresponding to the input arc functions) and are transformed into occurrence colours of the firing transitions, the second phase is the firing time period when the occurrence colours (created in the phase one) remain "within" the transition t, and in the last phase, occurrence colours are transformed into token colours of t's output places (in numbers corresponding to the output arc functions). As in timed nets [Z86,Z87], if a transition occurrence becomes enabled while the transition is firing, a new independent firing cycle begins. Moreover, it is assumed that the firing times of all occurrence colours of all transitions are independent random variables; this means that there is only one occurrence colour at a time that terminates its firing.

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

N is a coloured net, $\mathbf{N} = (P, T, A, C, z, w, m_0),$

u is a choice function which, for each marking m of \mathbf{N} , assigns the "choice" probability to each selection function g from the set Sel(m), $u:Sel(m)\to\mathbf{R}^{0,1}$, in such a way, that $\sum_{g\in Sel(m)}u(e)=1$,

f is a firing–rate function which assigns the (nonnegative) rate of exponentially distributed firing times to each colour and each transition of the net, $f:T\to C\to \mathbf{R}^+$, where \mathbf{R}^+ denotes the set of nonnegative real numbers.

The behaviour of a timed coloured net can be described by a sequence of states and state transitions. Any state description of a timed net must take into account the marking of a net (i.e., the distribution of token colours in places) as well as the distribution of occurrence colours in firing transitions.

Let any function n that maps T into multisets of occurrence colours, $n: T \to C \to \mathcal{N}$, such that for all $t \in T$ and all $c \in C$, n(t)(c) = 0 if $c \notin z(t)$, be called a firing of the net \mathbf{N} .

A state s of a net $\mathbf{T} = (\mathbf{N}, u, f)$ is a pair s = (m, n) where m is a marking of \mathbf{N} and n is a firing of \mathbf{N} .

A state $s_i=(m_i,n_i)$ is an initial state of $\mathbf{T}=(\mathbf{N},u,f),$ $\mathbf{N}=(P,T,A,C,z,w,m_0),$ if $n_i\in Sel(m_0),$ and

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

Moreover, a state $s_j = (m_j, n_j)$ is directly (t_k, c_ℓ, g_e) -reachable from a state $s_i = (m_i, n_i)$ if and only if:

- $n_i(t_k)(c_\ell) > 0$,
- $g_e \in Sel(m_{ij})$,
- $\forall p \in P : m_j(p) = m_{ij}(p) \sum_{t \in Out(p)} w(p, t)(g_e(t)),$
- $\forall t \in T : n_j(t) = n_i(t) + g_e(t) \begin{cases} \mathbf{1}(c_\ell), & \text{if } t = t_k, \\ 0, & \text{otherwise;} \end{cases}$
- $\forall p \in P : m_{ij}(p) = m_i(p) + \begin{cases} w(t_k, p)(\mathbf{1}(c_\ell)), \\ & \text{if } t_k \in Inp(p), \\ 0, \text{ otherwise.} \end{cases}$

where $\mathbf{1}(c_{\ell})$ is a mapping $C \to \{0,1\}$ which is equal to 1 for $c = c_{\ell}$ and is zero otherwise.

A state s_j is generally reachable from a state s_i in **T** if there exists a sequence of (intermediate) states s_{ℓ_0} , s_{ℓ_1} , ..., s_{ℓ_k} such that $s_i = s_{\ell_0}$, $s_j = s_{\ell_k}$, and for $0 \le h < k$, $s_{\ell_{h+1}}$ is directly reachable from s_{ℓ_h} .

The set of reachable states of a net T, S(T), is the set of all states that are (generally) reachable from initial states of T.

A timed coloured net T is bounded if and only if

$$\exists k > 0 \ \forall (m, n) \in S(\mathbf{T}) \ \forall c \in C :$$

$$(\forall p \in P : m(p)(c) \le k) \land (\forall t \in T : n(t)(c) \le k).$$

If a timed coloured net is bounded, its set of reachable states is finite. Only bounded nets are considered in this paper.

The behaviour of a bounded timed coloured net $\mathbf{T} = (\mathbf{N}, u, r)$ can be represented by a finite probabilistic labeled state graph $\mathbf{G}(\mathbf{T}) = (V, D, h, b)$ where

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

D is the set of directed arcs, $D \subseteq V \times V$, such that $(s_i, s_j) \in D$ if and only if s_j is directly reachable from s_i in \mathbf{T} ,

h is a vertex labeling function, $h:V\to \mathbf{R}^+$, which assigns the average holding time to each state of \mathbf{T} in such a way that if s=(m,n), then

$$h(s) = \frac{1}{\sum_{t \in T} \sum_{c \in C} n(t)(c) * f(t)(c)},$$

b is an arc labeling function, $b:D\to \mathbf{R}^{0,1}$, which assigns the probability of transitions from s_i to s_j to each arc $(s_i,s_j)\in D$ in such a way that if s_j is directly (t_k,c_ℓ,g_e) -reachable from s_i , then

$$b(s_i, s_j) = u(g_e) * \frac{f(t_k)(c_\ell)}{\sum_{t \in T} \sum_{c \in C} n(t)(c) * f(t)(c)}.$$

4. PERFORMANCE EVALUATION

Since in timed nets with exponentially distributed firing times, holding times of states are also exponentially distributed, the behaviour of a timed coloured net is a continuous–time, discrete–state homogeneous Markov process; for a bounded net, it is a finite–state process, so the stationary probabilities of states x(s), $s \in S(\mathbf{T})$, can be obtained by solving a system of simultaneous (linear) equilibrium equations [Klei]:

simultaneous (linear) equilibrium equations [Klei]:
$$\begin{cases} \sum_{1 \leq j \leq K} b(s_j, s_i) * x(s_j) / h(s_j) &= x(s_i) / h(s_i); \quad 1 \leq i < K, \\ \sum_{1 \leq i \leq K} x(s_i) &= 1 \end{cases}$$

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

Many performance, indices such as throughput, average waiting times, etc., can easily be derived from stationary probabilities of states, as shown in the following example.

Example. Timed nets can conveniently be used to model queueing systems; in such models places represent system queues, transitions servers, directed arcs model the flow of activities in the model as well as synchronization constraints for concurrent activities, and arc function are used to describe priorities of simultaneous events, queueing disciples, etc.

Fig.1(a) shows an M-timed inhibitor net [Z86] (as usual, places are represented by circles, transitions by bars, inhibitor

arcs have small circles instead of arrowheads, the initial marking function is indicated by a number of dots in corresponding places, and the firing-rates are given as additional description of transitions) that represents a closed network model of an interactive system with 2 classes of users (and jobs) and a nonpreemptive priority scheduling discipline. The system consists of a central server $(p_1, t_2 \text{ and } t_3)$ with two queues of waiting jobs, for class-1 (p_2) and class-2 (p_4) jobs, respectively, k_1 terminals in class-1 and k_2 terminals in class-2. The class-1 jobs have "higher" priority than the class-2 ones, i.e., they receive service before class-2 jobs (the inhibitor arc (p_2, t_3) disables t_3 if p_2 contains at least one token). It is assumed that all terminal and service times are exponentially distributed, that the average terminal times for class-1 and class-2 jobs are equal to 1 and 2 time units, respectively, and that the average service times are equal to 0.2 and 0.5 time units for class-1 and class-2 jobs, respectively (the numbers may look not very "realistic" but this is an illustrative example only).

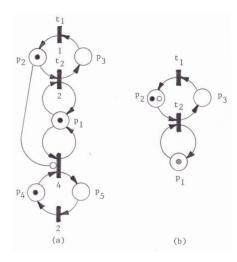


Fig.1. M-timed net (a) and coloured net (b).

An equivalent timed coloured Petri net is shown in Fig.1(b) which clearly illustrates "folding" of class–1 and class–2 subnets into one subnet (t_1, p_2, t_2, p_3) ; the "function" of the inhibitor arc is included into the function associated with the arc (p_2, t_2) . It should be observed that for each additional class of jobs, the net of Fig.1(a) must be extended by another subnet and a number of new inhibitor arcs, while the net of Fig.1(b) remains the same (although the set of colours and some of the functions must be modified).

For the net shown in Fig.1(b), the set C contains three colours, "h" for class-1 jobs ("high" priority), "l" for class-2 jobs ("low" priority), and "a" for available central processor(s). The colour function z maps p_1 into a one-element set {"a"}, and all other places and transitions into { "h", "l"}. The arc function w is the identity function for $(t_2, p_3), (p_3, t_1)$, and $(t_1, p_2); w(p_2, t_2) = \alpha_1 + \rho(m(p_2)(\text{"h"1})) * \alpha_2$, where α_1 and α_2 are projection functions

$$\alpha_1 = \lambda(x, y). \begin{cases} x(\text{``h''}), & \text{if } y = \text{``h''}, \\ 0, & \text{otherwise}; \end{cases}$$

$$\alpha_2 = \lambda(x, y). \begin{cases} x(\text{``l''}), & \text{if } y = \text{``l''}, \\ 0, & \text{otherwise}; \end{cases}$$

and $\rho(x)$ is equal to zero if $x \neq 0$, and is equal to 1 otherwise. Finally, $w(t_2, p_1)$ as well as $w(p_1, t_2)$ is a simple sum of "h" and "l" occurrence colours that is assigned to the colour "a"

$$w(t_2,p_1)=w(p_1,t_2)=\lambda(x,y).\left\{\begin{array}{l} x(\text{``h"})+x(\text{``l"}), \text{if }y=\text{``a"},\\ 0, \text{ otherwise}. \end{array}\right.$$

For $k_1 = k_2 = 1$, the initial marking function m_0 assigns one token "a" to p_1 , and a pair of tokens ("h","l") to p_3 (Fig.1).

The firing-rate function is as follows (Fig.1(a)), $f(t_1)(\text{"h"})=1$, $f(t_1)(\text{"l"})=2$, $f(t_2)(\text{"h"})=2$, and $f(t_2)(\text{"l"})=4$. It appears that the choice function u is always equal to 1 as the sets Sel(m) are singletons for this net.

For $k_1 = k_2 = 1$, there are 5 states of the system, as shown in Tab.1; Tab.1 also shows stationary probabilities of states, x(s).

		m_i	n_i				
		1 2 2 3 3	$1\ 1\ 2\ 2$				
i	$x(s_i)$	ahlhl	hlhl	$h(s_i)$	t_k	c_ℓ	j
1	0.157	00100	0 0 1 0	0.500	2	h	2
2	0.225	00000	1001	0.200	1	h	3
					2	1	4
3	0.056	01000	0001	0.250	2	1	5
4	0.404	10000	1100	0.333	1	h	5
					1	1	2
5	0.157	00000	0110	0.250	1	1	1
					2	h	4

Tab.1. Derivation of the set of states.

Since there are two classes of jobs, performance analysis can be done with respect to each class as well as for the whole system. The stationary probability that the system is idle is equal to $x(s_4)=0.404$ (since s_4 is the only state in which $m_i(p_1)$ ("a") $\neq 0$), and then the utilization of the system is equal to $1-x(s_4)=0.596$ which is composed of 0.314 for class-1 jobs $(x(s_1)+x(s_5))$ since $n_1(t_2)(h)=n_5(t_2)(h)=1$) and 0.282 for class-2 jobs $(x(s_2)+x(s_3))$. The throughput rates are equal to 0.314/0.5=0.628 for class-1 and 0.282/0.25=1.128 for class-2 jobs; for the whole system, the average throughput rate is thus 0.628+1.128=1.756 jobs per time unit. The average turnaround times are equal to 1.592 and 0.887 for class-1 and class-2 jobs, respectively, and then the average waiting times are equal to 0.092 and 0.137 time units for class-1 and class-2, respectively.

The differences between these two classes of jobs are more significant for increased "traffic" in the system. The same performance measures are shown below for $k_1 = k_2 = 1$ and $k_1 = k_2 = 3$:

	$k_1 = 1$	$k_1 = 3$
	$k_2 = 1$	$k_2 = 3$
probability that the system is idle	0.404	0.017
utilization of the system	0.596	0.983
average throughput rate	1.756	2.470
class-1 utilization of the system	0.314	0.731
average class–1 throughput rate	0.628	1.462
average class–1 turnaround time	1.592	2.052
average class-1 waiting time	0.092	0.552
class-2 utilization of the system	0.282	0.252
average class-2 throughput rate	1.128	1.008
average class–2 turnaround time	0.887	2.976
average class–2 waiting time	0.137	2.226

5. CONCLUDING REMARKS

It has been shown that for a class of timed coloured Petri nets, the state space can be derived systematically from net specifications, and than performance characteristics can be obtained from stationary probabilities of states. The derivations

can easily be automated (e.g., they can be performed by appropriate computer programs) and then performance analysis can be provided directly from model descriptions; all details of state descriptions and state transitions can be kept "invisible" to users.

In some cases the modeling nets are unbounded, which means that the state spaces are infinite; in fact, all open network models have infinite state spaces. Analysis of unbounded timed nets is needed not only for analysis of open network models; it can also provide a solution to the "state explosion" problem, i.e., analysis of nets in which the number a states increases very rapidly (e.g., exponentially) with some model parameters. In such cases an approximate analysis of an unbounded model may provide the results more efficiently (an even more accurately) than tedious evaluation of finite but huge state spaces.

The approach presented in this paper is based on analysis of the state space of a net; it belongs to the class of reachability analyses. In many cases properties of a net can be determined from structural analysis, based on net invariants [J81,R86]. Performance analysis based on "timed invariants" could eliminate the generation analysis of the state space of a net.

Acknowledgement

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

References

- [B82] J. Billington, "Specification of the transport service using numerical Petri nets"; in: "Protocol Specification, Testing and Verification II", C. Sunshine (ed.), pp.77-100, North-Holland 1982.
- [G87] H.J. Genrich, "Predicate/transition nets"; in: "Advanced Course on Petri Nets 1986" (Lecture Notes in Computer Science 254), G. Rozenberg (ed.), pp.207–247, Springer Verlag 1987.
- [GL81] H.J. Genrich, K. Lautenbach, "System modelling with high-level Petri nets"; Theoretical Computer Science, vol.13, pp.109–136, 1981.
- [J81] K. Jensen, "Coloured Petri nets and the invariant method"; Theoretical Computer Science, vol.14, pp.317– 336, 1981.
- [J83] K. Jensen, "High-level Petri nets"; in: "Applications and Theory of Petri Nets" (Informatik- Fachberichte 66), A. Pagnoni, G. Rozenberg (eds), pp.166-180, Springer Verlag 1983.
- [J87] K. Jensen, "Coloured Petri nets"; in: "Advanced Course on Petri Nets 1986" (Lecture Notes in Computer Science 254), G. Rozenberg (ed.), pp.248–299, Springer Verlag 1987.
- [Klei] L. Kleinrock, "Queueing systems", vol.1: "Theory", vol.2: "Computer applications"; J. Wiley & Sons 1975, 1976.
- [R86] W. Reisig, "Petri nets an introduction"; Springer Verlag 1985.
- [S80] F.J.W. Symons, "Introduction to numerical Petri nets"; Australian Telecommunications Research, vol.14, no.1, pp.28-32, 1980.
- [Z86] 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.
- [Z87] W.M. Zuberek, "D-timed Petri nets and modelling of timeouts and protocols"; Trans. of the Society for Computer Simulation, vol.4, no.4, pp.331–357, 1987.