MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Project MAC

Computation Structures Group Memo 95

Decision Problems for Petri Nets and Vector Addition Systems

by

Michel Hack

This work was supported by the National Science Foundation under research grant GJ-34671.

March 1974

Decision Problems for Petri Nets and Vector Addition Systems.

Table of contents:

0.	Introduction	3
1.	Definitions and Notations	5
2.	The Equivalence of the GPN and the RPN models	15
3.	Decidable Questions: Boundedness and Coverability	23
4.	An Undecidable Problem about Petri Nets	41
5.	The Liveness and Reachability Problems for Petri Nets	53
Appendix:	König's Infinity Lemma	77
References		78

0. Introduction

In 1968, R. Karp and R. Miller [10] introduced a formalism called Vector Addition Systems to discuss decidability questions about their Parallel Program Schemata. That same year, A. W. Holt introduced Petri Nets [8, 14] to model concurrent behavior in Systems. Both formalisms have been used to model and analyze the structural behavior of asynchronous and parallel systems [5, 6, 10, 11, 16, 17].

The similarity of these two formalisms has been recognized early, but had not been exploited until about 1972, when R. Keller [11] used a generalized form of Petri Nets as a convenient graphical representation for his Vector Replacement Systems, a generalization of Vector Addition Systems. Thus he translated Petri Net concepts such as Liveness -- which he needed to model Asynchronous Systems -- into Vector Replacement Systems terms.

In 1972 also, M. Rabin [2, 15] presented the Unsolvability of the Inclusion Problem for Reachability Sets in Vector Addition Systems in a talk at MIT. From this, two things appeared: (1) There are unsolvable problems about Petri Nets, and (2) The proof could be presented very clearly in Petri Net terms.

In this memo it is our purpose to establish the following results and observations:

- The four formalisms mentioned so far -- Vector Addition Systems, Petri Nets, Vector Replacement Systems, and Generalized Petri Nets -- are equivalent to each other, in the sense that any problem expressed in one formalism can be translated by a standard procedure into another formalism. Thus, the generalization of the original formalisms only buys convenience, not more generality.

- The graphical appeal of Petri Net methods permits a better grasp for intuitive arguments, which can help enormously to find rigorous proofs of various facts.

- Taking advantage of the above observation, we present new proofs of the major decidability results obtained for Vector Addition Systems by Karp and Miller, as well as of Rabin's Undecidability result.

- Finally, we apply our tools to several open questions, and prove the recursive reducibilities between various decidability questions. In particular, we prove the recursive equivalence of the Liveness Problem and the Reachability Problem, and explore some hypotheses which would imply the Undecidability or the Decidability of these problems.

1. Definitions and Notations

We begin by defining the most general concepts of which the earlier definitions are a restricted case.

1.1 Generalized Petri Nets

Definition 1.1: A Generalized Petri Net (GPN) $N = \langle \Pi, \Sigma, F, B, M_0 \rangle$ consists of the following:

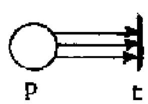
1. a finite set of places, $\Pi = \{p_1, \ldots, p_r\}$
2. a finite set of transitions, $\Sigma = \{t_1, \ldots, t_s\}$ disjoint from Π
3. a forwards incidence function $F: \Pi \times \Sigma \to \mathbb{N}$ ($\mathbb{N}$ is the set of non-negative integers)
4. a backwards incidence function $B: \Pi \times \Sigma \to \mathbb{N}$
5. an initial marking $M_0: \Pi \to \mathbb{N}$

It is represented graphically as follows:

1. places are represented by circles
2. transitions are represented by bars

Fig. 1.1

3. circles and bars are connected by bundles of arcs: if p is a place and t is a transition, and $F(p,t) = 3$, we have a bundle of 3 arcs going from p to t.



or

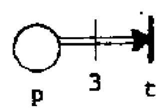
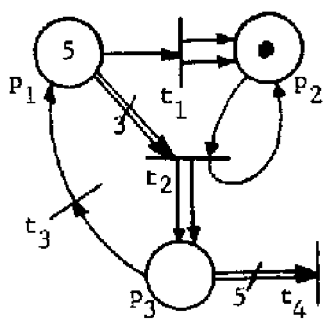


Fig. 1.2

4. a marking is represented by drawing a number of tokens into a place, or writing the number.

Example: $\Pi = \{p_1, p_2, p_3\}$

$\Sigma = \{t_1, t_2, t_3, t_4\}$



$$F = \left\{ \begin{array}{c|cccc} & t_1 & t_2 & t_3 & t_4 \\ \hline p_1 & 1 & 3 & 0 & 0 \\ p_2 & 0 & 1 & 0 & 0 \\ p_3 & 0 & 0 & 1 & 5 \end{array} \right.$$

i.e., $F(p_1, t_1) = 1$

$$B = \left\{ \begin{array}{c|cccc} & t_1 & t_2 & t_3 & t_4 \\ \hline p_1 & 0 & 0 & 1 & 0 \\ p_2 & 2 & 1 & 0 & 0 \\ p_3 & 0 & 2 & 0 & 0 \end{array} \right. \qquad M_0 = \left\{ \begin{array}{|cc} p_1 & 5 \\ p_2 & 1 \\ p_3 & 0 \end{array} \right.$$

Fig. 1.3

For the purpose of modeling Asynchronous Systems, a Petri Net is a dynamic object. The system starts in some initial configuration, and goes through a series of configurations by a sequence of actions. We study the set of possible configurations the System may assume, and the set of possible action sequences the system may go through. Configurations are modeled by markings, action sequences are modeled by firing sequences, and elementary configuration changes (actions) are modeled by the firing of a transition, which changes the marking by removing tokens from some places and adding tokens to some other places. A firing sequence is then denoted by a string over the alphabet of transition names. A Petri Net then represents the structure of the System with a given initial configuration, and the possible evolutions of the System are represented by the set of firing sequences and the set of reachable markings, also called reachability set or marking class.

Vector notation: We can interpret a marking M as a vector with r coordinates, where r is the number of places. Thus, the i^{th} coordinate of M is $M(p_i)$. The distinction will be clear from the context: M is a vector on $\mathbb{N}^r$ and $M(p_i)$ is a non-negative integer. For a given transition t_j, we similarly define an input vector $F(t_j)$ and an output vector $B(t_j)$ as follows:

the i^{th} coordinate of $F(t_j)$ is $F(p_i,t_j)$
$B(t_j)$ $B(p_i,t_j)$

Thus, $F(t_j)$ and $B(t_j)$ are also vectors on $\mathbb{N}^r$. When we look upon markings as r-dimensional vectors, it is sometimes useful to look upon B and F as $r \times s$-matrices, with $F_{i,j} = F(p_i, t_j)$. See definition 1.18 on page 13 for an application.

Firing Relation: We shall interpret a transition as a relation between markings:

Definition 1.2: We write $M[t\rangle M'$, and say that transition t is firable at marking M and leads to marking M', as follows:

$$M [t \rangle M' \equiv (M \geq F(t) \ \& \ M' - M = B(t) - F(t))$$

The relation $\geq$ for vectors is the componentwise greater-or-equal partial order relation on $\mathbb{N}^r$.

We then extend the Firing Relation to a sequence of firings $\sigma = t_i t_j \ldots t_k$ as the composition of the relations corresponding to $t_i, t_j \ldots t_k$. This composition of relations corresponds to the concatenation operation for strings. We thus define a firing sequence as follows:

Definition 1.3: A firing sequence from marking M to marking M' is a string $\sigma t \in \Sigma^*$ defined recursively as follows:

$$M[\sigma t\rangle M' \triangleq \exists M'' \in \mathbb{N}^r : M[\sigma\rangle M'' \ \& \ M''[t\rangle M'$$

If λ stands for the empty string (length zero), it is understood that, $\forall M \in \mathbb{N}^r : M[\lambda\rangle M$.

Now we are ready to define the two most important concepts: the set of firing sequences of a Petri Net, and the set of reachable markings, or marking class.

Definition 1.4: Given a GPN $N = \langle \Pi, \Sigma, F, B, M_0 \rangle$ with initial marking M_0, we define:

$S_N(M_0) = \{\sigma \in \Sigma^* \mid (\exists M \in \mathbb{N}^r) \ M_0[\sigma\rangle M\}$, the set of firing sequences starting at M_0.

$\vec{M}_o = R_N(M_0) = \{M \in \mathbb{N}^r \mid (\exists \sigma \in \Sigma^*) \ M_0[\sigma\rangle M\}$, the set of reachable markings from M_0, or the marking class, or the reachability set.

Note: The reachability set of a net N could of course be written R(N), since M_0 is part of the specification of N. But it is advantageous to show its dependence on M_0 in particular. This permits us to consider $S_N(M)$ and $R_N(M)$ for the same Net, except for the consideration of an arbitrary initial marking M.

1.2 Ordinary Petri Nets and Self-Loop-Free Petri Nets

Definition 1.5: An ordinary Petri Net is a GPN where the size of arc-bundles is restricted to one. Thus, the only possible values for the incidence functions F and B are zero and one:

$$\left.\begin{array}{l}\forall p\in\Pi\\ \forall t\in\Sigma\end{array}\right\} F(p,t) \le 1 \ \&\ B(p,t) \le 1$$

A Self-Loop-Free Petri Net is a GPN where no place-transition pair is both forwards and backwards connected:

$$\left.\begin{array}{l}\forall p\in\Pi\\ \forall t\in\Sigma\end{array}\right\} F(p,t)\cdot B(p,t) = 0$$

A Restricted Petri Net (RPN) is both ordinary and Self-Loop-Free.

Alternative representations:

For ordinary Petri Nets, the F and B incidence functions are often replaced by a relation $\cdot$ called the dot relation or arc relation over bipartite pairs of places and transitions. $\cdot \subseteq \Pi\times\Sigma \cup \Sigma\times\Pi$, and $\langle p,t\rangle\in\cdot$ is written as $p\cdot t$ and means that an arc goes from place p to transition t. Thus:

$$p\cdot t \equiv F(p,t)=1$$
$$t\cdot p \equiv B(p,t)=1$$

This is the definition used in MAC TR-94 [6] .

For Self-Loop-Free Petri Nets, the two incidence functions F and B can be replaced by a single incidence function T = B-F, where bundles from a transition to a place are represented by positive numbers (the number of tokens one firing adds to that place), and bundles from a place to a transition are indicated by negative numbers (the number of tokens a firing takes away from that place). It can be seen that firability is defined as follows:

$M[t\rangle M' \Leftrightarrow M\geq 0$ & $M'\geq 0$ & $M'-M = T(t)$ (where, of course, $T(t)$ stands for the vector whose components are $T(p_i,t)$.)

In particular, RPN's have a single incidence function whose range is $\{-1,0,+1\}$.

1.3 Vector Replacement Systems and Vector Addition Systems

We give Karp and Miller's original definition of a Vector Addition System below: ($\mathbb{N}$ = non-negative integers; $\mathbb{Z}$ = integers.) [10]

Definition 1.6: An r-dimensional Vector Addition System (VAS) is a pair $\mathcal{W} = \langle q, W\rangle$ in which q is an r-dimensional vector of non-negative integers, and W is a finite set of r-dimensional integer vectors: $q \in \mathbb{N}^r$, $W \subseteq \mathbb{Z}^r$. The reachability set $R(\mathcal{W})$ is the set of all vectors of the form

$$q + w_1 + w_2 + \ldots + w_n \qquad \text{such that, } \forall i \leq n:$$

$$w_i \in W \ \&\ q + \sum_{j=1}^{i} w_j \geq 0$$

Geometrically, in r-coordinate space, $R(\mathcal{W})$ is the set of points reachable from q by successive translations from the set W without ever leaving the first orthant.

Relation to Petri Nets: There is a one-to-one correspondence between VAS's and Self-Loop-Free Petri Nets:

VAS	Correspondence	Self-Loop-Free Petri Net
$\mathcal{W} = \langle q, W\rangle$		$N = \langle \Pi, \Sigma, T, M_0\rangle = \langle \Pi, \Sigma, F, B, M_0\rangle$
(the r "dimensions" of $\mathcal{W}$)		$\Pi = \{p_1, \ldots, p_r\}$
$q \in \mathbb{N}^r$	$q = M_0$	$M_0 \in \mathbb{N}^r$
$W = \{w_1, \ldots, w_s\}$ $W \subseteq \mathbb{Z}^r$	$w_i = T(t_i)$	$\Sigma = \{t_1, \ldots, t_s\}$ $T = B - F$ as defined before
w_i-translation without leaving the first orthant		t_i-firing: $M[t_i\rangle M'$ with $M \geq 0$ & $M' \geq 0$
reachability set $R(\mathcal{W})$	$R(\mathcal{W}) = R_N(M_0)$	reachability markings $R_N(M_0)$ or marking class $\vec{M}_0 = R_N(M_0)$

The isomorphism is quite apparent. Firing sequences were not explicitly defined for a VAS.

R. Keller defines Vector Replacement Systems in the wider context of Transition Systems, where a Transition System is anything having a possibly infinite set of distinguishable states or configurations, a finite set of transitions that describe elementary state changes, and an initial state. In this context, we have: [11]

Definition 1.7: A Vector Replacement System (VRS) $\langle q_0,\Sigma,U,V\rangle$ is a Transition System where:

1. The set of states is $Q \subseteq \mathbb{N}^r$, where r = dimension of the VRS.
2. q_0 is the initial state: $q_0 \in Q$.
3. Σ is the set of transitions: $\Sigma = \{t_1,\ldots t_s\}$.
4. U and V are functions from Σ to $\mathbb{Z}^r$, with the following properties: (let $t_i \in \Sigma$)
   a) $U(t_i)$ is called a test-vector for t_i
   b) $V(t_i)$ is called a replacement vector for t_i
   c) $U(t_i) \leq V(t_i)$
   d) t changes the state from q to q' iff $q + U(t_i) \geq 0$ and $q + V(t_i) = q'$.

The set of states Q is the reachability set of the VRS.

Thus, a VRS is like a VAS $\langle q_0, W = \{w_i \mid w_i = V(t_i)\}\rangle$ except that the condition restricting the application of some translation w_i to a point q depends on whether $q + U(t_i) \geq 0$, which is more restrictive than $q + V(t_i) = q + w_i \geq 0$.

Relation to Petri Nets:

VRS: $\langle q_0,\Sigma,U,V\rangle$		GPN: $\langle \Pi,\Sigma,F,B,M_0\rangle$
dim VRS = r		$\Pi = \{p_1,\ldots p_r\}$; $\lvert\Pi\rvert = r$
	$\Sigma = \{t_1,\ldots t_s\}$	
$q_0 \in \mathbb{N}^r$	$q_0 = M_0$	$M_0 \in \mathbb{N}^r$
$U,V:\Sigma \to \mathbb{Z}^r$ $\forall t_i \in \Sigma: U(t_i) \leq V(t_i)$ assume $U:\Sigma \to -\mathbb{N}^r$ (see note below)	 $U = -F$ $V = B-F$	$F,B:\Sigma \to \mathbb{N}^r$
set of states Q	$Q = R(M_0)$	reachability set $R(M_0)$

Note: As Keller himself points out, positive coordinates of a test vector $U(t_i)$ do not matter, i.e., we get exactly the same results if we set all positive coordinates of a test vector to zero.

1.4 Liveness, Boundedness, Reachability and Coverability

From now on, we shall use the language of GPN's, taking advantage however of the fact that markings are expressed as vectors, and the action of a transition firing can be expressed by the pair of vectors $F(t)$ and $B(t)$. Unless specified otherwise, we shall be talking about a Petri Net $N = \langle\Pi,\Sigma,F,B,M_0\rangle$, $\Pi = \{p_1,\dots,p_r\}$, $\Sigma = \{t_1,\dots,t_s\}$.

Definition 1.8: A marking M covers a marking M' iff $M \geq M'$. Two markings M and M' are incomparable iff neither covers the other. We write this:

$$M \not\asymp M' \Leftrightarrow M \ngeq M' \ \& \ M' \ngeq M$$

Definition 1.9: Two markings M and M' agree over a subset $P \subseteq \Pi$, which we write $M \equiv M'$ (mod P), iff the coordinates corresponding to places in P agree:

$$M \equiv M'(\mathrm{mod}\ P) \Leftrightarrow \forall p_i \in P: M(p_i) = M'(p_i).$$

The set of markings which agree over a given subset P with a given marking M is denoted by:

$$M/P = \{M' \mid M \equiv M'(\mathrm{mod}\ P)\}.$$

Instead of referring to the congruence class $m = M/P$, we often call it a submarking m of P; in this case we also say that M agrees with the submarking m.

Definition 1.10: A marking M is reachable iff $M \in R(M_0)$.
A submarking m of $P \subseteq \Pi$ is reachable iff $\exists M \in R(M_0)$: $m = M/P$, i.e. iff some marking M which agrees with m is reachable.

Definition 1.11: A marking M is coverable iff $\exists M' \in R(M_0)$ $M' \geq M$
A submarking is weakly coverable iff some marking which agrees with it is coverable.
A submarking is strongly coverable iff every marking which agrees with it is coverable.

Note that a reachable submarking is weakly coverable, but not necessarily strongly coverable.

<u>Definition 1.12</u>: A place p_i is <u>bounded</u> at M_0 iff there exists an integer $b_i \in \mathbb{N}$ such that:

$$\forall M \in R(M_0) \qquad M(p_i) \leq b_i$$

A subset $P \subseteq \Pi$ is bounded at M_0 iff every $p_i \in P$ is bounded at M_0. A GPN is bounded iff Π is bounded at M_0.

<u>Definition 1.13</u>: A place $p_i \in \Pi$ is <u>certainly unbounded</u> at M_0 iff it is unbounded (not bounded) at every $M \in R(M_0)$.

<u>Definition 1.14</u>: A set of places $P \subseteq \Pi$ is <u>simultaneously unbounded</u> iff any arbitrarily large submarking of P is weakly coverable, or equivalently iff the zero submarking of $\Pi - P$ is strongly coverable.

<u>Definition 1.15</u>: A transition t is <u>potentially firable</u>* at M_0 iff there exists a marking $M \in R(M_0)$ at which t is firable:

$$\exists M \in R(M_0): M \geq F(t)$$

<u>Definition 1.16</u>: A transition t is <u>live</u> at M_0 iff it is potentially firable at every $M \in R(M_0)$.

A subset of transitions is live iff every element is live; a Net is live iff Σ is live at M_0.

<u>Note</u>: For vectors, $\geq$ is a partial order. Thus, $\ngeq$ is not the same as $<$ ($\leq$ but not $=$). The order relationship between two vectors is either $\geq$ or $\leq$ or $\gtrless\!\!\!\!/$. Also, if we write $M < M'$ to indicate ($M \leq M'$ & $M \neq M'$), this does <u>not</u> mean that every coordinate of M is strictly less than the corresponding coordinate of M'. This latter requirement would be better indicated by writing $M \leq M'-1$, where 1 stands for the vector whose coordinates are all equal to one.

---

* R. Keller calls this property "pseudo-live", but various other live-like properties (such as infinitely often firable) have been called "pseudo-live", and we wish to avoid confusion. [11]

Definition 1.17: The marking change $\Delta\sigma$ associated with a firing sequence is precisely what it says: if $M_1[\sigma\rangle M_2$, then $\Delta\sigma = M_2 - M_1$.

Definition 1.18: The firing vector $\bar{\sigma}$ associated with a firing sequence is an s-dimensional vector ($s = |\Sigma|$, the number of transitions) whose i^{th} coordinate is the number of occurrences of t_i in σ.

This gives us an alternate way of defining $\Delta\sigma$, which is, like a marking, an r-dimensional vector: $\Delta\sigma = (B - F) \cdot \bar{\sigma}$, where B and F are viewed as r x s-matrices.

Definition 1.19: The hurdle $H\sigma$ of a firing sequence σ is the smallest marking which permits σ to be completely fired. We have:

$$H\sigma = -\underline{\text{glb}}\ \{V \mid V=0\ \ \underline{\text{or}}\ \ (\exists t_i,\ \sigma', \sigma'': \sigma' t_i \sigma'' = \sigma\ \&\ V = \Delta\sigma' - F(t_i))\}$$

(The greatest lower bound glb of a set of vectors is the largest vector (not necessarily in the set) which is covered ($\leq$) by all vectors in the set.) Also note that $\forall\sigma$: $H\sigma + \Delta\sigma \geq 0$.

Some useful properties of R(M) and S(M):

$M_1 \in R(M_0) \iff R(M_1) \subseteq R(M_0)$

$M_1 \geq M_0 \Rightarrow S(M_0) \subseteq S(M_1)$

(t is live at M_0 & $M_1 \in R(M_0)$) $\Rightarrow$ (t is live at M_1)

(p is bounded at M_0 & $M_1 \in R(M_0)$) $\Rightarrow$ (p is bounded at M_1)

(p is certainly unbounded at M_0 & $M_1 \in R(M_0)$) $\Rightarrow$ (p is certainly unbounded at M_1).

2. The Equivalence of the GPN Model and the RPN Model

Many systems can be naturally and easily represented by GPN's because in certain contexts the restrictions of RPN's seem to be arbitrary. On the other hand, certain analytical techniques that have been developed for RPN's could be very usefully applied to more general systems.

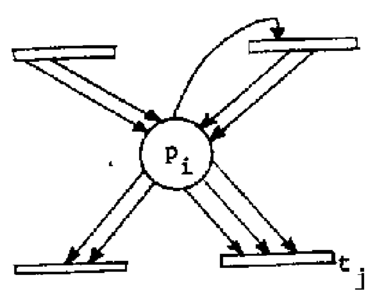
In this section we shall show how an arbitrary GPN can be represented by an RPN such that the two nets behave equivalently, in the following sense: Every firing sequence of one net can be translated into a corresponding firing sequence of the other net; every marking of one can be translated into a corresponding marking in the other net; and corresponding firing sequences yield corresponding markings. It will be seen that every question about the GPN can be answered by asking a corresponding question about the RPN used to represent the GPN.

2.1 The Construction of an RPN Equivalent to a Given GPN

Given a Generalized Petri Net $N = \langle \Pi, \Sigma, F, B, M_0 \rangle$, we shall construct a Restricted Petri Net $\hat{N} = \langle \hat{\Pi}, \hat{\Sigma}, \hat{F}, \hat{B}, \hat{M}_0 \rangle$ as follows: (let $\Pi = \{p_1 \ldots p_r\}$ and $\Sigma = \{t_1 \ldots t_s\}$)

a. for each place $p_i \in \Pi$, determine the maximum number of arcs (forwards or backwards) that go from p_i to each transition. Let this number be k_i:

$$k_i = \max_{1 \le j \le s} (F(p_i, t_j) + B(p_i, t_j))$$

Ex: 

$k_i = 3$

Fig. 2.1

b. for each place $p_i \in \Pi$, $\hat{\Pi}$ will contain a set of k_i places, which we denote:

$$\hat{p}_{i,1}, \hat{p}_{i,2}, \ldots \hat{p}_{i,k_i}$$

These are all the places in $\hat{\Pi}$. Thus, $|\hat{\Pi}| = \sum_{1 \le i \le r} k_i$

c. $\hat{\Sigma}$ will contain a transition $\hat{t}_j$ for each $t_j \in \Sigma$. But $\hat{\Sigma}$ will also contain additional λ-transitions,* which connect the k_i places $\hat{p}_{i,1}, \ldots, \hat{p}_{i,k_i}$ corresponding to p_i in Π into a ring: The example above thus transforms into:

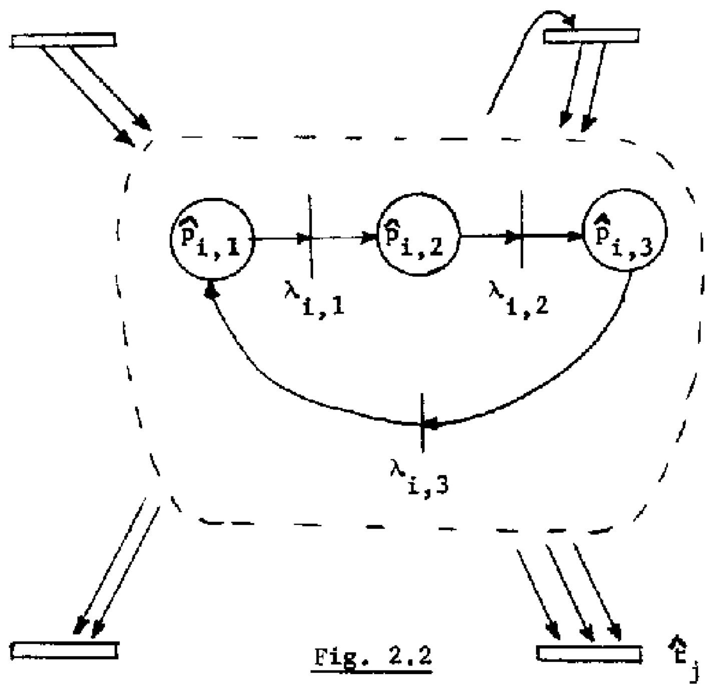


Fig. 2.2

Note: If for some p_i, $k_i = 1$, then there is no need for change, and for this place, no λ-transition need be introduced.

d. Now we generate $\hat{F}$ and $\hat{B}$ by distributing the arcs connected to a place p_i over the places in the corresponding ring in such a way as to create no self-loops and no multiple arcs. This is always possible, usually in many different (but equivalent) ways because of the choice of k_i:

*λ is the symbol of the empty string or the empty firing sequence. We talk about λ-transitions because, in a sense, their firings are invisible, i.e. the correspondence with firing sequences of the represented GPN is established by deleting the $\lambda_{i,j}$ occurences in the string corresponding to a firing sequence of the transformed net $\hat{N}$. Note that $\lambda_{i,j}$ itself is not the symbol of the empty string.

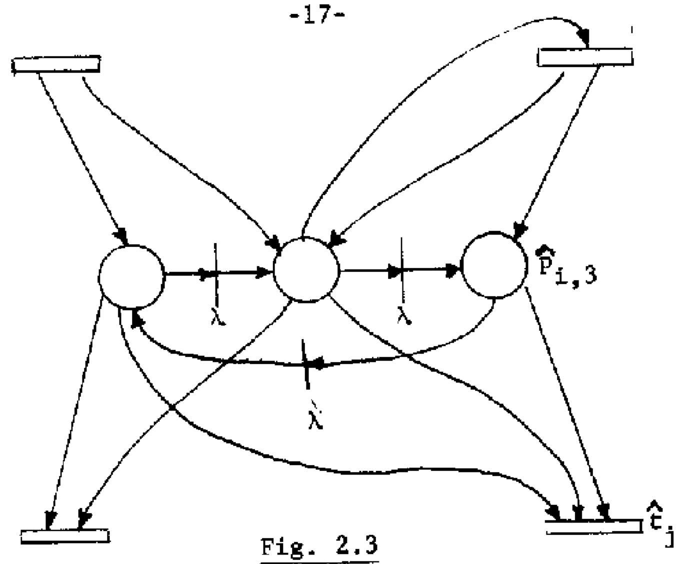


Fig. 2.3

e. Let $\hat{M}_o$ be defined as: $\begin{cases} \hat{M}_o(\hat{p}_{i,1}) = M_o(p_i) \\ \hat{M}_o(\hat{p}_{i,j}) = 0 \quad \text{for } j \neq 1 \end{cases}$

It should be clear that the tokens can always arrange themselves in the place rings in such a way as to permit exactly the same firing sequences as in the original net, if we disregard the additional firings of the λ-transitions, i.e., each firing sequence of $\hat{N}$ corresponds to the firing sequence of N obtained by deleting from the string in $\hat{\Sigma}^*$ all occurences of λ-transition, thus making it into a string in Σ^*, and $\hat{N}$ has no other firing sequences. (A firing sequence of $\hat{N}$ containing only λ-firings corresponds to the empty firing sequence of N, and in fact does not significantly change the marking of $\hat{N}$, because the sum of the tokens in any given place ring is not affected by λ-firings.) Also, to every marking $M \in R_N(M_0)$ (marking class) there will correspond a set of markings $\hat{\mathcal{M}} \subseteq R_{\hat{N}}(\hat{M}_0)$ such that:

$$(\forall \hat{M} \in \hat{\mathcal{M}})(\forall i) \boxed{M(p_i) = \sum_{1 \leq j \leq k_i} \hat{M}(\hat{p}_{i,j})}$$

We also readily convince ourselves that t_j is live in N at any reachable marking M if and only if $\hat{t}_j$ is live in $\hat{N}$ at any and all corresponding markings $\hat{M}$. The same applies to boundedness: p_i is bounded if and only if any (and all) places $\hat{p}_{i,j}$, $1 \leq j \leq k_i$, are bounded, and the bound is the same. Questions about reachability, coverability, firability, etc., can be answered in this very manner. We can therefore state:

Theorem 2.1: Generalized Petri Nets, Restricted Petri Nets, Vector Addition Systems, and Vector Replacement Systems are equivalent in modeling power for Asynchronous Systems.

Of course, this talks only about the modeling power, not the modeling convenience. But from an analytical point of view, it means that we can choose whichever form we like to prove our theorems. Karp and Miller's and Keller's decidability results for boundedness and coverability, Rabin's undecidability result for the inclusion of Reachability Sets, and the various results obtained by many authors for Petri Nets can be applied to any of the formalisms mentioned, and the proof uses the model most appropriate to the proof method. As an example, we shall present a Petri Net version of Rabin's proof in section 4.

2.2 Other even more restricted models of a Petri Net

a) Fan-in/Fan-out reduction: The fan-in and fan-out from every place and every transition can be reduced to 2.

It is easy to see that if we make the place-rings larger, we can generate an equivalent net where each place has at most one input and two outputs, or two inputs and one output:

$$\text{Just use} \quad k_i = \sum_{1 \le j \le s} (F(p_i, t_j) + B(p_i, t_j))$$

The example of figure 2.1 now becomes:

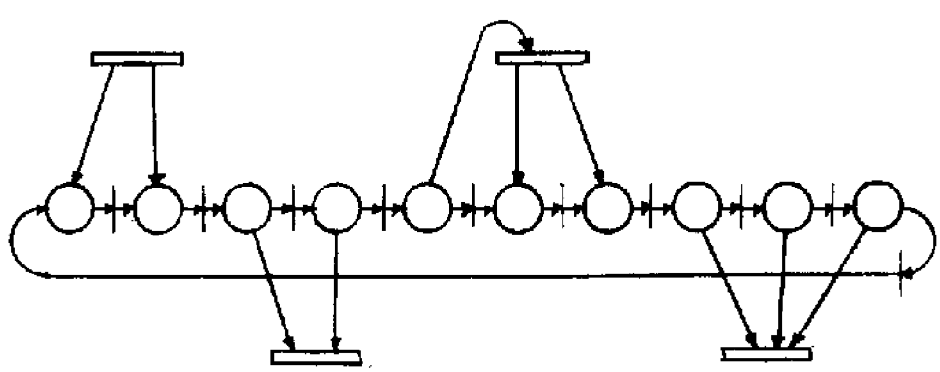


Fig. 2.4

But by extending the principle of a λ-transition ring, we can also reduce fan-in and fan-out of transitions, as we show by the following example:

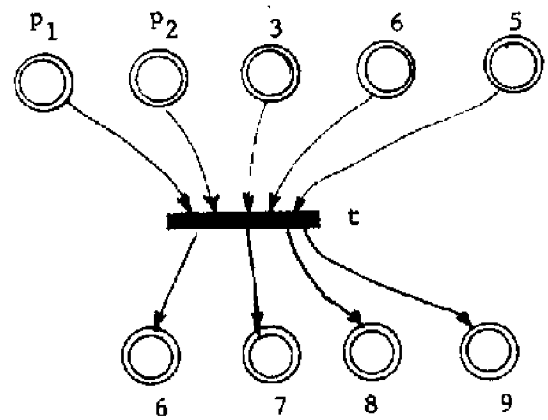


Figure 2.5

This can be replaced by the following:

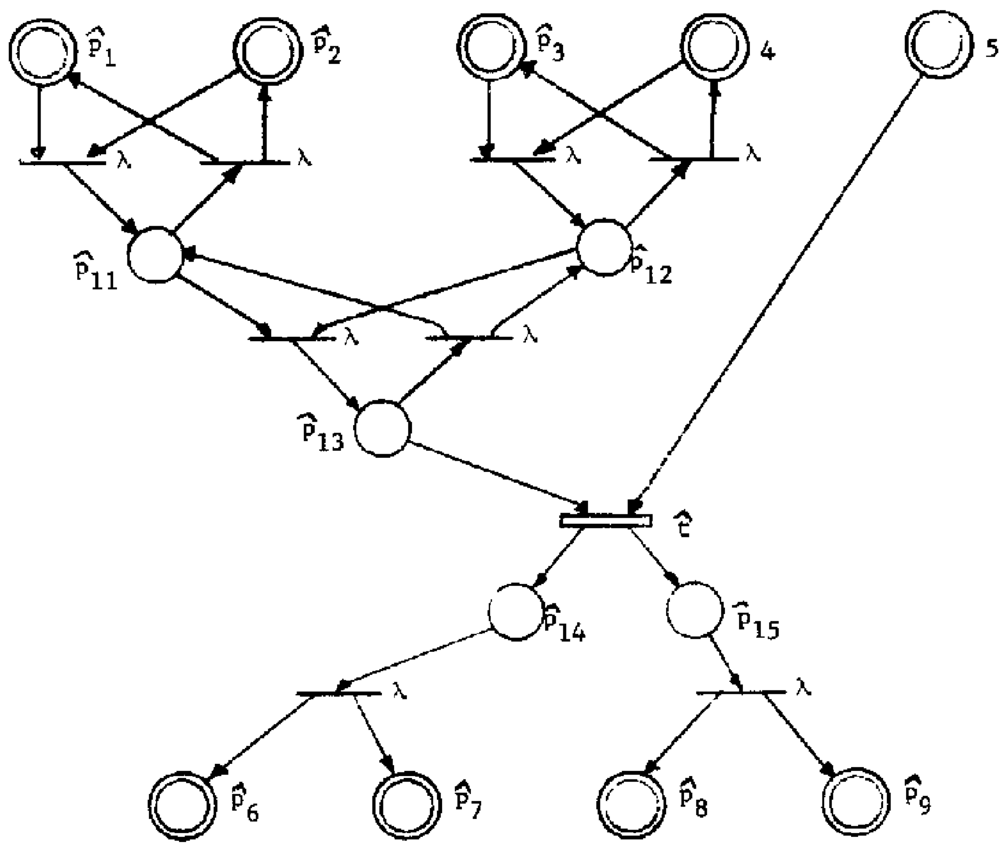


Figure 2.6

The net can be further transformed by reducing the fan-in and/or fan-out on the places; this only adds 1-in- 1-out λ-transitions.

The equivalence of firing sequences is as before: Same up to λ-firings.

The equivalence of markings is similar: We still have the linear function, but the sets of places over which the sums extend are not disjoint as before. In our example, we have:

$$
\begin{aligned}
M(1) &= \hat{M}(1) + \hat{M}(11) + \hat{M}(13) \\
M(2) &= \hat{M}(2) + \hat{M}(11) + \hat{M}(13) \\
M(3) &= \hat{M}(3) + \hat{M}(12) + \hat{M}(13) \\
M(4) &= \hat{M}(4) + \hat{M}(12) + \hat{M}(13) \\
M(5) &= \hat{M}(5) \\
M(6) &= \hat{M}(6) + \hat{M}(14) \\
M(7) &= \hat{M}(7) + \hat{M}(14) \\
M(8) &= \hat{M}(8) + \hat{M}(15) \\
M(9) &= \hat{M}(9) + \hat{M}(15)
\end{aligned}
$$

Thus, every generalized Petri Net is equivalent to a self-loop free Petri Net where the fan-in and fan-out is limited to 2 at every node.

We should note that the above constructions do not affect the following properties

- liveness
- boundedness
- decomposability into State Machines or Marked Graphs *)
- State Machine
- connectedness
- deadlock, trap *)

The constructions may affect the following properties:

- safeness
- conflict-freeness
- persistence
- Free Choice, Simple *)
- Marked Graph

But these concepts can usually be redefined. For example, F. G. Commoner [3] has liveness and deadlock results for multiple-arc Simple Nets.

*) These concepts are defined and used in [3,6,9].

b) <u>Almost-Euler Nets</u>: We can transform a Petri Net into an equivalent net where each transition is one-in-one-out or two-in-two-out, except for two non-Euler transitions, one of which is one-in-two-out and generates extra tokens as needed, the other is two-in-one-out and removes tokens from the net when needed.

We first reduce fan-in and fan-out: the only non-Euler transitions left are one-in-two-out, or two-in-one-out, or possibly zero-in or zero-out.

We successively use the following partial constructions

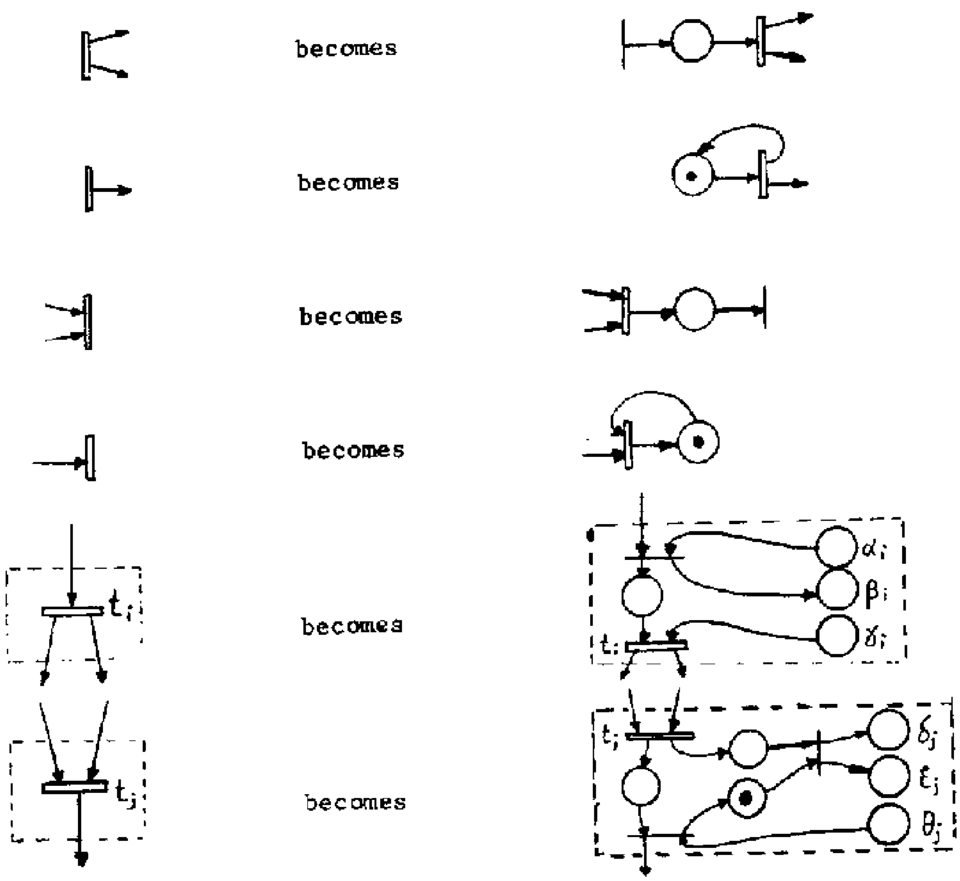
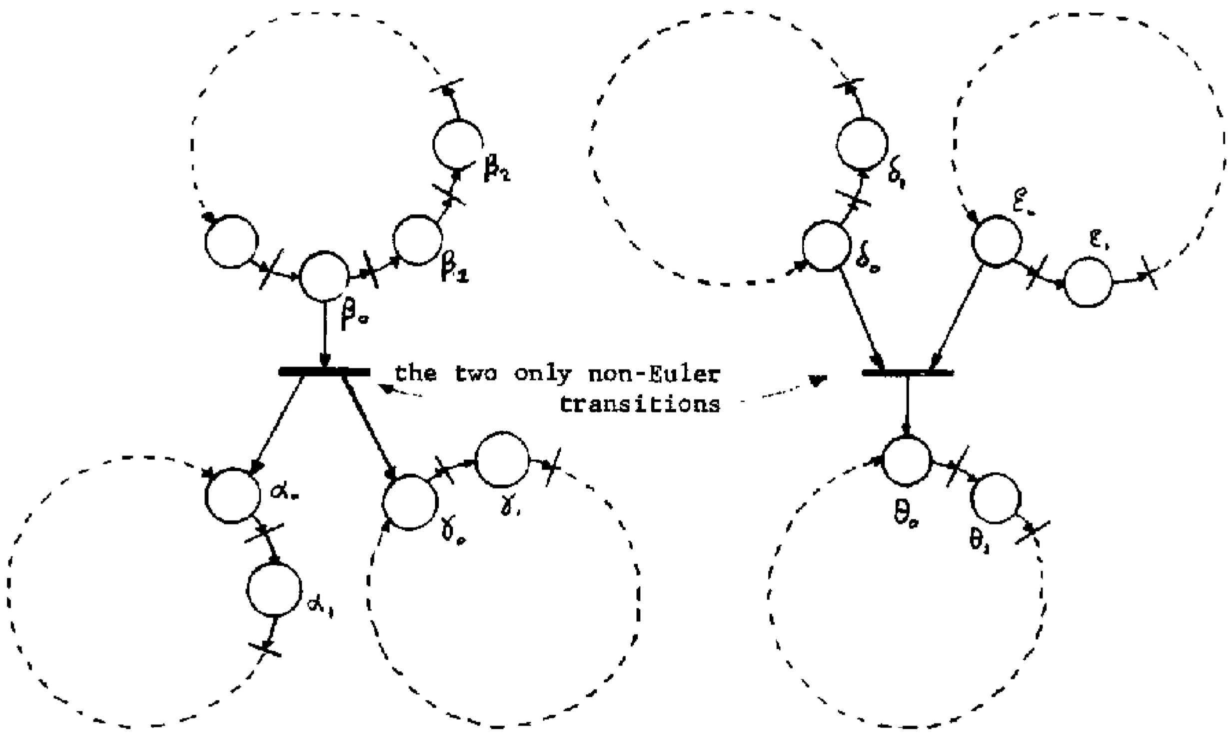


Figure 2.7

(all thin transitions are λ-transitions)

Finally, we connect all α places into a ring with one extra place α_0, all β places into a ring, etc., giving us 6 place-rings, which are then interconnected as follows:



<u>Figure 2.8</u>

We do not go into the detail of how this curiosity works. We only give it as an example of the kind of transformations one can make. We shall see another quite interesting transformation in Section 5.2.

3. Decidable Questions: Boundedness, Coverability

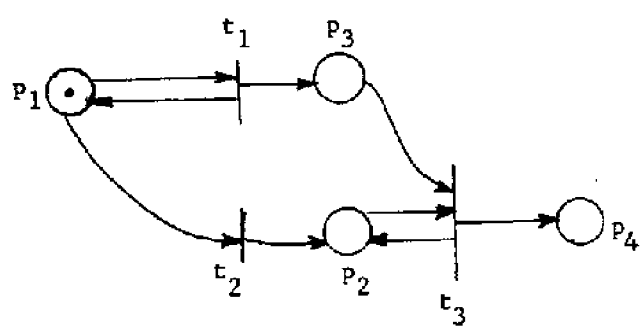
3.0 Introduction

One way a place p_i may become unbounded is the following:

Let M_0 be the original marking, and suppose there exists a firing sequence $\sigma_1\sigma_2$ such that:

$$M_0[\sigma_1\rangle M_1 \quad \& \quad M_1[\sigma_2\rangle M_2 \quad \& \quad M_2 \geq M_1 \quad \& \; M_2(p_i) > M_1(p_i)$$

Because of $M_2 \geq M_1$, every firing sequence possible from M_1 is also possible from M_2, in particular, σ_2 can be repeated, and therefore $\sigma_1(\sigma_2)^*$ is a legal set of firing sequences. But then it is clear that by repeating σ_2 arbitrarily often, the marking in p_i can grow without bounds. In particular, after the firing sequence $\sigma_1(\sigma_2)^n$, the marking will be $M_1 + n \cdot (M_2 - M_1)$. All places p_j for which $M_2(p_j) - M_1(p_j) > 0$ will be unbounded. But this is not the only way a place can become unbounded. Example:



$M_0 = \langle 1, 0, 0, 0\rangle$

Figure 3.1

p_4 is unbounded: given any number n, the firing sequence $(t_1)^n\, t_2(t_3)^n$ yields the marking $\langle 0, 1, 0, n\rangle$. But for no pair of reachable markings such that $M_2 \geq M_1$ do we also have $M_2(p_4) > M_1(p_4)$. This net incidentally has the interesting property that t_3 can fire any finite number of times, but cannot fire indefinitely.

However, in this case the unboundedness of p_4 follows from that of p_3, for which we do find two markings having the property described here: $M_0[t_1\rangle\, M_1$ and $M_1 \geq M_0$ and $M_0(p_3) > M_0(p_3)$: $M_1 = \langle 1, 0, 1, 0\rangle$.

Karp and Miller [10] have shown that there exists a finite construction which explicitly shows which places are unbounded, and which are not. We shall use basically the same construction, called a coverability tree.

3.1 Coverability Trees

A coverability tree is a rooted, labelled tree. The labels are chosen from the set $(\mathbb{N} \cup \{\omega\})^r$, where ω is a special symbol used to denote unboundedness. It means "arbitrarily many," and we will perform arithmetic with it as if it were a natural number larger than any other natural number. The greater or equal than relation ($\geq$) and the operations of addition (+) and subtraction (-), when applied to ω, satisfy the following rule:

$$\forall n \in \mathbb{N}: \ \omega \geq \omega \ \ \& \ \ \omega \geq n \ \ \& \ \ \omega + n = \omega \ \ \& \ \ \omega - n = \omega \ \ \& \ \ \omega \neq n$$

Thus indeed, "arbitrarily many" can exceed any given finite number, and is not affected by adding or subtracting a finite number.

The labels are thus r-dimensional vectors, where some coordinates may be ω, and the $\geq$ relation for vectors is defined as usual, taking into account the abovementioned rule for ω.

The arcs of this tree will also be labelled; the arc-labels will be transition names. In addition to the arcs of the tree, we will provide two kinds of backpointers, which can point from a node α to an antecedent of that node, i.e. a node β that lies on the (unique) path from the root node ρ to node α. These pointers are not considered to be arcs of the tree (it would not be a tree anymore) but are introduced for the purpose of record-keeping only.

If β is an antecedent of α, we write this $\beta \prec \alpha$, not to be confused with the relation $\leq$ for vectors or labels. The root node is an antecedent to every other node in the tree and has no antecedent; a leaf node is not antecedent to any node. The label of node α is denoted by L_α.

The label of the root node will be the initial marking vector, and the arcs of the tree will express transition firings. The node labels reflect the corresponding marking changes, but as soon as a node α is reached whose label L_α covers the label of some antecedent β, there is a possibility of unboundedness, and we introduce ω for those coordinates where arbitrarily many tokens can be generated if the firing sequence expressed by the arc labels along the path from β to α is repeated sufficiently often. To express this more conveniently, we include an ω-backpointer, labelled ω_i if we introduce ω in the i^{th} coordinate, from that node to the corresponding antecedent β.

If we reach a node α whose label equals that of some of its antecedents β, we make α a leaf node and introduce a loop backpointer, labelled λ, from α to β. The symbol λ stands for the empty string and suggests that, when one reaches the leaf node α, one has in fact also reached the interior node β and can continue tracing a path corresponding to a firing sequence, as we shall see. The reason for constructing a tree instead of the graph obtained by identifying nodes with identical labels is because the tree structure is more convenient for the proofs which will follow.

Definition 3.1: Given a Petri Net $N = \langle \pi, \Sigma, F, B, M_0 \rangle$, we define its coverability tree $T_N(M_0)$ recursively as follows:

basis: The label of the root node is the original marking:

$$L_o = M_0$$

step: Let α be a node in the coverability tree, with label L_α. There are several cases:

a. If no transition would be firable at a marking agreeing with L_α in its finite coordinates, i.e. if

$$\forall t \in \Sigma: L_\alpha \not\geq F(t)$$

then α is a leaf node called a dead-end.*

b. If some antecedent of α has a label equal to L_α, i.e. if

$$\gamma < \alpha \quad \& \quad L_\gamma = L_\alpha$$

then α is a leaf node called a loop-end, and there is a loop backpointer, labelled λ, from α to γ, written $\alpha[\lambda\rangle\gamma$.

c. If α is not a leaf node by (a) or (b), then it has a successor node for each transition which might be firable by a marking agreeing with L_α in its finite coordinates. If t is such a transition, an arc labelled t will go from α to a node β, which we write $\alpha[t\rangle\beta$.

---

*R. Keller calls this a null-end [11].

(This is not a firing relation for markings in $R_N(M_0)$, but a similar relation for nodes in the coverability tree $T_N(M_0)$. Thus we have: (assuming α is not a leaf node)

$$\forall t: L_\alpha \geq F(t) \Rightarrow \alpha[t\rangle\beta$$

Now we determine the label L_β, where $\alpha[t\rangle\beta$, as follows: Let A_β be the set of those antecedents of β (possibly including α) whose labels are covered by $L' = L_\alpha - F(t) + B(t)$:

$$A_\beta = \{\gamma \mid \gamma < \beta \ \ \& \ \ L_\gamma \leq L'\}$$

We consider two subcases:

c1. If $A_\beta = \emptyset$, let $L_\beta = L' = L_\alpha - F(t) + B(t)$.

c2. For every coordinate i in which L' is finite but strictly greater than the label L_γ of some $\gamma \in A_\beta$, we introduce an <u>ω-backpointer</u>, labelled ω_i, from β to γ, which we write as $\beta[\omega_i\rangle\gamma$:

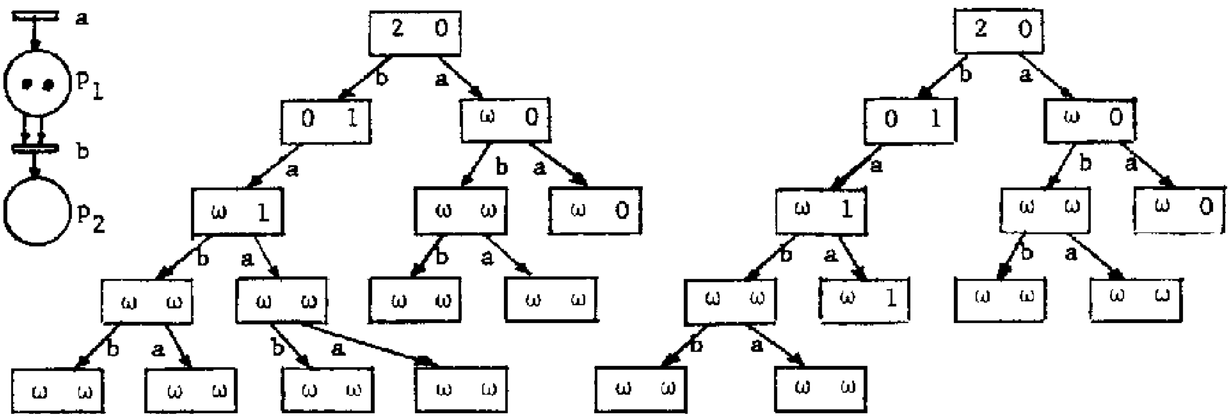
$$\forall i,\ 1 \leq i \leq r;\ \forall\gamma \in A_\beta:\ (L'(i) \neq \omega \ \&\ L'(i) > L_\gamma(i)) \Rightarrow \beta[\omega_i\rangle\gamma$$

The label L_β is then determined as follows:

$$\forall i,\ 1 \leq i \leq r:\ L_\beta(i) = \underline{\text{if}}\ (\exists\gamma:\ \beta[\omega_i\rangle\gamma)\ \underline{\text{then}}\ \omega\ \underline{\text{else}}\ L'(i)$$

We see that step (c2) is where ω-coordinates are introduced. The various ω_i-backpointers indicate which firing sequence can be used to increase the corresponding place marking beyond any bound -- provided that sequence can indeed be fired sufficiently often.

Disregarding the arc labels and the backpointers, this construction is exactly the same as Karp and Miller's [10]. It differs slightly from R. Keller's construction [11] in that Keller includes step (b) under step (c2) by checking whether A_β contains a node γ whose label L_γ is equal to L'. Figure 3.2 shows an example of a simple Self-Loop-Free Petri net, which thus directly corresponds to a Vector Addition System, where the two constructions yield different coverability trees.



Karp and Miller's construction, same as definition 3.1 without backpointers.

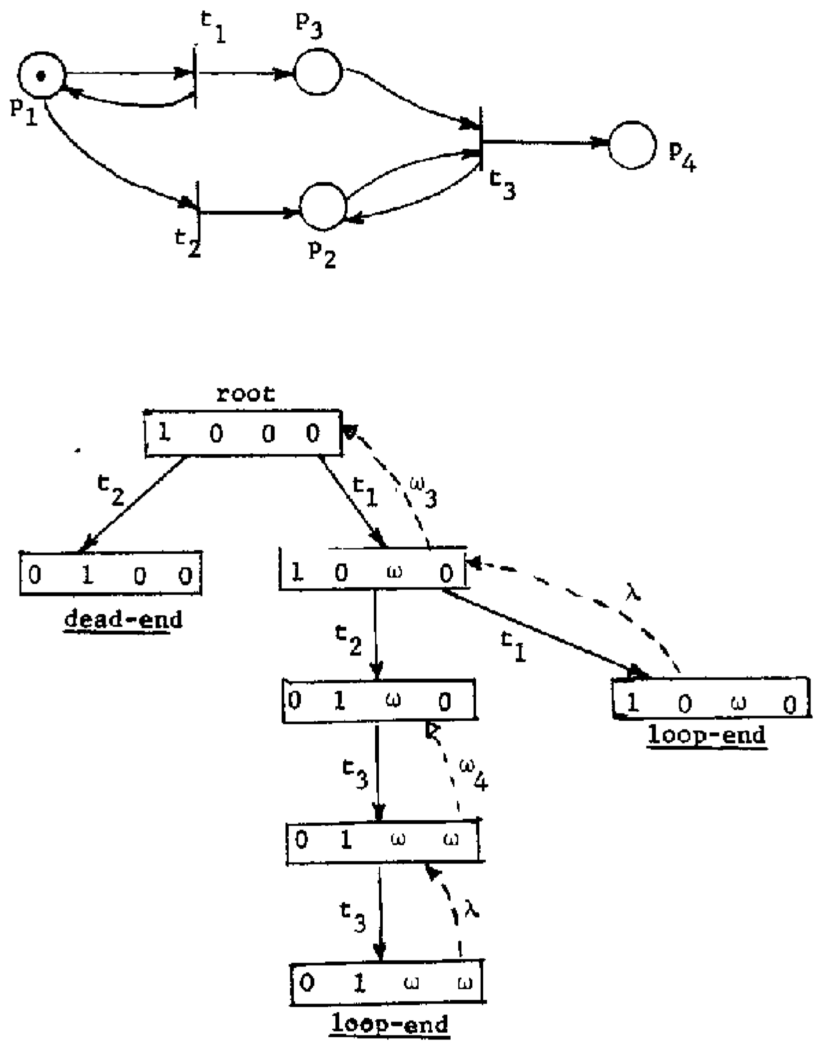
Keller's construction.

Figure 3.2

We will show that this coverability tree will be finite for any given Petri Net, and thus the recursive definition provides at the same time an algorithm for constructing the coverability tree of a Petri Net.

To illustrate this, we construct the coverability tree for the example shown in the beginning of this section, reproduced in Fig. 3.3 on the next page. The ω_3-backpointer shows us how to increase the third coordinate without bounds by repeating t_1. The ω_4-backpointer shows that we must repeat t_3 n times to get n tokens on p_4, by firing $(t_3)^n$. But this is possible only if p_3 has enough tokens, i.e. the fourth ω <u>depends</u> on the third ω. That is because the firing vector associated with t_3 is not positive. This does not mean that we cannot produce arbitrarily many tokens in p_4, but it does mean that to do so we must first produce enough tokens in p_3.

Our aim in constructing this coverability tree is to provide a decision procedure for deciding whether a given place is bounded, and whether a given marking can be covered by a reachable marking. For this, we need three theorems:



t1
P3
P1
P4
t3
t2
P2
root
1 0 0 0
t2
t1
ω3
0 1 0 0
dead-end
1 0 ω 0
λ
t2
t1
0 1 ω 0
1 0 ω 0
loop-end
t3
ω4
0 1 ω ω
t3
λ
0 1 ω ω
loop-end

Figure 3.3

Theorem 3.1: Every coverability tree is finite.

Theorem 3.2: A place is unbounded if and only if the coverability tree contains a label in which the corresponding coordinate is ω.

Theorem 3.3: There exists a reachable marking covering a given vector in $\mathbb{N}^r$ if and only if the coverability tree contains a label which covers that vector.

Theorem 3.3 also provides the justification of the name "coverability tree."

We shall now prove these theorems.

3.2 Finiteness

Lemma 3.4 Every infinite sequence of non-negative integers contains a non-decreasing subsequence.

Proof: If the sequence contains infinitely many mutually distinct elements, we can extract a strictly increasing subsequence starting with any element and scanning along the sequence until we find a larger element, and so on.

If the sequence does not contain infinitely many mutually distinct elements, some element must be repeated infinitely often, and there exists an infinite constant subsequence.

In any case, there is an infinite non-decreasing subsequence.

QED

Lemma 3.5 Every infinite sequence of r-dimensional vectors in $(\mathbb{N} \cup \{\omega\})^r$ contains an infinite non-decreasing subsequence.

Proof: Consider the first coordinate. If there are infinitely many vectors whose first coordinate is ω, they form an infinite subsequence non-decreasing in the first coordinate. Otherwise, disregarding those vectors whose first coordinate is ω, there exists an infinite subsequence of vectors whose first coordinate is non-decreasing, by Lemma 3.4. In any case, there exists an infinite subsequence non-decreasing in its first coordinate.

This infinite subsequence now contains another infinite subsequence non-decreasing in its second coordinate, and so on to the r^{th} coordinate. Thus there exists an infinite subsequence non-decreasing in each coordinate.

QED

<u>Corollary 3.6</u> There exists no infinite set of mutually incomparable vectors in $(\mathbb{N} \cup \{\omega\})^r$.

<u>Proof</u>: This infinite set, being denumerable, could be arranged in an infinite sequence where each element occurs exactly once. But then, by Lemma 3.5, any two elements of some infinite non-decreasing subsequence would be comparable, which contradicts the assumption of infinity.

QED

Note, however, that if $r \geq 2$, no a-priori bound exists on such sets of incomparable vectors: The set $\{\langle x, y\rangle \in \mathbb{N}^2 \mid x + y = k\}$ is such a set of mutually incomparable pairs of size k+1, arbitrarily large.

<u>Proof of Theorem 3.1</u>*) Every coverability tree is finite and can be effectively constructed.

Suppose some Petri Net has an infinite coverability tree. By construction, every node has at most as many immediate successors as there are transitions in the Petri Net, a finite number. Then, by König's Infinity Lemma for rooted trees, there must be an infinite path in the tree, i.e. a path which does not eventually end at a leaf node. But then, by Lemma 3.5, there must be an infinite non-decreasing subsequence of the sequence of node labels along that infinite path. In fact, it must be strictly increasing, otherwise the path would have to end in a loop-end leaf node at the first repetition of a label. But each time a label is reached which is strictly larger than some previous label, it will have, by construction, at least one more coordinate equal to ω than the smaller label. Since there can be at most r coordinates equal to ω, the existence of such an infinite increasing subsequence of labels along a path in the tree is contradictory. Now that we know that the tree is finite, we can convince ourselves that the recursive definition 3.1 also provides an algorithm for constructing the coverability tree.

QED

<u>Note</u>: König's Infinity Lemma for rooted trees can easily be prooved non-constructively. Assume the rooted tree is infinite, yet at each node there is a finite number of branches. Then at least one of the branches from the root node must point to the root of an infinite subtree. The path traced out by the root nodes of such sucessive infinite subtrees must be an infinite path -- QED. König's original Infinity Lemma [12] is more general. We provide a translation of his proof in appendix, page 77 .

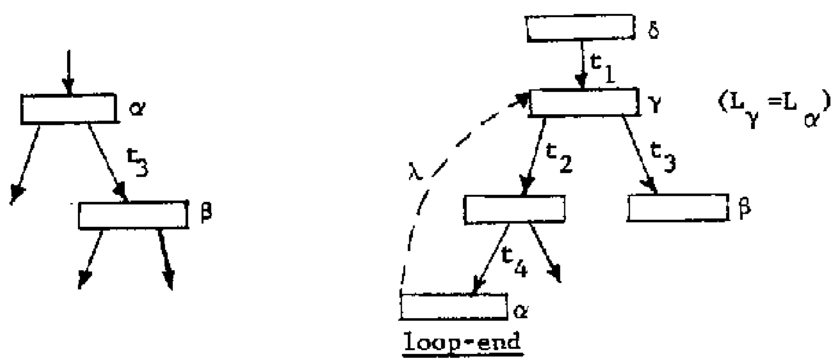
*) This is the same proof as in Karp and Miller [10].

3.3 Firing Sequences and Composite Paths in a Coverability Tree

Now that we know that coverability trees are finite objects, we can use them to answer certain questions about the corresponding Petri Net.

First, we show that every firing sequence can be folded on the coverability tree, in the sense that there exists a sequence of paths in the tree, linked by loop backpointers, such that the arc labels spell out the given firing sequence. This is why we write $\alpha[t\rangle\beta$ if an arc labelled t goes from node α to node β, and now we extend this to the case where α is a loop-end and $\alpha[\lambda\rangle\gamma$ and $\gamma[t\rangle\beta$. Indeed, as in the formation of a firing sequence, we have $\alpha[\lambda t\rangle\beta$, where λ is the symbol for the empty string. See Fig. 3.4. We then observe that the "firing rule" for labels is similar to that for markings, taking into account the rules for arithmetic with ω and the possible introduction of new ω-coordinates.

$$\alpha[t\rangle\beta \;\Rightarrow\; L_\alpha \geq F(t) \;\;\&\;\; L_\beta \geq L_\alpha - F(t) + B(t)$$



Two cases where $\alpha[t_3\rangle\beta$

Fig. 3.4

Definition 3.2: A λ-composite path σ from node α to node β in a coverability tree, written $\alpha[\sigma\rangle\beta$, is a concatenation of paths starting at α, ending at β, and linked by loop-backpointers.

An example of a λ-composite path in figure 3.4 (right) is $\delta[t_1 t_2 t_4 t_3\rangle\beta$.

<u>Lemma 3.7</u> If σ is a firing sequence of the Petri Net N leading from the original marking M_0 to some marking M, then σ is also a λ-composite path path in the coverability tree $T_N(M_0)$ from the root ρ to a node α such that $L_\alpha \geq M$, and such that L_α and M agree in the finite coordinates of L_α:

$$\forall\sigma \in S_N(M_0)\ \exists\alpha \in T_N(M_0) : M_0[\sigma\rangle M\ \&\ \rho[\sigma\rangle\alpha\ \&\ (\forall i,\ 1 \leq i \leq r : L_\alpha(i) \neq M(i) \Rightarrow L_\alpha(i) = \omega)$$

<u>Proof</u>: By induction on the length of σ.

<u>basis</u>: $L_\rho = M_0$ for the null sequence or path

<u>step</u>: assume $M_0[\sigma\rangle M_1$ and $M_1[t\rangle M_2$.
By induction, there is a node α such that $\rho[\sigma\rangle\alpha$ and $L_\alpha \geq M_1$ with L_α and M_1 agreeing in L_α's finite coordinates.

Since t is firable at M_1 we have $M_1 \geq F(t)$, and therefore $L_\alpha \geq F(t)$. Therefore, α cannot be a dead-end leaf node. If α is a loop-end leaf node, we follow the loop backpointer to α' and consider α', since we then also have $\rho[\sigma\rangle\alpha'$ and $L_{\alpha'} = L_\alpha$. Thus we can assume $\rho[\sigma\rangle\alpha$ and $L_\alpha \geq M_1 \geq F(t)$ for some interior node α. But then, by the construction of the coverability tree, there is an arc labelled t which goes from α to a node β such that $L_\beta \geq L_\alpha - F(t) + B(t)$. Since $M_2 = M_1 - F(t) + B(t)$, we have:

$$\alpha[t\rangle\beta \Rightarrow \rho[\sigma t\rangle\beta\ \&\ L_\beta \geq M_2$$

and the finite coordinates are transformed the same way for the labels as for the markings.

QED

<u>Corollary 3.8</u>: If place p_i is unbounded in a Petri Net, the corresponding coverability tree contains a label whose i^{th} coordinate is ω.

<u>Proof</u>: Suppose no label has ω as its i^{th} coordinate. Since the number of labels is finite, there is a largest value, $b_i \neq \omega$, of the i^{th} coordinate of all labels. Now, since every reachable marking is covered by the label reached by a corresponding composite path in the coverability tree, no reachable marking can exceed b_i tokens in p_i. Thus p_i must be bounded; in fact, b_i is a bound.

QED

<u>Corollary 3.9</u>: If a given marking M can be covered by a reachable marking M_1 in a Petri Net, then the coverability tree contains a label L_α which covers M.

<u>Proof</u>: By Lemma 3.7 there exists a label L_α which covers M_1, hence $L_\alpha \geq M_1 \geq M$

QED

Corollaries 3.8 and 3.9 are the "only if" parts of Theorems 3.2 and 3.3, respectively.

What remains to be shown is that ω indeed stands for "arbitrarily many tokens" as a coordinate in the coverability tree. To produce more than a given number of tokens in place p_i, we have to repeat the sequence of firings leading up the first occurence of the corresponding ω. That sequence is called an ω_i-loop:

<u>Definition</u>: If $\alpha[\sigma\rangle\beta$ and an ω_i-backpointer goes from β back to α, then σ is called an <u>ω_i-loop</u>, the vector $\Delta\sigma$ is the corresponding <u>loop change</u>, and $H\sigma$ is called the <u>loop hurdle</u>. (See Definitions 1.17 and 1.19 on page 13).

Note that there may be several different ω_i-loops for the same coordinate i, which means that sometimes the unboundedness of a place can be confirmed by different strategies.

It should also be pointed out that if σ is an ω_i-loop, as a firing sequence in the Petri Net it may not be firable the first time down the coverability tree, but only after certain other ω-coordinates have been made large enough to cover the loop hurdle by repeating other ω_j-loops before.

For this reason we must also not expect to find a reachable marking which equals any given marking agreeing in the finite coordinates with some label in the coverability tree. But we will show that we can <u>cover</u> any such marking in the ω-coordinates.

<u>Definition 3.4</u>: An <u>ω-composite path</u> in a coverability tree is a sequence of paths in the tree, linked by ω-backpointers.

<u>Lemma 3.10</u>: For every node α in the reachability tree $T_N(M_0)$ of a Petri Net, and for any target vector V_α agreeing with L_α in the latter's finite coordinates, there exists a firing sequence $\sigma(V_\alpha)$ which is also an ω-composite path from the root ρ to node α, such that the marking reached by $\sigma(V_\alpha)$ covers V_α:

$$M_0[\sigma(V_\alpha)\rangle M_\alpha \quad \& \quad M_\alpha \geq V_\alpha$$

<u>Proof</u>: By induction along the path from ρ to α.

<u>basis</u>: V_ρ must equal M_0 since L_ρ has only finite coordinates.

<u>step</u>: Suppose $\alpha[t\rangle\beta$, and suppose that, for every $V_\alpha \leq L_\alpha$, there exists a firing sequence $\sigma(V_\alpha)$ whose corresponding ω-composite path ends at α and leads to a marking $M_\alpha \geq V_\alpha$. We have $M_\alpha \leq L_\alpha$; $M_\alpha = M_0 + \Delta\sigma(V_\alpha)$, and the three vectors M_α, V_α, L_α agree in the finite coordinates of L_α.

We wish to find a firing sequence $\sigma(V_\beta)$ capable of reaching a marking M_β which covers a given target vector V_β, where V_β agrees with L_β in the finite coordinates of L_β.

If L_β has no more ω-coordinates than L_α, the situation is simple: $L_\beta = L_\alpha - F(t) + B(t)$, and to cover V_β, it is enough to choose $V_\alpha(i) = (V_\beta + F(t))(i)$ for ω-coordinates and $V_\alpha(i) = (V_\beta + F(t) - B(t))(i)$ for the finite coordinates (which must agree with the corresponding labels L_α and L_β), and then take $\sigma(V_\beta) = \sigma(V_\alpha)t$. It is clear that in this case, $M_\alpha \geq F(t)$ which makes t firable, and we have:

$M_\beta = M_\alpha - F(t) + B(t) \geq V_\alpha - F(t) + B(t) \geq V_\beta$.
Also, if the ω-composite path $\sigma(V_\alpha)$ ends at α, then clearly $\sigma(V_\beta)$ ends at β.

Now let us assume that L_β has one or several more ω-coordinates than L_α, say the i^{th} and j^{th}. We call these the "new" ω-coordinates, as opposed to the "old" ω-coordinates already present in L_α. The ω-loops corresponding to these new ω-coordinates are σ_i and σ_j, the corresponding loop changes are $\Delta\sigma_i$ and $\Delta\sigma_j$, the corresponding loop hurdles are $H\sigma_i$ and $H\sigma_j$. We shall try a firing sequence of the form:

$$\sigma(V_\beta) = \sigma(V_\alpha) \cdot t \cdot (\sigma_i)^x \cdot (\sigma_j)^y$$

Now we must prove that there exists a target V_α and two integers x and y such that this sequence is firable and leads to a marking $M_\beta \geq V_\beta$. We know that $\Delta\sigma_i(i) \geq 1$ and $\Delta\sigma_j(j) \geq 1$. Therefore, if we choose $x = V_\beta(i)$ and $y = V_\beta(j)$, the above choice for $\sigma(V_\beta)$ will produce a marking M_β which will cover V_β in coordinates i and j, provided we can find V_α such that this sequence be firable. In other words, given x and y as chosen above, we must find V_α such that $t(\sigma_i)^x(\sigma_j)^y$ be firable at M_α, where $M_0[\sigma(V_\alpha)\rangle M_\alpha$:

We need $V_\alpha \geq H(t(\sigma_i)^x \ (\sigma_j)^y) = H_\beta$

Let us look at the coordinates of this hurdle H_β. They are of three kinds: those for which L_β is still <u>finite</u>, the <u>new</u> ω-coordinates i and j (ω-coordinates in L_β but finite in L_α), and the <u>old</u> ω-coordinates (ω-coordinates in L_α).

The <u>finite</u> coordinates are transformed the same way by ω-composite paths for labels and by firing sequences for markings. The loop change for these coordinates is zero. Therefore, H_β does not exceed L_β in the <u>finite</u> coordinates.

The <u>new</u> ω-coordinates are also no problem. Indeed, they are finite in L_α, and in $L_\alpha - F(t) + B(t)$ they strictly exceed the hurdles of single firings of σ_i and σ_j respectively. That is because, if $\beta[\omega_i\rangle\gamma$, then $(L_\alpha - F(t) + B(t))(i) > L_\gamma(i)$, and as far as this coordinate is concerned, σ_i can be fired from γ back to β. For the second and subsequent firings, the hurdle coordinate i would be even less, and ultimately be zero, since the loop change is <u>positive</u> for this coordinate. If σ_i and σ_j are different sequences, then the loop change in one coordinate i is zero for the other loop σ_j and vice versa.

Since V_α must agree with L_α over the finite coordinates of V_α, i.e. those just discussed, we see that $V_\alpha \geq H_\beta$ is automatically satisfied in these coordinates.

For the old ω-coordinates, where by induction we can exceed any bound in a corresponding marking, we can choose a wildly exaggerated upper bound of H_β, like

$$\forall k,\ L_\alpha(k) = \omega:\quad V_\alpha(i) = V_\beta(i) + (F(t) + x \cdot H(\sigma_i) + y \cdot H(\sigma_j))(i)$$

Having thus established values for x, y, V_α, given V_β, we can now assert that:

- by induction, there exists $\sigma(V_\alpha)$ and M_α such that:

  $M_0[\sigma(V_\alpha)\rangle\ M_\alpha$

  $M_\alpha \geq V_\alpha$

  $\sigma(V_\alpha)$ is an ω-composite path to α

- at M_α, the following holds:

  $M_\alpha \geq H(t(\sigma_i)^x\ (\sigma_j)^y)$

  $M_\alpha[t(\sigma_i)^x\ (\sigma_j)^y\rangle\ M_\beta$

  $M_\beta \geq M_\alpha + \Delta(t(\sigma_i)^x\ (\sigma_j)^y) \geq V_\beta$

  $t(\sigma_i)^x\ (\sigma_j)^y$ is an ω-composite path from α to β.

Therefore, $\sigma(V_\beta) = \sigma(V_\alpha)\ t(\sigma_i)^x(\sigma_j)^y$ is an ω-composite path leading to β and a firing sequence leading to $M_\beta \geq V_\beta$.

QED

We have shown that in order to exceed a target vector $V_\beta \le L_\beta$, $V_\beta(k) = L_\beta(k)$ if $L_\beta(k) \ne \omega$, we compute a target vector V_α for the node preceding β in the coverability tree. Thus regressing along the path $\rho \to \ldots \to \alpha \to \beta$ we map a firing strategy to eventually exceed our target vector. As in Lemma 3.7 the finite coordinates of the labels change exactly like the markings. We note that in this strategy, the ω_i-loops are executed in the sequence in which the corresponding ω_i-coordinates are introduced, and that there is no embedding of the firing sequences corresponding to these loops even if the loops themselves are embedded. As an example, we show a coverability tree in Fig. 3.5 (on the next page), and a firing sequence for exceeding a given target vector.

The __if__ parts of Theorems 3.2 and 3.3 follow immediately from Lemma 3.10. We have thus proved Theorems 3.2 and 3.3.

And from Theorems 3.1, 3.2 and 3.3 follow the main results of this chapter:

__Theorem 3.11__: It is decidable whether a set of places is simultaneously unbounded.

__Proof__: We can check whether the coverability tree has a label in which the coordinates corresponding to these places are all ω.

QED

The meaning of "simultaneously unbounded" becomes clear if we look at Fig. 3.6. Also see Definition 1.14 on page 12.

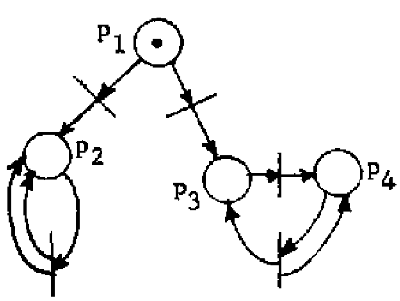
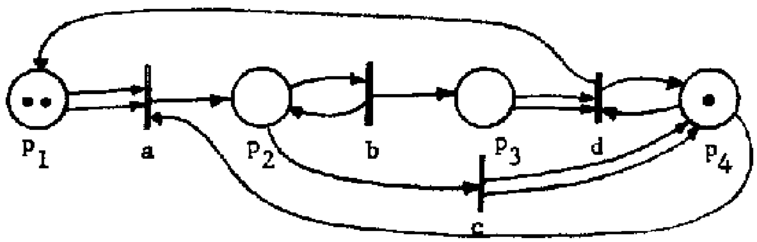
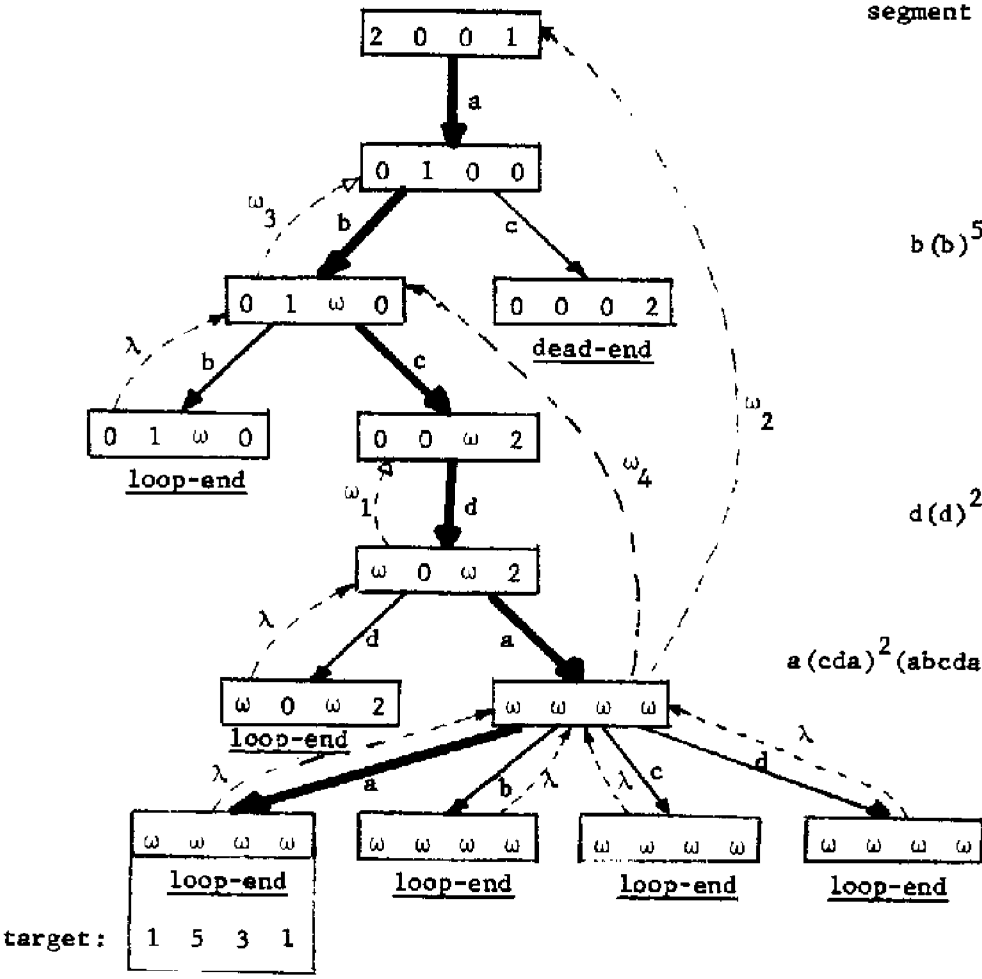


__Fig. 3.6__

P_3 and P_4 are simultaneously unbounded.

P_2 and P_3 are not simultaneously unbounded, even though each one is unbounded individually.





firing sequence segment	target	marking
	2 0 0 1	2 0 0 1
a		
	0 1 0 0	0 1 0 0
$b(b)^{56}$		
	0 1 56 0	0 1 57 0
c		
	0 0 56 2	0 0 57 2
$d(d)^{21}$		
	21 0 12 2	22 0 13 2
$a(cda)^2(abcda)^5$		
	3 5 3 2	3 6 5 3
a		
	1 5 3 1	1 7 5 2
	target	marking

$\sigma_3 = b$ $\qquad$ $\Delta\sigma_3 = \langle 0,0,1,0\rangle$

$\sigma_1 = d$; $H(\sigma) = \langle 0,0,2,1\rangle$ $\qquad$ $\Delta\sigma_1 = \langle 1,0,-2,0\rangle$

$\sigma_4 = cda$; $H(\sigma_4) = \langle 1,1,2,0\rangle$ $\qquad$ $\Delta\sigma_4 = \langle -1,0,-2,1\rangle$

$\sigma_2 = abcda$; $H(\sigma_2) = \langle 3,0,1,0\rangle$ $\qquad$ $\Delta\sigma_2 = \langle -3,1,-1,0\rangle$

Firing sequence to exceed the target: $a\ b^{37}\ c\ d^{22}\ a\ (cda)^2\ (abcda)^5\ a$

Fig. 3.5

Theorem 3.12: It is decidable whether a Petri Net can reach a marking which covers a given marking.

Proof: We can check whether the marking to be covered is covered by some label in the finitely constructible coverability tree.

QED

Note 1: The decidability results only depend on the set of labels in the coverability tree. As a matter of fact, we don't even need the coverability tree to find a firing sequence which leads to a covering marking, because if we know that it exists, we can find it by simply enumerating all possible firing sequences and their resulting markings until we find one whose resulting marking covers the target marking.

Note 2: By taking "totally unreasonable" upper bounds, we can establish a quick formula for finding a firing sequence which exceeds a given target vector. Let d be the distance of the target node from the root, let h be the largest coordinate of all loop hurdles and transition input vectors ($F(t)$), and let v be the largest coordinate of the target vector. We shall consider only those coordinates of a given intermediate target for which the corresponding label has an ω-coordinate.

Our first target vector will be replaced by one whose ω-coordinates (remember the restriction mentioned above) are all v. The closest antecedent where the last ω was introduced into the label is less than d arcs away, and so we choose the corresponding target to have all its ω-coordinates equal to $v + d \cdot h$, which should be large enough. A similar consideration applies to the path from one ω-introduction to the next.

Let there be k successive ω-introductions. The last ω-loop will be repeated $n_k = v + d \cdot h$ times, and thus the previous ω-coordinates can be required to exceed the target by $h \cdot n_k + d \cdot h$. We already see a recurrence relation in the making: $n_{k-1} = v + d \cdot h + h \cdot n_k$

Now let x be the largest of v, d and h. Clearly, we can use:

$$n_k = x + x^2$$

$$\vdots$$

$$n_1 = x + x^2 + \ldots\ldots + x^{k+1} = \frac{x^{k+2} - x}{x - 1}$$

Therefore, we simply follow an ω-composite path leading to the target node, and in the process we repeat the first ω-loop encountered, $\frac{x^{k+2} - x}{x - 1}$ times, the next one x^{k+1} times less, etc.

Applied to the example of Fig. 3.5, we have $d = 6$, $h = 3$, $v = 4$, and thus $x = 6$; we get the following sequence $(k = 4)$:

$$a\ b^{9331}\ c\ d^{1555}\ a\ (cda)^{258}\ (a\,b\,c\,d\,a)^{42}\ a$$

which results in the marking $\langle 1167, 44, 6263, 258\rangle$.

We can also use this approximation to show that in a GPN of r places with an upper bound h on the loop and transition hurdles, if a marking can be covered, there exists a firing sequence to cover it of a length proportional to the marking to be covered, the factor of proportionality being on the order of h^{r+1}.

Note 3: Although the principle of the proof of Lemma 3.10 is quite simple, we went to so much detail because, by our own experience, any firing strategy derived from an incomplete proof (which disregards loop hurdles, for example) has failed on some counterexamples to actually be firable without producing negative intermediate markings.

Also, while the language used in this section is mostly that of Vector Addition Systems, we found the graphical intuition provided by Petri Nets very useful to construct examples and counterexamples, and to test conjectures and unfinished proofs.

4. An Undecidable Problem About Petri Nets

When R. Karp and R. Miller [10] introduced Vector Addition Systems to answer certain decidability questions about their Parallel Program Schemata, M. Rabin showed that a particular problem about Vector Addition Systems was undecidable:* is the Reachability Set of one Vector Addition System a subset of the Reachability Set of some other given Vector Addition System. Rabin's first proof in 1967 used exponential polynomials [2] ; at that time Hilbert's 10^{th} Problem [7] had not yet been shown to be undecidable.

In 1970, Matijasevič [13] proved that Hilbert's 10^{th} Problem was undecidable, and thus permitted a technically simpler proof of Rabin's result. Rabin never published his proof, but in 1972 he presented his new proof in a talk at MIT, an account of which can be found in [2] .

Since Vector Addition Systems and Petri nets can fully represent each other, Rabin's result also gives us an undecidable problem about Petri nets. Furthermore, we believe that the graphical character of the Petri net model permits an easier exposition of the undecidability result.

Theorem 4.1: Given two Petri nets having the same number of places, each with a given initial marking, it is undecidable in general whether every marking reachable in one net is also reachable in the other.

Proof: We show that, given an arbitrary polynomial $P(x_1, \ldots, x_r)$ of r variables with integer coefficients, there exists a pair of Petri nets such that the set of reachable markings of one is a subset of the reachable markings of the other if and only if the polynomial P has an integral root. Thus, if we could decide for any two Petri nets whether in fact the set of reachable markings of one is a subset of the reachable markings of the other, we could also decide whether an arbitrary polynomial with integral coefficients has an integral root. But this is Hilbert's 10^{th} Problem, which has been shown to be undecidable by Matijasevič.

Actually, we use the following equivalent form of Hilbert's 10^{th} problem:

*Rabin was misquoted in [10] and [11] : Karp and Miller believed he had shown the Equality of Reachability Sets to be undecidable; to this author's knowledge, this question has not yet been resolved, as of 1973.

<u>Lemma</u> <u>4.2</u>: Given two polynomials of r variables with non-negative integer coefficients $P(\bar{x})$ and $Q(\bar{x})$ such that, $\forall \bar{x} \in \mathbb{N}^r$: $P(\bar{x}) \geq Q(\bar{x})$, it is <u>undecidable</u> whether there exists a solution $\bar{x} \in \mathbb{N}^r$ to $P(\bar{x}) = Q(\bar{x})$.

<u>Proof</u> <u>of</u> <u>Lemma</u> <u>4.2</u>: Let $R(\bar{x})$ be an arbitrary polynomial with r variables. Then $R(\bar{x}) = 0$ has a solution in $\mathbb{Z}^r$ if and only if one of the 2^r polynomials obtained from R by replacing some of the variables by their negative has a root in $\mathbb{N}^r$. Thus a finite number of tests for non-negative integer roots is enough to find any integer root of R.

Now, let $R_1(\bar{x})$ be a polynomial for which we check for roots in $\mathbb{N}^r$. Let $R_2(\bar{x}) = (R_1(\bar{x}))^2$. Then we have:

$\forall \bar{x} \in \mathbb{N}^r : R_2(\bar{x}) \geq 0$, and the roots of R_2 are clearly roots of R_1 and vice versa

Now, we separate positive and negative coefficients of R_2:

$$R_2(\bar{x}) = P(\bar{x}) - Q(\bar{x}) \geq 0$$

where P and Q are polynomials with non-negative coefficients and clearly satisfy the conditions of the Lemma.

First, we shall show how to get a Petri net to behave like a polynomial.

<u>Lemma</u> <u>4.3</u>: Given a polynomial with non-negative integer coefficients of r variables, $P(x_1, \ldots, x_r)$, there exists a Petri net with $r+1$ distinguished places such that the set of all markings reachable in these distinguished places is the set $\{(x_1, \ldots, x_r, z) \mid x_i \in \mathbb{N} \ \& \ 0 \leq z \leq P(x_1, \ldots, x_r)\}$

There may be many more places in this Petri net than just these distinguished places, but for the moment we disregard their markings.

As an example, consider the following net, which can be seen to correspond to the polynomial of one variable $P(x) = x+1$:

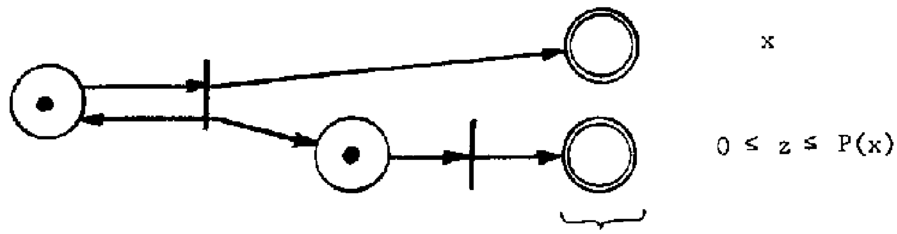


Fig. 4.1

The possible markings for the distinguished places are:

x z			
⟨0, 0⟩	⟨1, 0⟩	⟨2, 0⟩	
⟨0, 1⟩	⟨1, 1⟩	⟨2, 1⟩	
	⟨1, 2⟩	⟨2, 2⟩	
		⟨2, 3⟩	etc.

The relation to the graph of P(x) is obvious: The reachable markings can be represented by the integral points below or on the graph:

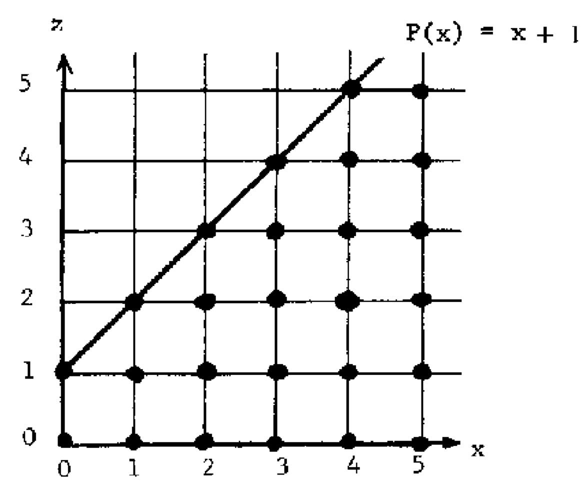


Fig. 4.2

Proof of Lemma 4.3: We shall show how to construct such a net, given a polynomial P with r variables $x_1, \ldots, x_r$.

The general structure is shown below:

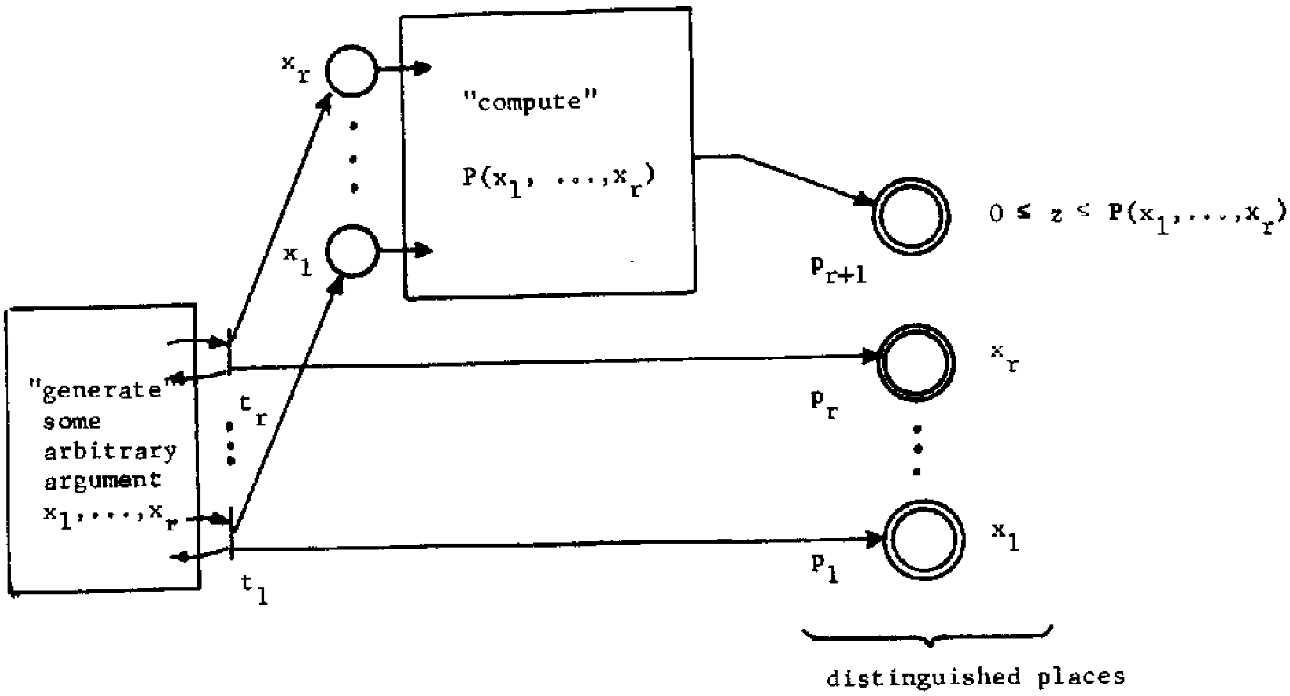


Fig. 4.3

The generation part is easy to build:

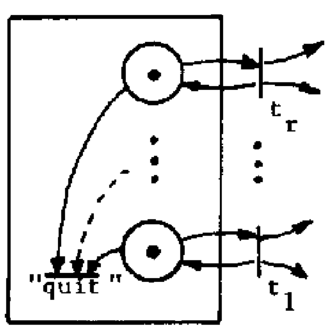


Fig. 4.4 "generate"

Each transition t_i fires some number (possibly zero) of times, generating a value for x_i in two copies (one for the "computer," one for the corresponding distinguished place), then the "generator" quits. The "argument" part of the distinguished marking is now established, and will not be altered.

The "computer" is a Petri net which, for a given "argument" $x_1, \ldots, x_r$, tries to compute $P(x_1, \ldots, x_r)$. However, for the marking z of its output place p_{r+1}, $P(x_1, \ldots, x_r)$ is only an upper bound: No firing sequence can possibly put more tokens on z, but there exists a firing sequence which does put $P(x_1, \ldots, x_r)$ tokens on z. It does not matter if some other firing sequence kills the net before the bound is reached.

Rabin calls such a computation by upper bounds "weak computation," and we are about to show that polynomials with non-negative integer coefficients are weakly computable by Petri nets.

Polynomials are computed by the operations of addition of two numbers, multiplication of two numbers, and substitution of previous results into one or several new additions or multiplications. Now, since, for positive integers, each of the operations add, multiply, copy is non-decreasing as a function of its arguments, if we substitute a reachable upper bound for its arguments, the result will also be a reachable upper bound.

Also, we shall make sure that the reachable upper bound can be approached one token at a time, so that the possible markings of the "result" place include all integers from zero to the bound included.

The add and copy operations can be represented by a Petri net as follows:

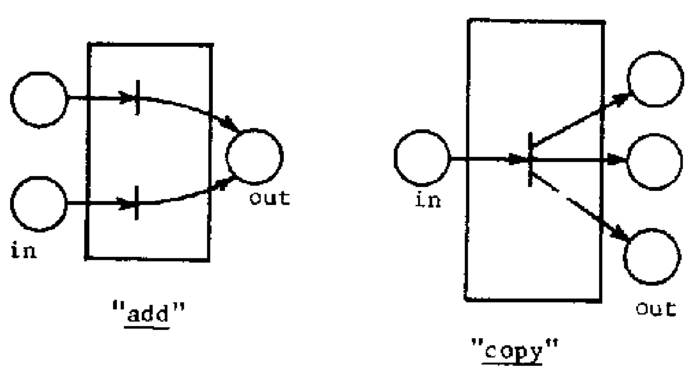


Fig. 4.5

And the following Petri net has a reachable upper bound of $x \cdot y$ in its output place:

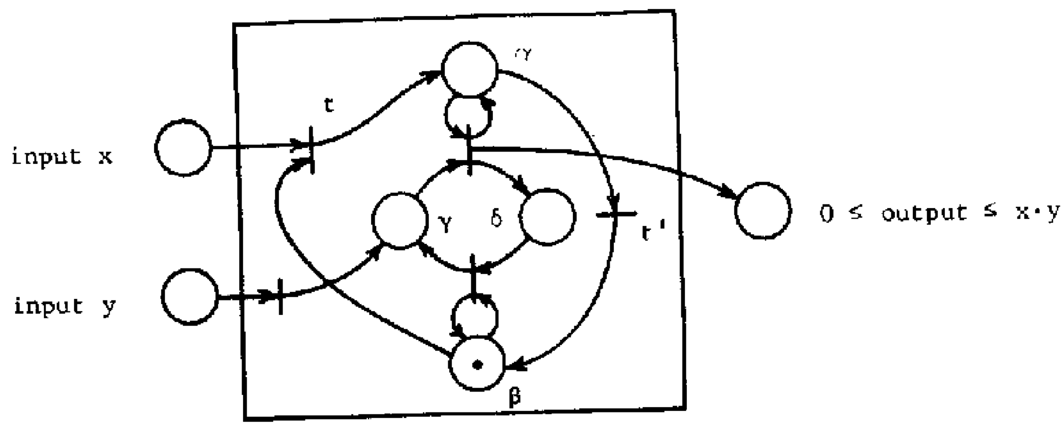


Fig. 4.6 "multiply"

It can be seen that the following strategy yields $x \cdot y$ tokens at the output, and that this cannot be exceeded, though it is possible to exhaust x and thus grind to a halt by firing only t and t', not producing any tokens at the output. The maximum output strategy is: Transfer all y tokens into γ, fire t, transfer all of γ into δ (at this point we have y tokens at the output, $x - 1$ at the input), then fire t' and bring all y tokens back to γ, and repeat this for the remaining $x - 1$ tokens. t can fire only x times, and at most y tokens can be transferred to the output between firings of t.

Having thus shown that addition, multiplication and substitution are weakly computable by Petri nets (and argued that substitution in fact preserves weak computability), we can now construct a Petri net that weakly computes a polynomial, say $3x^2 + 2xy + y^3$, by interconnecting the Petri nets weakly computing add, copy, and multiply, as shown in Fig. 4.7.

Example:

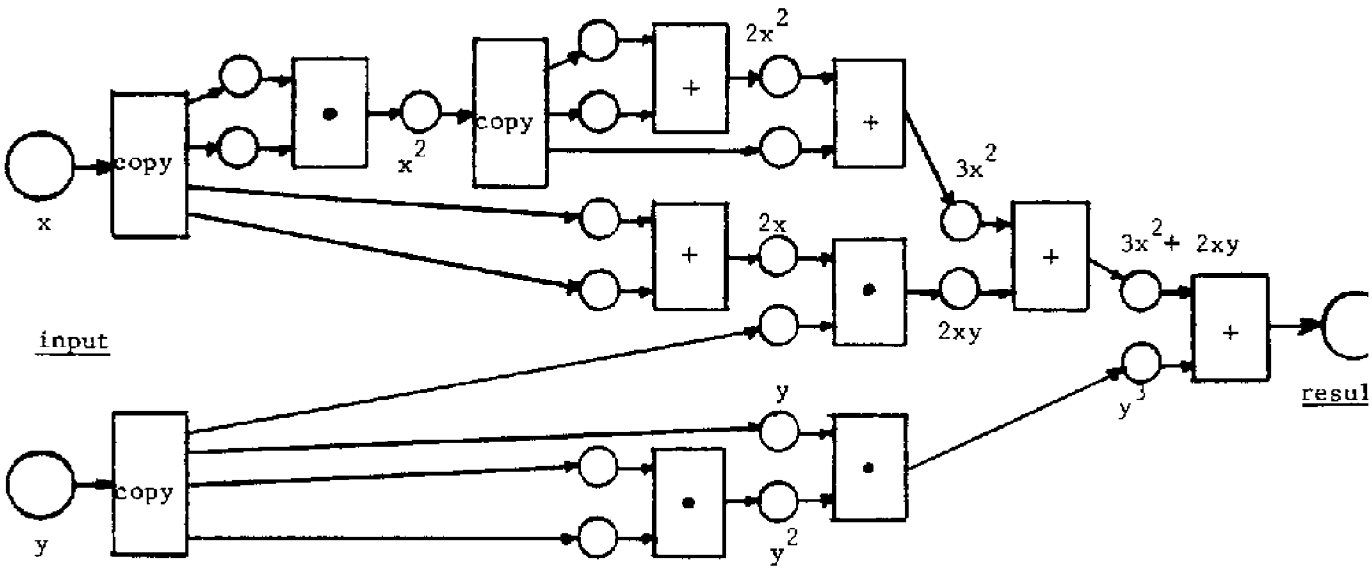


Fig. 4.7 "Compute" $3x^2 + 2xy + y^3$

QED

Now we will show how to construct two Petri nets, A and B, such that every marking reachable by A is also reachable by B if and only if there exists a collection of non-negative integers $x_1, \ldots, x_r$ such that, for two given polynomials P and Q as described in Lemma a, we have:

$$P(x_1, \ldots, x_r) = Q(x_1, \ldots, x_r)$$

Since $P(\bar{x}) \geq Q(\bar{x})$ and since the polynomials only take integral values for integral arguments, we have:

$$(\forall \bar{x} \in \mathbb{N}^r) \quad (P(\bar{x}) = Q(\bar{x}) \Longleftrightarrow P(\bar{x}) < Q(\bar{x}) + 1)$$

As far as the graphs of P and Q+1 in (r+1)-space are concerned, it means that the graph of P "dips under"* the graph of Q+1 if and only if P = Q has a solution:

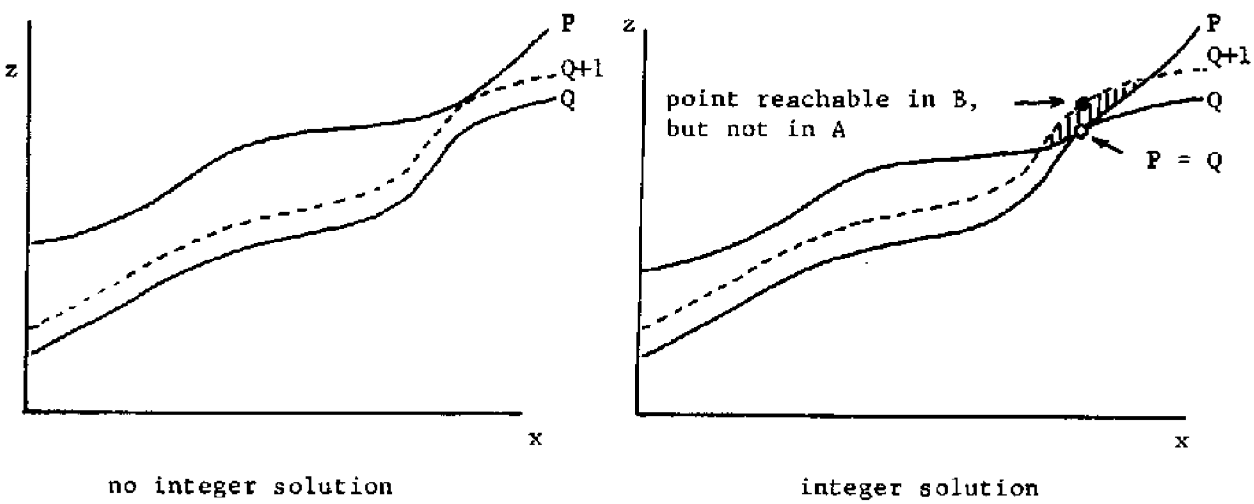


Fig. 4.8

Now let A' and B' be Petri nets corresponding to the polynomials P and Q + 1 according to Lemma b. Every marking of the set of r+1 distinguished places of B' is reachable as a marking of the distinguished set of r+1 places of A' except if the graph of P "dips under" the graph of Q, i.e. if there is an integer solution to the equation P = Q. Yet we want to have two Petri nets A and B where every marking of B is reachable by A if and only if there is no solution to P = Q; we want to compare the markings of two complete nets, not just for a subset of the places.

What remains to be done is to modify A' and B' into two nets A and B of same number of places n, such that every marking of B is reachable in A except if the

*Enough for the "dip" (shaded area in Fig.4.8) to contain an integral point.

marking of the distinguished places of B' cannot be reached by the distinguished places of A'.

As a first step, we add enough extra blank places, not connected to any existing transition, to one of the nets, in order to get two nets of the same number of places $n-2$, then we add two more places α, β to each net. These are all the places in A resp. B. In B, let α be blank and β be marked with one token; neither place is connected to any transition. This completes B, which thus differs from B' in only a few disconnected places. In A, however, we insert a transition from α to β, and we let place α be in a self-loop on every transition of A. We let α be originally marked with one token, and β be initially blank. Thus, as long as the token is in α, A behaves just like A', but when the token transfers to β, all transitions become permanently disabled, and in particular, the marking in the $r+1$ distinguished places will be frozen.

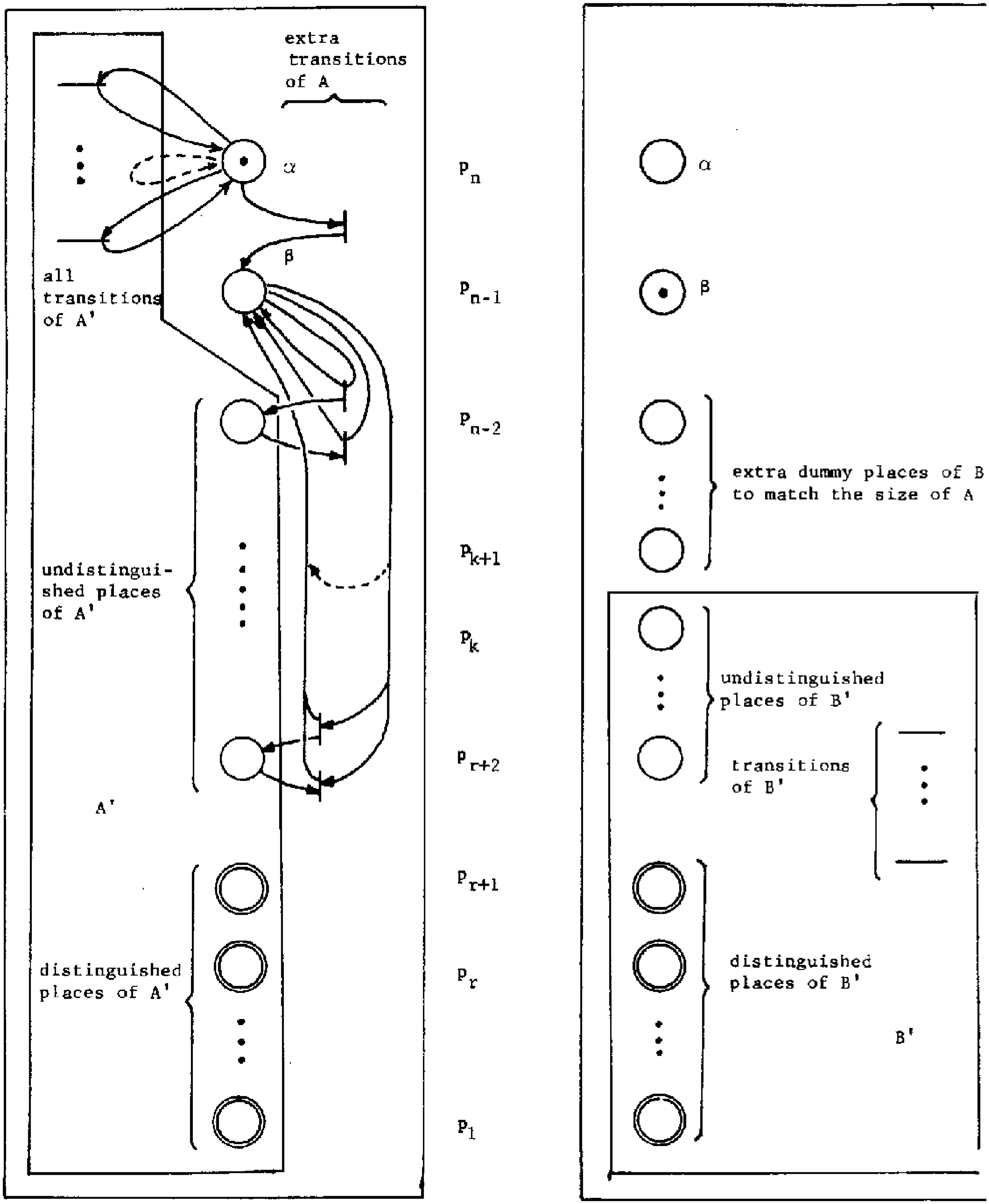
Now, for each of the $n-2-(r+1)$ undistinguished places of A, we add two transitions, one of which puts a token on the place, the other removes a token from it; then we put all these new transitions in self-loops on place β. Thus, after the token from α is transferred to β, any marking can be reached in the undistinguished places of A by firing these extra transitions a suitable number of times.

To see how this construction works, let us see under what conditions every marking reachable in B can also be reached in A.

Let us label the places as follows: $p_1, \ldots, p_r$ are the places containing the argument for the polynomial, p_{r+1} contains a partial result of the computation. These are the $r+1$ distinguished places. For the sake of argument, let the number of places of B' be the smaller number k, and the number of places of A' be $n-2 > k$. We add $n-2-k$ undistinguished places to B'. Let us label the undistinguished places of A and B $p_{r+2} \ldots p_{n-2}$, and let us label α and β, p_n and p_{n-1}, respectively. (See Fig. 4.9)

For comparing markings in A and B, we pair the places according to their labels p_i. Now, any marking of B will be, by construction, of the following form, where $z \leq Q(x_1, \ldots, x_n) + 1$:

$$\begin{array}{ccccccccc} p_1 & \cdots & p_r & p_{r+1} & p_{r+2} & \cdots & p_{n-2} & p_{n-1} & p_n \\ \langle x_1, & \ldots, & x_r, & z, & y_1, & \ldots, & y_{n-r-3}, & 1, & 0\rangle \end{array}$$


extra transitions of A
all transitions of A'
α
β
undistinguished places of A'
A'
distinguished places of A'
p_n
p_{n-1}
p_{n-2}
p_{k+1}
p_k
p_{r+2}
p_{r+1}
p_r
p_1
α
β
extra dummy places of B to match the size of A
undistinguished places of B'
transitions of B'
distinguished places of B'
B'

A B

Fig. 4.9

To reach this marking in A, we must first try to match $p_1, \ldots, p_{r+1}$, since after we match p_{n-1} and p_n, we will have frozen the marking of the distinguished places of A. Therefore, we first generate the argument $x_1, \ldots, x_r$ for polynomial P, then partially compute $P(x_1, \ldots, x_r)$ in a way that, if completed, would actually yield $P(x_1, \ldots, x_n)$ tokens in p_{r+1} of A. But we stop as soon as we reach z, the marking we try to match in p_{r+1} of B. This is possible if and only if $P(x_1, \ldots, x_r) \geq z$, which in turn could fail only if $z = Q(x_1, \ldots, x_r) + 1$ and in fact $P(x_1, \ldots, x_r) = Q(x_1, \ldots, x_r)$. Suppose we could reach z in p_{r+1} of A. As soon as we do, we switch off all transitions of A' by transferring the token from $\alpha(p_n)$ to $\beta(p_{n-1})$, at the same time matching the marking in these two places to the one in B. But now, we can reach any marking we wish in $p_{r+2}, \ldots, p_{n-2}$ of A, by firing the extra transitions of A a suitable number of times; in particular, we can match $y_1, \ldots, y_{n-r-3}$, thus reaching in A the proposed marking of B. As we pointed out, this can be carried out for all markings of B except one where we have:

$$z = Q(x_1, \ldots, x_r) + 1 = P(x_1, \ldots, x_r) + 1$$

But such a marking is reachable in B if and only if the above equation does have a solution in non-negative integers. Thus:

$$(\forall \bar{x} \in \mathbb{N}^r)\, P(\bar{x}) \neq Q(\bar{x}) \iff \text{every marking reachable in B is also reachable in A}$$

QED

5. The Liveness and Reachability Problems for Petri Nets

5.0 Introduction

In this section we study the recursive reducibilities of several related decision problems about Petri Nets, and therefore also about Vector Addition Systems, in view of Section 2.

The main problems are the Liveness Problem and the Reachability Problem. Both have been conjectured to be undecidable, and the first has been conjectured by R. Keller [11] to be reducible to the second.

Liveness Problem: Given a Petri Net and an initial marking, is it live?

Reachability Problem: Given a Petri Net N, an initial marking M_0, and a marking M, is M reachable from M_0? (Is M in the marking class of M_0: $M \in \vec{M}_0$, or $M \in R_N(M_0)$?)

We shall prove Keller's reducibility conjecture, as well as the reducibility of the Reachability Problem to the Liveness Problem:

Theorem 5.1: The Liveness Problem and the Reachability Problem are recursively equivalent.

Finally, we shall discuss some sufficient conditions for these problems to be undecidable or to be decidable.

5.1 The Sub-Problems

We shall prove our result by showing various recursive reducibilities between the following problems:

LP: The Liveness Problem.

SLP: The Liveness Problem for a subset of the transitions of a Petri Net: Is every transition in a given subset live? (In particular, is a given transition live?)

RP: The Reachability Problem: Are Marking Classes recursive?

SRP: The Reachability Problem for a subset of the places of a Petri Net: Given a marking M, does there exist a marking M' reachable from the initial marking such that M and M' coincide on the given subset of places?

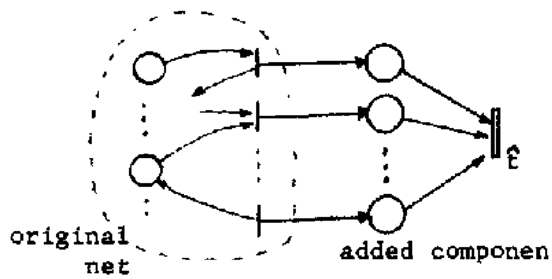
ZRP: The Reachability of the zero (empty) marking. (In Vector Addition

Systems language: Does the Reachability Set contain the Origin?)
SZRP: The Reachability of the Zero marking for a subset of the places.

5.2 The Recursive Equivalence of LP and SLP

The reducibility of LP to SLP is trivial, since LP is a special case of SLP. More to the point, if we know how to test for the liveness of a given transition, we can determine the liveness of a subset of transitions by repeating the test for each transition of the subset.

In fact, we can also construct from a given Petri Net a new net containing an extra transition $\hat{t}$ such that $\hat{t}$ is live if and only if the original net is live.



Clearly, $\hat{t}$ will die if and only if at least one original transition dies.
(Each transition of the original net is connected to one of the additional places)

Fig. 5.1

Now we shall show that if we can decide the liveness of a whole Net, we can decide whether a given subset is live. (Just knowing that a Net is dead does not tell us which transitions are dead; a non-live Net can certainly contain live transitions.) Example:

Figure 5.2

We shall first prove the following remarkable result:

Lemma 5.2: Any Restricted Petri Net N can be simulated by a live Petri Net $\hat{N}$.

That is, we can construct a net $\hat{N}$ such that to every firing sequence of N there corresponds a distinct set of firing sequences of $\hat{N}$; to every marking of N there corresponds a distinct set of markings of $\hat{N}$; the markings reached by corresponding firing sequences always correspond; and if two markings correspond to each other, they can be reached by corresponding firing sequences or not at all. Moreover, the translation is straightforward both ways.

Proof: Let us first dispel the mystery. The following Net is clearly non-live:

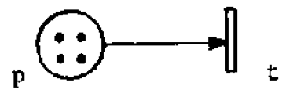
N: 

Figure 5.3

Its firing sequences are:	λ	(the empty string)
	t	
	tt	
	ttt	
	tttt	

The corresponding live net $\hat{N}$ must have infinitely many and unboundedly long firing sequences. Thus, clearly there must be a set of firing sequences of $\hat{N}$ for each sequence of N. We achieve this by having 4 transitions a,b,c,d correspond to t, and certain patterns of firings will correspond to a firing of t; others will correspond to a non-firing of t. In particular, the correspondence will be: (represented as regular expressions)

N	$\hat{N}$
λ	(acbd)*
t	(acbd)* (ab + acbabd) (acbd)*
tt ttt tttt	(ab + acb(ab)*d)* where the number of occurences of substring ab is 2,3 and 4, respectively.

In other words, every firing of t is represented by the occurence of the substring ab against a background pattern of acbdacbdacbd...
The arrows show where the substring may occur (singly or multiply).
The background pattern fires all four transitions arbitrarily often, without possibility of deadlock: It is live.

The graph of $\hat{N}$ is shown below:

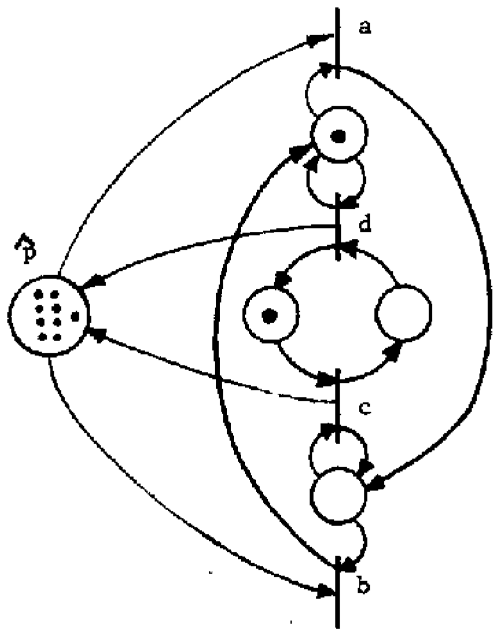


Figure 5.4

Place $\hat{p}$ corresponds to p; its initial marking is twice the marking of p plus one extra token. This extra token is what keeps the net alive when p is empty. It can be seen that a and b remove tokens from $\hat{p}$, whereas c and d put a token back. The "empty" pattern acbd thus jiggles the extra token back and forth, whereas the pattern ab removes two tokens from $\hat{p}$ (corresponding to one token removed by a firing of t in N) and restores the state of the four additional places. These four places can be in any of four "phases":

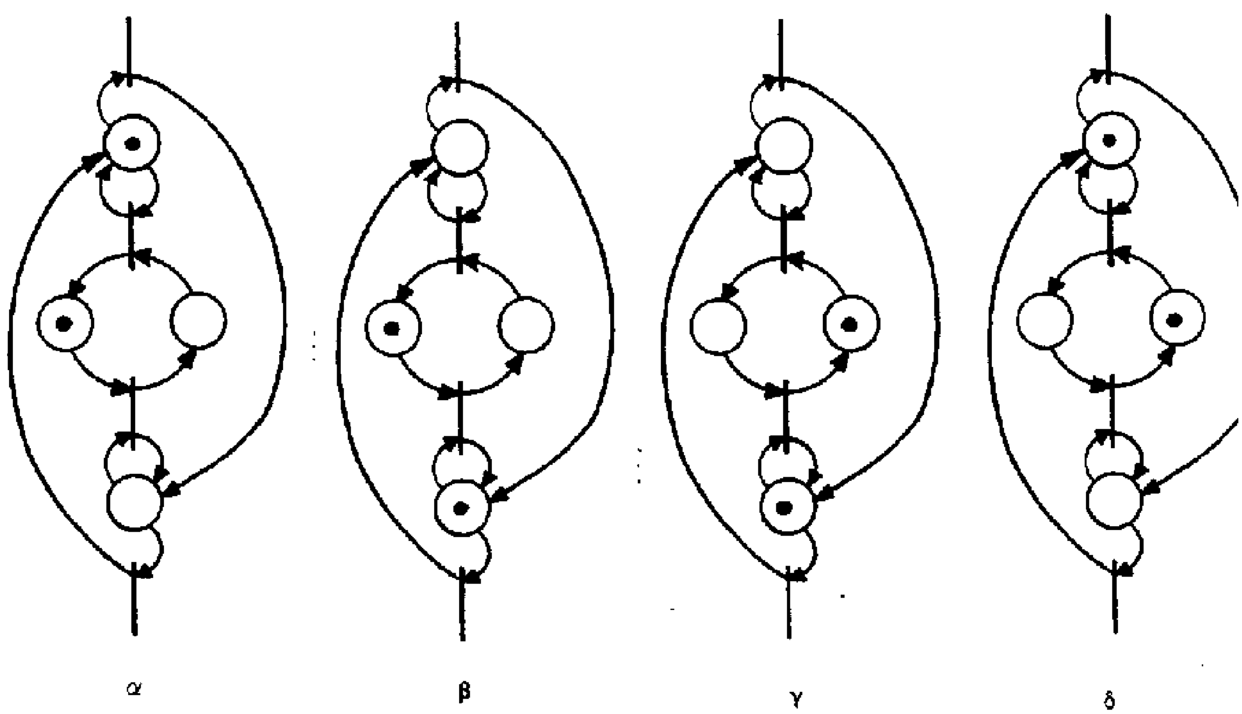
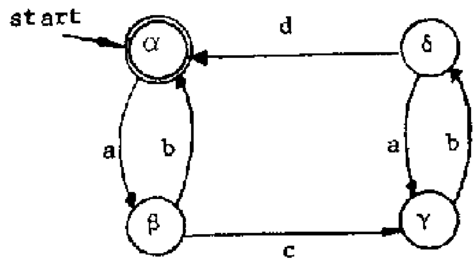


Figure 5.5

The action of this net can be represented by the following state diagram:



$$\begin{cases} ((ab)^* \; acb(ab)^*d)^* \\ (ab)^* \end{cases}$$

Figure 5.6

We now proceed to the general construction of $\hat{N}$, given a Petri Net N whose places are $p_1 \ldots p_r$ and whose transitions are $t_1 \ldots t_s$. $\hat{N}$ will contain one place $\hat{p}_i$ for each place p_i in N, plus, for each transition t_j, four places π_{j1}, π_{j2}, π_{j3}, π_{j4} and six transitions $t_{ja}, \ldots t_{jf}$, plus one additional place which we call the hub. Each transition t_j is replaced by a construction like that shown before, which is connected to the hub by means of t_{je} and t_{jf}, and to the appropriate $\hat{p}_i$ places as follows: if there is an arc from p_i to t_j in N, there will be two arcs from $\hat{p}_i$ to t_{ja} and t_{jb} and two arcs from t_{jc} and t_{jd} to $\hat{p}_i$. If there is an arc from t_j to p_k, there will be two arcs from t_{ja} and t_{jb} to $\hat{p}_k$, and two arcs from $\hat{p}_k$ to t_{jc} and t_{jd}. Thus, the effect of firing t_j can be modeled as in the example before.

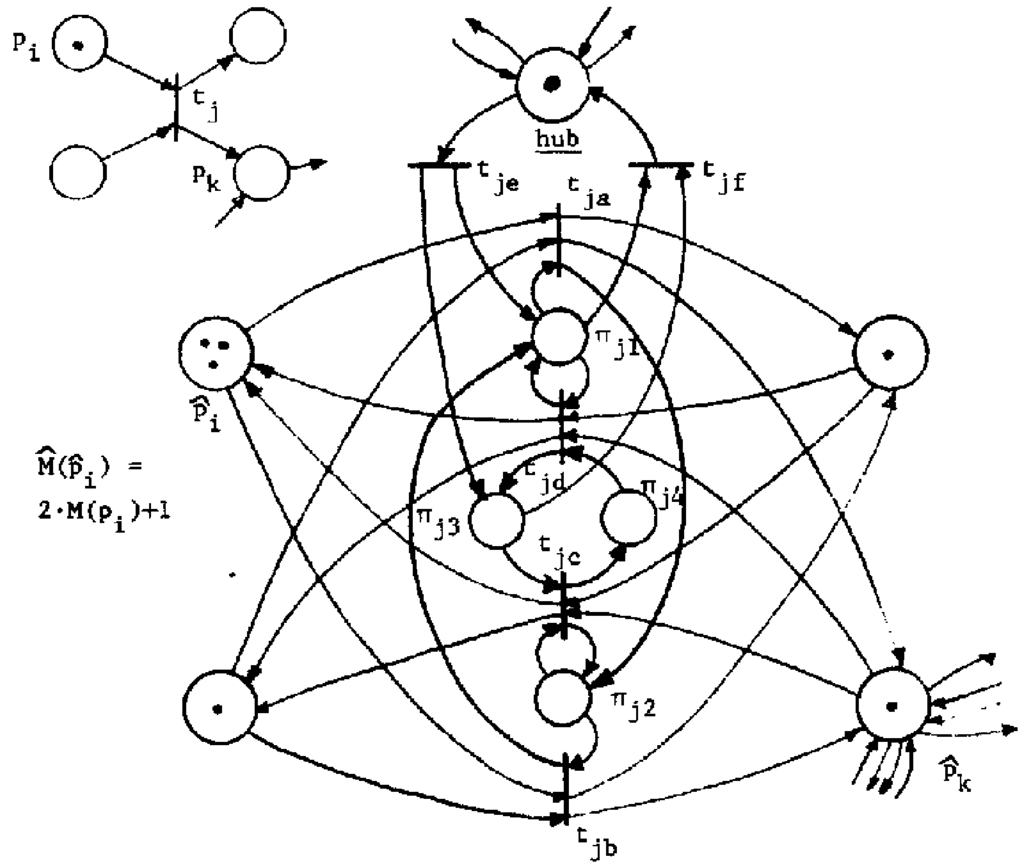


Figure 5.7

The initial marking of $\hat{N}$ consists of twice the marking of N in p_i for $\hat{p}_i$ plus the "steady-state" background marking of one extra token for each $\hat{p}_i$, and a token in the hub. Whenever the hub is marked, we say that the net $\hat{N}$ is at rest. Otherwise, it is active, and is in some phase α_j, β_j, γ_j or δ_j as illustrated before. Note that each transition cycle will start and end in phase α; t_{je} starts the cycle for t_j, and t_{jf} returns the token to the hub, thus permitting some other transition firing to be simulated. This guarantees that all steady state tokens have been returned to where they were before, switching to some other transition complex. The only effective marking changes are those due to an ab firing of t_j, such as $t_{ja}t_{jb}$, which transfers a pair of tokens from an input place (as seen in the original net N) to an output place of t_j.

The correspondence between markings is simple. If $\hat{N}$ is at rest or in some phase α or γ, we have $\hat{M} = 2M + 1$. If $\hat{N}$ is in phase β_j or δ_j the input places to t_j lack the steady state token, and the output places of t_j have two steady state tokens in addition to 2M. The firing sequences of $\hat{N}$ are clearly of the form: $(w_1 + w_2 + \ldots w_m)^*$

where

$$w_j = t_{je}(\underline{t_{ja}t_{jb}} + t_{ja}t_{jc}t_{jb}(\underline{t_{ja}t_{jb}})^* t_{jd})^* t_{jf}$$

and each occurence of $\underline{t_{ja}t_{jb}}$ corresponds to a firing of t_j in N, and has the corresponding effect on the marking of $\hat{N}$.

<u>QED</u>

It can be seen that this construction would fail if there were multiple arcs or self-loops, since it would be impossible to manage the steady-state tokens. But if we have an arbitrary generalized Petri Net, we can always transform it into an equivalent self-loop free single-arc Petri Net by using the construction shown in Section 2, which does not affect liveness. The translation would then be a two-step procedure, but we observe that the combined translation still has the properties outlined under Lemma 5.2. This gives the following

<u>Corollary 5.3</u>: Any Generalized Petri Net can be simulated (in the sense of Lemma 5.2) by a live Petri Net.

<u>Remark</u>: In [1] , Baker objects to the fact that the constructions used in Rabin's proof are neither live nor consistent, whereas all "nice" systems should only be represented by live consistent nets. (A Petri Net is said to be consistent if there exists a firing sequence which fires each transition at least once and returns to the original marking, i.e., there exists a steady-state behaviour involving all transitions in the net). But we can easily apply the method just presented to construct from the two nets A and B of Section 4, two new nets $\hat{A}$ and $\hat{B}$ which are live and consistent (our construction certainly provides for a consistent steady-state firing - the one corresponding to no "real" firings at all) whose reachable markings agree if and only if those of A and B agree:

<u>Corollary 5.4</u>: Given two live, consistent GPN's, it is undecidable whether every marking reachable by one is also reachable by the other.

Now we are ready to use the construction of Lemma 5.2 to prove the following Lemma:

<u>Lemma 5.5</u>: LP and SLP are recursively equivalent.

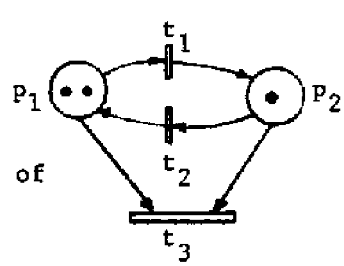
<u>Proof</u>: We have to show that SLP can be reduced to LP.

Suppose we wish to test the liveness of a certain subset of transitions $T \subseteq \{t_1,\ldots t_m\}$, say $T = \{t_1,\ldots,t_k\}$, in a given Petri Net N.

We construct a new net $\hat{N}'$ by using the construction of the live equivalent $\hat{N}$ for the transitions not to be tested for liveness, i.e., for $\{t_{k+1},\ldots t_n\}$. Remembering that the marking of $\hat{N}$ is double that of N (plus steady-state tokens), we replace the single arcs leading to or from the transitions to be tested ($\{t_1,\ldots t_k\}$) with double arcs in $\hat{N}'$, and call the transitions $\{\hat{t}_1,\ldots\hat{t}_k\}$. Thus, the effect of firing $t_{ja}t_{jb}$ or $\hat{t}_i$ in $\hat{N}'$ affects the marking of $\hat{N}'$ similarly by moving <u>pairs of tokens</u> in $\hat{N}'$ for each token moved correspondingly in N. But we have to make sure that the steady-state tokens do not interfere. As long as there is only one, it will not be noticed by the double arcs. But if the net is in phase β or δ, there may be two steady-state tokens in some place, which could cause a false firing of some $\hat{t}_i$. To prevent this, we put each $\hat{t}_i$ in a self-loop through the hub, as shown in the following example. Now these transitions can fire only when the net $\hat{N}$ is "at rest".

<u>Example</u>:

Test the liveness of $\{t_1, t_2\}$

N = 

<u>Figure 5.8</u>

The corresponding net $\hat{N}'$ is:

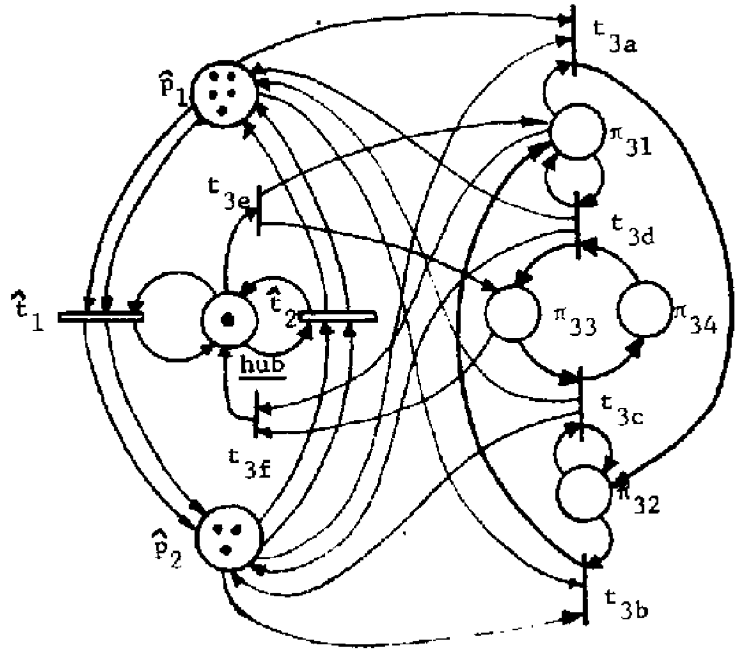


Figure 5.9

Now, in $\hat{N}'$, all transitions except possibly $\hat{t}_1$ and $\hat{t}_2$ are live by construction. Thus, the whole net is live if and only if $\{\hat{t}_1,\hat{t}_2\}$ is live, which gives us a liveness test for $\{t_1,t_2\}$ in N as soon as we can test the liveness of the net $\hat{N}'$.

QED.

5.3 The Recursive Equivalence of RP, SRP, ZRP and SZRP

Let us first establish the trivial reducibilities: Both RP and SZRP are particular cases of SRP, and ZRP is a particular case of RP and SZRP. Thus ZRP is reducible to both RP and SZRP, each of which is reducible to SRP.

Lemma 5.6: RP, SRP, ZRP and SZRP are recursively equivalent.

Proof: We have to show that SRP is reducible to ZRP to complete the proof.

Suppose we wish to test for the reachability of the submarking $(m_1, m_2, \ldots m_k)$ of the subset of places $\{p_1, \ldots p_k\} \subseteq \{p_1, \ldots p_r\}$ of some Petri Net N with a given initial marking. We shall construct a net $\tilde{N}$ obtained from N by adding

1. an extra transition θ_i for each place $p_i \in \{p_{k+1}, \ldots, p_r\}$ in whose marking we are not interested.
2. two extra transitions θ_a and θ_b.
3. two extra places π_a and π_b, where π_a is initially marked with one token and π_b is blank.

Now we connect these extra elements to a copy of N as shown below, where the size of the bundle from p_i to θ_a is m_i, i.e., the firing of θ_a removes exactly the submarking whose reachability we wish to test.

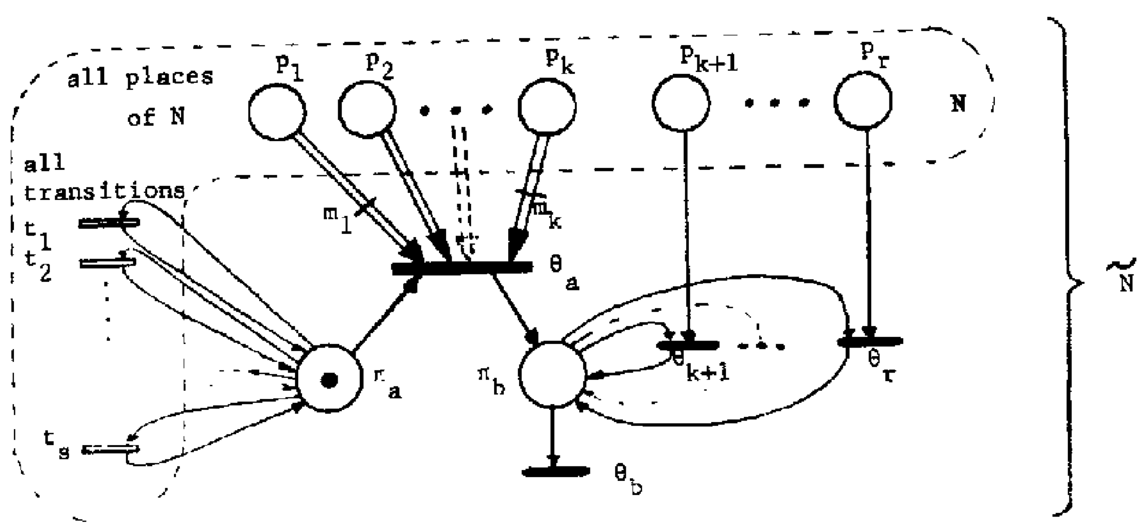


Figure 5.10

π_a self-loops on <u>every</u> transition in N. θ_a transfers the token from π_a to π_b, which self-loops on every θ_j, $k + 1 \le j \le n$. Now θ_a can fire if and only if a marking can be reached which covers the one we are testing for reachability. θ_a can fire at most once; if it does, it freezes all activity in N by removing the token from π_a, thus disabling every transition of N. The marking of $\{p_1, \ldots p_k\}$ is now zero if and only if the tested submarking was reachable. Now the token in π_b can pump all other tokens from $\{p_{k+1}, \ldots p_n\}$ via transitions $\theta_{k+1}, \ldots, \theta_n$, and finally exit π_b via θ_b, reaching the all-zero marking of $\tilde{N}$ if and only if the tested submarking was reachable in N.

QED.

5.4 The Reducibility of RP to LP

<u>Lemma 5.7</u>: RP is recursively reducible to LP.

<u>Proof</u>: Actually, we prove that ZRP is reducible to SLP. Lemma 5.7 then follows from the equivalences proved in Lemmas 5.5 and 5.6

We wish to test whether in a given Petri Net N (with its initial marking), the zero marking is reachable. We construct from it a new net N* in which a certain transition θ_a is live if and only if the zero marking is not reachable in N. Then a test for the liveness of θ_a in N* will be a reachability test for N.

We construct N* as follows, starting with a copy of N, to which we add:

1. two places π_a and π_b, where π_a self-loops on every transition in N. π_a is initially marked with one token; π_b is blank.
2. a transition θ_a from π_a to π_b.
3. for every place p_i in N, a transition θ_i which self-loops on p_i and transfers a token from π_b to π_a.

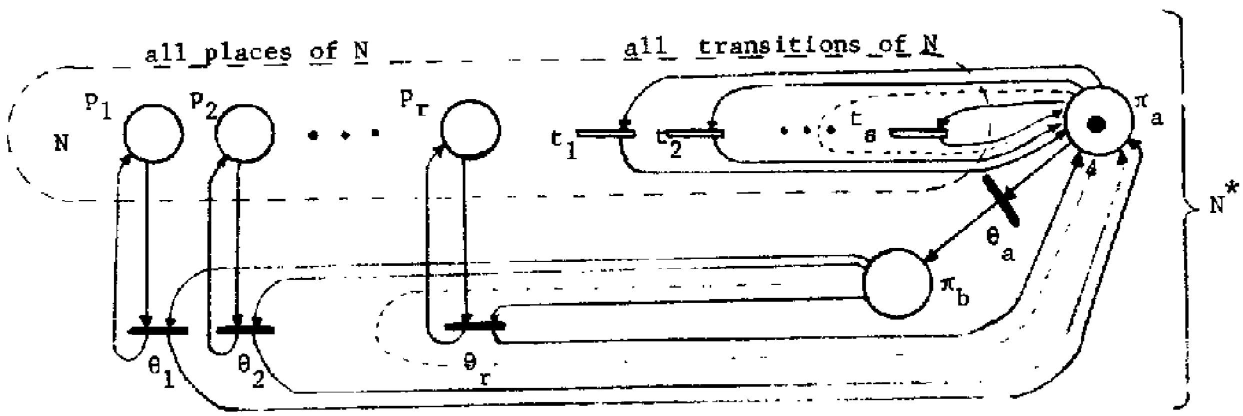


Figure 5.11

The token in π_a permits N* to fire exactly like N and generates the same markings in $p_1 \cdots p_r$. Once in a while θ_a fires and thus freezes N by removing this token. The token can get back to π_a if and only if at least one θ_i is enabled, i.e., the present marking of N is not zero. It is thus clear that θ_a is live if and only if that zero marking of N is not reachable.

QED.

5.5 The Equivalence of LP and RP

What remains to be proved is that some form of the Liveness Problem can be reduced to some form of the Reachability Problem. In particular, we show that the liveness of a transition t (SLP) can be decided by testing the reachability of a finite number of submarkings (SRP) which we call t-dead submarkings.

Definition 5.1: Given a Petri Net $N = \langle \Pi, \Sigma, F, B, M_0 \rangle$ with $\Pi = \{p_1, \ldots, p_r\}$ and a transition $t \in \Sigma$:

a. A marking $M \in \mathbb{N}^r$ is said to be t-dead iff, starting from M, there does not exist a firing sequence which eventually fires t:

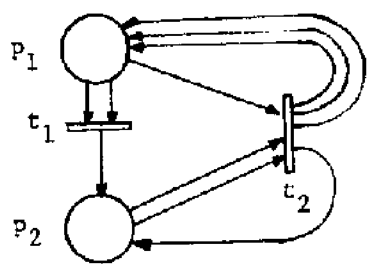
$$T \text{ is t-dead} \overset{\Delta}{\Leftrightarrow} t \text{ is not potentially firable at } M$$

b. A submarking m of a subset $P \subseteq \Pi$ is said to be <u>t-dead</u> iff every marking M which agrees with m (i.e. $m = M/P$) is t-dead.

(See Definitions 1.9 and 1.10, page 11.)

From the definition of liveness (Definition 1.16), it follows that t is live if and only if no t-dead marking is reachable. Now there may be an infinite number of t-dead markings, but by checking the reachability of a submarking, we are in fact checking the reachability of an infinite number of markings in one step: if the submarking is <u>not</u> reachable, no marking agreeing with it is reachable. Therefore, if every t-dead marking agrees with at least one t-dead submarking from a finite set D_t of t-dead submarkings, then transition t is live if and only if no submarking in D_t is reachable: Checking liveness reduces to checking the reachability of a finite number of submarkings.

The following example shows in what context submarkings are considered for t-deadness. In the net of Fig. 5.12, if p_1 is blank, no amount of



<u>Fig. 5.12</u>

tokens will make t potentially firable; if p_2 is blank, it must receive a token via a firing of t_1, to fire t_2, and therefore we can see that the only t_1-dead markings are $\langle 1, 0\rangle$, $\langle 2, 0\rangle$, and all markings of the form $\langle 0, x\rangle$, where $x \in \mathbb{N}$. But these markings $\langle 0, x\rangle$ are precisely all markings which agree with the submarking $p_1 = 0$, or, more formally, the submarking $\langle 0, 0\rangle/\{p_1\}$. If we are given an initial marking, say $M_0 = \{5, 0\}$, it is therefore enough to check the reachability of one submarking $p_1 = 0$ and two markings $\langle 1, 0\rangle$ and $\langle 2, 0\rangle$. As it turns out, neither of the two markings $\langle 1, 0\rangle$ and $\langle 2, 0\rangle$ are reachable, since if t_1 does not fire, there will always be more than 4 tokens in p_1, and after t_1 fires, p_2 will always contain at least one token. The submarking $p_1 = 0$ is also not reachable since no firing of t_1 or t_2 changes the parity of the marking in p_1. Since $M_0(p_1)$ is odd, we cannot reach a marking with zero tokens in p_1. The conclusion is that t_2 is live at $M_0 = \langle 5,0\rangle$.

An important property of t-dead markings is that any marking covered by a t-dead marking must also be t-dead. That is because any firing sequence starting at the smaller marking is also firable at the larger marking.

Now we adopt the following convention for representing a submarking M/P, where $P \subseteq \Pi$, by a vector $V \in (\mathbb{N} \cup \{\omega\})^r$.

Definition 5.2: A vector $V \in (\mathbb{N} \cup \{\omega\})^r$ is said to be a submarking M/P iff the finite coordinates of V are those of the places in P, and they agree with M:

$$\forall i,\ 1 \le i \le r: \begin{cases} p_i \in P \Rightarrow V(i) = M(p_i) \\ p_i \in \Pi - P \Rightarrow V(i) = \omega \end{cases}$$

Example: The t-dead submarkings of the Petri net in Fig. 5.12 are $\langle 1, 0\rangle$, $\langle 2, 0\rangle$, and $\langle 0, \omega\rangle$.

Now we can compare t-dead markings and submarkings by means of the $\le$ relation on $(\mathbb{N} \cup \{\omega\})^r$, as defined in Section 3.1 page 24.

Lemma 5.8: If V is a t-dead marking or submarking, and $V' \le V$, then V' is also t-dead. $(V, V' \in (\mathbb{N} \cup \{\omega\})^r)$.

Proof: If V is a marking, i.e. has no ω-coordinates, then V' is also a marking. V' is covered by a t-dead marking and hence must be t-dead. If V is a submarking, then any marking V" which agrees with V in its finite coordinates is t-dead, by definition. If V' is smaller than V, then every marking which agrees with V' is covered by some marking V" which agrees with V, hence must also be t-dead. Then V' is t-dead by Definition 5.1.

QED

This lemma justifies our convention for representing submarkings as vectors in $(\mathbb{N} \cup \{\omega\})^r$. In fact, this convention also permits us to give a stronger form to Theorem 3.3: A submarking $V \in (\mathbb{N} \cup \{\omega\})^r$ is <u>strongly coverable</u> (see Definition 1.11) in a Petri Net N iff there exists a label $L \geq V$ in the coverability tree of N.

Now we are ready to look for a finite set of t-dead submarkings D_t which is sufficient to decide the liveness of t.

Let $\hat{D}_t$ be the (possibly infinite) set of <u>all</u> t-dead markings and submarkings. For example, $\hat{D}_t$ of the Petri Net in Fig. 5.12 would be the set $\{\langle 1, 0\rangle, \langle 2, 0\rangle, \langle 0, \omega\rangle, \langle 0, 0\rangle, \langle 0, 1\rangle, \langle 0, 2\rangle, \langle 0, 3\rangle, \ldots\}$.

<u>Definition</u> <u>5.3</u>:

a. A submarking $V \in (\mathbb{N} \cup \{\omega\})^r$ is said to be <u>superseded</u> by a proper submarking V' of V iff every finite coordinate of V' is equal to the corresponding coordinate of V"

$$V' \text{ \underline{supersedes} } V \overset{\Delta}{\Leftrightarrow} V' \neq V \;\&\; (\forall i: V'(i) \neq \omega \Rightarrow V(i) = V'(i))$$

b. The set of <u>unnecessary t-dead submarkings</u> is the set

$$U_t = \{V \in (\mathbb{N} \cup \{\omega\})^r \,|\, \exists V' \in \hat{D}_t: V' \text{ \underline{supersedes} } V\}$$

c. The <u>reduced t-dead set</u> is defined as

$$D_t = \hat{D}_t - U_t$$

From Lemma 5.8 it is clear that $U_t \subseteq \hat{D}_t$. The submarkings in U_t are unnecessary for the purpose of testing the liveness of t. Indeed, if V' supersedes V and V' is reachable, then some t-dead V of which V' is a submarking will be reachable, hence t is not live. If V' is not reachable, then no V of which V' is a submarking is reachable. In either case, V is not needed explicitly to establish the liveness of t. Therefore, it is enough to check the non-reachability of all submarkings in the reduced t-dead set D_t to establish the liveness of t.

Now we show that D_t must always be finite.

Lemma 5.9: For a given Petri net and a given transition t, the reduced t-dead set D_t is finite.

Proof: Assume D_t is infinite. It is certainly denumerable, so let us arrange it into an infinite sequence of distinct vectors in $(\mathbb{N} \cup \{\omega\})^r$. By Lemma 3.5, there must be an infinite strictly increasing (all elements being distinct) subsequence of this sequence. Some coordinates in this sequence may be bounded, others may eventually reach ω, after which they must remain at ω. After some finite initial segment, there remains an infinite tail where some coordinates are constant, the others increase without bound. Thus an infinite D_t must contain an infinite subset of t-dead submarkings W whose elements all agree in some set of coordinates, and take on arbitrarily large finite (non-ω) values in the others. Let V be a vector which agrees with all vectors in W in the "constant" coordinates, and whose remaining coordinates are ω. Clearly, V denotes a submarking which is not reachable in N only if no submarking in W is reachable in N. V must also be t-dead, because if it were not, then some marking which agrees with V in its finite coordinates would not be t-dead, and yet it would be exceeded by some vector V' in W, since the coordinates of V' which correspond to ω-coordinates in V are either ω or can be made arbitrarily large in W. But this contradicts Lemma 5.9. But this vector V, which is t-dead and covers all of W, is in $\hat{D}_t$, and is a proper submarking of every element of W. Hence $W \subseteq U_t$, which is incompatible with $D_t = \hat{D}_t - U_t$. Thus D_t must be finite.

QED

What we have shown so far is:

Corollary 5.10: The liveness of a transition t can be established by checking the reachability of a finite set D_t of t-dead submarkings.

What remains to be proved is that D_t can be effectively constructed.

<u>Lemma</u> <u>5.11</u>: Given a Petri Net $N = \langle \Pi, \Sigma, F, B \rangle$ and a transition $t \in \Sigma$, it is decidable whether a submarking $V \in (\mathbb{N} \cup \{\omega\})^r$ is t-dead.

<u>Proof</u>: (No initial marking is mentioned for N, since the concept of t-deadness is independent of the initial marking.) For a marking $M \in \mathbb{N}^r$, transition t is potentially firable iff a marking can be reached which covers $F(t)$, i.e. iff some label in $T_N(M)$ covers $F(t)$. (Theorem 3.3). The argument can be adapted for submarkings as follows. We extend the definition of a coverability tree $T_N(V)$, where $V \in (\mathbb{N} \cup \{\omega\})^r$, by allowing the label of the root node to already contain some ω-coordinates, without ω-backpointers to be sure. The definition (Definition 3.1, page 25) need not be changed; the label of the root node of $T_N(V)$ will be $L_\rho = V$, and the construction proceeds without modification.

Suppose that V is not t-dead. Then there exists a marking M which agrees with V in the finite coordinates, from which a firing sequence leads to a marking which covers $F(t)$, of course, $M \leq V = L_\rho$. By repeating the argument used in the proof of Lemma 3.7 (page 32), we can see that the firing sequence is also a λ-composite path to a node α such that $L_\alpha \geq F(t)$.

Suppose that there exists a node α such that $L_\alpha \geq F(t)$. Let $V_\alpha \geq F(t)$ be a target for α, i.e. V agrees with L_α in its finite coordinates. By using the proof of Lemma 3.10 (page), we compute a target V_ρ for the root node of $T_N(V)$; we have $V_\rho \leq V$ and V_ρ agrees with V in the finite coordinates of V, i.e. V_ρ is a marking which agrees with submarking V, and from V_ρ a marking M_α can be reached such that $M_\alpha \geq V_\alpha \geq F(t)$, i.e. t is potentially firable at V_ρ, and thus V_ρ is not t-dead. Therefore, submarking V cannot be t-dead.

Thus, V is t-dead if and only if $T_N(V)$ does not contain any label which covers $F(t)$, which is decidable since the construction of $T_N(V)$ is finite and effective. There is no change to the finiteness proof in Theorem 3.1 for $T_N(V)$.

QED

<u>Lemma</u> <u>5.12</u>: For a given Petri Net N and a given transition t, the reduced t-dead set D_t, as defined in Definition 5.3, can be effectively constructed.

<u>Proof</u>: We show how to effectively find an upper bound on the finite coordinates of all vectors in D_t. Then there will be a known finite set from which all vectors in D_t are taken. Since this selection is itself an effective procedure, by virtue of Lemma 5.11, the construction of D_t is effective.

We shall use the following property of a strict upper bound -- call it a; $a \in \mathbb{N}$ -- on the finite coordinates in D_t. Because of Lemma 5.8 this means that no finite coordinate of any vector in D_t is equal to a. Now suppose there is a t-dead marking $V \in \hat{D}_t$, and some of its coordinates are equal to a. Then $V \notin D_t$, which implies $V \in U_t$, and hence is superseded by a submarking V' of V which has ω-coordinates where V has a-coordinates. We write this as:

$$\forall i,\ 1 \leq i \leq r:\ V_{a\to\omega}(i) = \underline{\text{if}}\ V(i) \geq a\ \underline{\text{then}}\ a\ \underline{\text{else}}\ V(i)$$

$$\left.\begin{matrix}(\underline{a} \text{ is a strict upper bound for} \\ \text{the finite coordinates in } D_t)\end{matrix}\right\} \Rightarrow \quad (\forall V:\ V \in \hat{D}_t \Rightarrow V_{a\to\omega} \in \hat{D}_t)$$

Now we show that the converse is true in a stronger form:

Let $A \in \mathbb{N}^r : \forall i,\ A(i) = a$.

Suppose $a \in \mathbb{N}$ is such that:

$$\forall V \leq A:\quad V \in \hat{D}_t \Rightarrow V_{a\to\omega} \in \hat{D}_t$$

Let $V_1 \in \mathbb{N}^r$ be a t-dead marking with possibly some coordinates greater than $\underline{a}$. Let V_2 be the vector whose coordinates are those of V_1 if they are less than $\underline{a}$, and which are set equal to $\underline{a}$ otherwise:

$$\forall i,\ 1 \leq i \leq r:\ V_2(i) = \underline{\text{if}}\ V_1(i) \geq a\ \underline{\text{then}}\ a\ \underline{\text{else}}\ V_1(i)$$

Clearly we have $V_2 \leq V_1$, and hence $V_2 \in \hat{D}_t$. But we also have $V_2 \leq A$, and hence, from our assumption about $\underline{a}$:

$$(V_2 \leq A\ \ \&\ \ V_2 \in \hat{D}_t) \Rightarrow {V_2}_{a\to\omega} \in \hat{D}_t$$

But $V_{2_{a\to\omega}}$ has ω-coordinates where V_1 has coordinates greater than or equal to $\underline{a}$, and if V_1 has such coordinates, then $V_{2_{a\to\omega}}$ supersedes V_1, hence $V_1 \in U_t$ and $V_1 \notin D_t$. This shows that no vector in D_t has finite coordinates greater than or equal to $\underline{a}$:

$$\left.\begin{array}{l}\underline{a}\text{ is a strict upper bound on}\\ \text{the finite coordinates in } D_t\end{array}\right\} \Leftrightarrow \forall V \le A:\ V \in D_t \Rightarrow V_{a\to\omega} \in D_t$$

For a given $\underline{a}$, only a finite number of questions of the form $V \overset{?}{\in} D_t$ and $V_{a\to\omega} \overset{?}{\in} D_t$, for $V \le A$, have to be answered to establish $\underline{a}$ as a strict upper bound. By trying for successively higher values for $\underline{a}$, this bound will be effectively found, since it exists (D_t is finite by Lemma 5.9).

QED

To summarize this lengthy section:

We have shown that by testing the reachability of the elements of an effectively constructible finite set D_t of submarkings, we can decide the liveness of t. Thus SLP is reducible to SRP. Together with Lemmas 5.5, 5.6 and 5.7 this proves Theorem 5.1: LP and RP are recursively equivalent to each other and to SLP, SRP, ZRP and SZRP.

5.6 A Sufficient Condition for the Undecidability of the Liveness and Reachability Problems

In Section 4, we use Petri Nets that behave like polynomials in the sense that there is a place which has a reachable upper bound expressed by a polynomial whose argument is the initial marking of some distinguished set of places. Now we introduce the concept of a reachable lower bound as a function of the initial marking in some distinguished set of places.

Definition 5.4: In a Petri Net with a given initial marking M_0, a place p_i is said to have a reachable guaranteed minimum (rgm) b_i iff, from every marking M_1 in the marking class, we can reach a marking M_2 such that $M_2(p_i) \geq b_i$, and there also exists a marking $M_3 \in R(M_0)$ such that $M_3(p_i) = b_i$ and $\forall M_4 \in R(M_3)$, $M_4(p_i) \leq b_i$.

Basically, this means that no matter what firing sequence has already happened, it can always be continued until b_i is reached or exceeded, but there also exists a firing sequence after which b_i cannot be exceeded anymore (but can still be reached).

A reachable guaranteed minimum is of course not a bound, but it is potentially a lower bound. This is in contrast to the reachable upper bound (rub); this latter of course is a bound. No firing sequence can exceed the rub, but there exists a firing sequence which reaches it.

Now suppose instead of a rub-polynomial "computer" as in Section 4, we had an rgm-polynomial "computer". Then, given two polynomials with non-negative coefficients P and Q satisfying the conditions of Lemma 4.1, let us construct a rub-computer for Q and an rgm-computer for P (assuming this can be done, which is by no means certain), and then let us connect them together in the following way:

(If A is a quantity, we indicate that it is a reachable upper bound by writing $\ulcorner A \urcorner$; if it is a reachable guaranteed minimum, we write $\llcorner A \lrcorner$)

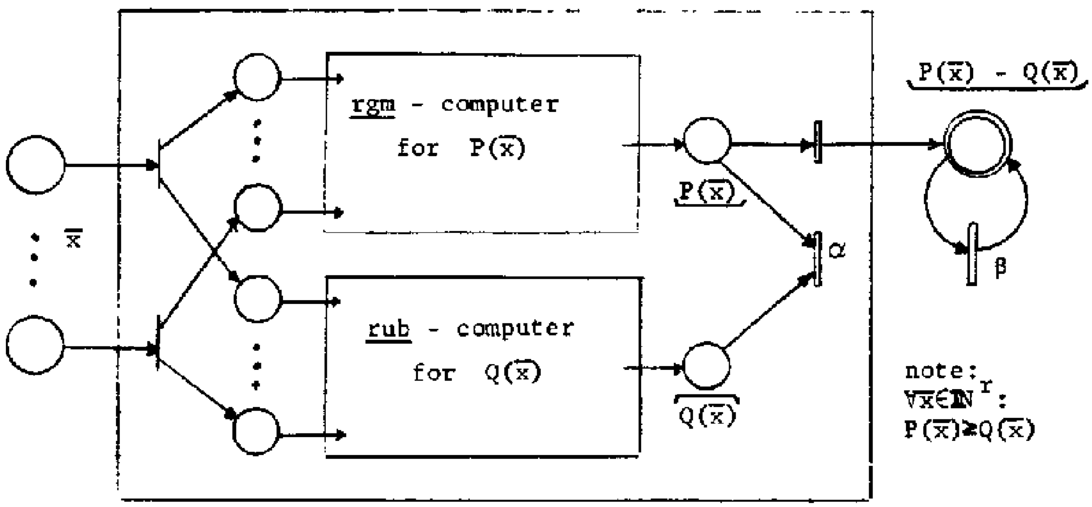


Figure 5.13

It can be seen that transition α can remove at most $Q(\overline{x})$ tokens from the output of the rgm-computer for $P(\overline{x})$, and thus the whole construction above is a rgm-computer for $P(\overline{x}) - Q(\overline{x})$. Therefore, if there is no integer root to the equation $P(\overline{x}) - Q(\overline{x}) = 0$, the output place is guaranteed to eventually get a token, and therefore transition β will be live, whereas, if there exists such a root $\overline{x}$, for that input there will be a firing sequence such that both the rgm-computer for $P(\overline{x})$ and the rub-computer for $Q(\overline{x})$ will actually reach their bound (and the rgm-computer will be unable to exceed it), and since they are equal, repeated firings of transition α will exhaust all tokens in the output place of the P-"computer", thus effectively killing transition β.

Thus, if we connect a "generator" to the input of the rgm-computer for P-Q, (See Fig.5.14), we get a Petri Net where a given transition β is live iff there exists an integer root for $P(\overline{x}) - Q(\overline{x}) = 0$, which is undecidable according to Lemma 4.2 We state this as:

Lemma 5.13 If it is possible to construct a Petri Net which admits in one of its places a reachable guaranteed minimum which is a polynomial function (with non-negative coefficients) of the initial marking in a subset of its places, then the Liveness and Reachability Problems are undecidable.

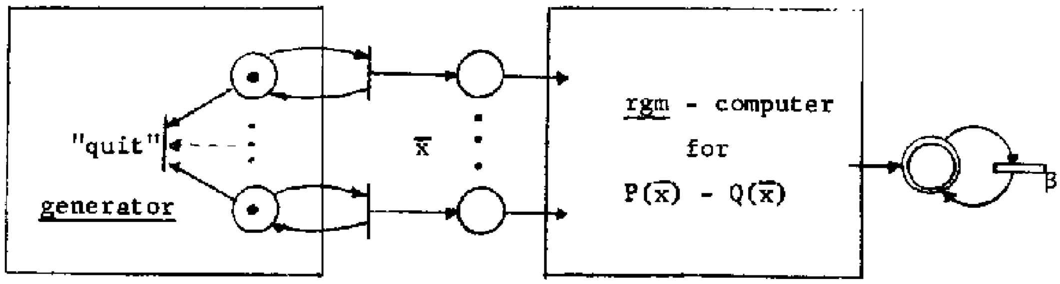


Figure 5.14

We can actually restrict the condition for undecidability as follows:

Theorem 5.14: If there exists a Petri Net with two distinguished places a and b and an initial marking which places x tokens in a and zero tokens in b such that a reachable guaranteed minimum for place b is x^2, then the Liveness and Reachability Problems are undecidable.

Proof: We show that, given an rgm-computer for x^2, we can construct an rgm-computer for any polynomial with non-negative coefficients. As in the proof of Lemma 4.3, we show that we could construct such an rgm-computer out of the operations add, copy and multiply. For each of these operations, if the inputs are reachable lower bounds, the output will be too. We can use the same construction for add and copy. Then we would construct our multiplier as follows:

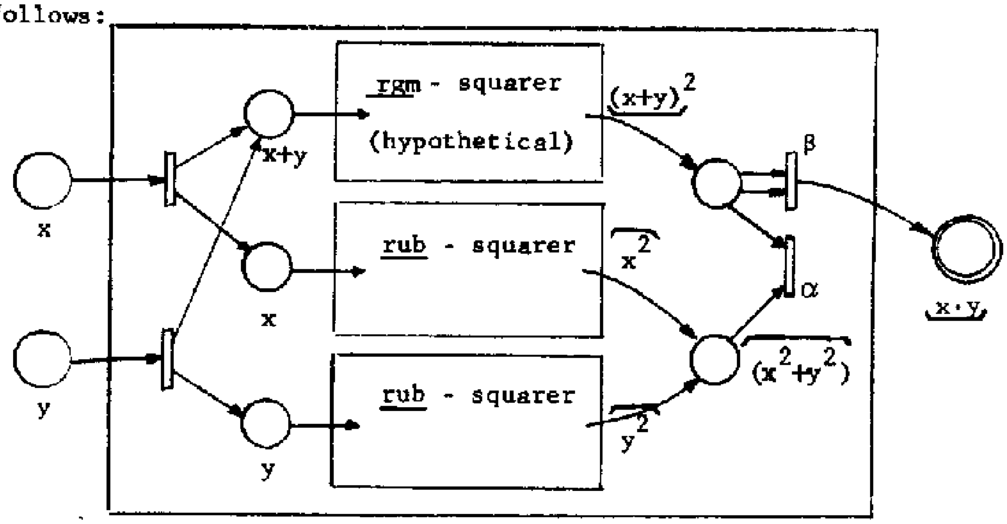


Figure 5.15: An rgm - multiplier

The rgm-square computer generates (x^2+y^2+2xy) as a reachable guaranteed minimum; from this, α subtracts at most (x^2+y^2) tokens, leaving a reachable guaranteed minimum of $2xy$, thus producing the desired rgm of $x.y$ at the output.

QED

5.7 Conclusion: Decidable or Undecidable?

Consider the length of the shortest firing sequence to kill a transition t or to reach the zero marking. This length can also be interpreted as the reachable guaranteed minimum number of tokens in an additional place which collects one token from each transition firing.

In the light of the preceding subsection (5.6), we see that the liveness and reachability problems would be undecidable if the length of such a shortest firing sequence increased like the square of the initial marking for some net.

This suggests that the decidability might follow from the fact that the length of, say, the shortest killing sequence as a function of the initial marking, is a linear function. After some preliminary analysis of this question it is this author's belief that this is the case, and that the constant of proportionality is bounded by a factor of the order of the product, over all transitions, of the number of input arcs of transitions. This is similar to the bound indicated for firing sequences used to cover a given marking, as shown at the end of Section 3, page 40 .

APPENDIX

König's Infinity Lemma: Let Π_1, Π_2, Π_3 ... ad infinitum be a denumerably infinite sequence of mutually distinct finite sets of points. Let these points be the vertices of a graph G. If G has the property that each point of Π_{n+1} is connected to one point of Π_n by an edge of G, then G possesses a forwards infinite path $P_1P_2P_3 \cdots$ ad infinitum where, for each n, P_n is a point of Π_n.

Proof: A finite path in G will be called an S-path if its successive vertices belong to Π_1, Π_2, $\cdots$, Π_k. There are infinitely many S-paths in G since every vertex which is not a point of Π_1 is the second end-point of an S-path. Each S-path begins with an edge which connects a point P_1 of Π_1 to a point X_2 of Π_2. Since there are only finitely many such edges, there must be one such edge, say P_1P_2, which occurs in infinitely many S-paths. All these S-paths now have as a second edge one of the finitely many edges P_2X_3, where X_3 belongs to Π_3, hence there must exist in Π_3 a point P_3 such that infinitely many S-paths that start with P_1P_2 also contain P_2P_3. Continuing in this manner we define a point P_4 of Π_4, P_5 of Π_5, etc. This procedure does not terminate and generates an infinite path $P_1P_2P_3 \cdots$ with the desired property.

QED

(König points out that this proof requires the Axiom of Choice.)

(Translated by the author from pp. 81-82 of [12].)

References

1. Baker, H. G., Jr. Equivalence Problems of Petri Nets. S.M. Thesis, Department of Electrical Engineering, M.I.T., June 1973.

2. Baker, H. G., Jr. Rabin's Proof of the Undecidability of the Reachability Set Inclusion Problem of Vector Addition Systems. CSG Memo 79, Project MAC, M.I.T., July 1973.

3. Commoner, F. Deadlocks in Petri Nets. Report CA-7206-2311, Applied Data Research, Wakefield, Mass., June 1972.

4. Davis, M., H. Putnam, and J. Robinson. The decision problem for exponential diophantine equations. Annals of Math 74, 3 (November 1961), 425-436.

5. Dennis, J. B. Modular, asynchronous control structures for a high performance processor. Record of the Project MAC Conference on Concurrent Systems and Parallel Computation, ACM, New York 1970, 55-80.

6. Hack, Michel H. Analysis of Production Schemata by Petri Nets. Technical Report TR-94, Project MAC, M.I.T., February 1972.

7. Hilbert, D. Mathematische Probleme. Vortag, gehalten auf dem internationalen Mathematiker-Kongress zu Paris 1900. Nachr. K. Ges. Wiss. Göttingen, Math.-Phys. Kl. 1900, pp 253-297. Translation: Bull. Amer. Math. Soc., 8 (1901-1902), 437-479.

8. Holt, A. W. Final Report of the Information System Theory Project. Technical Report RADC-TR-68-305, Rome Air Development Center, Griffiss Air Force Base, New York, 1968.

9. Holt, A. W., and F. Commoner, Events and Conditions. Applied Data Research, New York 1970.

10. Karp, R. M. and R. E. Miller. Parallel program schemata: A mathematical model for parallel computation. IEEE Conference Record, Eighth Annual Symposium on Switching and Automata Theory, October 1967, 55-61.

11. Keller, R. M. Vector Replacement Systems: A Formalism for Modeling Asynchronous Systems. Technical Report 117, Princeton University, Computer Science Laboratory, December 1972.

12. König, D. Theorie der Endlichen und Unendlichen Graphen. Akademische Verlagsgesellschaft, Leipzig 1936.

13. Matijasevič, Ju. V. Enumerable sets are diophantine. Soviet Math. Dokl. 11, 2 (1970), 354-357.

14. Petri, C. A. Communication With Automata. Supplement 1 to Technical Report RADC-TR-65-377, Vol. 1, Griffiss Air Force Base, New York, 1966. Originally published in German: Kommunikation mit Automaten, University of Bonn, 1962.

15. Rabin, M. Private communication, Fall 1972.

16. Ramchandani, C. Analysis of Asynchronous Concurrent Systems by Timed Petri Nets. Technical Report TR-120, Project MAC, M.I.T., February 1974.

17. Slutz, D. R. The Flow Graph Schemata Model of Parallel Computation. Technical Report TR-53, Project MAC, M.I.T., September 1968.