Deadlock Prevention Based on Structure Reuse of Petri Net Supervisors for Flexible Manufacturing Systems

ZhiWu Li
School of Electro-Mechanical Engineering
Xidian University
No.2 South Taibai Road
Xi’an 710071, China
e-mail: systemscontrol@gmail.com

Hans-Michael Hanisch
Institute of Computer Science
Martin Luther Universität of Halle–Wittenburg
Kurt-Mothes Straße 1
06120 Halle, Germany
e-mail: hans-michael.hanisch@informatik.uni-halle.de

MengChu Zhou
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
Newark, NJ 07102, USA
e-mail: zhou@njit.edu

Contact Information:
Office Location: A1111, New Sci-Tech Building, North-Campus, Xidian University
Access ID: SCATR201008001
Tel: 86-29-88-20-24-41
Fax: 86-29-88-20-43-38
Website: http://sca.xidian.edu.cn/

2010.08.13
Deadlock Prevention Based on Structure Reuse of Petri Net Supervisors for Flexible Manufacturing Systems

Zhiwu Li, Hans-Michael Hanisch, and MengChu Zhou

Abstract

Deadlocks are an undesirable situation in automated flexible manufacturing systems. Their occurrences often deteriorate the utilization of resources and may lead to catastrophic results. The solution of a maximally permissive, i.e., optimal, supervisor in general belongs to the class of NP-hard problems. A computationally efficient policy, however, often implies a behaviorally restricted supervisor. This paper develops a deadlock prevention policy that establishes a trade-off between the behavioral permissiveness and computational tractability for a class of Petri nets, which can model many flexible manufacturing systems. The theory of regions guides our efforts towards the development of near-optimal solutions for deadlock prevention. Given a plant net structure, a minimal initial marking is first decided by structural analysis, and an optimal live controlled system is computed. Then, a set of inequality constraints is derived with respect to the markings of monitors and the places in the plant model such that no siphon can be insufficiently marked. A method is proposed to identify the redundancy condition for a constraint. For a new initial marking of the plant net structure, a deadlock prevention controlled system can be obtained by regulating the markings of the monitors such that the inequality constraints are satisfied, without changing the structure of the controlled system obtained previously. The behavioral performance of a controlled net system via the proposed method is shown to be nearly optimal by experimental studies.

Index Terms: Petri net, deadlock prevention, flexible manufacturing system, the theory of regions, siphon

1 Introduction

The survivability of traditional mass production is challenged by the quick changes of market requirements, leading to the emergence of flexible manufacturing systems (FMS) that can produce multiple product types with a small batch. An FMS is a conglomeration of computer numerically controlled machine tools, robots, AGV, buffers and fixtures, served by a material-handling system. It usually exhibits a high degree of resource sharing in order to increase flexibility such that it can quickly respond to market changes. Resource sharing may lead to circular wait conditions, the cause of deadlocks in which each of a set of two or more jobs keeps waiting indefinitely for the other jobs in the set to relinquish resources that they hold. Deadlocks and related blocking phenomena can give rise to unnecessary productivity loss, and even catastrophic results in some highly automated systems such as semiconductor manufacturing and safety-critical distributed databases.

It is thereof necessary to explore an effective and computationally efficient mechanism to properly allocate resources such that deadlocks can never occur. With the wide application of FMS, their deadlock control problem was extensively studied over the last two decades, leading to significant results in theory and successful industrial applications [8], [9], [14], [17], [19], [28], [26], [29], [33], [46].

Digraphs, automata, and Petri nets are three major mathematical tools to investigate deadlock problems in FMS [12], [16]. Due to their inherent characteristics, Petri nets as well as their extended versions are increasingly becoming a fully fledged formalism and have received particular attention from academic and industrial communities [24], [25], [39], [40], [41], [42], [45], [49].
The major strategies using Petri net techniques to cope with deadlocks in FMS are deadlock detection and recovery [12], deadlock avoidance [15], [1], [3], and deadlock prevention [4], [6], [9], [20], [46], [26], [28]. Deadlock detection and recovery is an optimistic strategy that grants a resource to a request as long as the resource is available. A deadlock detection algorithm is used to detect the occurrence of deadlocks. Once a deadlock is detected, a recovery mechanism is initialized by aborting one or more processes involved in the deadlock and the resources held by the aborted processes are relinquished. A deadlock detection and recovery strategy is often used in the case where deadlocks are infrequent and resulting consequence is not serious or does not cause much damage.

In a deadlock avoidance strategy, a resource is granted to a process only if the resulting state is not a deadlock. In order to decide whether the forthcoming state is safe if a resource is allocated to a process, every cell controller and global controller need to keep track of the global system state. Some aggressive deadlock avoidance policies do not eliminate all deadlock states [52], [53], [54], [55].

Deadlock prevention is a well-defined problem in FMS, which is usually achieved either by designing an effective system [58] or by using an off-line mechanism to control the requests for resources to ensure that deadlocks never occur. In the Petri net framework, control places (monitors) as well as related arcs can be used to implement this mechanism [1], [9], [26], [27], [28], [51]. Recent effective and computationally efficient deadlock prevention policies are proposed by Piroddi et al. in [36], [37], and [38].

The deadlock prevention policies underlying Petri net formalisms in the the literature are developed on the basis of either state space or structural analysis, e.g., siphon control. Falling into the first category, the theory of regions [2] that can derive Petri nets from automaton-based models is an important method for supervisory control of discrete event systems. Its appearance is followed by a large amount of research aiming to develop optimal deadlock prevention policies [13], [46], [47], [48]. Ghaffari et al. [13] explore the condition on the existence of a monitor-based liveness-enforcing Petri net supervisor that is optimal, i.e., maximally permissive, and develop a methodology to synthesize such a supervisor. The most attractive advantage of the approach is that an optimal Petri net supervisor can always be obtained, by adding monitors that are used to separate events from unsafe states, when such a supervisor exists. However, it bears much computational cost. In such an approach, one first needs to generate the reachability graph given a plant Petri net model. Then, the set of marking/transition separation instances is calculated, whose number is in theory exponential with respect to the size of a net and its initial marking. Finally, for each instance, a monitor is computed by solving a linear programming problem in which the number of constraints is approximately equal to that of nodes in the reachability graph. For a system under any change in its initial marking, the aforementioned steps have to be repeated. In such an approach, no information of previously computed supervisors is reused.

Deadlock prevention based on siphon control is a typical application of structural analysis techniques of Petri nets. The approaches derived from siphon control are in general not optimal, or even overly restrictive and conservative, from the behavioral permissiveness standpoint. However, they can usually provide a computationally tractable supervisor reconfiguration for a system with a new configuration of resource capacity and processing instances, whose supervisor structure has been computed by a siphon-based approach [4], [9], [10], [17], [18], [19], [26], [27], [28], [44], [50], [57].

To inherit and reserve the advantages of two classes of approaches, this work proposes a novel design method of deadlock prevention supervisors based on Petri nets, which does not guarantee maximal permissiveness but empirical results show its superiority over other suboptimal approaches derived from siphon control. Given the plant Petri net model of
an FMS, one first designs an optimal liveness-enforcing controlled system for the model at a minimal initial marking by utilizing the theory of regions. Then, we calculate all strict minimal siphons in the controlled system. Such a siphon does not contain a trap. For each strict minimal siphon, an algebraic inequality with respect to the markings of monitors and resource places in the controlled system, also called a liveness constraint, is established in terms of the concept of max-controlled or invariant-controlled siphons. The satisfaction of such a constraint implies the absence of dead transitions in the postset of the siphon. Consequently, given initial markings that satisfy all the liveness inequality constraints, all siphons can be max-controlled, and the resulting controlled system is live.

After a controlled system structure is found, one can reallocate the initial markings according to the inequality constraints. No matter how large the initial markings and the number of states are, the liveness constraints remain unchanged. Their satisfaction ensures the absence of uncontrolled siphons. This implies that, for a plant model with a fixed net structure, we only need to compute its reachability graph and the siphons of the controlled system once. Whenever the number of process instances and the capacity of manufacturing resources change, a Petri net supervisor can be determined easily via these algebraic inequality constraints.

The paper proceeds as follows. Section 2 provides the necessary background on Petri nets and the concept of multisets. Section 3 formulates the considered problem through a motivation example and generalizes a class of manufacturing-oriented Petri nets. Section 4 elaborates a method that properly allocates initial markings for monitors to prevent siphons from being uncontrolled. Section 5 proposes an algorithm to identify the redundant constraints. A deadlock prevention policy is developed in Section 6. A real-world example is given in Section 7 to demonstrate the proposed method. Experimental studies are conducted in Section 8, showing the near optimality of the proposed deadlock prevention policy. A problem of the proposed method is discussed in Section 9. Section 10 concludes this work.

2 Preliminaries

2.1 Basics of Petri Nets

A Petri net [34] is a four-tuple $N = (P,T,F,W)$ where $P$ and $T$ are finite and nonempty sets. $P$ is the set of places and $T$ is the set of transitions with $P \cup T \neq \emptyset$ and $P \cap T = \emptyset$. $F \subseteq (P \times T) \cup (T \times P)$ is called flow relation of the net, represented by arcs with arrows from places to transitions or from transitions to places. $W : (P \times T) \cup (T \times P) \rightarrow \mathbb{N}$ is a mapping that assigns a weight to an arc: $W(f) > 0$ if $f \in F$ and $W(f) = 0$ otherwise, where $\mathbb{N} = \{0,1,2,\ldots\}$. $N = (P,T,F,W)$ is called an ordinary net, denoted as $N = (P,T,F)$, if $\forall f \in F, W(f) = 1$. A marking $M$ of $N$ is a mapping from $P$ to $\mathbb{N}$. $(N,M_0)$ is called a marked net. We usually use $\sum_{p \in P} M(p)p$ to denote vector $M$. For example, $M = (2,0,1,0)^T$ in a net with four places $p_1 \cdots p_4$ can be re-written as $M = 2p_1 + p_3$.

The preset of $x \in P \cup T$ is defined as $\bullet x = \{y \in P \cup T | (y,x) \in F \}$. While the postset of $x$ is defined as $\circ x = \{y \in P \cup T | (x,y) \in F \}$. This notation can be extended to a set of nodes as follows: given $X \subseteq P \cup T$, $\bullet X = \cup_{x \in X} \bullet x$, and $\circ X = \cup_{x \in X} \circ x^\bullet$.

A transition $t \in T$ is enabled at a marking $M$ iff $\forall p \in \bullet t, M(p) \geq W(p,t)$; this fact is denoted as $M[t]$; when fired, this gives a new marking $M'$ such that $\forall p \in P$, $M'(p) = M(p) - W(p,t) + W(t,p)$; it is denoted as $M[t]M'$. Marking $M'$ is said to be reachable from $M$ if there exist a sequence of transitions $\sigma = t_0t_1 \cdots t_n$ and markings $M_1,M_2,\ldots,$ and $M_n$ such that $M(t_0)M_1(t_1)M_2 \cdots M_n(t_n)M'$ holds. The set of markings...
Let \( \Omega \) be an integer \( N \) denoted as \( R(N, M) \). \([N]\) is called the incidence matrix of \( N \).

It is a \( |P| \times |T| \) integer matrix with \([N](p,t) = W(t,p) - W(p,t) \). For a place \( p \) in net \( N \), its incidence vector is denoted as \([N](p,:)\).

A transition \( t \in T \) is live at \( M_0 \) iff \( \forall M \in R(N, M_0) \), \( \exists M' \in R(N, M) \), \( M'[t] \). \( N \) is dead at \( M_0 \) iff \( \not\exists t \in T, M_0[t] \). \( (N, M_0) \) is live if \( \forall t \in T, t \) is live at \( M_0 \). \( (N, M_0) \) is live if \( \exists t \in T, M[t] \).


A P-vector is a column vector \( I : P \rightarrow \mathbb{Z} \) indexed by \( P \) and a T-vector is a column vector \( J : T \rightarrow \mathbb{Z} \) indexed by \( T \), where \( \mathbb{Z} \) is the set of integers. A \((P|T)\)-vector \( I(J) \) is denoted by \( \sum_{p \in P} I(p)p \ (\sum_{t \in T} J(t)t) \) for economy of space. P-vector \( I \) is a P-invariant (place invariant) iff \( I \neq 0 \) and \( I^T[N] = 0^T \). P-invariant \( I \) is said to be a P-semiflow if no element of \( I \) is negative. \( ||I|| = \{p \in P|I(p) \neq 0\} \) is called the support of \( I \). \( ||I||^+ = \{p|I(p) > 0\} \) denotes the positive support of P-invariant \( I \), while \( ||I||^- = \{p|I(p) < 0\} \) denotes the negative support of \( I \). An invariant is called minimal when its support is not a strict superset of the support of any other, and the greatest common divisor of its elements is one. If \( I \) is a P-invariant of \( (N, M_0) \) then \( \forall M \in R(N, M_0) \), \( I^TM = I^TM_0 \).

Let \( S \) be a non-empty subset of \( P \) and \( S \subseteq P \) is a siphon (trap) iff \( \bullet S \subseteq S^* \) \((S^* \subseteq \bullet S)\).

A marked trap can never be emptied. A siphon is said to be minimal iff it contains no other siphons as its proper subsets. A minimal siphon is strict if it contains no marked trap. Siphon \( S \) is called uncontrolled in \( (N, M_0) \) iff \( \exists M \in R(N, M_0), \forall t \in S^*, t \) is dead at \( M \). It is said to be invariant-controlled by P-invariant \( I \) if \( I^T M_0 > 0 \) and \( ||I||^+ \subseteq S \). A siphon \( S \) in \( (N, M_0) \) is said to be uncontrolled if \( \exists t \in S^*, \forall M \in R(N, M_0), t \) is dead at \( M \).

A siphon \( S \) is said to be max-marked at a marking \( M \in R(N, M_0) \) iff \( \exists p \in S \) such that \( M(p) > max_{p^*} \bullet S \), where \( max_{p^*} = max \{W(p,t)|t \in p^*\} \). \( S \) is max-controlled iff it is max-marked at any reachable marking. \( (N, M_0) \) satisfies the maximal cs-property (maximal controlled-siphon property) iff each minimal siphon of \( N \) is max-controlled. A siphon satisfying the max-controlled property can be always marked sufficiently to allow firing a transition once at least. The work in [5] shows that a Petri net is deadlock-free if it satisfies the cs-property. The concept of invariant-controlled siphons is a special case of max-controlled siphons.

**Proposition 1** [5] Let \( (N, M_0) \) be a Petri net and \( S \) be a siphon of \( N \). If there exists a P-invariant \( I \) such that \( \forall p \in (||I||^- \cap S) \), \( max_{p^*} = 1 \), \( ||I||^+ \subseteq S \), and \( I^TM_0 > \sum_{p \in S} I(p)(max_{p^*} - 1) \), then \( S \) is max-controlled.

As for the concepts of plants, supervisors, and controlled systems, the reader is referred to [31]. A supervisor is said to be maximally permissive, i.e., optimal, if its resulting controlled system comprise all safe states of a plant and every reachable state in the controlled system is a safe state of the plant. Such a controlled system is said to be optimal.

### 2.2 Multisets

**Definition 1** A multiset \( \Omega \) over a non-empty set \( A \), is a mapping \( \Omega : A \rightarrow \mathbb{N} \), which we represent as a formal sum \( \sum_{a \in A} \Omega(a).a \).

A multiset is a generalization of the concept of a set. In multiset \( \Omega \), non-negative integer \( \Omega(a) \) is the coefficient of element \( a \in A \), indicating the number of occurrences of \( a \) in \( \Omega \). It is said that \( a \in A \) belongs to \( \Omega \), denoted by \( a \in \Omega \), if \( \Omega(a) > 0 \). It does not
belong to Ω, denoted by a /∈ Ω, if Ω(a) = 0. Let Ω₁ and Ω₂ be two multisets. The basic operations on multisets are union, intersection, addition, difference and comparison, which are defined, respectively, as follows:

\[ Ω₁ ∪ Ω₂ := \sum_{a ∈ A} max(Ω₁(a), Ω₂(a)).a, \]
\[ Ω₁ ∩ Ω₂ := \sum_{a ∈ A} min(Ω₁(a), Ω₂(a)).a, \]
\[ Ω₁ + Ω₂ := \sum_{a ∈ A} (Ω₁(a) + Ω₂(a)).a, \]
\[ Ω₁ − Ω₂ := \sum_{a ∈ A} (Ω₁(a) − (Ω₁ ∩ Ω₂)).a, \]
\[ Ω₁ ≤ Ω₂ ⇔ ∀a ∈ Ω₁, Ω₁(a) ≤ Ω₂(a), \]
\[ Ω₁ < Ω₂ ⇔ ∀a ∈ Ω₁, Ω₁(a) < Ω₂(a). \]

Let Ω₁ = a + b, Ω₂ = 2a + b + c, and Ω₃ = 3a + 2c be three multisets over A = {a, b, c}. We have Ω₁ ∪ Ω₂ = 2a + b + c, Ω₁ ∩ Ω₂ = 3a + b + 2c, Ω₁ ∩ Ω₃ = a, Ω₂ ∩ Ω₃ = 2a + c, Ω₁ + Ω₂ = 3a + 2b + c, Ω₁ − Ω₂ = 0, Ω₃ − Ω₁ = 2a + 2c, Ω₃ − Ω₂ = a + c, and Ω₁ < Ω₂.

Later, we can say that multiset Ω₁ is less than Ω₂ if Ω₁ < Ω₂ and Ω₂ is greater than or equals to Ω₁ if Ω₁ ≤ Ω₂. A multiset without any element is denoted by ∅ as an empty set. A multiset becomes a set if the multiplicity of every element is one.

3 Structure Design of a Petri Net Supervisor

3.1 Motivation and Problem Formulation

Let us recall the steps to use the theory of regions to design a supervisor for a Petri net model. First, generate the reachability graph of the model. Then, find all marking/transition separation instances. For each instance, a monitor is computed by solving a linear programming problem (LPP). In theory, the number of marking/transition separation instances grows exponentially with the net size and initial marking. So is the number of constraints in each LPP. Moreover, the size of a reachability graph is rather sensitive to the size and initial marking of a net. These facts make it infeasible for the theory of regions to be applied to real-world problems.

We formulate the problem and illustrate the proposed method through a small example from [58]. Consider a net (N, M₀) in Figure 1(a) with its reachability graph shown in Figure 1(b). Its optimal controlled system (Nᵥ, M₀ᵥ) can be found by the theory of regions [46], [13], as shown in Figure 1(f). Now we consider the deadlock control in (Nᵐ, M₀ᵐ) as shown in Figure 1(c). It has the same topology structure as (N, M₀) in Figure 1(a), but has a small initial marking. Its reachability graph is shown in Figure 1(d). Figure 1(e) shows its controlled system (Nⁿmc, M₀ⁿmc) obtained by using the theory of regions. The computation of (Nⁿmc, M₀ⁿmc) is obviously more tractable than that of (Nᵥ, M₀ᵥ) since (Nᵐ, M₀ᵐ) has a small reachability space.

Now we investigate the relationship between controllability of siphons in (Nⁿmc, M₀ⁿmc) and its initial marking. Specifically, Nⁿmc has five minimal siphons: S₁ = {{p₁, p₂, p₃, p₄}}, S₂ = {{p₃, p₅}}, S₃ = {{p₂, p₄, p₆}}, S₄ = {{p₂, p₃, p₅}}, and S₅ = {{p₄, p₅, p₆}}. The first four siphons are also traps, implying that they cannot be unmarked once p₁, p₅, and p₆ are initially marked. From the original net model (N, M₀), p₁, p₅, and p₆ are initially marked. As a monitor, p₆ must be initially marked. Otherwise there exist dead transitions at the initial marking. Next we give a marking relationship at which S₅ is controlled.

Note that \( I_{p₅} = p₃ + p₅ \), \( I_{p₆} = p₂ + p₄ + p₆ \), and \( I_{p₆} = p₂ + p₃ + p₆ \) are P-invariants in Figure 1(e). Let \( I = I_{p₅} + I_{p₆} - I_{p₆} \). Clearly, \( I = p₄ + p₅ + p₆ - p₆ \) is a P-invariant. Since \( |I| + S₅ \), S₅ is controlled if \( I_{M₀ⁿmc} > 0 \) holds, i.e., S₅ is controlled if \( M₀ⁿmc(p₄) + M₀ⁿmc(p₅) > M₀ⁿmc(p₆) \). The above results indicate that each siphon in the net shown in Figure 1(e) is controlled if p₁, p₅, and p₆ are initially marked, and \( M₀ⁿmc(p₄) + M₀ⁿmc(p₃) > M₀ⁿmc(p₆) \) is true.
Figure 1: (a) A plant model \((N, M_0)\), (b) the reachability graph of \((N, M_0)\), (c) a modified model \((N^m, M_0^m)\), (d) the reachability graph of \((N^m, M_0^m)\), (e) a controlled system \((N^{mc}, M_0^{mc})\) for \((N^m, M_0^m)\), and (f) a controlled system \((N^c, M_0^c)\) for \((N, M_0)\).
Now we consider the deadlock prevention problem for \((N, M_0)\) by using the net structure of the controlled system in Figure 1(e). In \((N, M_0)\), \(M_0 = 4p_1 + 2p_5 + p_6\), i.e., \(p_1, p_5,\) and \(p_6\) are initially marked. If \(p_c\) is initially marked and \(M_0'(p_5) + M_0'(p_6) > M_0'(p_c)\) is true, then a controlled system for \((N, M_0)\) can be obtained. Since \(M_0'(p_5) = M_0(p_5) = 2\) and \(M_0'(p_6) = M_0(p_6) = 1\), \(M_0'(p_c) = 2\) means the truth of \(M_0'(p_5) + M_0'(p_6) > M_0'(p_c)\), as shown in Figure 1(f).

In summary, by using the structure of the controlled system of a net with a small initial marking, we can compute a controlled system for the same net structure with a large initial marking, in terms of algebraic inequality constraints with respect to markings. That is to say, once the structure of a controlled system is computed, the initial marking of monitors is determined by a set of inequality constraints. For a plant model \((N, M_0')\) with its structure shown in Figure 1(a) and a new initial marking \(M_0' = 4p_1 + 3p_5 + p_6\), one can easily find a controlled system with its structure shown in Figure 1(f) and \(M_0'^m (p_c) = 3\) such that \(M_0'(p_5) + M_0'(p_6) > M_0'(p_c)\) is true. It is not necessary to apply the theory of regions afresh to \((N, M_0')\).

Let \((N, M_0)\) be a plant net model with place set \(P\) and \(P_V\) be the set of monitors in its controlled system \((N^c, M_0^c)\). The deadlock prevention procedure proposed in this study contains the following steps:

1. Find a controlled system \((N^{mc}, M_0^{mc})\) for \((N^m, M_0^m)\) by using the theory of regions, where \(N^m = N\) and \(M_0^m \leq M_0\).
2. Derive the controllability conditions of siphons in \(N^{mc}\), which are represented by algebraic inequalities of markings of the places in the plant model and monitors in \(N^{mc}\).
3. Decide an initial marking \(M_0^c\) such that \(\forall p \in P, M_0^c (p) = M_0 (p)\) and \(\forall p \in P_V, M_0^c (p)\) satisfies its corresponding inequality constraints. \((N^c, M_0^c)\) is a controlled system for \((N, M_0)\), where \(N^c = N^{mc}\).

3.2 M-nets

This study considers deadlock problems for a class of manufacturing-oriented Petri nets, M-nets for short. It is a generalization of the existing net classes that model FMS. In what follows, an M-net is referred to as a net system unless otherwise stated.

**Definition 2** An M-net denoted by \((N, M_0)\) satisfies the following statements:

1. \(N = \bigcap_{i=1}^n N_i = (P^0 \cup P_A \cup P_R, T, F, W)\) is composed of \(n\) nets \(N_1, N_2, \ldots,\) and \(N_n,\) where \(\forall i \in N_n, N_i = (\{p_i^0\} \cup P_A, \cup P_R, T_i, F_i, W_i)\) is called a subnet of \(N\).
2. \(P^0 = \bigcup_{i=1}^n \{p_i^0\}\) is called a set of idle process places with \(p_i^0 \neq p_j^0, \forall i, j \in N_n, i \neq j;\)
3. \(P_A = \bigcup_{i=1}^n P_A,\) is called a set of operation places with \(P_A \cap P_A = \emptyset, \forall i, j \in N_n, i \neq j;\) and \(P_R = \bigcup_{i=1}^n P_R,\) is called a set of resource places.
4. \(\forall i, j \in N_n, i \neq j, T_i \cap T_j = \emptyset.\)
5. \(\forall r \in P_R, r\) is associated with a minimal \(P\)-semiflow \(I_r\) such that \(I_r(r) = 1, \forall p \in P_A, I_r(p) \geq 0,\) and \(\forall p \in P^0, I_r(p) = 0.\)
6. \((N_i, M_0)\) is quasi-live, bounded, and conservative.
7. \((N_i', M_0')\) with \(N_i' = (\{p_i^0\} \cup P_A, T_i, F_i', W_i')\) is live, bounded, and reversible, where \(N_i'\) is the resulting net from removing resource places in \((N_i, M_0)\).
8. Let \((N_i, M_{i0})\) \((i = 1, 2)\) be two subnets with \(N_i = (\{p_i^0\} \cup P_i \cup P_{R_i}, T_i, F_i, W_i)\). Their composition, denoted by \((N_{12}, M_{12})\) with \(N_{12} = N_1 \circ N_2 = (P_{12}^0 \cup P_{A_{12}} \cup P_{R_{12}}, T_{12}, F_{12}, W_{12})\), is defined as follows:

- \(P_{12}^0 = \{p_{1}^0\} \cup \{p_{2}^0\} = \{p_1^0, p_2^0\}\), \(P_{A_{12}} = P_{A_1} \cup P_{A_2}\), and \(P_{R_{12}} = P_{R_1} \cup P_{R_2}\)

- \(T_{12} = T_1 \cup T_2\)

- \(F_{12} = F_1 \cup F_2\)

- \(\forall f \in F_1, W(f) = W_1(f)\) and \(\forall f \in F_2, W(f) = W_2(f)\)

- \(\forall p \in \{p_1^0\} \cup P_{A_1}, M_{12}(p) = M_{01}(p)\)

- \(\forall p \in \{p_2^0\} \cup P_{A_2}, M_{12}(p) = M_{02}(p)\)

- \(\forall r \in P_{R_1} \setminus P_{R_2}, M_{12}(r) = M_{01}(r)\)

- \(\forall r \in P_{R_2} \setminus P_{R_1}, M_{12}(r) = M_{02}(r)\)

9. The net \(N\) resulting from the composition of \(n\) subnets \(N_1, N_2, \ldots, N_n\) is defined as follows: if \(n = 1\), then \(N = N_1\); if \(n > 1\), then \(N = \bigcirc_{i=1}^{n} N_i = (\bigcirc_{i=1}^{n-1} N_i) \circ N_n\).

10. \(\forall p \in P^0, M_0(p) > 0; \forall p \in P_A, M_0(p) = 0\); and \(\forall r \in P_R, M_0(r) \geq \max\{I_r(p)|p \in \|I_r|\}\). Such a marking is said to be an admissible initial marking.

11. An uncontrolled siphon in \((N, M_0)\) contains at least one resource place and an operation place. No idle process place belongs to an uncontrolled siphon.

12. \((N, M_0)\) is live if no siphon is uncontrolled.

13. Liveness can be enforced to \((N, M_0)\) by adding monitors whose addition leads to a controlled system.

14. Let \((N^c, M_0^c)\) be a controlled system for \((N, M_0)\). \((N^c, M_0^c)\) is live if it is ordinary and no siphon is unmarked. \((N^c, M_0^c)\) is live if it is generalized and satisfies the cs-property.

15. Let \(P_V\) be the set of monitors in \((N^c, M_0^c)\). \(\forall v \in P_V\), there exists a minimal \(P\)-semiflow \(I_v\) such that \(I_v(v) \geq 1\) and \(\forall p \in \|I_v\| \setminus \{v\}, p \in P_A\).

For example, the net shown in Figure 1(a) is an M-net, where \(p_1\) is an idle process place, \(p_2, p_3,\) and \(p_4\) are operation places, and \(p_5\) and \(p_6\) are resource places. It is quasi-live, bounded, and conservative. It is live if no siphon is uncontrolled. For example, the net at initial marking 2\(p_1 + 2p_5 + p_6\) is live since every siphon can never be emptied.

It is easy to verify that M-nets are more general than almost all manufacturing-oriented Petri net subclasses in the literature such as PPN, augmented marked graphs [8], S3PR [9], L-S3PR [10], S4PR [11], S3PR [44], ES3PR [18], WS3PSR [43], S2PR [11], S3PMR [19], PNR [22], RCN-merged nets [21], ERCN-merged nets [56], ERCN*-merged nets [23], S3PGR [35], G-tasks [6], and well-formed G-systems [59]. In an appendix, a formal proof is presented to show that an M-net is more general than a well-formed G-system.

### 3.3 Minimal Initial Marking

Let \((N, M_0)\) be a plant M-net in which \(M_0\) is admissible. This section finds a minimal initial marking \(M_0^m\) at which M-net \(N\) contains deadlocks and any strict minimal siphon of \(N\) can become uncontrolled at a marking \(M \in R(N, M_0^m)\).
Algorithm 1 finding a minimal initial marking for $N$

Input: a plant model $(N, M_0)$ with $N = (P^0 \cup P_A \cup P_R, T, F, W)$
Output: $M^m_0$, “$(N, M_0)$ is live”, or “$(N, M_0)$ cannot be handled by the proposed method”

begin
the MIP-based deadlock detection method [8], [35] is applied to $(N, M_0)$
if {there are uncontrolled siphons in $(N, M_0)$} then
compute the set $\Pi_u$ of strict minimal siphons in $N$
$\forall p \in P^0 \cup P_A, M^m_0(p) := M_0(p)$
$\forall r \in P_R, M^m_0(r) := max\{I_r(p) | p \in ||I_r||\}$
the MIP-based deadlock detection method is applied to $(N, M^m_0)$
if {there are no uncontrolled siphons in $(N, M^m_0)$} then
flag:=2
else
compute the set of reachable markings of $(N, M^m_0)$
Find the set of dead markings $R_D(N, M^m_0)$
if {\forall S \in \Pi_u, \exists M \in R_D(N, M^m_0), S is uncontrolled at M} then
flag:=0
else
flag:=2
end if
end if
else
flag:=1
end if
if {flag==2} then
output “$(N, M_0)$ cannot be handled by the proposed method”
else
if {flag==1} then
output “$(N, M_0)$ is live”
else
output $M_0^m$
end if
end if
end of the algorithm

First, Algorithm 1 decides whether the original plant model $(N, M_0)$ is live. By the definition of M-nets, it is live if there is no uncontrolled siphon, which can determined be by an MIP-based deadlock detection method. If it is live (flag=1), the algorithm exits. If it is not live, then the set of strict minimal siphons and the minimal admissible initial marking $M^m_0$ are computed. If $(N, M^m_0)$ is live (flag=2), then we cannot find a minimal initial marking for it and its deadlock problems can be handled by any existing approach. If $(N, M^m_0)$ is not live and for any strict minimal siphon, there exists a dead marking in $R(N, M^m_0)$ at which the siphon is uncontrolled, then the minimal admissible marking $M^m_0$ is obtained. Otherwise (flag=2), we cannot find a minimal initial marking for it and its deadlock problems can be handled by any existing approach, e.g., the theory of regions and siphon-based ones. Note that this algorithm needs to compute the state space and the set of strict minimal siphons of $(N, M^m_0)$ if the original plant model is not live. However, if the algorithm outputs the minimal initial marking $M^m_0$ for $N$, the information of the state space and siphons of $(N, M^m_0)$ will be useful later. Take the net $(N, M_0)$ shown in Figure 1(a) as an example. $(N, M_0)$ is not live and $S = \{p_4, p_5, p_6\}$ is unique strict
3.4 Derivation of the Structure of a Controlled System

Let \((N, M_0)\) be an M-net with \(N = (P^0 \cup P_A \cup P_R, T, F, W)\). To find its controlled system, as stated in Section 3.1, we first design a controlled system for \((N^m, M^m_0)\), where \(N^m = N\) and \(M^m_0\) is the minimal initial marking.

**Algorithm 2** structure design of a controlled system for \((N, M_0)\)

**Input:** a plant model \((N, M_0)\)

**Output:** \((N^{mc}, M^{mc}_0)\)

begin

\(N^m := N\)

find the minimal initial marking \(M^m_0\) by Algorithm 1

if \{there exists an optimal controlled system for \((N^m, M^m_0)\)\} then

design a controlled system \((N^{mc}, M^{mc}_0)\) for \((N^m, M^m_0)\) by the theory of regions

else

design a controlled system \((N^{mc}, M^{mc}_0)\) for \((N^m, M^0_m)\) by the method in [47]

end if

\(m := (N^{mc}, M^{mc}_0)\)
end of the algorithm

The motivation to design a controlled system \((N^{mc}, M^{mc}_0)\) for \((N^m, M^m_0)\) is that \(M^m_0\) is not greater than \(M_0\) and it is more tractable by using the theory of regions to design a controlled system for \((N^m, M^m_0)\) than that for \((N, M_0)\).

The net shown in Figure 1(a) with an initial marking \(M_0 = 150p_1 + 100p_5 + 50p_6\) has more than \(1.3 \times 10^5\) states. One can imagine the computational overhead if the theory of regions is applied to such a net. However, Algorithm 2 considers \((N^m, M^m_0)\) as shown in Figure 1(c), which has five reachable markings only. As a result, it is easy to compute a controlled system for \((N^m, M^m_0)\) by using the theory of regions, as shown in Figure 1(e).

4 Siphon Controllability Constraints

This section presents a set of algebraic inequality constraints with respect to the markings of resource places and monitors at which \((N^{mc}, M^{mc}_0)\) is live, where \((N^{mc}, M^{mc}_0)\) is a controlled system for \((N^m, M^m_0)\) with \(N^{mc} = (P^0 \cup P_A \cup P_R \cup P_v, T, F, W, V)\) and \(N^m = (P^0 \cup P_A \cup P_R, T, F, W)\), which has been computed in Algorithm 1. Let \(\Pi_u\) denote the set of uncontrolled minimal siphons in \((N^{mc}, M^{mc}_0)\). \(\Pi_u\) can be further divided into three disjoint subsets \(\Pi_G, \Pi_H, \) and \(\Pi_V\), called the sets of plant, hybrid, and monitor siphons, respectively, such that \(\forall S \in \Pi_G, \forall p \in S, p \in P_A \cup P_R; \forall S \in \Pi_H, \forall v \in S \setminus (S \cap P_A), v \in P_v;\) and \(\forall S \in \Pi_V, \forall r \in P_v\) such that \(\{r, v\} \subseteq S\).

The Petri net shown in Figure 2(b) is an optimal controlled system for a plant model \((N, M_0)\) depicted in Figure 2(a) that is an M-net with \(P^0 = \{p_1, p_3\}, P_R = \{p_9, p_{10}, p_{11}\}\), and the others are operation places. There are 15 minimal siphons: \(S_1 = \{p_3, p_8, p_9, p_{10}\}\), \(S_2 = \{p_4, p_7, p_{10}, p_{11}\}\), \(S_3 = \{p_4, p_8, p_9, p_{10}, p_{11}\}\), \(S_4 = \{p_3, p_7, v_1, v_2\}\), \(S_5 = \{p_3, p_8, p_9, v_1, v_2\}\), \(S_6 = \{p_4, p_7, p_{11}, v_1, v_2\}\), \(S_7 = \{p_4, p_8, p_9, p_{11}, v_1, v_2\}\), \(S_8 = \{p_1, p_2, p_3, p_4\}\), \(S_9 = \{p_5, p_6, p_7, p_8\}\), \(S_{10} = \{p_2, p_8, p_9\}\), \(S_{11} = \{p_3, p_7, p_{10}\}\), \(S_{12} = \{p_4, p_6, p_{11}\}\), \(S_{13} = \{p_2, p_7, v_1\}\), \(S_{14} = \{p_3, p_6, v_2\}\), and \(S_{15} = \{p_2, p_6, v_3\}\). Note that \(S_8, S_9, \ldots,\) and \(S_{15}\) are also traps that
are marked at the minimal initial marking by the definition of M-nets. As a result, they
do not contribute to deadlocks. We have $\Pi_u = \{S_1, S_2, \ldots, S_7\}$ with $\Pi_G = \{S_1, S_2, S_3\}$,$\Pi_V = \{S_4\}$, and $\Pi_H = \{S_5, S_6, S_7\}$. From the definition of an M-net, a plant siphon $S$
can be represented by $S_A \cup S_R$, where $S_A \subseteq P_A$ and $S_R \subseteq P_R$.

![Figure 2](image_url)

**Figure 2:** (a) A plant model $(N, M_0)$, (b) a controlled system $(N^{mc}, M_0^{mc})$ for $(N, M_0)$.

**Definition 3** Let $S$ be a plant siphon in $N^{mc} = (P^0 \cup P_A \cup P_R \cup P_V, T, F^{mc}, W^{mc})$ and $I_r$
the minimal P-semiflow associated with $r \in P_R \cup P_V$. $I_r$ and $S$ are considered as multisets.
Let $\Omega_S = \sum_{r \in S \cap P_A} I_r$ and $\Omega'_S = \sum_{p \in S} \Omega_S(p)p$. As a multiset, $Th_S = \Omega_S - \Omega'_S$ is called
the complementary set of siphon $S$.

For example, $S_1 = \{p_3, p_8, p_9, p_{10}\}$ is a plant siphon in Figure 2(b). We have $\Omega_{S_1} =$
$p_2 + 2p_8 + p_9 + p_3 + 2p_7 + p_{10}$ and $\Omega'_{S_1} = p_3 + 2p_8 + p_9 + p_{10}$. Thus, $Th_{S_1} = p_2 + p_7$.$S_2 = \{p_4, p_7, p_{10}, p_{11}\}$ is also a plant siphon with $\Omega_{S_2} = p_3 + p_7 + p_{10} + p_4 + p_6 + p_{11}$ and
$\Omega'_{S_2} = p_4 + p_7 + p_{10} + p_{11}$. Therefore, we have $Th_{S_2} = p_3 + p_6$.

**Definition 4** Let $r \in P_R \cup P_V$ be a resource or monitor in $N^{mc}$. $H(r) = I_r - r$ is called
the set of holders of $r$, where $H(r)$ is a multiset.

**Corollary 1** $\forall r \in P_R \cup P_V$, $\{p|r \in H(r)\} \subseteq P_A$.

For example, $H(p_8) = I_{p_0} - p_9 = p_2 + 2p_8 + p_9 - p_9 = p_2 + 2p_8$ and $H(p_{10}) = p_3 + p_7$. It
is easy to verify that $\{p_2, p_8\} \subseteq P_A$ and $\{p_3, p_7\} \subseteq P_A$.

**Definition 5** Let $S$ be a plant siphon. $\alpha_S \subseteq P_R \cup P_V$ is called a minimal set of constraint
places of $S$ with respect to $A_S = \{d_{\alpha} \in N^+ | i = 1, 2, \ldots, |\alpha_S|\}$ if
(1) $\alpha_S \neq S \setminus P_A$; (2) $Th_S \subseteq \sum_{v_i \in \alpha_S} d_{v_i}H(v_i)$; (3) $\exists p \in Th_S$, $Th_S(p) > \Omega(p)$, where $\Omega = \sum_{v_i \in \alpha_S}(d_{v_i} - 1)H(v_i)$;
and (4) $\exists \alpha' \subseteq P_R \cup P_V$ such that (1)–(3) are satisfied and $|\alpha'_S| < |\alpha_S|$.

The addition of the sets of the holders of the monitors in $\alpha_S$ is greater than or equals
to multiset $Th_S$. For example, $S_1 = \{p_3, p_8, p_9, p_{10}\}$ is a plant siphon in Figure 2(b)
with $Th_{S_1} = p_2 + p_7$. Note that $H(v_1) = p_2 + 2p_7$. We have $\alpha_{S_1} = \{v_1\}$ and $d_{v_1} = 1$
since $Th_{S_1} < p_2 + 2p_7$. Consider a plant siphon $S_3 = \{p_4, p_8, p_9, p_{10}, p_{11}\}$ with $Th_{S_3} =$
p_2 + p_3 + p_6 + p_7. We have $\alpha_{S_3} = \{v_1, v_2\}$ with $d_{v_1} = 1$ and $d_{v_2} = 1$. 

11
Proposition 2 Let $S$ be a plant siphon and $\alpha_S$ be its minimal set of constraint places. Let $I_S = \sum_{r \in S} I_r - \sum_{v \in \alpha_S} d_v I_v$. $S$ is max-controlled if $M_0^{mc}(S R) - \sum_{v \in \alpha_S} d_v M_0^{mc}(v) > \sum_{p \in S} I_S(p)(\max_p - 1)$. 

Proof: Both $\sum_{r \in S} I_r$ and $\sum_{v \in \alpha_S} d_v I_v$ are P-semiflows. As a result, $I_S$ is a P-invariant. By the definition of a minimal set of constraint places for a plant siphon $S$, we have $S\Theta_S \leq \sum_{v \in \alpha_S} d_v H(v)$. Therefore, $\|I_S\| - \sigma S = 0, \|I_S\|^+ \subseteq S$, and $\forall p \in \|I_S\| \setminus (S \cup \alpha_S) = 0$. $M_0^{mc}(S R) - \sum_{v \in \alpha_S} d_v M_0^{mc}(v) > \sum_{p \in S} I_S(p)(\max_p - 1)$ implies the truth of $I_S M_0^{mc} > \sum_{p \in S} I_S(p)(\max_p - 1)$. By Proposition 1, $S$ is max-controlled. 

Note that the condition in Proposition 2 is rather conservative. This is due to Proposition 1. Consider $S_3 = \{p_4, p_8, p_9, p_{10}, p_{11}\}$ in Figure 2(b). We have $\alpha_{S_3} = \{v_1, v_2\}$ with $d_{v_1} = 1$ and $d_{v_2} = 1$. As a result, $S_3$ is max-controlled if $M_0^{mc}(p_6) + M_0^{mc}(p_{10}) + M_0^{mc}(p_{11}) - M_0^{mc}(v_1) - M_0^{mc}(v_2) > 1$. It is easy to verify that the above inequality does not hold in Figure 2(b). However, transitions in $S_3^\bullet$ are still live.

Corollary 2 Let $S$ be a minimal siphon in an ordinary net $N$. $\forall p \in S$, $\max_p - 1 = 0$.

Proof: By contradiction, suppose that $\exists p \in S$, $\max_p = 0$. It means that $\star(S \setminus \{p\}) \subseteq (S \setminus \{p\})$. This contradicts the minimality of siphon $S$.

Corollary 3 Let $S$ be a plant siphon in $(N^{mc}, M_0^{mc})$ that is ordinary. $S$ is controlled if $M_0^{mc}(S R) - \sum_{v \in \alpha_S} d_v M_0^{mc}(v) > 0$.

Proof: It immediately follows from Corollary 2.

For example, $S_5 = \{p_4, p_5, p_6\}$ is a plant siphon in Figure 1(e) that is ordinary. We have $S\Theta_{S_5} = p_2 + p_3$ and $H(p_c) = p_2 + p_3$. Thus, $\alpha_{S_5} = \{p_c\}$. Note that $p_4 \in P_A$ and $\{p_5, p_6\} \subseteq P_R$. $S_5$ is controlled if $M_0^{mc}(p_5) + M_0^{mc}(p_6) - M_0^{mc}(p_c) > 0$.

Definition 6 Let $S$ be a monitor siphon. $\alpha_S \subseteq P_R \cup P_V$ is called a minimal set of constraint places of $S$ with respect to $A_S = \{d_v \in \mathbb{N}^+ | i = 1, 2, \ldots, |\alpha_S|\}$ if (1) $\alpha_S \not\subseteq S \setminus P_A$; (2) $S\Theta_S \leq \sum_{v \in \alpha_S} d_v H(v_i)$; (3) $\exists p \in S\Theta_S, T_S(p) > \Omega(p)$, where $\Omega = \sum_{v \in \alpha_S} (d_v - 1)H(v_i)$; and (4) $\exists \alpha_S \subseteq P_R \cup P_V$ such that (1)-(3) are satisfied and $|\alpha_S| < |\alpha_S|$.

Proposition 3 Let $S$ be a monitor siphon and $\alpha_S$ its minimal set of constraint places. Let $I_S = \sum_{v \in S \setminus P_V} I_v - \sum_{r \in \alpha_S} d_r I_r$. $S$ is max-controlled if $M_0^{mc}(S \cap P_V) - \sum_{r \in \alpha_S} d_r M_0^{mc}(r) > \sum_{p \in S} I_S(p)(\max_p - 1)$.

For instance, $S_4 = \{p_3, p_7, v_1, v_2\}$ is a monitor siphon in Figure 2(b). Its minimal set of constraint places is $\alpha_{S_4} = \{p_9, p_{11}\}$ with $H(p_9) = p_2 + 2p_8, d_{p_9} = 1, H(p_{11}) = p_4 + p_6$, and $d_{p_{11}} = 1$. Thus, $S_4$ is max-controlled if $M_0^{mc}(v_1) + M_0^{mc}(v_2) - M_0^{mc}(p_9) - M_0^{mc}(p_{11}) > 1$.

Definition 7 Let $S$ be a hybrid siphon. $\alpha_S \subseteq P_R \cup P_V$ is called a minimal set of constraint places of $S$ with respect to $A_S = \{d_v \in \mathbb{N}^+ | i = 1, 2, \ldots, |\alpha_S|\}$ if (1) $\alpha_S \not\subseteq S \setminus P_A$; (2) $S\Theta_S \leq \sum_{v \in \alpha_S} d_v H(v_i)$; (3) $\exists p \in S\Theta_S, T_S(p) > \Omega(p)$, where $\Omega = \sum_{v \in \alpha_S} (d_v - 1)H(v_i)$; and (4) $\exists \alpha_S \subseteq P_R \cup P_V$ such that (1)-(3) are satisfied and $|\alpha_S| < |\alpha_S|$.

Proposition 4 Let $S$ be a hybrid siphon and $\alpha_S$ its minimal set of constraint places. Let $I_S = \sum_{v \in S \setminus P_V} I_v - \sum_{r \in \alpha_S} d_r I_r$. $S$ is max-controlled if $M_0^{mc}(S \setminus P_A) - \sum_{r \in \alpha_S} d_r M_0^{mc}(r) > \sum_{p \in S} I_S(p)(\max_p - 1)$.
For example, \( S_6 = \{p_4, p_7, p_{11}, v_1, v_2\} \) is hybrid with \( Th_{S_6} = p_2 + p_3 + 2p_6 \). We have \( \alpha_{S_6} = \{p_{10}, v_3\} \) with \( d_{p_{10}} = 1 \) and \( d_{v_3} = 1 \). As a result, \( S_6 \) is max-controlled if \( M_0^{mc}(p_{11}) + M_0^{mc}(v_1) + M_0^{mc}(v_2) - M_0^{mc}(p_{10}) - M_0^{mc}(v_3) > 1 \).

**Algorithm 3** controllability constraint generation for siphons in \((N^{mc}, M_0^{mc})\)

**Input:** a set of uncontrolled siphons \( \Pi_u \)

**Output:** a set of inequality constraints \( C \)

begin

divide \( \Pi_u \) into \( \Pi_G, \Pi_H, \) and \( \Pi_V \)

derive constraints for siphons in \( \Pi_G, \Pi_H, \) and \( \Pi_V \) by Propositions 2, 3, and 4, respectively

denote the set of constraints for siphons in \( \Pi_G (\Pi_H; \Pi_V) \) by \( C_G (C_H; C_V) \) and let \( C := C_G \cup C_H \cup C_V \)

Reorder the variables in each constraint in \( C \) such that no resource place is in the left side and no monitor is in the right side

output \( C \)

end of the algorithm

Let \( C = \{c_i | i \in \mathbb{N}_{|\Pi_u|}\} \) be the set of constraints of siphons in a net with \( P_V = \{v_1, v_2, \ldots, v_m\} \) and \( P_R = \{r_1, r_2, \ldots, r_k\} \). A constraint \( c_i \in C \) can be represented by

\[
c_i \equiv \sum_{j=1}^{m} \beta^{S_i}_j M_0^{mc}(v_j) < \sum_{j=1}^{k} \delta^{S_i}_j M_0(r_j) - \omega^{S_i} \tag{1}
\]

where \( \beta^{S_i}_j \) and \( \delta^{S_i}_j \) are integers and \( \omega^{S_i} = \sum_{p \in S_i} I_S(p)(\max_{r_i} - 1) \).

Eq. (1) can be re-written as

\[
c_i \equiv \sum_{j=1}^{m} \beta^{S_i}_j M_0^{mc}(v_j) \leq \sum_{j=1}^{k} \delta^{S_i}_j M_0(r_j) - \omega^{S_i} - 1 \tag{2}
\]

Let \( L = [l_{ij}]_{n \times m} = [\beta^{S_i}_j]_{n \times m}, x = [M_0^{mc}(v_1), M_0^{mc}(v_2), \ldots, M_0^{mc}(v_m)]^T, \) and \( \bar{B} = [\bar{b}_{ij}]_{n \times k}, \omega = [-\omega^{S_1} - 1, -\omega^{S_2} - 1, \ldots, -\omega^{S_n} - 1]^T, \) \( B = [\bar{B}\omega], \) and \( y = [M_0^{mc}(r_1), M_0^{mc}(r_2), \ldots, M_0^{mc}(r_k), 1]^T. \) The controllability constraints of uncontrolled siphons can be written to be

\[
Lx \leq By \tag{3}
\]

Eq. (3) is called a (liveness) constraint equation. To find a controlled system given a plant net model, the marking of monitors can be decided by solving the following LPP:

\[
z = \max \{\sum_{i=1}^{m} M_0^{mc}(v_i)\}
\]

s.t.

\[
Lx \leq By
\]

Take the net shown in Figure 2(b) as an example. The constraints of uncontrolled siphons are as follows:
Let \( \theta_1 = M^{mc}_0(v_1) \), \( \mu_i = M^{mc}_0(p_i) \), and \( \mu_0 = 1 \). The matrix form of the controllability constraints is as follows:

\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
1 & 1 & 0 \\
-1 & -1 & 0 \\
0 & -1 & 1 \\
-1 & -1 & 1 \\
0 & -1 & 1 
\end{pmatrix}
\begin{pmatrix}
\theta_1 \\
\theta_2 \\
\theta_3 
\end{pmatrix}
\leq
\begin{pmatrix}
1 & 1 & 0 & -2 \\
0 & 1 & 1 & -1 \\
1 & 1 & 1 & -2 \\
-1 & 0 & -1 & -2 \\
0 & -1 & 1 & -2 \\
1 & -1 & 1 & -3 
\end{pmatrix}
\begin{pmatrix}
\mu_0 \\
\mu_{10} \\
\mu_{11} \\
\mu_0 
\end{pmatrix}
\tag{4}
\]

Suppose that \( M_0 = h_1 p_1 + h_5 p_5 + 5 p_9 + 4 p_{10} + 3 p_{11} \) is an initial marking for the net structure in Figure 2(a), where \( h_1 \) and \( h_{10} \) are integers and big enough such that no transition is disabled due to deficiency of tokens in idle places. At initial marking \( M_0 = h_1 p_1 + h_5 p_5 + 5 p_9 + 4 p_{10} + 3 p_{11} \), the above constraints can be re-written as

\[
\begin{cases}
\theta_1 \leq 7 \\
\theta_2 \leq 6 \\
\theta_1 + \theta_2 \leq 10 \\
-\theta_1 - \theta_2 \leq -10 \\
\theta_3 - \theta_2 \leq 2 \\
\theta_3 - \theta_1 - \theta_2 \leq -3 \\
\theta_3 - \theta_2 \leq 1
\end{cases}
\]

Taking \( z = \max\{\sum_{i=1}^{3} \theta_i\} \) as an objective function, the LPP with the above constraint equation has an optimal solution \( z^* = 17 \) with \( \theta_1 = 4, \theta_2 = 6, \) and \( \theta_3 = 7 \). That is to say, the net structure shown in Figure 2(b) at initial marking \( M_0 = h_1 p_1 + h_5 p_5 + 5 p_9 + 4 p_{10} + 3 p_{11} + 4 r_1 + 6 v_2 + 7 v_3 \) is a controlled system with liveness for the net structure shown in Figure 2(a) at the given initial marking \( M_0 \).

Eq. (4) shows the liveness requirements for the net in Figure 2(a). Next section focuses on reducing the size of a constraint equation.

### 5 Redundancy Condition Identification of Constraints

Due to a large number of siphons in a controlled system, the size of a constraint equation is in theory exponential with respect to the structural scale of a Petri net model. In the standard form of an LPP, i.e., \( Ax \leq b \), constraint matrix \( A \) is usually assumed to have full row rank, where \( b \) is right-hand-side vector. It is easy to see that in this study, matrix \( L \), in a general case, does not have full row rank, which can be shown by the existing experimental results in the literature. On the other hand, a designer always hopes to have a concise constraint equation. This section focuses on identifying redundant constraints in order to find a small set of siphon controllability conditions.
Let \((N, M_0)\) be a plant net model with minimal initial marking \(M_0^m\) and \((N^{mc}, M_0^{mc})\) be a controlled system for \((N, M_0^m)\). The set of siphons to be controlled in \((N^{mc}, M_0^{mc})\) is denoted by \(\Pi_q\). As stated in the previous section, the controllability of siphons can be represented by a set of constraints \(\mathcal{C}\), taking the form of \(Lx \leq By\). A single constraint is denoted by \((l_i, b_i)\) that represents \(l_i^T \mathbf{x} \leq b_i^T \mathbf{y}\). Alternatively, a constraint \(c_i \in \mathcal{C}\) can be written as \((l_i, b_i)\) \(\equiv l_i^T \mathbf{x} \leq b_i^T \mathbf{y}\).

**Definition 8** A constraint \((l, b)\) is said to be redundant with respect to \((l_i, b_i)\) if the truth of \((l_i, b_i)\) implies that of \((l, b)\).

**Definition 9** A constraint \((l, b)\) is said to be redundant with respect to a set of constraints \(\mathcal{C}_S\) if the truth of one or more constraints in \(\mathcal{C}_S\) implies that of \((l, b)\).

**Definition 10** Let \(t_\alpha, t_\beta, \ldots, t_\gamma\) \(\subseteq \Pi(q)\) be a linearly independent maximal set of matrix \(L\). Then \(\mathcal{C}_E = \{(l_\alpha, a_\alpha), (l_\beta, a_\beta), \ldots, (l_\gamma, a_\gamma)\}\) is called a set of elementary constraints in \(\mathcal{C}\).

**Definition 11** \((l, b) \notin \mathcal{C}_E\) is called a strongly dependent constraint if \(\exists a_i \geq 0\) and \((l_i, b_i) \in \mathcal{C}_E\) such that \(l = \sum (l_i, b_i) \in \mathcal{C}_E a_i l_i\).

**Definition 12** \((l, b) \notin \mathcal{C}_E\) is called a weakly dependent constraint if \(\exists a_i > 0\) and \(\exists\) non-empty \(A, B \subseteq \mathcal{C}_E\) such that \(A \cap B = \emptyset\), and \(l = \sum (l_i, b_i) \in A a_i l_i - \sum (l_i, b_i) \in B a_i l_i\).

**Theorem 1** \(|\mathcal{C}_E| = \text{rank}(L)\), where \(\text{rank}(L)\) denotes the rank of \(L\).

**Proof:** It follows immediately from the definition of elementary constraints.

Since the rank of \(L\) is bounded by the smaller of \(|P_V|\) and \(|\Pi_q|\), Theorem 1 leads to the following important conclusion.

**Theorem 2** \(|\mathcal{C}_E| \leq |P_V|\).

**Proof:** \(|\mathcal{C}_E| = \text{rank}(L) \leq \min\{|P_V|, |\Pi_q|\} \leq |P_V|\).

**Theorem 3** Let \((l, b)\) be a strongly dependent constraint with \(l = \sum (l_i, b_i) \in \mathcal{C}_E a_i l_i (a_i \geq 0)\).

\(l^T \mathbf{x} \leq b^T \mathbf{y}\) is true, i.e., \((l, b)\) is redundant with respect to \(\mathcal{C}_E\), if \(\sum (l_i, b_i) \in \mathcal{C}_E a_i l_i \leq \sum (l_i, b_i) \in \mathcal{C}_E a_i b_i^T \mathbf{y} \leq b^T \mathbf{y}\) that results from \(l = \sum (l_i, b_i) \in \mathcal{C}_E a_i l_i\).

**Proof:** It immediately follows due to \(l^T \mathbf{x} = \sum (l_i, b_i) \in \mathcal{C}_E a_i l_i \leq \sum (l_i, b_i) \in \mathcal{C}_E a_i b_i^T \mathbf{y} \leq b^T \mathbf{y}\) that results from \(l = \sum (l_i, b_i) \in \mathcal{C}_E a_i l_i\).

**Theorem 4** Let \((l, b)\) be weakly dependent with \(l = \sum (l_i, b_i) \in \mathcal{C}_E a_i l_i - \sum (l_i, b_i) \in \mathcal{C}_E a_i l_i \leq \sum (l_i, b_i) \in \mathcal{C}_E a_i^T \mathbf{y} \leq b^T \mathbf{y}\).

\(l^T \mathbf{x} \leq b^T \mathbf{y}\) is true (redundant with respect to \(\mathcal{C}_E\)) if \(\sum (l_i, b_i) \in \mathcal{C}_E a_i b_i^T \mathbf{y} - \sum (l_i, b_i) \in \mathcal{C}_E a_i l_i \leq b^T \mathbf{y}\).

**Proof:** This result is trivially true.

For example, \(\mathcal{C}_E = \{(l_1, b_1), (l_2, b_2), (l_6, b_6)\}\) is a set of elementary constraints of Eq. (4). We have \(l_3 = l_1 + l_2, l_4 = -l_1 - l_2, l_5 = l_1 + l_6,\) and \(l_7 = l_1 + l_6\). By Definitions 11 and 12, \((l_3, b_3)\), \((l_5, b_5)\), and \((l_7, b_7)\) are strongly dependent constraints, and \((l_4, b_4)\) is weakly dependent. Specifically, the redundancy condition of \((l_3, b_3)\) is

\[M_0^{mc}(p_9) + M_0^{mc}(p_{10}) - 2 + M_0^{mc}(p_{11}) + M_0^{mc}(p_{11}) - 1 \leq M_0^{mc}(p_9) + M_0^{mc}(p_{10}) + M_0^{mc}(p_{11}) - 2,\]

i.e., \(M_0^{mc}(p_{10}) \leq 1\). The redundancy conditions of the dependent constraints in Eq. (4) are summarized in Table 1.
Table 1: Redundancy conditions of the dependent constraints in Figure 2(b).

<table>
<thead>
<tr>
<th>dependent constraint</th>
<th>redundancy condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l_3 = l_1 + l_2$</td>
<td>$M_0^{mc}(p_{10}) \leq 1$</td>
</tr>
<tr>
<td>$l_4 = -l_1 - l_2$</td>
<td>$M_0^{mc}(p_{10}) \geq 2.5$</td>
</tr>
<tr>
<td>$l_5 = l_1 + l_6$</td>
<td>$M_0^{mc}(p_{11}) \leq 2$</td>
</tr>
<tr>
<td>$l_7 = l_1 + l_6$</td>
<td>$M_0^{mc}(p_{10}) \leq 2$</td>
</tr>
</tbody>
</table>

6 Deadlock Prevention Policy

This section proposes a deadlock prevention policy based on the results obtained in the previous sections. Its computational complexity is also discussed.

Let $\Pi_{ST}$ denote the set of minimal siphons that are also traps marked at the minimal initial marking of an M-net. They are called st-siphons. In fact, $\Pi_{ST}$ can be easily computed from $(N_{mc}, M_{0}^{mc})$ by its structural properties. For example, $\{p_3, p_5\}$, $\{p_2, p_4, p_6\}$, and $\{p_1, p_2, p_3, p_4\}$ are st-siphons since they are marked at any admissible initial marking.

By using the MIP-based deadlock detection method [8] or complete siphon enumeration [9], it can be verified whether a plant M-net contains deadlocks. We assume that there are deadlocks in a plant M-net.

Algorithm 4 controlled system design for $(N, M_0)$

Input: an M-net $(N, M_0)$ with $N = (P^0 \cup P_A \cup P_R, T, F, W)$

Output: controlled system $(N_c, M_0^c)$

begin
find $(N^m, M_0^m)$ for $(N, M_0)$, where $N^m = N$
design a controlled system $(N^{mc}, M_0^{mc})$ for $(N^m, M_0^m)$ by Algorithm 2
compute the set $\Pi_A$ of all minimal siphons in $N^{mc}$
find $\Pi_{ST}$
$\Pi_u := \Pi_A \setminus \Pi_{ST}$
derive controllability constraints for siphons in $\Pi_u$ by Algorithm 8
find redundancy conditions for constraints
decide $M_0^c(v)$ by reduced constraint set and $M_0$, $\forall v \in P_V$
$\forall p \in P^0 \cup P_A \cup P_R$, $M_0^c(p) := M_0(p)$
$N_c := N^{mc}$
output $(N_c, M_0^c)$
end of the algorithm

Theorem 5 $(N_c, M_0^c)$ resulting from Algorithm 4 is a live controlled system for plant model $(N, M_0)$.

Proof: All siphons in $(N_c, M_0^c)$ are max-controlled, implies that it satisfies cs-property. From Definition 2, $(N_c, M_0^c)$ is live. □

The complexity of this deadlock control algorithm is exponential with respect to the size of a plant net model since both the theory of regions and complete siphon enumeration are exponential. However, the fact underlying this deadlock control policy is that, given a plant model with any admissible initial marking, its controlled system can be easily decided by the controllability constraints of siphons once the structure of a controlled system is computed. That is to say, given a plant net model, we just need to use the theory of regions and compute all minimal siphons once to find the structure of a controlled system. Even if the initial marking of the plant model changes, the structure of the controlled system
obtained previously can be reused as the structure of a new controlled system. This means that we just need to find the markings of the monitors in the new controlled system, which can be decided by satisfying the controllability constraints of siphons. Figure 3 shows the flowchart underlying the deadlock control strategy, where the computation involved in the steps above the dotted line is carried out only once for a net structure $N$.

Figure 3: Flowchart of the deadlock prevention policy.

One may wonder the superiority of the proposed method since many deadlock prevention approaches existing in the literature need to compute siphons only once given a net structure, e.g., the ones in [9], [26], [1], [17], [18], [19], and [28]. However, they are usually overly conservative. For example, a study on a typical FMS shows that the liveness-enforcing controlled system resulting from the deadlock prevention policy developed by Ezpeleta et al. usually has 30% or so permissive behavior of an optimally controlled system [30]. Section 8 indicates the behavioral permissiveness of the proposed method, which is close to the optimal one.

7 A Real-world Example

Fig. 4 shows a real-world part cleaning and storage system that consists of an automatic storage and retrieval system (ASRS), an elevator, three conveyors C1, C2, and C3, two automatic cleaning machines M1 and M2, two buffers B1 and B2, and a robot R. B1, B2, and C1 can store five, five, and ten parts, respectively. Machine 1 can process three parts, machine 2 can process two parts and the robot can move one part at a time. The robot is used to load and unload the cleaning machines. After a part is cleaned, it is then moved to C1 by the robot. The elevator can carry one part at a time to ASRS. As shown in Fig. 5(a), the Petri net model of the system is an M-net, where the place set has a partition $P_0$, $P_A$, and $P_R$. Specifically, we have $P_0 = \{p_1\}$, $P_R = \{p_{13}, p_{14}, p_{15}, p_{16}, p_{17}, p_{18}, p_{19}\}$, and $P_A = \{p_2, p_3, p_4, p_5, p_7, p_8, p_9, p_{10}, p_{11}, p_{12}\}$.

In order to find its controlled system by using the theory of regions with overhead as
inexpensive as possible, we first convert the plant model \((N, M_0)\) into a net that has the
same structure and minimal initial marking that is not greater than the initial marking.
Such a revised model \((N^m, M_0^m)\) is shown in Figure 5(b). Its controlled system is de-
picted in Figure 5(c), which is found by the theory of regions, where \(v_1, v_2\), and \(v_3\) are
monitors. In \((N^m, M_0^m)\), we have \(\Pi_A = \{S_1, S_2, \ldots, S_{11}\}\), where \(S_1 = \{p_4, p_{16}, p_{18}\},
S_2 = \{p_9, p_{17}, p_{18}\}, S_3 = \{p_4, p_9, p_{16}, p_{17}, p_{18}\}, S_4 = \{p_2, p_{13}\},
S_5 = \{p_7, p_{14}\}, S_6 = \{p_4, p_{16}\}, S_7 = \{p_9, p_{17}\}, S_8 = \{p_3, p_5, p_8, p_{10}, p_{18}\},
S_9 = \{p_{11}, p_{15}\}, S_{10} = \{p_{12}, p_{19}\},\) and \(S_{11} = \{p_1, p_2, p_3, p_4, p_5, p_7, p_8, p_9, p_{10}, p_{11}, p_{12}\}\).

From Figure 5(c), we have \(\Pi_{ST} = \{S_4, S_5, \ldots, S_{11}\}\). As a result, \(\Pi_u = \{S_1, S_2, S_3\}\) and
all the siphons in \(\Pi_u\) are plant siphons. Consider \(S_1 = \{p_4, p_{16}, p_{18}\}\) with \(Th_{S_1} = p_3 + p_4\).
Note that \(H(v_1) = p_3 + p_4\). As a result, the controllability of siphon \(S_1\) can be represented
by \(M_0^{mc}(p_{16}) + M_0^{mc}(p_{18}) - M_0^{mc}(v_1) > 0\). Similarly, the controllability constraints of
siphons \(S_2\) and \(S_3\) can be found, which are \(M_0^{mc}(p_{17}) + M_0^{mc}(p_{18}) - M_0^{mc}(v_2) > 0\) and
\(M_0^{mc}(p_{16}) + M_0^{mc}(p_{17}) + M_0^{mc}(p_{18}) - M_0^{mc}(v_1) - M_0^{mc}(v_2) > 0\), respectively. Let \(\mu_0 = 1\).
Their matrix form is as follows:

\[
\begin{pmatrix}
1 & 0 \\
0 & 1 \\
1 & 1
\end{pmatrix}
\begin{pmatrix}
v_1 \\
v_2 \\
v_3
\end{pmatrix}
\leq
\begin{pmatrix}
1 & 0 & 1 & -1 \\
0 & 1 & 1 & -1 \\
1 & 1 & 1 & -1
\end{pmatrix}
\begin{pmatrix}
p_{16} \\
p_{17} \\
p_{18} \\
\mu_0
\end{pmatrix}
\]

(5)

Constraint \(M_0^{mc}(p_{16}) + M_0^{mc}(p_{17}) + M_0^{mc}(p_{18}) - M_0^{mc}(v_1) - M_0^{mc}(v_2) > 0\) is redundant with respect to \(M_0^{mc}(p_{16}) + M_0^{mc}(p_{18}) - M_0^{mc}(v_1) > 0\) and \(M_0^{mc}(p_{17}) + M_0^{mc}(p_{18}) -
M_0^{mc}(v_2) > 0\) if \(M_0^{mc}(p_{18}) \leq 1\).

The controlled system of the original plant model shown in Figure 5(a) can be decided
by Eq. (5) and by considering \(M_0(p_{16}) = 3, M_0(p_{17}) = 2, M_0(p_{18}) = 1,\) and \(\mu_0 = 1\). It
is shown in Figure 5(d) in which \(v_3\) is redundant with respect to \(v_1\) and \(v_2\) and can be
removed.

For the net structure shown in Figure 5(a) at any admissible initial marking, it can be
verified that Eq. (5) leads to an optimal liveness-enforcing controlled system.

Figure 4: A real-world part cleaning system.
Figure 5: (a) An M-net $(N, M_0)$, (b) $(N^m, M^m_0)$, (c) controlled system $(N^{mc}, M^{mc}_0)$, and (d) controlled system for $(N, M_0)$. 
8 Experimental Studies

This section considers two typical examples that are widely investigated in the literature, indicating that the proposed deadlock prevention policy is nearly optimal.

Figure 6(a) shows an FMS that consists of two robots R1 and R2 and three machine tools M1, M2, and M3. Each robot can handle one part at a time and each machine tool can process one part at a time. Parts enter and leave the system through two loading and unloading buffers I1/O1 and I2/O2. The system can produce three part types P1, P2, and P3 whose processing routings are shown in 6(b).

![Figure 6(a)](image)

Figure 6: (a) The layout of an FMS, (b) part routings of the FMS

The Petri net model of the FMS is shown in Figure 7(a). It is an M-net, where \( p_1 \) and \( p_{10} \) are idle places, \( p_{11}, p_{12}, \ldots \), and \( p_{15} \) are resource places, and the others are operation places. The net \((N^{mc}, M_0^{mc})\) shown in Figure 7(b) is the controlled system for the plant net with the minimal initial marking. Seven uncontrolled siphons in \((N^{mc}, M_0^{mc})\) are \( S_1 = \{p_6, p_8, p_{13}, p_{14}\} \), \( S_2 = \{p_5, p_8, v_2, v_3\} \), \( S_3 = \{p_5, p_7, p_{12}, p_{13}\} \), \( S_4 = \{p_6, p_7, p_{12}, p_{14}, v_2, v_3\} \), \( S_5 = \{p_6, p_8, p_{14}, v_2, v_3\} \), \( S_6 = \{p_6, p_7, p_{12}, p_{13}, p_{14}\} \), and \( S_7 = \{p_5, p_7, p_{12}, v_2, v_3\} \). Let \( \theta_i = M_0^{mc}(v_i), \mu_j = M_0^{mc}(p_j) \), and \( \mu_0 = 1 \), where \( i = 1, 2, 3 \) and \( j = 12, 13, 14 \). The matrix form of the controllability constraints is as follows:

![Figure 7(a)](image)

Figure 7: (a) An M-net \((N, M_0)\), and (b) controlled system \((N^{mc}, M_0^{mc})\).
Table 2 shows the permissive behavior of controlled systems at different initial markings, where the initial markings of the monitors are decided by Eq. (6). In this table, $B_p$ is the number of reachable states of a plant model $(N, M_0)$, $B_L$ represents the number of states that an optimal liveness-enforcing controlled system for $(N, M_0)$ has, $B_m$ indicates the number of states of controlled system $(N^{mc}, M_0^{mc})$, and $B_m/B_L$ implies the optimality degree of the controlled systems.

<table>
<thead>
<tr>
<th>$p_1$, $p_{10}$, $p_{11}$, $p_{12}$, $p_{13}$, $p_{14}$, $p_{15}$</th>
<th>$v_1$, $v_2$, $v_3$</th>
<th>$B_p$</th>
<th>$B_L$</th>
<th>$B_m$</th>
<th>$B_m/B_L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1</td>
<td>1, 1, 1</td>
<td>73</td>
<td>54</td>
<td>54</td>
<td>100%</td>
</tr>
<tr>
<td>4, 4, 2, 2, 2, 2, 2</td>
<td>3, 3, 2</td>
<td>1093</td>
<td>1047</td>
<td>941</td>
<td>89.88%</td>
</tr>
<tr>
<td>5, 5, 3, 3, 3, 3, 3, 3</td>
<td>5, 5, 3</td>
<td>5767</td>
<td>5705</td>
<td>5151</td>
<td>90.29%</td>
</tr>
<tr>
<td>6, 6, 4, 4, 4, 4, 4, 4</td>
<td>7, 7, 4</td>
<td>20324</td>
<td>20263</td>
<td>18517</td>
<td>91.38%</td>
</tr>
<tr>
<td>7, 7, 5, 5, 5, 5, 5, 5</td>
<td>9, 9, 5</td>
<td>57450</td>
<td>57390</td>
<td>52995</td>
<td>92.25%</td>
</tr>
<tr>
<td>8, 8, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6</td>
<td>11, 11, 6</td>
<td>140703</td>
<td>140643</td>
<td>131000</td>
<td>93.14%</td>
</tr>
<tr>
<td>9, 9, 7, 7, 7, 7, 7</td>
<td>13, 13, 7</td>
<td>310783</td>
<td>310723</td>
<td>291363</td>
<td>93.77%</td>
</tr>
<tr>
<td>10, 10, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8</td>
<td>15, 15, 8</td>
<td>634173</td>
<td>634113</td>
<td>597853</td>
<td>94.28%</td>
</tr>
<tr>
<td>11, 11, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9</td>
<td>17, 17, 9</td>
<td>1214679</td>
<td>1214619</td>
<td>1150189</td>
<td>94.70%</td>
</tr>
<tr>
<td>12, 12, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10</td>
<td>19, 19, 10</td>
<td>2208445</td>
<td>2208385</td>
<td>2098887</td>
<td>95.04%</td>
</tr>
</tbody>
</table>

The second flexible manufacturing cell is shown in Figure 8(a). It has two robots R1 and R2, each of which can hold one product at a time. The cell also contains four machine tools M1–M4, and each of them can hold one part. Parts enter the cell through two automatic loading buffers I1 and I2, and leave the cell through two unloading buffers O1 and O2. The robots deal with the movements of parts. Two part types P1 and P2 are produced. Their respective production routings are shown in Figure 8(b).

Figure 8: (a) The layout of an FMS, (b) The routings of part types P1 and P2.

Figure 9(a) shows the Petri net model of the FMS, which is an M-net with $P_0 =$
\[ \{p_1, p_8\}, P_R = \{p_{15}, p_{16}, p_{17}, p_{18}, p_{19}\}, \] and the others are operation places. It can be easily verified that the current initial marking is minimal. The controlled system of such a plant model is depicted in Figure 9(b), which can be obtained by the theory of regions [46], [29]. In Figure 9(b), there are 54 uncontrolled minimal siphons. For economy of space, the liveness constraint equation for the system is not shown. Table 3 shows the performance of the controlled systems at different initial markings. From this table, we conclude that the proposed method for this example is near-optimal.

**Figure 9:** (a) Petri net model for an FMS, and (b) the structure of the controlled system.

**Table 3:** Behavioral permissiveness of the proposed deadlock prevention policy

<table>
<thead>
<tr>
<th></th>
<th>(p_1, p_8, p_{14} - p_{19})</th>
<th>(p_{20} - p_{25})</th>
<th>(B_p)</th>
<th>(B_L)</th>
<th>(B_m)</th>
<th>(B_m/B_L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3, 3, 1, 3, 1, 2, 3, 2</td>
<td>5, 4, 5, 8, 7, 7</td>
<td>2946</td>
<td>2945</td>
<td>2842</td>
<td>96.50%</td>
</tr>
<tr>
<td>2</td>
<td>3, 4, 1, 4, 1, 1, 3, 3</td>
<td>4, 4, 7, 9, 8, 9</td>
<td>5235</td>
<td>5233</td>
<td>4730</td>
<td>90.35%</td>
</tr>
<tr>
<td>3</td>
<td>4, 4, 1, 5, 1, 2, 4</td>
<td>6, 7, 6, 14, 14, 14</td>
<td>6877</td>
<td>6868</td>
<td>6861</td>
<td>99.90%</td>
</tr>
<tr>
<td>4</td>
<td>4, 5, 1, 5, 2, 2, 4, 5</td>
<td>5, 9, 9, 5, 10, 18</td>
<td>31759</td>
<td>31578</td>
<td>29129</td>
<td>92.24%</td>
</tr>
<tr>
<td>5</td>
<td>5, 5, 1, 6, 1, 3, 5</td>
<td>8, 9, 7, 10, 12, 17</td>
<td>28243</td>
<td>28233</td>
<td>28177</td>
<td>99.80%</td>
</tr>
<tr>
<td>6</td>
<td>5, 6, 1, 7, 1, 6, 3, 1</td>
<td>9, 10, 8, 13, 13, 13</td>
<td>24448</td>
<td>24438</td>
<td>24384</td>
<td>99.78%</td>
</tr>
<tr>
<td>7</td>
<td>6, 6, 3, 7, 3, 3, 5, 7</td>
<td>8, 8, 10, 17, 13, 12</td>
<td>298725</td>
<td>298724</td>
<td>290187</td>
<td>97.14%</td>
</tr>
</tbody>
</table>

**9 Discussion**

For a generalized Petri net model, the proposed deadlock prevention policy does not provide a controlled system as permissive as the case in which \(N^{mc}\) is ordinary. From the theoretical point of view, this is due to the conservativeness of Proposition 1, which can be
verified by the following example. Figure 10(a) shows a generalized net and Figure 10(b) depicts its controlled system derived from Propositions 1 and 2. For the sake of simplicity, \( r_1 \) (or \( r_2; v \)) is used to denote a place as well as the number of tokens in it.

![Diagram](image)

Figure 10: (a) an example net, and (b) its controlled system.

The only siphon in Figure 10(a) is \( S = \{ p_2, p_4, r_1, r_2 \} \). Its controllability is ensured by the addition of monitor \( v \), as shown in Figure 10(b) [6]. By Proposition 2, \( S \) is max-controlled if \( v < r_1 + r_2 - \sum_{p \in S} I_S(p)(\max_i - 1) \), where \( I_S = p_2 + p_4 + r_1 + r_2 - v - 9p_3 \). When \( r_1 = r_2 = 10 \), we have \( v \leq 10 \). However, \( v \) can be obtained by solving the following LPP:

\[
\begin{align*}
\max & \quad v = x + y \\
\text{s.t.} & \quad x \leq r_1 \\
& \quad (1/W(v, t_4))y \leq r_2 \\
& \quad (r_1 + r_2) - x - (1/W(v, t_4))y \geq \sum_{p \in S} I_S(p)(\max_i - 1)
\end{align*}
\]

This problem has an optimal solution \( v = 100 \) with \( x = 0 \) and \( y = 100 \). It can be verified that the net shown in Figure 10(b) at initial marking \( M_0 = 10r_1 + 10r_2 + 100v \) is live. Compared with \( v = 10 \) resulting from Propositions 1 and 2, this novel siphon control method achieves much better control effects. For the net shown in Figure 2(b), when \( M_0(p_9) = M_0(p_{10}) = M_0(p_{11}) = 6 \), we have \( M_0^{mc}(v_1) = 5, M_0^{mc}(v_2) = 11 \), and \( M_0^{mc}(v_3) = 14 \) due to Eq. (4). However, by the novel siphon control approach, we can have \( M_0^{mc}(v_1) = 16, M_0^{mc}(v_2) = 11, \) and \( M_0^{mc}(v_3) = 16 \). The net in Figure 2(a) at \( M_0(p_9) = M_0(p_{10}) = M_0(p_{11}) = 6 \) has 12,495 reachable states. An optimal live controlled system should have 12,415 states. The controlled system at \( M_0^{mc}(v_1) = 16, M_0^{mc}(v_2) = 11, \) and \( M_0^{mc}(v_3) = 16 \) has 12,374 reachable states. The optimality ratio is 12,374/12,425=99.67%. The proposed deadlock prevention policy can also lead to a nearly optimal controlled system for a generalized plant net model if the concept of max’-controlled or max”-controlled siphons [7], [32] is employed to derive siphon controllability condition.

## 10 Concluding Remarks

Deadlocks are a threat to the safety and high productivity of a highly automated flexible manufacturing system. Deadlock prevention is a well defined problem in resource allocation systems. The deadlock prevention approaches in the literature are usually developed by either structural or state space analysis. The policies based on structural analysis such as siphons cannot in general lead to optimal supervisors [20], [30], while the policies combining state space analysis can lead to optimal or at least suboptimal supervisors [37], [38],
[46], [47], [48]. For example, the theory of regions is a technique that can find an optimal liveness-enforcing supervisor in general cases when it exists. However, its computation is notoriously expensive since the complete state enumeration is necessary and the number of LPP to be solved is exponential with respect to the structural size of a plant model and its initial marking.

The proposed deadlock prevention policy aims to felicitously trade off behavioral optimality for computational tractability. To achieve this, we first derive, by using the theory of regions, a supervisor for the plant model with the minimal initial marking. The controllability of siphons is expressed as a set of inequality constraints with respect to the markings of resource places and monitors. For a fixed net structure with different initial markings, the theory of regions is used and siphon enumeration is computed once only. Then, the supervisor can be decided by satisfying the constraints of siphon controllability via adjusting the initial marking of monitors only. Once an optimally controlled system’s structure is found, a nearly optimal system can be obtained without using the theory of regions. Experimental studies show that the proposed method is nearly optimal. Future efforts will be guided to a near-optimal supervisor with low computational overhead and an efficient method to find a deadlock-liable minimal initial marking for an M-net.

Acknowledgments

This work was supported in part by the Natural Science Foundation of China under Grant No 60773001 and 61074035, the Fundamental Research Funds for the Central Universities under Grant No. 72103326, the National Research Foundation for the Doctoral Program of Higher Education, the Ministry of Education, P. R. China, under Grant No. 20090203110009, the Research Fellowship for International Young Scientists, National Natural Science Foundation of China, under Grant No. 61050110145, the Chang Jiang Scholar Program, Ministry of Education, P. R. China, “863” High-tech Research and Development Program of China under Grant No 2008AA04Z109, and Alexander von Humboldt Foundation.

References


Appendix

Proposition A1 There exists an M-net that is not a G-system.

We claim that the net shown in Figure A1 is an M-net that is more general than a G-system. In $(N, M_0)$, we have $P^0 = \{p_1\}$, $P_A = \{p_2, p_3, p_4, p_5, p_6\}$, $P_R = \{M, R\}$, and $T = \{t_1, t_2, t_3, t_4, t_{rf}, t_{rr}, t_{mf}, t_{mr}\}$, where M models a machine tool and R models a robot that uploads and downloads the machine tool. Place $p_1$ models the availability of
raw materials, and $p_2$ and $p_4$ are used to model uploading and downloading operations of the robot for the machine tool, respectively. Place $p_3$ indicates that M is performing an operation on a raw part. Note that there is only one process. It is clear that the net satisfies items 1), 2), and 3) of the definition of M-nets (Definition 2).

![Figure A1: An M-net ($N, M_0$).](image)

Let $I_p$ be a minimal P-invariant of the net associated with place $p$. From its structural analysis, we have $I_{p1} = I_{p2} = I_{p3} = I_{p4} = I_{p5} = p_1 + p_2 + p_3 + p_4 + p_5$, $I_{p6} = R + p_2 + p_4 + p_6$, $I_M = M + p_3 + p_5$, and $I_R = R + p_2 + p_4 + p_6$. As a result, the net satisfies items 4) and 5) of Definition 2.

It is verified that $\sigma = t_1t_2tmft_mrtrftrrt_3t_4$ is a feasible firing sequence such that $M_0[\sigma]M_0$ is true. Thus, $(N, M_0)$ is quasi-live. Note that each place is covered by a positive P-semiflow. The net system is conservative. The bound of place $p_1$ is four and that of other places is one. The net system is bounded. That is to say, $(N, M_0)$ in Figure A1 is quasi-live, bounded, and conservative. Item 6) of Definition 2 is satisfied.

The net system resulting from removing resources in $(N, M_0)$ is shown in Figure A2. It is obviously live, bounded, and reversible. Consequently, net system $(N, M_0)$ satisfies item 7) of Definition 2. Also, items 8) and 9) are true for $(N, M_0)$.

![Figure A2: An M-net ($N', M'_0$).](image)

Note that $p_1$ is an idle place with $M_0(p_1) = 4$. $p_2$, $p_3$, $p_4$, $p_5$, and $p_6$ are operation places that are unmarked at $M_0$. Places M and R are resources that initially marked.
Hence, \((N, M_0)\) satisfies item 10) of Definition 2.

In the net shown in Figure A1, the unique minimal siphon is
\[ S = \{p_4, p_6, M, R\} \]
that contains two resource places and no idle place. If \(S\) is always marked at any reachable marking, the net system is live. Therefore it satisfies items 11) and 12) of Definition 2.

The net system shown in Figure A3 is a controlled system resulting from adding a monitor to the original net in Figure A1. The monitor is used to make the unique siphon
\[ S = \{p_4, p_6, M, R\} \]
controlled. It is verified that \((N^c, M^c_0)\) is live. That is to say, liveness is enforced by the addition of a monitor. Item 13) of Definition 2 is hence satisfied, which also implies the truth of item 14).

![Figure A3: An M-net \((N^c, M^c_0)\).](image)

From the net system shown in Figure A3, Monitor \(v\) is associated with a minimal P-semiflow
\[ I_v = v + p_2 + p_3 + p_5. \]
Note that \(I_v(v) = 1\) and \(\forall p \in |I_v| \setminus \{v\}, p \in P_A.\) Thus item 15) of Definition 2 is true. We conclude that net system \((N, M_0)\) in Figure A1 is an M-net.

We claim that net system \((N, M_0)\) in Figure A1 is not a G-system. First, let us present the definitions and results with respect to a G-system, which are primarily from [59].

**Definition A1** A G-task is a Petri net \(GT = (N, M_0, M_F)\), where
- \(N = (P, T, F, W)\) is a circuit-free Petri net with two special places \(i\) and \(o.\) The former is called a source place with \(*i = 0\) and the latter is called a sink place with \(*o = 0.\)
- The augmented net \(N^*\) obtained from \(N\) by adding a transition \(t^*\) with \(*t^* = \{o\},\)
  \([t^*] = \{i\},\) and \(W(i, t^*) = W(o, t^*) = n\) is strongly connected.
- \(M_0 = n. i\) and \(M_F = n.o\)
- \((N, n.i)\) is quasi-live.

**Definition A2** A G-task \(GT\) is said to be sound if
(i) \(\forall M \in R(N, n.i),\) \(n.o \in R(N, M);\)
(ii) \(\forall M \in R(N, n.i),\) \(M(o) \geq n\) means \(M = n.o.

**Proposition A2** Let \(GT\) be a G-task. \(GT\) is sound iff \((N^*, n.i)\) is live and bounded.

**Definition A3** A G-task \(GT\) is well-formed if \(\exists M_0 = n.i\) such that \(GT\) is sound.
Theorem A1 Let $GT$ be a $G$-task. $GT$ is well-formed if $\exists M_0 = n.i$ such that $(N^*, n.i)$ is bounded and satisfies the max-cs property.

Definition A4 A $G$-task system with resources $GTR$ is a Petri net $(NR, MR_0, MR_F)$, where

- $NR = (P \cup P_R, T, F \cup F_R, W \cup W_R)$ is a net structure. $P \setminus \{i, o\}$ is called the set of operation places.
- $P_R \neq \emptyset$ is a set of resources with $P \cap P_R = \emptyset$.
- $F_R \subseteq (P_R \times T) \cup (T \times P_R)$ is the flow relation of resources.
- $\forall f \in F_R$, $W_R(f) \geq 1$.
- $\forall r \in P_R$, $r$ is associated with a $P$-semiflow $f_r$ such that $\|f_r\| \cap P_R = \{r\}$.
- $MR_0 = n.i + \sum_{j=1}^{P_R} k_j r_j$ and $MR_F = n.o + \sum_{j=1}^{P_R} k_j r_j$ ($\forall j \in \{1, 2, \ldots, |P_R|\}, k_j \geq 1$).
- Subsystem $G = (N, M_0, M_F)$ with $N = (P, T, F, W)$ is a $G$-task, where $M_0$ and $M_F$ are the restrictions of $MR_0$ and $MR_F$ to $P$, respectively.

Definition A5 A $G$-system $GS$ is recursively defined.

- A $GTR$ is a $G$-system
- Let $GS_i = (NS_i, MS_{0i}, MS_{Fi})$ ($i = 1, 2$) be two $G$-systems such that $P_1 \cap P_2 = T_1 \cap T_2 = \emptyset$. We denote the set of shared resources by $P_{R1}P_{R2} = P_{R1} \cap P_{R2}$. The system $GS = (NS, MS_0, MS_F) = GS_1 \circ GS_2$ with $NS = (P \cup P_R, T, F \cup F_R, W \cup W_R)$, resulting from the fusion of systems $GS_1$ and $GS_2$ over the set $P_{R1}P_{R2}$, is a $G$-system if (1) $\forall r \in P_{R1} \setminus P_{R1}P_{R2}$, $MS_0(r) = MS_{01}(r)$; (2) $\forall r \in P_{R2} \setminus P_{R1}P_{R2}$, $MS_0(r) = MS_{02}(r)$; and (3) $\forall r \in P_{R1}P_{R2}$, $MS_0(r) = \max\{MS_{01}(r), MS_{02}(r)\}$.

Definition A6 A $G$-system $GS$ is well-formed if there exists an initial marking $MS_0$ such that $\forall M \in R(NS, MS_0), MS_F \in R(NS, M)$.

Proposition A3 Let $S$ be a minimal siphon in the augmented net $NS^*$. There exists an initial marking $M_0$ at which $S$ is max-controlled.

Let $g_S = \sum f(r), r \in S \cap P_R; Out(S) = ||g_S|| \setminus S; h_S = \sum f(p), p \in Out(S); \lambda_S = \max\{g(p)|p \in Out(S) \cap ||h_S||\}; and z_S = g_S - \lambda_S h_S$. Siphon $S$ is max-controlled

$$z_S^T M_0 > \sum_{p \in S} z_S(p)(max_{r^*} - 1)$$

Therefore, there exists an initial marking at which siphon $S$ is controlled. Also its controllability can be ensured by a monitor [59].

Theorem A2 Let $GS$ be a $G$-system. $GS$ is well-formed iff there exists an initial marking $MS_0$ at which $(N^*, MS_0)$ is bounded and satisfies max-cs property.
From Definitions A1, A4, and A5, a G-system can be composed of a number of G-task systems with resources. While, a G-task system with resources is obtained by adding resources to a G-task. Note that a G-task is circuit-free, i.e., it is inadmissible to have loops in a G-task. However, a loop in an M-net is allowed. For example, \( p_3t_{mf}p_5t_{mr}p_3 \) is a loop in Figure A1, which can model the fault occurrence and recovery of machine M. Specifically, \( t_{mf} \) models that a fault occurs when machine M is performing some operation on a raw material and transition \( t_{mr} \) means that a fault is recovered.

Next we show that a G-system is an M-net. Shown in Figure A4(a) is a well-formed G-system and its augmented system by adding a transition \( t^* \) is depicted in Figure A4(b). The net system shown in Figure A4(c) is obtained by merging \( o, i, \) and \( t^* \) in \( (N^*, MS_0) \) into an idle place \( p^0 \). Let \( u \) and \( v \) be two transition sequences consisting transitions in \( T \). Let \( \#(\sigma, t) \) denote the number of occurrences of transition \( t \) in \( \sigma \), where \( \sigma \) is a transition sequence. The languages of a net system \( (N, M_0) \) is defined as \( L = \{ \sigma | \sigma \) is a feasible transition sequence in \( (N, M_0) \} \).

![Figure A4: (a) A G-system \((N, M_0)\), (b) augmented system \((N^*, MS_0)\), and (c) an equivalent net \((N, M_0)\) of \((N^*, MS_0)\).](image)

**Definition A7** Let \((N_1, M_2)\) and \((N_2, M_2)\) be two net systems. They are said to be behaviorally equivalent if they have the same languages, i.e., \( L_1 = L_2 \).

Let \( T \) be the set of transitions in \( N^* \) and \( \overline{N} \), and \( \sigma = u_1t^*...u_2t^*...u_{n-1}t^*u_n \) be a feasible transition sequence in \( (N^*, MS_0) \), where \( \forall i \in \{1, 2, ..., n\} \), \( \#(u_i, t^*) = 0 \). Let \( \sigma|_T = u_1u_2...u_n \) be the transition sequence obtained by removing \( t^* \) from \( \sigma \). Let \( v \) be a feasible transition sequence in \( (\overline{N}, \overline{M}_0) \) and \( v^* \) be a transition sequence in \( (N^*, MS_0) \) such that \( v = v^*|_T \), where \( \#(v^*, t^*) \geq 0 \).

**Definition A8** \((N^*, MS_0)\) is said to be behaviorally equivalent to \((\overline{N}, \overline{M}_0)\) with respect to transition \( t^* \) if

1. \( \forall \sigma \in L(N^*, MS_0), \sigma|_T \in L(\overline{N}, \overline{M}_0); \)
2. \( \forall v \in L(\overline{N}, \overline{M}_0), \) there exists \( v^* \in L(N^*, MS_0) \) such that \( v = v^*|_T \) is true, where \( \#(v^*, t^*) \geq 0 \).
Proposition A4 \((N, M_0)\) is behaviorally equivalent to \((N^*, MS_0)\) with respect to transition \(t^*\).

**Proof:** It is easy for us to conclude that \(\sigma|_{\mathcal{F}}\) is feasible in \((N, M_0)\) if \(\sigma\) is feasible in \((N^*, MS_0)\). If \(v\) is feasible in \((N, M_0)\), then there exists a transition sequence \(v^*\) with \(v = v^*|_{\mathcal{F}}\) and \(#(v^*, t^*) \geq 0\) that is feasible in \((N^*, MS_0)\). By Definition A8, the proposition holds.

Proposition A5 A well-formed G-system \((N, M_0)\) is an M-net.

**Proof:** From Definitions A1 and A4, items 1), 2), 3), 4), and 5) in the definition of an M-net (Definition 2 in the mainbody of the manuscript) are satisfied.

Now we show that \((N, M_0)\) satisfies item 6) in Definition 2. By Theorem A2, \((N, M_0)\) is bounded. By Definition A6, it is quasi-live. The conservativeness of \((N, M_0)\) can be ensured by the fact that each place is associated with a P-semiflow. Item 7) is satisfied due to Definitions A1, A2, and A3, and Theorem A1. Items 8), 9), and 10) are true due to Definition A5.

Deadlocks in an automated manufacturing system result from the existence of shared resources. That is to say, deadlocks do not exist if there is no shared resources in a system. Siphons are a structural object in a Petri net model of an automated manufacturing system, whose uncontrollability is the root of deadlock occurrences. In fact, if all siphons are controlled or max-controlled in a Petri net, it is deadlock-free. Without an external agent, an uncontrolled siphon can lead to deadlock transitions, making loss of system liveness. From a technical point of view, deadlocks are caused by shared resources and from a Petri net point of view, deadlocks are caused by uncontrolled siphons. Since a siphons is a set of places that can represent operations and resources, we conclude that item 11) holds. In the proof of Proposition 4.1 in [59], it is also shown that if \(S\) is an uncontrolled siphon in a G-system, \(S \cap P_R \neq \emptyset\) is true.

From Theorem A2, items 12) and 14) are satisfied. Siphon control can be considered as a GMEC (Generalized Mutual Exclusive Constraints) problem. Thus, it can be implemented by a monitor. Item 13) is true in a well-formed G-system. Monitors are used to limit the number resource units to be used by some operation places. From the structural theory (P-invariants) of Petri nets, item 15) is true. In summary, a well-formed G-system satisfies the definition of M-net.

Theorem A3 An M-net is general than a well-formed G-system.

**Proof:** It follows from Propositions A1 and A5.