Int J Adv Manuf Technol (2009) 45:397–405
DOI 10.1007/s00170-009-1967-1

ORIGINAL ARTICLE

Direct minimal empty siphon computation using MIP

Daniel Y. Chao

Received: 8 October 2008 / Accepted: 4 February 2009 / Published online: 13 March 2009

Abstract We propose a revised mixed-integer programming (MIP) method to directly compute unmarked siphons with a minimal number of places. This eliminates the need to deduce a minimal siphon from an unmarked maximal siphon obtained from the traditional MIP method proposed by Chu et al. The revised MIP test reports smaller siphons earlier than larger siphons and adds monitors to basic siphons before compound siphons. This results in adding fewer monitors and reaching more states.

Keywords Petri nets · Siphons · Deadlocks · Integer programming

1 Introduction

Deadlock occurs due to mutual waiting for resources among processes in flexible manufacturing systems (FMS). Petri nets (PN) can model and analyze such deadlocks. Siphons are a structural object in PN that, once unmarked, remain so forever [1–6].

It is well-known that the set of unmarked or empty places in a dead net forms a siphon. Output transitions of these empty places are permanently dead since the siphon stays unmarked forever. This implies that a net is deadlock-free if all its siphons are never empty.

To avoid deadlocks, every siphon must never be empty of tokens by adding monitors and control arcs [3]. Unfortunately, the number of siphons grows quickly with the size of nets; so does the number of monitors required. The mixed-integer programming (MIP) method avoids the complete enumeration of siphons.

The MIP method proposed by Chu and Xie [7] detects deadlocks quickly for structurally bounded nets whose deadlocks are tied to unmarked siphons. Requiring no explicit enumeration of siphons, it opens a new avenue for checking deadlock-freeness of large systems. Its computational efficiency is relatively insensitive [7] to the initial marking and more efficient than classical state enumeration methods.

The MIP method is able to find an unmarked maximal siphon in an ordinary Petri net (OPN) (all arcs unit weighted). It has been applied to design liveness-enforcing supervisors such as systems of simple sequential processes with resources (S^3PR) [3] for FMS. The MIP method has been applied to detect deadlocks [8, 9] and remove redundant monitors [10] (more efficiently than the method by Uzam et al. [11] based on reachability analysis). Park and Reveliotis [12] and Huang [13] modified the MIP test for general Petri nets (GPN) such as systems of simple sequential processes with general resources requirement.

This method is an iterative approach consisting of two main stages. At each iteration, find an unmarked maximal siphon S using a fast deadlock detection technique based on MIP. From the maximal siphon, efficiently obtain an unmarked minimal siphon using an algorithm. The first stage, called siphon control, adds, for each unmarked siphon, a control place to the original net with its output arcs to the sink transitions of the minimal siphon. This is to prevent a minimal siphon from being unmarked. The second, called augmented siphon control stage, adds a control place to the modified net with its output arcs to the source transitions of the net. This is required since adding control places in the first stage may generate new unmarked siphons. In addition, the second stage assures no new unmarked siphons.

D. Y. Chao (✉)
National Chengchi University,
Taipei, Taiwan
e-mail: yaw@mis.nccu.edu.tw

An extra step is required to locate an unmarked minimal siphon from S by repetitively removing dangling transitions/places without input and/or output nodes. This is because any minimal siphon and its set of input transitions always generate a strongly connected (SC) subnet.

For a deadlocked net, unmarked places form an empty siphon S. This implies that, if all siphons never get emptied, the net is deadlock-free. We suggest a way to eliminate the need to deduce a minimal siphon from an unmarked maximal siphon S after reviewing the MIP method.

The rest of paper is organized as follows: Section 2 presents the preliminaries to understand the paper. The MIP method and its revision are introduced in Section 3, followed by discussions in Section 4. Finally, Section 5 concludes the paper.

2 Preliminaries

In this paper, we assume that all nets to which we refer, unless otherwise mentioned, are OPN. Here, we present only the definitions used in this paper. A PN (or place/transition net) is a three-tuple $N=(P, T, \boldsymbol{F})$ where $P=\{p_1, p_2, \ldots, p_a\}$ is a set of places; $T=\{t_1, t_2, \ldots, t_b\}$ is a set of transitions, with $P \cup T \neq \emptyset$ and $P \cap T=\emptyset$; and $\boldsymbol{F}$ a mapping from $(P\times T)\cup(T\times P)$ to nonnegative integers indicating the weight of directed arcs between places and transitions. $\boldsymbol{M}_0$: $P\rightarrow\{0, 1, 2, \ldots\}$ denotes an initial marking whose ith component, $\boldsymbol{M}_0(p_i)$, represents the number of tokens in place p_i. $(N, \boldsymbol{M}_0)$ is called a marked net or a net system.

A node x in $N=(P, T, \boldsymbol{F})$ is either a $p\in P$ or a $t\in T$. The postset of node x is $x^\bullet = \{y \in P \cup T | F(x,y) > 0\}$, and its preset ${}^\bullet x = \{y \in P \cup T | F(y,x) > 0\}$.

t_i is fireable if each place p_j in ${}^\bullet t_i$ holds no fewer tokens than the weight $\boldsymbol{w}_j=\boldsymbol{F}(p_j,t_i)$. Firing t_i under $\boldsymbol{M}_0$ removes $\boldsymbol{w}_j$ tokens from p_j and deposits $\boldsymbol{w}_k=\boldsymbol{F}(t_i,p_k)$ tokens into each place p_k in $t_i^\bullet$; moving the system state from $\boldsymbol{M}_0$ to $\boldsymbol{M}_1$. Repeating this process, it reaches $\boldsymbol{M}'$ by firing a sequence of transitions. $\boldsymbol{M}'$ is said to be reachable from $\boldsymbol{M}_0$; i.e., $\boldsymbol{M}_0[\sigma> \boldsymbol{M}'$. t is potentially fireable if there is a firing sequence to reach $\boldsymbol{M}$ from $\boldsymbol{M}_0$ such that t is fireable under $\boldsymbol{M}$. Place p is said to be $\boldsymbol{M}_0$ if there is a firing sequence to reach $\boldsymbol{M}$ from *potentially marked* such that p is marked under $\boldsymbol{M}$.

A marked PN is pure iff $(x,y)\in(P\times T)\cup(T\times P)$, $\boldsymbol{F}(x, y)> 0\Rightarrow\boldsymbol{F}(y, x)=0$. For a pure net, the flow relation can be represented by the flow matrix $\boldsymbol{A}=\boldsymbol{A}^+ + \boldsymbol{A}^-$, where $\boldsymbol{A}^+[p, t]= \boldsymbol{F}(t, p)$ and $\boldsymbol{A}^-[p, t]=\boldsymbol{F}(p, t)$.

OPN are those for which $\boldsymbol{F}$: $(P\times T)\cup(T\times P)\rightarrow\{0,1\}$. An OPN is called a state machine if $\forall t\in T$, $|t^\bullet| = |{}^\bullet t| = 1$. GPN are those for which $\exists j$, $\boldsymbol{w}_j>1$, or $\exists k$, $\boldsymbol{w}_k>1$.

$R(N, \boldsymbol{M}_0)$ is the set of markings reachable from $\boldsymbol{M}_0$. A transition $t\in T$ is live under $\boldsymbol{M}_0$ iff $\forall \boldsymbol{M}\in R(N, \boldsymbol{M}_0)$, $\exists \boldsymbol{M}'\in R(N, \boldsymbol{M}_0)$, t is fireable under $\boldsymbol{M}'$. A transition $t\in T$ is dead under $\boldsymbol{M}_0$ iff $\nexists \boldsymbol{M}\in R(N, \boldsymbol{M}_0)$, where t is fireable. A PN $(N, \boldsymbol{M}_0)$ is live under $\boldsymbol{M}_0$ iff $\forall t\in T$, t is live under $\boldsymbol{M}_0$. It is weakly live under $\boldsymbol{M}_0$ iff N is not live and $\exists t\in T$, t is live under $\boldsymbol{M}_0$. It is bounded if $\forall \boldsymbol{M}\in R(N, \boldsymbol{M}_0)$, $\forall p\in P$, the marking at p, $\boldsymbol{M}(p)\leq k$, where k is a positive integer.

For a PN, a nonempty subset τ of places is called a trap if $\tau^\bullet \subseteq {}^\bullet\tau$. A nonempty subset D of places is called a siphon if ${}^\bullet D \subseteq D^\bullet$. That is, every transition having an output place in D has an input place in D. If $M_0(D) = \sum_{p\in D} M_0(p) = 0$, D is called a token-free or empty siphon at $\boldsymbol{M}_0$. Otherwise, D is said to be filled. A minimal siphon D_m does not contain a siphon as a proper subset. It is called a strict minimal siphon (SMS), denoted by S, if it does not contain a trap.

An S^3PR is defined [3] as the union of a set of nets $N_i = \left(P_i \cup \{p_i^0\} \cup P_{Ri}, T_i, F_i\right)$, sharing common places, where the following statements are true:

1. p_i^0 is called the process idle place of N_i. Places in P_i and P_{Ri} are called operation and resource places, respectively.
2. $P_{Ri}\neq\emptyset$; $P_i\neq\emptyset$; $p_i^0\notin P_i$; $\left(P_i \cup \{p_i^0\}\right) \cap P_{Ri} = \emptyset$; $\forall p\in P_i$, $\forall t'\in p^\bullet$, $\exists r_p\in P_{Ri}$, ${}^\bullet t \cap P_{Ri} = t'^\bullet \cap P_{Ri} = \{r_p\}$; $\forall r\in P_{Ri}$, ${}^{\bullet\bullet} r \cap P_i = r^{\bullet\bullet} \cap P_i \neq \emptyset$; $\forall r \in P_{Ri}$, ${}^\bullet r \cap r^\bullet = \emptyset$. ${}^{\bullet\bullet}(p_i^0) \cap P_{Ri} = (p_i^0)^{\bullet\bullet} \cap P_{Ri} = \emptyset$.
3. N_i' is a SC state machine, where $N_i' = \left(P_i \cup \{p_i^0\}, T_i, F_i\right)$ is the resultant net after the places in P_{Ri} and related arcs are removed from N_i.
4. Every circuit of N contains the place p_i^0.
5. Any two nets N_i and N_j are composable, denoted by N_i o N_j, if they share a set of common resource places. Every shared place must be a resource place.
6. Transitions in ${}^\bullet p_i^0$ and $p_i^{0\bullet}$ are called source and sink transitions, respectively.

An example of S^3PR is shown in Fig. 2. The following siphon properties are useful to understand the theory developed.

Property 1 The set of unmarked places in a dead net forms a siphon.

Property 2 An OPN is deadlock-free if no minimal siphon eventually becomes unmarked.

Property 3 A siphon free of tokens at a marking remains token-free whatever the transition firings. A trap marked by a marking remains marked.

Property 4 Let S be a minimal siphon. Then, the subnet induced by S and ${}^\bullet S$ is SC.

Property 5 Let $\boldsymbol{M}\in R(N, \boldsymbol{M}_0)$ and siphon S be empty of tokens under $\boldsymbol{M}$, and $\boldsymbol{M}(p)>0$, then (1) $p\notin S$. (2) If $\forall p \in {}^\bullet t$,

$\boldsymbol{M}(p)>0$, then t is enabled and not an output transition of any place in S; i.e., $t \notin S^\bullet$.

3 MIP and its revision

This section introduces the MIP basics via Subsection 3.1 on the constraints in an MIP and Subsection 3.2 on objective functions. The tracing method to propagate and obtain variables in an MIP is addressed in Subsection 3.3, and the revision to obtain minimal unmarked siphons is addressed in the final subsection 3.4. Examples at the end of Subsection 3.3 may help the readers understand the theory.

3.1 Constraints

Definition 1 Let S be a siphon. (1) ν_p is a binary variable for p, and $\nu_p=1$ if $p \notin S$. (2) z_t is a binary variable for t, and $z_t=1$ if $t \notin S^\bullet$. Clearly, any p with $\nu_p=1$ and any t with $z_t=1$ will be removed if the classical algorithm is used. Note that $S=\{p|\nu_p=0\}$.

Lemma 1 *Let S be a siphon and $p \in S$. Then, $\forall t \in {}^\bullet p$, (1) $z_t=0$. (2) $\nu_p \geq z_t$.*

Proof (1) $\forall t \in {}^\bullet p$, $t \in S^\bullet$ and $z_t=0$. (2) ν_p is a binary variable; hence, $\nu_p=0$ or 1. Thus, $\nu_p \geq 0 = z_t$. ■

Based on this lemma, we have

Theorem 1 $\forall p \in P$, $\forall t \in {}^\bullet p$, $\nu_p \geq z_t$.

Proof If $p \in S$, then, by Lemma 1.2, $\nu_p \geq z_t$. If $p \notin S$, then also $\nu_p=1 \geq z_t$ (=0 or 1). ■

The inequality $\nu_p \geq z_t$ is a constraint used in the MIP test in [7].

Theorem 2 *Let $\boldsymbol{M} \in R(N, \boldsymbol{M}_0)$ and siphon S be empty of tokens under $\boldsymbol{M}$, and $\boldsymbol{M}(p)>0$, then (1) $\nu_p=1$. (2) $\forall t \in T$, if $\forall p \in {}^\bullet t$, $\boldsymbol{M}(p)>0$, then (a) $z_t=1$. (b) $z_t \geq \sum_{p \in {}^\bullet t} \nu_p - |{}^\bullet t| + 1$.*

Proof (1) $\boldsymbol{M}(p)>0$ implies that $p \notin S$ and $\nu_p=1$. (2.a) t is enabled and $t \notin S^\bullet$; $z_t=1$. (2.b) Since $\forall p \in {}^\bullet t$, $\boldsymbol{M}(p)>0$, we have $\nu_p=1$ by (1) $\Rightarrow \sum_{p \in {}^\bullet t} \nu_p = |{}^\bullet t| \Rightarrow \sum_{p \in {}^\bullet t} \nu_p - |{}^\bullet t| + 1 = 1 = z_t$ (by 2.a) $\Rightarrow z_t \geq \sum_{p \in {}^\bullet t} \nu_p - |{}^\bullet t| + 1$. On the other hand, if $\exists p \in {}^\bullet t$, $\boldsymbol{M}(p)=0$,then t is disabled, $\sum_{p \in {}^\bullet t} \nu_p - |{}^\bullet t| < 0$ and $\sum_{p \in {}^\bullet t} \nu_p - |{}^\bullet t| + 1 < 1 = z_t$ since z_t is a binary variable. Thus, $z_t \geq \sum_{p \in {}^\bullet t} \nu_p - |{}^\bullet t| + 1$, whether t is enabled or not. ■

The above inequality $z_t \geq \sum_{p \in {}^\bullet t} \nu_p - |{}^\bullet t| + 1$ in Theorem 2.2.b is another constraint used in the MIP test in [7]. Theorem 2.1 implies that, if a place is marked, then $\nu_p=1$. Similarly, if a transition is enabled, then $z_t=1$.

Note that, even if t is not enabled, but all of its input places carry $\boldsymbol{\nu}_p=1$, $z_t=1$ by Theorem 2.2.b. We say that transition t is pseudoenabled.

Corollary 1 *$z_t=1$ for any enabled transition.*

Proof When $\forall p \in {}^\bullet t$, $\nu_p=1$, $\sum_{p \in {}^\bullet t} \nu_p = |{}^\bullet t|$. Theorem 2.2.b implies that $z_t=1$ for the enabled transition. ■

Theorem 3 $\forall p \in t^\bullet$, $\boldsymbol{\nu}_p=1$, if $z_t=1$.

Proof If $\nu_p=0$, then $\nu_p<z_t$ against the fact that $\nu_p \geq z_t$ in Theorem 1. ■

This theorem helps to propagate binary variables in a forward fashion. Another constraint comes from the state equation; i.e., $\boldsymbol{M}=\boldsymbol{M}_0+\boldsymbol{AX}$ where $\boldsymbol{M} \in R(N, \boldsymbol{M}_0)$, $\boldsymbol{A}$ is the incidence matrix of net N, and X is a firing vector. The constraints so far obtained are summarized as follows:

$$z_t \geq \sum_{p \in t^\bullet} \boldsymbol{\nu}_p - |{}^\bullet t| + 1 \tag{1}$$

$$\nu_p \geq z_t,\ \forall (t\ p)\ \in \boldsymbol{F} \tag{2}$$

$$\boldsymbol{\nu}_p,\ z_t \in \{0,\ 1\} \tag{3}$$

$$\boldsymbol{\nu}_p = 1,\ \text{if}\ \boldsymbol{M}(p) > 0 \tag{4}$$

$$\boldsymbol{M} = \boldsymbol{M}_0 + \boldsymbol{AX},\ \boldsymbol{M} \geq 0,\ \boldsymbol{X} \geq 0, \tag{5}$$

where $\boldsymbol{A}$ is the incidence matrix of net N.

Note that Eq. (4) is nonlinear; it can be linearized by the following

$$\boldsymbol{\nu}_p \geq \boldsymbol{M}(p)/\mathrm{SB}(p),\ \forall p \in P \tag{4'}$$

for a structurally bounded net, where the structural bound (SB (p)) is defined as $\mathrm{SB}(p) = \max\{\boldsymbol{M}(p)|\boldsymbol{M} = \boldsymbol{M}_0 + \boldsymbol{AX}, \boldsymbol{M} \geq 0,\ \boldsymbol{X} \geq 0\}$.

When $\boldsymbol{M}(p)>0$, $\boldsymbol{\nu}_p = 1 \geq \boldsymbol{M}(p)/\mathrm{SB}(p) > 0$. When $\boldsymbol{M}(p) = 0 = \boldsymbol{M}(p)/\mathrm{SB}(p)$, $\boldsymbol{\nu}_p=0$ or 1 and again $\boldsymbol{\nu}_p \geq \boldsymbol{M}(p)/\mathrm{SB}(p)$.

3.2 Objective functions

There are only two possible objective functions: (1) min-objective function: $N^{\mathrm{MIP}}(\boldsymbol{M}) = \min \sum_{p \in P} \boldsymbol{\nu}_p$, and (2) max-objective function: $N^{\mathrm{MIP}}(\boldsymbol{M}) = \max \sum_{p \in \boldsymbol{P}} \boldsymbol{v_p}$. The following theorems and lemmas help to find the correct objective function to find empty or unmarked siphons.

Lemma 4 *If t is potentially fireable from $\boldsymbol{M}_0$, then $\boldsymbol{v}_p = 1\ \forall p \in t^\bullet$, if $\boldsymbol{z}_t$=1.*

Proof If ν_p=0, then ν_p<$\boldsymbol{z}_t$ against the fact that $\nu_p \geq \boldsymbol{z}_t$ in Theorem 1. ∎

This lemma also follows from Theorem 3. The following theorem confirms that $\boldsymbol{z}_t$=1, and hence, ν_p=1 in the above lemma.

Theorem 4 (1) *If t∈T is potentially fireable under $\boldsymbol{M}_0$, then $\boldsymbol{z}_t$=1.* (2) *If p∈P is potentially marked, then ν_p=1.*

Proof Prove by induction, with relation to $|\sigma|$, the length of firing sequence σ. First, prove for $|\sigma|$=1, σ=*t*. *t* is enabled under $\boldsymbol{M}_0$. By Theorem 2.1, $\forall p \in {}^\bullet t$, $\boldsymbol{M}_0(p)$>0, and ν_p=1. Thus, $\sum_{p \in {}^\bullet t} v_p = |{}^\bullet t| \Rightarrow \sum_{p \in t_\bullet} v_p - |{}^\bullet t| + 1 = 1 = z_t$. After firing *t*, $\forall p' \in t^\bullet$, $\boldsymbol{M}_0$ becomes $\boldsymbol{M}$, $\boldsymbol{M}(p')$>0, and hence, $v'_p = 1$. Assume the theorem holds for any σ' such that $|\sigma'|$<$|\sigma|$=*m*; now, prove it also holds for σ, where transition *t′* is enabled; i.e., potentially fireable under $\boldsymbol{M}_0$. $\forall p \in {}^\bullet t'$, let σ'' be the firing sequence such that *p* is potentially marked. Thus, ν_p=1 by the assumption and $\sum_{p \in {}^\bullet t} v_p = |{}^\bullet t| \Rightarrow \sum_{p \in {}^\bullet t} v_p - |{}^\bullet t| + 1 = 1 = z_t$. This proves (1). By Lemma 4, $\forall p' \in t'^\bullet$, $v'_p = 1$. This proves (2). ∎

Corollary 2 *∀t∈T in a SC net, if t is potentially fireable under $\boldsymbol{M}_0$, then $\sum_{p \in P} v_p = |P|$.*

Proof By Theorem 4, $\boldsymbol{z}_t$=1. $\forall p \in t^\bullet$, $\boldsymbol{v}_p$=1. Since the net is SC, ∀*p*∈*P*, ∃*t*∈*T*, such that $p \in t^\bullet$. Thus, $\boldsymbol{v}_p$=1 for every place *p* in the net and $\sum_{p \in P} v_p = |P|$. ∎

Thus, if all transitions are potentially fireable from the initial marking, then all $\boldsymbol{v}_p$=1 and *S*={*p*|$\boldsymbol{v}_p$=0}=∅. One cannot find any unmarked siphon under $\boldsymbol{M}_0$. This remains so for any reachable marking $\boldsymbol{M}$ as shown by the next theorem.

Theorem 5 *If every t is potentially fireable from an $\boldsymbol{M}$∈R(N, $\boldsymbol{M}_0$), then max $\sum_{p \in P} v_p = |P|$.*

Proof From Corollary 2, $\sum_{p \in P} v_p = |P|$, which is the maximal of $\sum_{p \in P} v_p$ corresponding to an empty siphon since *S*={*p*|$\boldsymbol{v}_p$=0}. ∎

This theorem immediately leads to the following:

Corollary 3 (1) *If there is an $\boldsymbol{M}$∈R(N, $\boldsymbol{M}_0$) such that every t is potentially fireable from $\boldsymbol{M}$, then the MIP method with max-objective function produces an incorrect solution of empty siphon.* (2) *The MIP method is able to produce an empty siphon only if ∀$\boldsymbol{M}$∈R(N, $\boldsymbol{M}_0$), there is a transition that is not potentially fireable.*

Proof (1) By Theorem 5, max $\sum_{p \in P} v_p = |P|$ and *S*={*p*|$\boldsymbol{v}_p$=0}=∅, which may not always hold. (2) Assume contrarily that all transitions are potentially fireable, and then by (1), only a null siphon can be obtained—contradiction. ∎

Therefore, if the objective function $N^{\mathrm{MIP}}(\boldsymbol{M}) = \max \sum_{p \in P} v_p$ is chosen, one would not be able to obtain unmarked siphons if all transitions are potentially fireable under the initial marking (true for a well-designed FMS). On the other hand, the maximal siphon unmarked at a given marking can be determined by the following MIP problem, and there exist siphons unmarked at $\boldsymbol{M}$ iff $N^{\mathrm{MIP}}(\boldsymbol{M})$<|*P*|:

$$N^{\mathrm{MIP}}(M) = \min \sum_{p \in P} v_p$$

under constraints 1–4 and

$$\boldsymbol{M} = \boldsymbol{M}_0 + \boldsymbol{AX},\ \boldsymbol{M} \geq 0,\ \boldsymbol{X} \geq 0,$$

where $\boldsymbol{A}$ is the incidence matrix of net *N*.

Note that Eq. 4 implies that ν_p=1 if $\boldsymbol{M}(p)$>0 and ν_p=1 or 0 if $\boldsymbol{M}(p)$=0. To minimize $\sum_{p \in P} v_p$, ν_p=0 if $\boldsymbol{M}(p)$=0. Also, from (1), $\boldsymbol{z}_t \geq 0$ when $\sum_{p \in {}^\bullet t} v_p - |{}^\bullet t| + 1 \leq 0$; $\boldsymbol{z}_t$=0 or 1. We pick $\boldsymbol{z}_t$=0 to have more ν_p=0 since $\nu_p \geq \boldsymbol{z}_t$=0, ∀(*t p*)∈$\boldsymbol{F}$ by (2). If $\boldsymbol{z}_t$=1, then ν_p=1, and the objective function becomes bigger.

Based on the above results, we can conclude that there is no unmarked siphon in a net system *N*=(*P*, *T*, $\boldsymbol{F}$, $\boldsymbol{M}$) if $N^{\mathrm{MIP}}(\boldsymbol{M})$=min $\sum_{p \in P} v_p = |P|$ is true, which implies *S*={*p*|ν_p=0}=ϕ. To find maximal unmarked siphons, the method (in the next subsection) to trace a net to find all ν_p is helpful for finding $N^{\mathrm{MIP}}(\boldsymbol{M})$.

3.3 Manual tracing

Instead of relying on a computer to compute ν_p and $\boldsymbol{z}_t$, it is desirable to manually trace the net to propagate binary variables given a marking $\boldsymbol{M}$. When the values of ν_p or $\boldsymbol{z}_t$ cannot be precisely determined by the constraints 1–5; i.e., (1) ν_p=0 or 1 or (2) $\boldsymbol{z}_t$=0 or 1, this uncertainty is called binary ambiguity. Careless assignments of ν_p or $\boldsymbol{z}_t$ in case of binary ambiguity may result in conflicting binary values of ν_p or $\boldsymbol{z}_t$. For instance, in Fig. 1a, $\boldsymbol{M}_0(p_3)$=0. If one assigns ν_{p3}=0, it would conflict with the ν_{p3}=1 obtained using Theorem 4 since p_3 is potentially marked from $\boldsymbol{M}_0$.

There are two cases where one can assign ν_p∀*p*∈*P* and $\boldsymbol{z}_t$∀*t*∈*T* easily. First, if all transitions are potentially fireable,

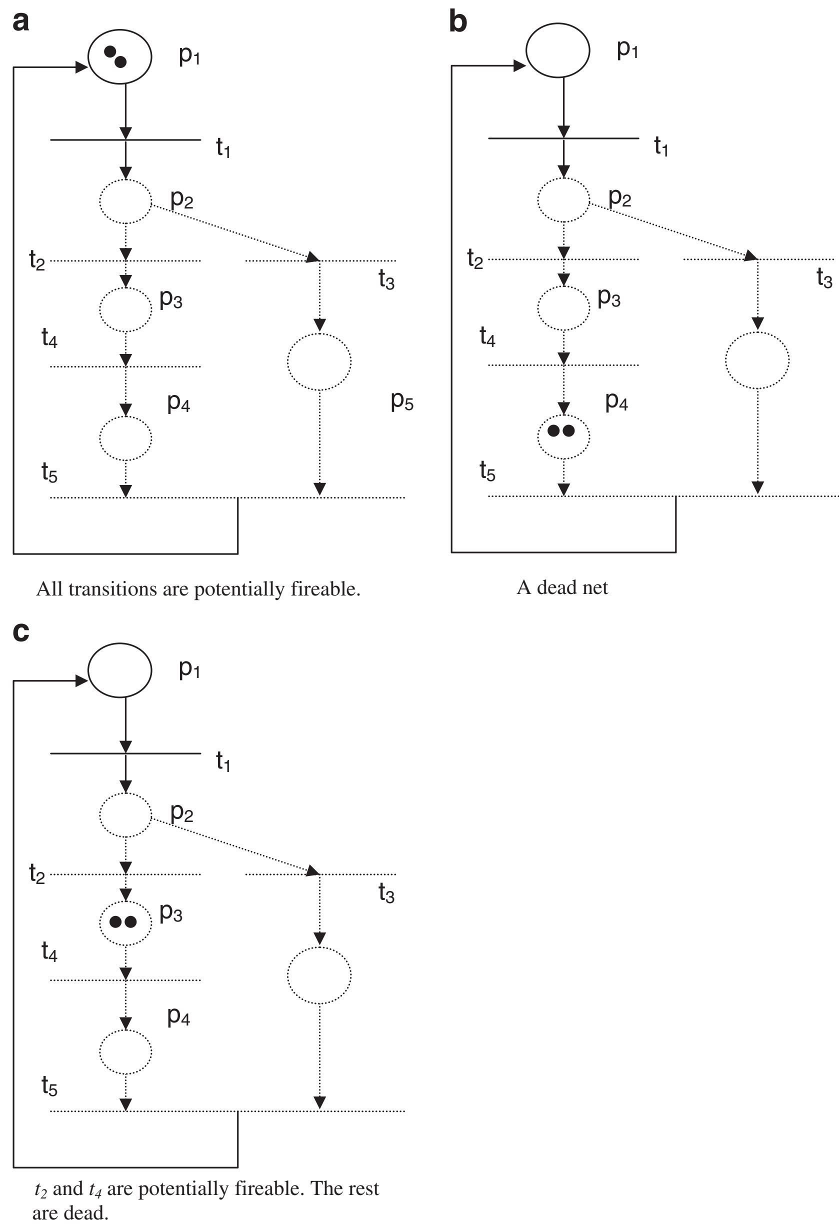

Fig. 1 **a** All transitions are potentially fireable. **b** A dead net. **c** t_2 and t_4 are potentially fireable. The rest are dead

then all $\nu_p = z_t = 1$. Second, if the net is deadlocked, then all $\nu_p = 0$ if $\boldsymbol{M}(p) = 0$ and $\nu_p = 1$ otherwise. There is no need to propagate and update binary variables—called immediate assignment contrasting to that (called propagating assignment) below where some initial guesses are made. In this case, use constraints 1–5 to propagate and update other binary variables.

For other cases, it is unclear how to assign all ν_p and z_t. First, assign (1) $\nu_p = 1$ if $\boldsymbol{M}(p) = 1$ and (2) $z_t = 1$ if t is enabled. Then, propagate assigned binary variables to those unassigned under no binary ambiguity. That is, if a transition is enabled, then assign $z_t = 1$ (Corollary 1). If $\nu_p = 0$, then $\forall t \in {}^{\bullet}p$, $z_t = 0$ (constraint 2). For the unassigned ν_p or z_t, one has to resolve the binary ambiguity to assign z_t for a disabled transition or assign ν_p for an unmarked place p whose input $z_t = 0[\nu_p \geq z_t = 0 \ (t\ p) \in \boldsymbol{F}]$.

Depending on the values assigned, one of the objective functions may be achieved. For instance, if one assigns $z_t = 1$ to a disabled transition, then $\forall p \in t^{\bullet}$, $\nu_p = 1$. If one assigns $z_t = 0$ to a disabled transition, then $\forall p \in t^{\bullet}$, $\nu_p = 0$ or 1 based on constraint 2, which needs to be resolved. If an unresolved ν_p is assigned with 1, the objective function value is increased by 1. Otherwise, if it is assigned with 0, the objective function value is unchanged. As a result, a uniform assignment of 1 to unresolved ν_p may result in an objective function to reach its maximum. On the other hand, a uniform assignment of 0 to unresolved ν_p may result in an objective function to reach its minimum.

The following rules are useful for manual tracing:

Tracing Rules

1. $p \in P$, if $\boldsymbol{M}(p)>0$, then $\nu_p=1$.
2. $t \in T$, if t is enabled, then $z_t=1$.
3. $\forall t \in {}^{\bullet}p$, (a) if $\nu_p=0$, then $z_t=0$, (b) if $\nu_p=1$, then $z_t=0$ or 1.
4. $\forall p \in t^{\bullet}$, (a) if $z_t=1$, $\nu_p=1$, (b) if $z_t=0$, then $\nu_p=0$ or 1.
5. $t \in T$, if t is disabled, then (a) $z_t=0$ if $\exists p \in t^{\bullet}$, $\nu_p=0$. (b) $z_t=$ 0 or 1, if $\forall p \in t^{\bullet}$, $\nu_p=1$.
6. $p \in P$, if $\boldsymbol{M}(p)=0$, then (a) $\nu_p=1$ if $\exists t \in {}^{\bullet}p$, $z_t=1$. (b) $\nu_p=0$ or 1, if $\forall t \in {}^{\bullet}p$, $z_t=0$.

Examples The three nets in Fig. 1a–c have the same structure but different initial markings. All transitions in Fig. 1a are potentially fireable (even though they are not live); hence, all $\nu_p=z_t=1$. $\sum_{p \in P} \nu_p = |P|$. The net in Fig. 1b is dead and all $\nu_p=0$ if $\boldsymbol{M}(p)=0$ (only $\nu_{p4}=1$) and all $z_t=0$. In Fig. 1c, $\sum_{p \in P} \nu_p = 1$. Only t_4 is potentially fireable ($z_{t4}=1$), and the rest of the transitions t are dead and disabled ($z_t=0$ or 1). Resolving the binary ambiguity by setting $z_t=1$ is equivalent to pseudoenabling the disabled transition t and by Lemma 4, $\nu_p=1$ for all their output places $p \in t^{\bullet}$. As a result, $\Sigma_{p \in P}\ \nu_p=|P|$ and the siphon obtained is an empty set. On the other hand, resolving the binary ambiguity by setting $z_t=0$ sets $\nu_p=0$ for all their output places $p \in t^{\bullet}$. As a result, the set of places with $\nu_p=0$ is an empty siphon. Note that the obtained siphon is larger than the minimal $S=\{p_1,\ p_2,\ p_5\}$. This example clearly indicates that, to find unmarked siphons, one should always resolve the binary ambiguity by setting $z_t=0$ for disabled transitions.

3.4 Revised MIP test

Since there are only two objective functions and the min-objective function needs an extra step to extract minimal siphons, the only choice is to maximize (instead of minimize) $\sum_{p \in P} \nu_p$. $\sum_{p \in P} \nu_p$ is maximized if we set $\nu_p=1$ if $\boldsymbol{M}(p)=0$ $\forall p \in P$. However, we will always get $N^{\mathrm{MIP}}(M) = \max \sum_{p \in P} \nu_p = |P|$ with no unmarked siphons. To prevent this, we add one more constraint:

$$\sum_{p \in P} \nu_p < |P| \tag{6}$$

so that $N^{\mathrm{MIP}}(\boldsymbol{M})<|P|$. This way, $\nu_p=1$ iff $p \notin S$. Maximizing $\sum_{p \in P} \nu_p$ minimizes the number of places in S and, thus, leads to a minimal siphon.

To maximize $\sum_{p \in P} \nu_p$, $\nu_p=1$ if $\boldsymbol{M}(p)=0$ and p is not in S. Also, from constraint 1, $z_t \geq 0$ when $\sum_{p \in \bullet t} \nu_p - |{}^{\bullet}t| + 1 \leqq 0$; $z_t=0$ or 1. We pick $z_t=1$ to have more $\nu_p=1$ since $\nu_p \geq z_t$, $\forall (t\ p) \in \boldsymbol{F}$ by constraint 2.

For instance, for the net in Fig. 2, we may get maximal unmarked siphon $S=\{p_4, p_6, p_7, p_8, p_9, p_{10}, p_{11}\}$, which is not minimal. Note that $z_{t7}=0$ implies $\nu_{p7} \geq z_{t7}=0$ by constraint 2. If we pick $\nu_{p7}=1$ to maximize, p_7 would no longer be in S since $\nu_{p7}=1$ $\{p_7 \notin S\}$. Additionally, $S/\{p_7\}$ is a minimal siphon.

If the net never has any unmarked siphon, then $\sum_{p \in P} \nu_p = |P|$ and violates inequality 6 and the MIP test would result in no feasible solutions. Experimental results indicate that such an infeasibility conclusion can be obtained quickly.

The revised MIP test is summarized below.

$$N^{\mathrm{MIP}}(\boldsymbol{M}) = \max \sum_{p \in P} \nu_p$$

$$z_t \geq \sum_{p \in t\bullet} \nu_p - |{}^{\bullet}t| + 1 \tag{1'}$$

$$\nu_p \geq z_t,\ \forall (t\ p) \in \boldsymbol{F} \tag{2'}$$

$$\nu_p,\ z_t \in \{0,\ 1\} \tag{3'}$$

$$\nu_p \geq \boldsymbol{M}(p)/\mathrm{SB}(p),\ \forall p \in P \tag{4'}$$

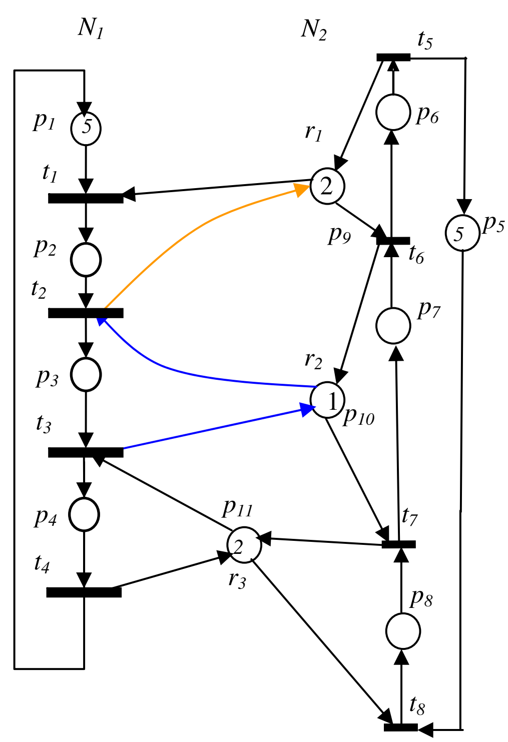

Fig. 2 Example

$$\sum_{p\in P} v_p < |P| \tag{5'}$$

$$\boldsymbol{M} = \boldsymbol{M}_0 + \boldsymbol{A}\boldsymbol{X},\ \boldsymbol{M} \geq 0,\ \boldsymbol{X} \geq 0 \tag{6'}$$

The net is deadlock-free if no solution is declared under the MIP test. Otherwise, a random minimal unmarked siphon S with a minimal number of places in S is reported. The following theorem proves the correctness.

Theorem 6 Any feasible solution from the above revised MIP test is a minimal unmarked siphon.

Proof Note that the feasible solution corresponds to an unmarked siphon S under a certain reachable marking $\boldsymbol{M}$

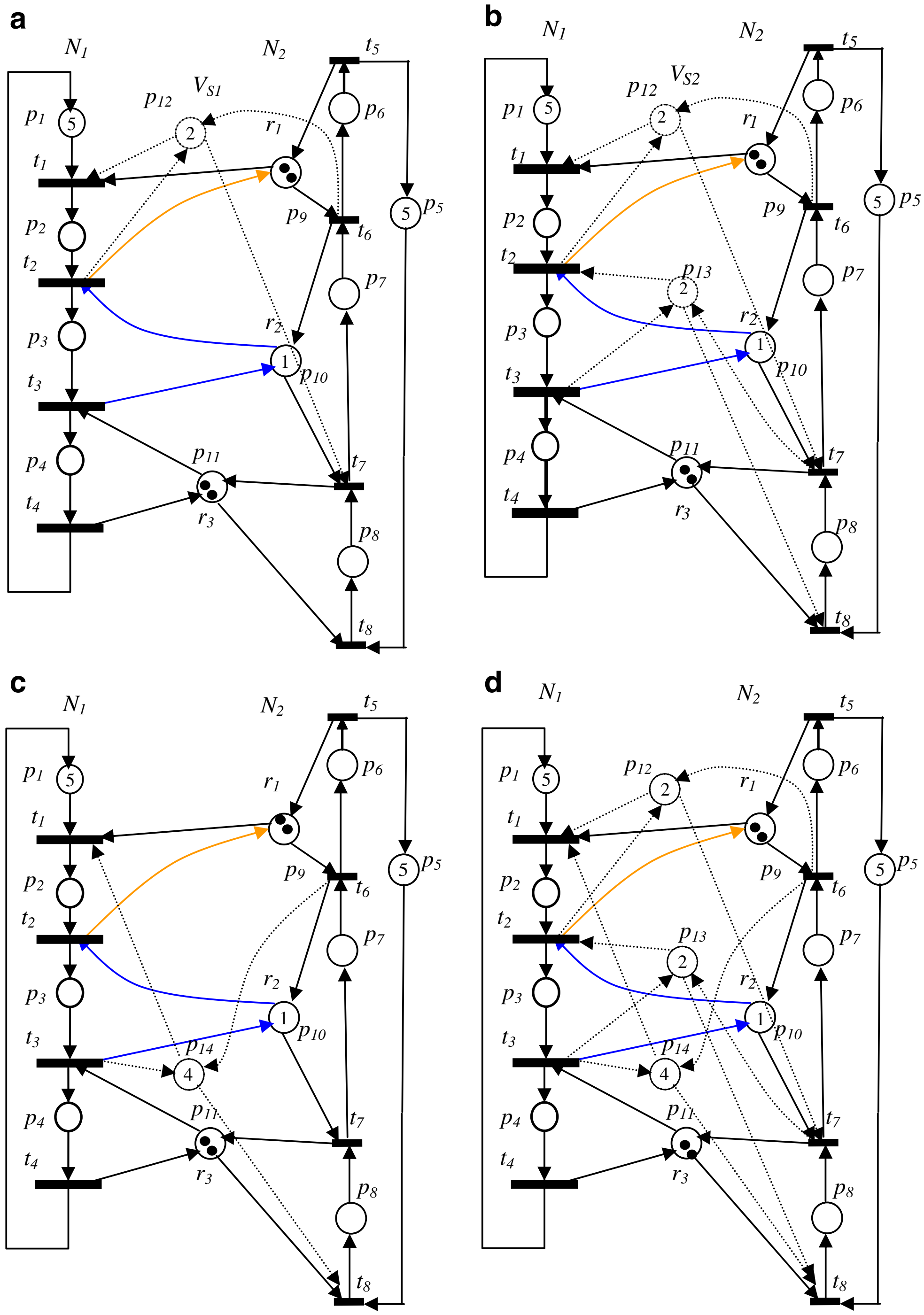

Fig. 3 **a** Adding V_{S1}. **b** Adding V_{S2}. **c** Adding V_{S3}. **d** Final controlled model

since inequalities 1′–4′ and 5′ are the same as those of traditional MIP. Assume that it is not minimal; then S contains a minimal siphon S' as a proper subset. Then there exists a reachable dead marking $\boldsymbol{M}'$ such that $\nu_p=0$ iff $p\in S'$. Obviously, it is a new feasible solution where $N^{\mathrm{MIP}}(\boldsymbol{M}')>N^{\mathrm{MIP}}(\boldsymbol{M})$, violating the fact that $N^{\mathrm{MIP}}(\boldsymbol{M})$ is the maximal among all solutions meeting inequalities 1′–4′ and 5′. ∎

Example For the net in Fig. 2, the revised MIP test reports minimal $S_1=\{p_9, p_{10}, p_3, p_6\}$ unmarked at the first run. We then add monitor V_{S1} and control arcs as shown in Fig. 3a. The second run reports minimal $S_2=\{p_{10}, p_{11}, p_4, p_7\}$ unmarked. There is no need to run a second program to get minimal siphons. We then add monitor V_{S2} and control arcs as shown in Fig. 3b. If we set $\boldsymbol{M}_0(p_1)=\boldsymbol{M}_0(p_{11})=1$, no siphons can become empty, and the MIP test declares no solution since max $\sum_{p\in P} v_p = |P|$.

On the other hand, following the traditional MIP method, one may find $S_3' = \{p_9, p_{10}, p_{11}, p_3, p_4, p_6\}$ as a maximal unmarked siphon, from which we obtain $S_3=\{p_9, p_{10}, p_{11}, p_4, p_6\}$ as an unmarked minimal siphon and add a monitor V_{S3} and some control arcs as shown in Fig. 3c. Next, we obtain $S_1' = \{p_9, p_{10}, p_3, p_4, p_6, p_8\}$ as a maximal unmarked siphon, from which we obtain $S_1=\{p_9, p_{10}, p_3, p_6\}$ as an unmarked minimal siphon and add a monitor V_{S1} and some control arcs. Finally, the MIP test reports $S_2' = \{p_{10}, p_{11}, p_4, p_7, p_3, p_6\}$ as a maximal unmarked siphon and add a monitor V_{S3}, from which we obtain $S_2=\{p_{10}, p_{11}, p_4, p_7\}$ as an unmarked minimal siphon and add a monitor V_{S2} and some control arcs as shown in Fig. 3d. It is easy to see that the traditional method results in one additional monitor V_{S3}, which is redundant and can be removed. In addition, it takes extra steps and time to find unmarked minimal siphons S_1–S_3 from maximal unmarked siphons S_1'–S_3'.

This can be explained below and is true in general. Let $S_p^1 = S_1 \cap P = \{p_3, p_6\}$, $S_p^2 = S_2 \cap P = \{p_4, p_7\}$, and $S_p^3 = S_3 \cap P = \{p_4, p_6\}$. $S_p^1 \cap S_p^3 = \{p_6\}$, $S_p^2 \cap S_p^3 = \{p_4\}$. To be marked, min $\boldsymbol{M}(S_1)$=min $\boldsymbol{M}(S_2)$=min $\boldsymbol{M}(S_3)$=1. Let $\boldsymbol{M}(S_1)=\boldsymbol{M}(p_3)=\boldsymbol{M}(S_2)=\boldsymbol{M}(p_7)=1$, then $\boldsymbol{M}(S_3)=0$. Note $\boldsymbol{M}(p_3)+\boldsymbol{M}(p_7)+\boldsymbol{M}(p_{10}) = M_0(p_{10})$. Thus, if $\boldsymbol{M}_0(p_{10})=1$, then it is not true that $\boldsymbol{M}(p_3)=\boldsymbol{M}(p_7)=1$. If $\boldsymbol{M}(S_1)=\boldsymbol{M}(p_3)=1$ and $\boldsymbol{M}(p_7)=0$, then $\boldsymbol{M}(p_4)=\boldsymbol{M}(S_3)=1$ in order for $\boldsymbol{M}(S_2)=1$. If $\boldsymbol{M}(S_2)=\boldsymbol{M}(p_7)=1$ and $\boldsymbol{M}(p_3)=0$, then $\boldsymbol{M}(p_6)=\boldsymbol{M}(S_3)=1$ in order for $\boldsymbol{M}(S_1)=1$. In both cases, $\boldsymbol{M}(S_3)=1$. Thus, if $\boldsymbol{M}_0(p_{10})=1$, then S_3 is controlled as long as both S_1 and S_2 are controlled. As a result, only two monitors are required using the revised MIP technique.

On the other hand, if we add monitor V_{S3} first, then $\boldsymbol{M}(S_3) = 1 = \boldsymbol{M}(p_4) + \boldsymbol{M}(p_6)$. If $\boldsymbol{M}(S_3)=\boldsymbol{M}(p_6)=1$ and $\boldsymbol{M}(p_4)=0$, then it is possible that $\boldsymbol{M}(p_7)=1$, $\boldsymbol{M}(S_1)=\boldsymbol{M}(p_3)=0$. Thus, we need to add a monitor (V_{S1}) for S_1. Similarly, if $\boldsymbol{M}(S_3)=\boldsymbol{M}(p_6)=1$ and $\boldsymbol{M}(p_4)=0$, then it is possible that $\boldsymbol{M}(p_3)=1$, $\boldsymbol{M}(S_2)=\boldsymbol{M}(p_7)=0$. Thus, we need to add a monitor (V_{S2}) for S_2. As a result, there are three monitors using the traditional MIP technique.

4 Discussion

The proposed approach helps to reduce the number of MIP iterations [14] and to reach more states, as explained below. To maximize the number of good states, the original uncontrolled model should be disturbed as little as possible and each SMS S should be allowed to reach its limit state; i.e., $\boldsymbol{M}(S)$ reaches its minimal value while $\boldsymbol{M}(S)>0$. A monitor V_S and control arcs are added to achieve this minimal value.

There are two kinds of new SMS. If there are no resource places, the new SMS is called a control siphon since all places that are shared between processes are monitor or control places. Otherwise, it is called a mixture siphon.

If one carefully selects a sequence of unmarked siphons to add monitors, the number of monitors required can be reduced. This also reduces the number of MIP iterations required to add all monitors to complete the control.

For instance, in an earlier paper [15], we propose to synthesize elementary (dependent) siphons for an S^3PR from resource (compound) circuits. They are also called basic (compound) siphons. Several basic siphons make up a compound siphon. We show that, if all such basic siphons are controlled under a certain condition, so is some compound siphon, which then needs no monitor. The converse is not true; even though a compound siphon is controlled, all basic siphons remain uncontrolled and monitors are needed for each of them.

The size of basic siphons is obviously smaller than that of compound siphons. Hence, the controlled system gets less disturbed and reaches more states for basic siphons than for compound siphons. Thus, the revised MIP test reports smaller siphons earlier than larger siphons and adds monitors to basic siphons before compound siphons. This results in adding few monitors and reaching more states. This work (1) relieves the problem of siphon enumeration, which grows exponentially, and (2) reduces the number of subsequent time-consuming MIP iterations.

5 Conclusions

The classical MIP efficiently finds unmarked maximal siphons. An extra step is required to obtain minimal unmarked siphons, upon which monitors are added to prevent the siphon from becoming unmarked.

We propose a direct MIP method to eliminate the extra step to compute minimal unmarked siphons. The revised MIP method does not assume the net is an S^3PR, and hence, the method works for arbitrary OPN, which is more general than S^3PR.

References

1. Abdallah IB, ElMaraghy HA (1998) Deadlock prevention and avoidance in FMS: a Petri net based approach. Int J Adv Manuf Technol 14(10):704–715. doi:10.1007/BF01438223
2. Huang YS (2007) Design of deadlock prevention supervisors for FMS using Petri nets. Int J Adv Manuf Technol 35(3–4):349–362. doi:10.1007/s00170-006-0708-y
3. Ezpeleta J, Colom JM, Martinez J (1995) A Petri net based deadlock prevention policy for flexible manufacturing systems. IEEE Trans Robot Autom 11:173–184. doi:10.1109/70.370500
4. Uzam M, Wonham WM (2006) A hybrid approach to supervisory control of discrete event systems coupling RW supervisors to Petri nets. Int J Adv Manuf Technol 28(7–8):747–760
5. Li ZW, Wei N (2007) Deadlock control of flexible manufacturing systems via invariant-controlled elementary siphons of Petri nets. Int J Adv Manuf Technol 33(1–2):24–35
6. Uzam M (2002) An optimal deadlock prevention policy for flexible manufacturing systems using Petri net models with resources and the theory of regions. Int J Adv Manuf Technol 19(3):192–208
7. Chu F, Xie XL (1997) Deadlock analysis of Petri nets using siphons and mathematical programming. IEEE Trans Robot Autom 13:793–804
8. Li ZW, Zhou MC (2006) Two-stage method for synthesizing liveness-enforcing supervisors for flexible manufacturing systems using Petri nets. IEEE Trans Indust Infor 2(4):313–325
9. Li ZW, Hu HS, Wang AR (2007) Design of liveness-enforcing supervisors for flexible manufacturing systems using Petri nets. IEEE Trans Sys Man Cyb C 37(4):517–526
10. Li ZW, Hu HS (2008) On systematic methods to remove redundant monitors from liveness-enforcing net supervisors. Comput Ind Eng. doi:10.1016/j.cie.2008.03.013
11. Uzam M, Li ZW, Zhou MC (2007) Identification and elimination of redundant control places in Petri net based liveness enforcing supervisors of FMS. Int J Adv Manuf Technol 35(1–2):150–168
12. Park J, Reveliotis SA (2001) Deadlock avoidance in sequential resource allocation systems with multiple resource acquisitions and flexible routings. IEEE Trans Automat Contr 46(10):1572–1583
13. Huang YS (2007) Deadlock prevention for sequence resource allocation systems. J Inf Sci Eng 23:215–231
14. Chao DY (2008) MIP iteration-reductions for deadlock prevention of flexible manufacturing systems. Int J Adv Manuf Technol. doi:10.1007/s00170-008-1473-x
15. Chao DY (2007) An incremental approach to extract minimal bad siphons. J Inf Sci Eng 23(1):203–214