A Polynomial-Complexity Deadlock Avoidance Policy for Sequential Resource Allocation Systems with Multiple Resource Acquisitions and Flexible Routings

Jonghun Park and Spyros A. Reveliotis

School of Industrial & Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332

Abstract

The need for effective and efficient deadlock avoidance policies (DAP's) is ever increasing due to the higher demand for system automation. This paper considers the deadlock avoidance problem for the class of Conjunctive/Disjunctive (sequential) resource allocation systems (C/D-RAS), in which multiple resource acquisitions and flexible routings are allowed. A new siphon-based characterization of deadlocks arising in C/D-RAS is developed, and subsequently, this characterization facilitates the development of a polynomial complexity deadlock avoidance policy for the considered RAS class. The developed policy can be perceived as a generalization of RUN DAP, originally developed for sequential RAS with unit resource allocations and no routing flexibility. The proposed approach is demonstrated by an example.

1 Introduction

Deadlock avoidance in sequential resource allocation systems (S-RAS) is a well-defined problem in Discrete Event System literature. From a theoretical standpoint, the problem arises in any resource sharing system where a set of concurrently executing sequential processes can get permanently blocked – i.e., *deadlocked* – due to the fact that each process in that set is allocated and holds resource(s) requested by some other process in the set for its further advancement. From a practical standpoint, the problem is experienced in the operational control of many contemporary technological systems, including the material flow control of flexibly automated production systems, the traffic management of unmanned discrete material handling systems like automated and/or rail guided vehicle systems, and the traffic control of railway and monorail transport systems. In all of these systems, the resolution of a developed deadlock can imply a major disruption of the normal system operation, while during its occurrence, the utilization of the involved resources is driven to zero. It is, therefore, desirable that the controller supervising the real-time operation of these systems incorporates a control function that foresees and effectively prevents the occurrence of the problematic deadlock states, by appropriately restricting the allocation of the system resources to the various requesting processes.

Ideally, due to performance considerations, the aforementioned deadlock avoidance function should be carried out in the *least* restrictive way. However, it has been formally established [1] that, in the general case, implementation of the *optimal* – i.e., least restrictive – deadlock avoidance policy (DAP) is an *NP-Hard* [7] problem for the considered class of S-RAS. Hence, in the light of this result, one should aim at the development of suboptimal, but computationally efficient – i.e., polynomial complexity – deadlock avoidance policies, which also maintain a significant level of operational flexibility.[1] It has also been found that the complexity of the synthesis of effective and efficient policies for the considered class of resource allocation systems strongly depends on the underlying system structure and the particular assumptions regarding the admissible resource allocation requests. A taxonomy that attempts to manage the problem complexity by classifying S-RAS on the basis of the resource allocation request structure is presented in [14]. According to that taxonomy, S-RAS are classified in four major classes: (i) *Single-Unit (SU)* RAS, in which every process stage requires only a *single* unit of a *single* resource for its successful execution, (ii) *Single-Type (ST)* RAS, in which every process stage requires an *arbitrary* number of units of a *single* resource for its successful execution, (iii) *Conjunctive (C)* RAS, in which every process stage requires an *arbitrary* number of units from an *arbitrary* resource (sub-)set for its successful execution, and (iv) *Conjunctive / Disjunctive (C/D)* RAS, in which every process stage poses a finite number of *alternative* conjunctive-type resource requests.

Past research has addressed successfully the aforestated deadlock avoidance problem primarily in the context of SU-RAS. Some indicative results can be found in [2, 13, 10, 6]. Of particular interest to this work are the developments presented in [5, 4, 11], which also leverage recently obtained results in the area of Petri Net (PN) structural analysis. Specifically, the results presented in [5, 4, 11] have established that for the case of single-unit resource allocation, the occur-

[1] Another interesting research line in the area of Deadlock Avoidance for S-RAS has sought to identify special system structure that admits polynomial complexity deadlock avoidance; c.f. [9] for an overview of these results.

rence of deadlock can be structurally explained in the PN formalism through the concept of an *empty siphon*. Recently, the work of [15] has also shown that the empty siphon can be the deadlock interpreting mechanism, even in the case of C/D-RAS, in which, however, resources are acquired one unit at a time. An attempt to generalize the notion of empty siphon towards the interpretation of deadlock occurring in the general C/D-RAS has been presented in [3], but that work fails to provide a polynomial-complexity solution to the problem of deadlock avoidance. Motivated by these remarks, this work: (i) characterizes deadlock arising in the general C/D-RAS through a new pertinent siphon construct, and (ii) employs this characterization towards the development of a polynomial complexity deadlock avoidance policy for the considered S-RAS class. The developed policy can be perceived as a generalization of the RUN DAP, originally developed in [13] for SU-RAS. The paper is organized as follows: Section 2 first provides a formal (set-theoretic) characterization of the C/D-RAS structure and the underlying deadlock avoidance problem, and subsequently it proceeds to their further modeling in the PN formalism. Some important properties related to the liveness of the resulting PN class is presented in Section 3. Section 4 presents a polynomial-complexity DAP for C/D-RAS and establishes its correctness. An example is given to demonstrate the implementation of the proposed DAP. Finally Section 5 concludes the paper and suggests further research directions. A more extensive discussion of the paper results and some additional developments can be found in [12].

2 C/D-RAS and their Petri Net-based Modeling

C/D-RAS and polynomial-complexity deadlock avoidance C/D-RAS is formally defined by a set of *resource types* $\mathcal{R} = \{R_i, i = 1, \ldots, m\}$, and a set of *job types* $\mathcal{J} = \{J_j, j = 1, \ldots, n\}$. Every resource type R_i is further characterized by its *capacity* $C_i \in Z^+$, where Z^+ is the set of positive integers. Job type J_j is defined by a sequence of stages $< J_{jk}, k = 1, \ldots, \lambda_j >$, where λ_j represents the number of job stages. Each *job stage* J_{jk} is further characterized by a set of m-dimensional vectors $A_{jk} = \{J_{jkl} \mid l = 1, \ldots, \alpha_{jk}\}$, of cardinality α_{jk}, with component $J_{jkl}[i], i = 1, \ldots, m$, indicating how many units of resource R_i are required for the execution of the l-th processing *alternative* of the k-th stage of job type J_j.[2] The entire set of resource allocation schemes according to which a job instance of type J_j can be executed, is given by the cartesian product $\mathcal{JR} = A_{j1} \times \ldots \times A_{j\lambda_j}$. Hence the number of possible *routes* for a certain job type J_j is $\alpha_{j1} \cdots \alpha_{j\lambda_j}$, which is of exponential size with respect to λ_j. The C/D-RAS *state* is uniquely determined by the currently executed processing alternatives by all active job instances. Since each resource possesses finite capacity, the set of distinct RAS states is finite. Hence, the state space of the RAS can be modeled by an finite state automaton (FSA) [8]. Naturally, the initial and final states in this automaton correspond to the RAS idle and empty state, $\mathbf{q}_0$, and thus, in order to avoid deadlock in the least restrictive way, the system operation must be confined to the maximal communicating class of the underlying state transition diagram (STD) that contains the FSA initial state, $\mathbf{q}_0$. Let us denote this class by Q_s. Then every state $\mathbf{q} \in Q_s$ is characterized as *safe* while states $\mathbf{q}$ belonging to the complement of Q_s, denoted by Q_u, are characterized as *unsafe*. It has been shown in the literature that the decision problem $\mathbf{q} \in Q_u$ is NP-complete [1]. Due to this result, it has been proposed [13] that, in the general case, the development of a *polynomial-complexity* DAP should be based on the identification of a property $\mathcal{H}(\mathbf{q}), \mathbf{q} \in Q$, such that: (i) $\mathcal{H}(\mathbf{q}) \Rightarrow \mathbf{q} \notin Q_u, \forall \mathbf{q} \in Q$, and (ii) $\mathcal{H}(\cdot)$ is polynomially computable on any given system state. Then, by allowing only transitions to states satisfying $\mathcal{H}(\cdot)$, it can be ensured that the visited states will also satisfy the condition $\mathbf{q} \notin Q_u$, and therefore the system will never enter the forbidden region Q_u. In addition, the subspace $Q(\mathcal{H})$ admitted by the resulting policy must be a single communicating class containing $\mathbf{q}_0$ in order to be *induced* deadlock-free.

[2] It should be noted that the conjunctive resource requirements must be satisfied *simultaneously* for the initiation of a job stage alternative. Nevertheless, the situation in which multiple resources are allowed to be allocated one by one until the requirements are satisfied, can be modeled by defining a series of job stages.

Petri Net preliminaries A marked Petri Net (PN) is defined by a quadruple $N = (P, T, W, M_0)$, where P is the set of *places*, T is the set of *transitions*, $W : (P \times T) \cup (T \times P) \rightarrow Z^+$ is the *flow relation*, and $M_0 : P \rightarrow Z^+$ is the net *initial marking*, assigning to each place $p \in P$, $M_0(p)$ *tokens*. The set of input (resp., output) transitions of a place p is denoted by ${}^\bullet p$ (resp., $p^\bullet$). Similarly, the set of input (resp., output) places of a transition t is denoted by ${}^\bullet t$ (resp., $t^\bullet$). This notation is also generalized to any set of places or transitions, X, e.g. ${}^\bullet X = \bigcup_{x \in X} {}^\bullet x$. The ordered set $X =< x_1 \ldots x_n > \subseteq P \cup T$ is a *path*, iff $x_{i+1} \in x_i^\bullet, i = 1, \ldots, n-1$. Furthermore, a path X is characterized as a *circuit* iff $x_1 \equiv x_n$. Given a marking M, a transition t is *enabled* iff $\forall p \in {}^\bullet t$, $M(p) \geq W(p,t)$, and this is denoted by $M[t\rangle$. $t \in T$ is said to be *disabled* by $p \in {}^\bullet t$ at M iff $M(p) < W(p,t)$. Furthermore, a place $p \in P$ for which $\exists t \in p^\bullet$ s.t. $M(p) < W(p,t)$ is said to be a *disabling* place at M. Firing an enabled transition t results in a new marking M', which is obtained by removing $W(p,t)$ tokens from each place $p \in {}^\bullet t$, and placing $W(t,p')$ tokens in each place $p' \in t^\bullet$. The set of markings reachable from M_0 through any firable sequence of transitions is denoted by $R(N, M_0)$. If a marked PN is *pure* (i.e., $\forall (x,y) \in (P \times T) \cup (T \times P)$, $W(x,y) > 0 \Rightarrow W(y,x) = 0$), the flow relation can be represented by the *flow matrix* $\Theta = \Theta^+ - \Theta^-$ where $\Theta^+[p,t] = W(t,p)$ and $\Theta^-[p,t] = W(p,t)$. A *p-semiflow* y is a $|P|$-dimensional vector satisfying $y\Theta = 0$ and $y \geq 0$, and a *t-semiflow* x is a $|T|$-dimensional vector satisfying $\Theta x = 0$ and $x \geq 0$. A p-semiflow y (t-semiflow x, resp.) is said to be *minimal* iff $\nexists$ a p-semiflow y' (t-semiflow x', resp.) such that $\|y'\| \subset \|y\|$ ($\|x'\| \subset \|x\|$, resp.), where $\|y\| = \{p \in P \mid y(p) > 0\}$ ($\|x\| = \{t \in T \mid x(t) > 0\}$, resp.). Given a marked PN $N = (P, T, W, M_0)$, a transition $t \in T$ is *live* iff

$\forall M \in R(N,M_0), \exists M' \in R(N,M)$ s.t. $M'[t\rangle$, and $t \in T$ is *dead* at $M \in R(N,M_0)$ iff $\nexists M' \in R(N,M)$ s.t. $M'[t\rangle$. A marking $M \in R(N,M_0)$ is a (total) *deadlock* iff $\forall t \in T$, t is dead. A marked PN N is *quasi-live* iff $\forall t \in T, \exists M \in R(N,M_0)$ s.t. $M[t\rangle$, it is *weakly live* iff $\forall M \in R(N,M_0), \exists t \in T$ s.t. $M[t\rangle$, and it is *live* iff $\forall t \in T$, t is live. Of particular interest for the liveness analysis of marked PN is a structural element known as *siphon* which is a set of places $S \subseteq P$ such that ${}^\bullet S \subseteq S^\bullet$. A siphon S is minimal iff $\nexists$ a siphon S' s.t. $S' \subset S$. A siphon S is said to be *empty* at marking M iff $M(S) \equiv \sum_{p \in S} M(p) = 0$.

While the notion of empty siphon has been very useful for characterizing the existence of a (partial) deadlock in RAS classes where acquisition of each resource type is limited to one unit at a time, it fails to effectively characterize deadlock in the considered C/D-RAS class. The operational characteristic of multiple resource acquisitions allows a partial deadlock to occur even if there is no empty siphon. Hence, next we propose a new siphon construct that is subsequently shown (i) to effectively characterize partial deadlock in C/D-RAS, and (ii) to facilitate the further development of an efficient deadlock avoidance policy for this class of RAS. This new siphon construct is formally defined as follows:

Definition 1 *Consider a marked PN $N = (P,T,W,M_0)$. A siphon $S \subseteq P$ is said to be* deadly marked *at $M \in R(N,M_0)$ iff $\forall t \in {}^\bullet S$, t is disabled by some $p \in S$.*

The next lemma follows directly from Definition 1.

Lemma 1 *Consider a marked PN $N = (P,T,W,M_0)$, and let $S \subseteq P$ be a deadly marked siphon at $M \in R(N,M_0)$. Then, (i) $\forall t \in {}^\bullet S$, t is a dead transition in M, and (ii) $\forall M' \in R(N,M)$, S is deadly marked.*

PN-based modeling of C/D-RAS The PN class that can effectively model the C/D-RAS class considered in this paper is called S^2PGR^2 nets (System of Sequential Processes with General Resource Requirements), and it is formally defined as follows :

Definition 2 *A* well-marked *S^2PGR^2 net is a marked PN $N = (P,T,W,M_0)$ such that*

1. *$P = P_S \cup P_0 \cup P_R$, where $P_S = \bigcup_{i=1}^n P_{S_i}$ s.t. $P_{S_i} \cap P_{S_j} = \emptyset, \forall i \neq j$, $P_0 = \bigcup_{i=1}^n \{p_{0_i}\}$ s.t. $P_0 \cap P_S = \emptyset$, and $P_R = \{r_1, \ldots, r_m\}$ s.t. $(P_S \cup P_0) \cap P_R = \emptyset$.*
2. *$T = \bigcup_{i=1}^n T_i$.*
3. *$W = W_S \cup W_R$, where $W_S : ((P_S \cup P_0) \times T) \cup (T \times (P_S \cup P_0)) \to \{0,1\}$ s.t. $\forall j \neq i, ((P_{S_j} \cup P_{0_j}) \times T_i) \cup (T_i \times (P_{S_j} \cup P_{0_j})) \to \{0\}$, and $W_R : (P_R \times T) \cup (T \times P_R) \to Z^+$ s.t. $\forall r \in P_R, \sum_{t \in r^\bullet} W_R(r,t) = \sum_{t \in {}^\bullet r} W_R(t,r)$.*
4. *$\forall i$, $i = 1, \ldots, n$, the subset N_i generated by $P_{S_i} \cup \{p_{0_i}\} \cup T_i$ is a strongly connected state machine such that every cycle contains $\{p_{0_i}\}$.*
5. *$\forall r \in P_R$, $\exists$ a unique minimal p-semiflow y_r s.t. $\|y_r\| \cap P_R = \{r\}$, $\|y_r\| \cap P_0 = \emptyset$, $\|y_r\| \cap P_S \neq \emptyset$, and $y_r(r) = 1$. Furthermore, $P_S = \bigcup_{r \in P_R} (\|y_r\| - P_R)$.*
6. *N is pure and strongly connected.*
7. *$\forall p \in P_S$, $M_0(p) = 0$; $\forall r \in P_R$, $M_0(r) \geq max_{p \in \|y_r\|} y_r(p)$; and $\forall p_{0_i} \in P_0$, $M_0(p_{0_i}) \geq 1$.*

We note that the proposed S^2PGR^2 net is a weighted generalization of the ES^3PR net, proposed in [15]. As in [15], let the set of *holders* of a resource place r be defined by $H(r) = \|y_r\| - \{r\}$. Then, given an S^2PGR^2 net representing a C/D-RAS, it follows that $\forall r_i \in P_R$, $\sum_{p \in \{r_i\} \cup H(r_i)} y_{r_i}(p) \cdot M(p) = M_0(r_i) \equiv C_i$.

Finally, the subsequent theoretical developments involve also the notion of the *modified S^2PGR^2* markings, formally defined as follows:

Definition 3 *Given a well-marked S^2PGR^2 net $\mathcal{N} = (P_S \cup P_0 \cup P_R, T, W, M_0)$ and $M \in R(\mathcal{N}, M_0)$, the* modified marking *$\overline{M}$ is defined by*

$$\overline{M}(p) = \begin{cases} M(p) & \text{if } p \notin P_0 \\ 0 & \text{otherwise} \end{cases}$$

Furthermore, the set of all modified markings induced by the reachable markings is defined by $\overline{R(\mathcal{N}, M_0)} = \{\overline{M} \mid M \in R(\mathcal{N}, M_0)\}$

3 Liveness Analysis of the S^2PGR^2 Net

In this section we derive some important properties related to the liveness of the S^2PGR^2 net. First we establish some basic properties that characterize the net behavior under the existence of deadlock. Then, we show that the non-existence of a special type of reachable deadly marked siphons in the space of modified net markings is a necessary and sufficient condition for the S^2PGR^2 net to be live. We start with the following lemma, which is a straightforward implication of Definition 2; its formal proof can be found in [12]:

Lemma 2 *[12] Let $N = (P,T,W,M_0)$ be a well-marked S^2PGR^2 net. Then, N is quasi-live.*

A straightforward implication of Lemma 2 is the following:

Lemma 3 *Let $N = (P,T,W,M_0)$ be a well-marked S^2PGR^2 net. If there exists a dead transition at $M \in R(N,M_0)$, then $M_0 \notin R(N,M)$.*

The next lemma relates the notion of total deadlock in well-marked S^2PGR^2 nets to the notion of deadly marked siphon.

Lemma 4 *Let $N = (P, T, W, M_0)$ be a well-marked S^2PGR^2 net. If marking $M \in R(N, M_0)$ is a total deadlock, then there exists a deadly marked siphon at M.*

Proof : Since all transitions are dead at M, $\forall t \in T$, t is disabled by some $p \in P$. Let S be the set of disabling places. Since $S^\bullet = T$, ${}^\bullet S \subseteq S^\bullet$. Therefore, S is a siphon. It is clear that S is deadly marked from the construction. $\diamond$

Lemma 5 [3] *Let $N = (P, T, W, M_0)$ be a well-marked S^2PGR^2 net. If there exists a dead transition at $M \in R(N, M_0)$, then there exists a marking $M' \in R(\mathcal{N}, M)$ such that the modified marking $\overline{M'}$ contains a deadly marked siphon, S, with at least one disabling resource place $r \in S$.*

Proof: From the fact that there exists a dead transition at M, it is easy to see that, by firing only transitions in $T - P_0^\bullet$, $\mathcal{N}$ can reach $M' \in R(\mathcal{N}, M)$ at which $\forall p \in P_S$ s.t. $M'(p) \geq 1$, $\forall t \in p^\bullet$, t is dead (otherwise, M_0 can be reached from M by firing non-dead transitions, which results in a contradiction, due to Lemma 3). Let $B = \{p \in P_S \mid \overline{M'}(p) \geq 1\}$. $B \neq \emptyset$ (otherwise, $M' = M_0$). Consider the set of places $S = S_R \cup S_{P1} \cup S_{P2}$, where $S_R = \bigcup_{p \in B} \bigcup_{t \in p^\bullet} \{r \in {}^\bullet t \cap P_R \mid \overline{M'}(r) < W(r,t)\}$, $S_{P1} = \{p \in P_S \mid p \in H(S_R) \wedge \overline{M'}(p) = 0\}$, and $S_{P2} = \bigcup_{i: \exists t \in {}^\bullet p, p \in (P_{S_i} \cap S_{P1}) \wedge {}^\bullet t \cap S_{P1} = \emptyset \wedge \forall r \in {}^\bullet t \cap S_R, \overline{M'} \geq W(r,t)\}} P_{S_i} \cup \{p_{0_i}\}$ The definitions of M', B and S_R, together with the fact that $\forall t$, $|{}^\bullet t \cap P_s| \leq 1$, imply that the set of disabling resource places $S_R \neq \emptyset$. Next we show that S is a deadly marked siphon containing at least one disabling resource place.

Case 1: $t \in {}^\bullet S_R$. Let $r \in t^\bullet \cap S_R$. The net purity and the selection of t also imply that ${}^\bullet t \cap P_0 = \emptyset$. Let $\{p\} = {}^\bullet t \cap P_S$. Obviously, $p \in H(r)$. We consider two sub-cases: (i) $\overline{\mathbf{M'}}(\mathbf{p}) = \mathbf{0}$: Then, $p \in S_{P1}$, $t \in S_{P1}^\bullet \subseteq S^\bullet$, and t is disabled by $p \in S$. (ii) $\overline{\mathbf{M'}}(\mathbf{p}) \geq \mathbf{1}$: From the definition of $\overline{M'}$, t is dead. It follows that $\exists r'$ s.t. $\overline{M'}(r') < W(r', t)$. But then, $r' \in S_R$, $t \in S_R^\bullet \subseteq S^\bullet$, and t is disabled by $r' \in S$.

Case 2: $t \in {}^\bullet S_{P1}$. Let $p \in t^\bullet \cap S_{P1}$. We consider three sub-cases: (i) $\nexists \mathbf{r} \in \mathbf{S_R}$ **s.t.** $\mathbf{t} \in \mathbf{r}^\bullet$: $\exists p'$ s.t. $\overline{M'}(p') = 0 \wedge t \in p'^\bullet$ (otherwise, $\overline{M'}(p') \geq 1$, which contradicts the deadness of t). Furthermore, $p \in S_{P1}$ implies $\exists r'$ s.t. $p \in H(r') \wedge r' \in S_R$, and by the sub-case assumptions, $t \notin r'^\bullet$. Therefore, $p' \in H(r')$, which implies that $p' \in S_{P1}$. It follows, then, that $t \in S_{P1}^\bullet \subseteq S^\bullet$, and t is disabled by $p' \in S$. (ii) $\exists \mathbf{r} \in \mathbf{S_R}$ **s.t.** $\mathbf{t} \in \mathbf{r}^\bullet \wedge \overline{\mathbf{M'}}(\mathbf{r}) < \mathbf{W}(\mathbf{r}, \mathbf{t})$: $t \in S_R^\bullet \subseteq S^\bullet$, and t is disabled by $r \in S$. (iii) $\forall \mathbf{r} \in {}^\bullet \mathbf{t} \cap \mathbf{S_R}, \overline{\mathbf{M'}}(\mathbf{r}) \geq \mathbf{W}(\mathbf{r}, \mathbf{t})$: If $\exists p' \in S_{P1}$ s.t. $t \in p'^\bullet$, $t \in S_{P1}^\bullet \subseteq S^\bullet$, and t is disabled by $p' \in S$. Otherwise, $\exists p' \in S_{P2}$ s.t. $t \in p'^\bullet$ (from the definition of S_{P2}). Furthermore, $\overline{M'}(p') = 0$ (otherwise, $\exists r \in S_R$ s.t. $t \in r^\bullet$ which contradicts the sub-case assumption). Therefore, $t \in S_{P2}^\bullet \subseteq S^\bullet$, and t is disabled by $p' \in S$.

Case 3: $t \in {}^\bullet(S_{P2} - S_{P1})$. Let $p \in t^\bullet \cap S_{P2}$. We consider two sub-cases: (i) $\overline{\mathbf{M'}}(\mathbf{p}) = \mathbf{0}$: $p \notin H(S_R)$. Therefore, $\exists p' \in S_{P2}$ s.t. $\overline{M'}(p') = 0 \wedge t \in p'^\bullet$ (otherwise, $\exists r \in S_R$ s.t. $t \in r^\bullet \wedge \overline{M'}(r) < W(r,t)$, which implies $p \in H(S_R)$). It follows that $t \in S_{P2}^\bullet \subseteq S^\bullet$, and t is disabled by $p' \in S$. (ii) $\overline{\mathbf{M'}}(\mathbf{p}) \geq \mathbf{1}$: If $\exists p' \in S_{P2}$ s.t. $t \in p'^\bullet \wedge \overline{M'}(p') = 0$, $t \in S_{P2}^\bullet \subseteq S^\bullet$, and t is disabled by $p' \in S$. Otherwise, $\exists r \in S_R$ s.t. $t \in r^\bullet \wedge \overline{M'}(r) < W(r,t)$. It follows that $t \in S_R^\bullet \subseteq S^\bullet$, and t is disabled by $r \in S$. $\diamond$

Theorem 1 *Let $N = (P, T, W, M_0)$ be a well-marked S^2PGR^2 net. The net is* live *iff the space of modified reachable markings, $\overline{R(\mathcal{N}, M_0)}$, contains no deadly marked siphon with a disabling resource place.*

Proof: (i) To show the necessity part, suppose that there exists a marking $M \in R(\mathcal{N}, M_0)$, such that the modified marking $\overline{M}$ contains a deadly marked siphon, S, with at least one disabling resource place. Let $r \in S \cap P_R$ be a disabling resource place, and consider $t \in r^\bullet$ s.t. $\overline{M}(r) < W(r,t)$. Lemma 1 implies that $\forall t' \in {}^\bullet r$, t' is dead in $R(\mathcal{N}, \overline{M})$. From the definition of $\overline{M}$, it follows that $\forall M' \in R(\mathcal{N}, M)$, $M'(r) \leq M(r)$. Therefore, t is a dead transition at M, which contradicts the assumption of the net liveness.

(ii) To show the sufficiency for liveness of the condition stated in Theorem 1, suppose $\mathcal{N}$ is not live. Then, $\exists\ M \in R(\mathcal{N}, M_0)$ and $t \in T$ s.t. t is dead at M. But then, Lemma 5 implies that there exists a marking $\overline{M'} \in \overline{R(\mathcal{N}, M)} \subseteq \overline{R(\mathcal{N}, M_0)}$, containing a deadly marked siphon with at least one disabling resource place. $\diamond$

Finally, we remark that in the class of S^2PGR^2 nets, weak liveness does not imply liveness. This is due to the existence of partial deadlocks in the underlying C/D-RAS, which might not lead to total deadlock.

4 A Polynomial-Complexity DAP for C/D-RAS

As it was mentioned in the introductory section, it has been established in [1] that, in the considered RAS class, state safety is an NP-complete decision problem, and therefore, the implementation of the *optimal* – i.e., least restrictive – DAP is NP-hard [7]. In this section, we proceed to the development of a suboptimal, but polynomial-complexity DAP, which is an effective generalization to C/D-RAS of the RUN DAP, originally

[3] We would like to thank Drs Tricas and Ezpeleta for pointing out a problem in the original version of this result.

developed in [13] for SU-RAS configurations; for that reason, we call the policy C/D-RUN. Next, we first provide the formal policy definition, then we establish its correctness, and finally, we show that its implementation presents polynomial complexity.

C/D-RUN: Let $o_i \equiv O(R_i), O() : \mathcal{R} \rightarrow \{1, \ldots, m\}$ be any partial order imposed on $\mathcal{R}$. Also, let $\rho^{max}_{jkl} = \max\{o_i \mid J_{jkl}[i] > 0, i = 1, \ldots, m\}$ and $\rho^{min}_{jkl} = \min\{o_i \mid J_{jkl}[i] > 0, i = 1, \ldots, m\}$. Job stage alternative $J_{j,k-q,l}, q > 0$ is said to be in the *reservation span* of J_{jk}, RS_{jk}, iff

$$\rho^{min}_{j,k-q,l} \leq \max_{u=0,\ldots,q-1} \min_{l} \rho^{max}_{j,k-u,l}$$

and $\forall U \in \{\{0, \ldots, w\} : w = 0, \ldots, q-2\}$,

$$\min_{l} \rho^{min}_{j,k-w-1,l} \leq \max_{u \in U} \min_{l} \rho^{max}_{j,k-u,l}$$

Let $NS_{jkl} = \{J_{ju} \mid J_{jkl} \in RS_{ju}\}$ be the *neighborhood* (stage set) of alternative J_{jkl}. Note that $J_{jkl} \notin NS_{jkl}$. The (resource) *reservation scheme*, π_{jkl}, for J_{jkl}, under partial ordering $O()$, is the subset of $(\bigcup_{J_{ju} \in NS_{jkl}} A_{ju} \cup \{J_{jkl}\})$ such that: (i) $J_{jkl} \in \pi_{jkl}$, and (ii) $\forall J_{ju} \in NS_{jkl}$, $\pi_{jkl} \cap A_{ju} \equiv \{J_{juv} : \rho^{max}_{juv} = \min_w \rho^{max}_{juw}\}$. The *nominal* resource allocation requirement for J_{jkl} (under reservation scheme π_{jkl}) is defined by $J^{\pi}_{jkl}[i] = \max\{J_{juv}[i] \mid J_{juv} \in \pi_{jkl}\}$, if $o_i \geq \rho^{min}_{jkl}$; $J^{\pi}_{jkl}[i] = 0$, o.w.; $\forall i = 1, \ldots, m$. $\diamond$

Since a job stage alternative J_{jkl} corresponds to a place $p \in P_S$, and the input transition to p represents a resource allocation / de-allocation for J_{jkl}, given an S^2PGR^2 net, the controlled net, CS^2PGR^2, under a C/D-RUN implementation, is constructed as follows: (i) Introduce the set of control places $P_W = \{w_1, w_2, \ldots, w_m\}$ with $M_0(w_i) = C_i$, $\forall i$. (ii) $\forall t \in T$, let J_{jkl} (resp., $J_{j,k-1,l'}$) denote the corresponding job stage alternative for $t^{\bullet} \cap (P_S \cup P_0)$ (resp., ${}^{\bullet}t \cap (P_S \cup P_0)$). Then, $\forall w_i \in P_W$, introduce $W(w_i, t) = J^{\pi}_{jkl}[i] - J^{\pi}_{j,k-1,l'}[i]$, if $J^{\pi}_{jkl}[i] - J^{\pi}_{j,k-1,l'}[i] > 0$; $W(t, w_i) = J^{\pi}_{j,k-1,l'}[i] - J^{\pi}_{jkl}[i]$, if $J^{\pi}_{jkl}[i] - J^{\pi}_{j,k-1,l'}[i] < 0$; nothing otherwise. It should be easy to see that the constructed CS^2PGR^2 net still belongs to the class of S^2PGR^2, with control places w_i playing the role of additional resources. We state this effect in the following lemma.

Lemma 6 *Let $N = (P_S \cup P_0 \cup P_R \cup P_W, T, W, M_0)$ be a (well-marked) CS^2PGR^2 net corresponding to a C/D-RUN implementation. Then, $\forall w_i \in P_W$, $\exists$ a unique minimal p-semiflow y_{w_i} s.t. $\|y_{w_i}\| \cap P_W = \{w_i\}$, $\|y_{w_i}\| \cap P_R = \emptyset$, $\|y_{w_i}\| \cap P_0 = \emptyset$, $\|y_{w_i}\| \cap P_S \neq \emptyset$, and $y_{w_i}(w_i) = 1$. Furthermore, $\sum_{p \in \{w_i\} \cup H(w_i)} y_{w_i}(p) \cdot M(p) = M_0(w_i) \equiv C_i$, where $H(w_i)$ extends the notion of resource holders to control places.*

Next, we prove the correctness of the proposed C/D-RUN implementations, by showing that the policy negates the development of deadly marked siphons containing at least one disabling resource place, in the space of modified reachable markings, $\overline{R(\mathcal{N}, M_0)}$.

Lemma 7 *Let $\mathcal{N} = (P_S \cup P_0 \cup P_R \cup P_W, T, W, M_0)$ be a (well-marked) CS^2PGR^2 net corresponding to a C/D-RUN implementation. Then, $\overline{R(\mathcal{N}, M_0)}$ contains no deadly marked siphon with at least one disabling resource place.*

Proof: We prove the above result by contradiction. Hence, for the sake of the argument, suppose that there exists $\overline{M'} \in \overline{R(\mathcal{N}, M_0)}$ containing a deadly marked siphon S with at least one disabling resource place. Then, from Lemma 1, and working as in the proof of Lemma 5, we can construct a reachable marking $M \in R(\mathcal{N}, M')$, s.t. (i) $\overline{M}$ contains the deadly marked siphon S, (ii) $\{p \in P_S : \overline{M}(p) \geq 1\} \neq \emptyset$, (iii) $\forall p \in P_S$ s.t. $\overline{M}(p) \geq 1$, $\forall t \in p^{\bullet}$, t is dead in $R(\mathcal{N}, \overline{M})$, and (iv) $(S_R \cup S_W) \neq \emptyset$, where $S_R = S \cap P_R$ and $S_W = S \cap P_W$. Consider $q_1 \in S_R \cup S_W$.

From the construction of S, $\exists p_1 \in P_S$ s.t. $M(p_1) \geq 1 \wedge p_1 \in H(q_1)$, and $\forall t \in p_1^{\bullet}$, t is dead. Let J_{juv} be the job stage alternative associated with p_1. We select $t_1 \in p_1^{\bullet}$ s.t. $t_1 \in {}^{\bullet}s_1$, where $s_1 \in P_S$ corresponds to job stage alternative $J_{j,u+1,v}$ with $\rho^{max}_{j,u+1,v} = \min_w \rho^{max}_{j,u+1,w}$. Since t_1 is dead, and $M(p_1) \geq 1$, $\exists q_2 \in S_R \cup S_W$ disabling t_1. Repeating the above argument on place q_2, and considering the finiteness of the set $S_R \cup S_W$, we conclude that there exists a set $\{q_1, q_2, \ldots, q_k\} \subseteq S_R \cup S_W$, and a corresponding set $\{p_1, p_2, \ldots, p_k\} \subseteq P_S$ s.t. $p_i, i = 1, \ldots, k$, is a marked place belonging to $H(q_i)$. Furthermore, $t_i \in p_i^{\bullet}$ is disabled by q_{i+1}, $i = 1, \ldots, k-1$, and t_k is disabled by some q_i, $i \in \{1, \ldots, k\}$. Next, consider a place $q_{i^*} \in \{q_1, q_2, \ldots, q_k\}$ with $o_{i^*} = \min_{i=1,\ldots,k} o_i$, where o_i is the partial ordering used in the C/D-RUN implementation, extended to $P_R \cup P_W$ by imposing the same order to a resource and its corresponding control place. Let J_{jkl} and $J_{j,k+1,l'}$ be the job stage alternatives respectively associated with places p_{i^*} and s_{i^*}. We consider four cases:

Case 1: $p_{i^*} \in H(r_{i^*})$ and $s_{i^*} \in H(r_{i^*+1})$. Since $\rho^{min}_{jkl} \leq o_{i^*} \leq o_{i^*+1} \leq \rho^{max}_{j,k+1,l'} = \min_w \rho^{max}_{j,k+1,w}$, $J_{j,k+1} \in NS(J_{jkl})$. Furthermore, since $\rho^{min}_{jkl} \leq o_{i^*+1}$, $p_{i^*} \in H(w_{i^*+1})$, where w_{i^*+1} is the control place corresponding to resource r_{i^*+1}.

Case 2: $p_{i^*} \notin H(r_{i^*})$ and $s_{i^*} \in H(r_{i^*+1})$. Then, there exists $J_{j,k+\mu}, \mu \geq 1$ s.t. $J_{j,k+\mu} \in NS(J_{jkl})$, and exists $J_{j,k+\mu,u}$ such that $J_{j,k+\mu,u}[i^*] > 0$. Also, from the definition of C/D-RUN DAP, $\rho^{min}_{jkl} \leq o_{i^*}$. Hence, it follows that $\rho^{min}_{jkl} \leq o_{i^*} \leq o_{i^*+1} \leq \rho^{max}_{j,k+1,l'} = \min_w \rho^{max}_{j,k+1,w}$. Therefore, $J_{j,k+1} \in NS(J_{jkl})$. Since $\rho^{min}_{jkl} \leq o_{i^*+1}$, $p_{i^*} \in H(w_{i^*+1})$, where w_{i^*+1} is the control place corresponding to resource r_{i^*+1}.

Case 3: $p_{i^*} \in H(r_{i^*})$ and $s_{i^*} \notin H(r_{i^*+1})$. There exists $J_{j,k+\mu}, \mu > 1$ such that $J_{j,k+\mu} \in NS(J_{j,k+1,l'})$, and exists $J_{j,k+\mu,u}$ such that $J_{j,k+\mu,u}[i^* + 1] > 0$ and $\rho^{max}_{j,k+\mu,u} = \min_w \rho^{max}_{j,k+\mu,w}$. Furthermore, from the definition of C/D-RUN DAP, $\rho^{min}_{j,k+1,l'} \leq o_{i^*+1}$. Since $J_{j,k+\mu} \in NS(J_{j,k+1,l'})$, and $\rho^{min}_{jkl} \leq o_{i^*} \leq o_{i^*+1} \leq \rho^{max}_{j,k+\mu,u} = \min_w \rho^{max}_{j,k+\mu,w}$, it follows that $J_{j,k+\mu} \in$

$NS(J_{jkl})$. Since $\rho^{min}_{jkl} \leq o_{i^*+1}$, $p_{i^*} \in H(w_{i^*+1})$, where $w_{i^*+1} \equiv q_{i^*+1}$.

Case 4: $p_{i^*} \notin H(r_{i^*})$ and $s_{i^*} \notin H(r_{i^*+1})$. There exists $J_{j,k+\mu}, \mu > 1$ such that $J_{j,k+\mu} \in NS(J_{j,k+1,l'})$, and exists $J_{j,k+\mu,u}$ such that $J_{j,k+\mu,u}[i^*+1] > 0$. Also, from the definition of C/D-RUN DAP, $\rho^{min}_{jkl} \leq o_{i^*}$, and $\rho^{min}_{jkl} \leq o_{i^*} \leq o_{i^*+1} \leq \rho^{max}_{j,k+\mu,u} = \min_w \rho^{max}_{j,k+\mu,w}$. The last set of inequalities, combined with the fact that $J_{j,k+\mu} \in NS(J_{j,k+1,l'})$, imply that $J_{j,k+\mu} \in NS(J_{jkl})$. Since $\rho^{min}_{jkl} \leq o_{i^*+1}$, it follows that $p_{i^*} \in H(w_{i^*+1})$, where $w_{i^*+1} \equiv q_{i^*+1}$.

So, we have established that in all four cases, $p_{i^*} \in H(w_{i^*+1})$. By the policy definition, $J^{\pi}_{jkl}[i^*+1] \geq J^{\pi}_{j,k+1,l'}[i^*+1]$, which further implies that $W(w_{i^*+1}, t_{i^*}) \equiv 0$. This establishes the contradiction for Cases (3) and (4), above. For the remaining Cases (1) and (2), we have $q_{i^*+1} \in S_R$, which combined with the fact that $p_{i^*} \in H(w_{i^*+1})$, imply that $y_{w_{i^*+1}}(p_{i^*}) - y_{r_{i^*+1}}(p_{i^*}) \geq W(r_{i^*+1}, t_{i^*})$. Furthermore, since r_{i^*+1} disables t_{i^*} in marking M, $0 \leq M(r_{i^*+1}) < W(r_{i^*+1}, t_{i^*})$. The last three inequalities, combined with the facts that $M(w_{i^*+1}) \geq 0$, $M(p_{i^*}) \geq 1$, $H(w_{i^*+1}) \supseteq H(r_{i^*+1})$ and $\forall p$, $y_{w_{i^*+1}}(p) \geq y_{r_{i^*+1}}(p)$, imply that

$$\begin{aligned} &\sum_{p \in \{w_{i^*+1}\} \cup H(w_{i^*+1})} y_{w_{i^*+1}}(p) M(p) \\ &- \sum_{p \in \{r_{i^*+1}\} \cup H(r_{i^*+1})} y_{r_{i^*+1}}(p) M(p) \quad \geq \\ &\left(y_{w_{i^*+1}}(p_{i^*}) - y_{r_{i^*+1}}(p_{i^*})\right) M(p_{i^*}) \\ &\qquad + M(w_{i^*+1}) - M(r_{i^*+1}) \quad > \quad 0 \end{aligned}$$

But from Lemma 6,

$$\begin{aligned} &\sum_{p \in \{w_{i^*+1}\} \cup H(w_{i^*+1})} y_{w_{i^*+1}}(p) M(p) \; = \\ &\sum_{p \in \{r_{i^*+1}\} \cup H(r_{i^*+1})} y_{r_{i^*+1}}(p) M(p) \; = \; C_{i^*+1} \end{aligned}$$

which establishes the contradiction for Cases (1) and (2), above, and concludes, thus, the proof. $\diamond$

Theorem 2 *Any (well-marked) CS^2PGR^2 net corresponding to a C/D-RUN implementation, is live.*

Proof : Since $CS^2PGR^2 \subseteq S^2PGR^2$, this follows directly from Lemma 7 and Theorem 1. $\diamond$

Example: (Modified from [15]) As an example of the policy implementation, consider a C/D-RAS consisting of three resource types R_1, R_2, and R_3, with capacities $C_1 = 3$, $C_2 = 4$, $C_3 = 2$, and supporting two job types J_1 and J_2, with corresponding job stage sequences $\{(1,0,0)\} \rightarrow \{(2,0,0)\} \rightarrow \{(1,0,1),(3,0,0)\} \rightarrow \{(0,1,0)\}$, and $\{(0,1,0)\} \rightarrow \{(0,3,0)\} \rightarrow \{(1,0,0),(0,1,1)\}$. The resource ordering employed in the implementation of the C/D-RUN DAP is $o_1 = 1$, $o_2 = 2$, and $o_3 = 3$. Then, the policy definition requires that $NS(J_{111}) = \{J_{12}, J_{13}, J_{14}\}$, $NS(J_{121}) = \{J_{13}, J_{14}\}$, $NS(J_{131}) = \{J_{14}\}$, $NS(J_{132}) = \{J_{14}\}$, $NS(J_{141}) = \emptyset$, $NS(J_{211}) = \{J_{22}\}$, and $NS(J_{221}) = NS(J_{231}) = NS(J_{232}) = \emptyset$. This neighborhood definition implies that $\pi_{111} = \{J_{111}, J_{121}, J_{132}, J_{141}\}$, $\pi_{121} = \{J_{121}, J_{132}, J_{141}\}$, $\pi_{131} = \{J_{131}, J_{141}\}$, $\pi_{132} = \{J_{132}, J_{141}\}$, $\pi_{141} = \{J_{141}\}$, $\pi_{211} = \{J_{211}, J_{221}\}$, $\pi_{221} = \{J_{221}\}$, $\pi_{231} = \{J_{231}\}$, and $\pi_{232} = \{J_{232}\}$ Therefore, $J^{\pi}_{111} = (3,1,0)$, $J^{\pi}_{121} = (3,1,0)$, $J^{\pi}_{131} = (1,1,1)$, $J^{\pi}_{132} = (3,1,0)$, $J^{\pi}_{141} = (0,1,0)$, $J^{\pi}_{211} = (0,3,0)$, $J^{\pi}_{221} = (0,3,0)$, and $J^{\pi}_{231} = (1,0,0)$, and $J^{\pi}_{232} = (0,1,1)$. The resulting CS^2PGR^2 is depicted in Figure 1. $\diamond$

It should be clear from the previous example, that, in the derived CS^2PGR^2 net, the size of the control subnet is polynomially related to the size of the original S^2PGR^2 net. Moreover, the complexity of the derivation of this control structure is polynomial with respect to the original (uncontrolled) S^2PGR^2 net size. A formal analysis of the policy complexity can be found in [12].

5 Conclusion

This paper studied the deadlock avoidance problem for the class of C/D-RAS, that allows for multiple resource acquisitions and flexible routings. A new class of PN, S^2PGR^2, was presented to model the C/D-RAS effectively, and some important liveness properties were derived. In particular, we were able to establish that in the considered PN class, non-liveness can be interpreted through the development of a special siphon construct, characterized as a deadly marked siphon, in a modified marking space, which constitutes the projection of the original net reachability space to a pertinently selected subset of the marking components. This construction generalized the notion and the role of the empty siphon in the deadlock analysis of SU-RAS to the class of C/D-RAS, and provided the basis for the development of C/D-RUN, an efficiently computable deadlock avoidance policy for this broader class of RAS. Future work will develop techniques for enhancing the flexibility of the proposed C/D-RUN implementations, and will seek to identify interesting special subclasses of C/D-RAS that admit polynomial-time optimal deadlock avoidance.

References

[1] T. Araki, Y. Sugiyama, and T. Kasami. Complexity of the deadlock avoidance problem. In *2nd IBM Symp. Math. Found. Computer Sci.*, pages 229–257, 1977.

[2] Z. A. Banaszak and B. H. Krogh. Deadlock avoidance in flexible manufacturing systems with concurrently competing process flows. *IEEE Transactions on Robotics & Automation*, 6(6):724–734, 1990.

[3] K. Barkaoui, A. Chaoui, and B. Zouari. Supervisory control of discrete event systems based on structure theory of petri nets. In *IEEE International Con-*

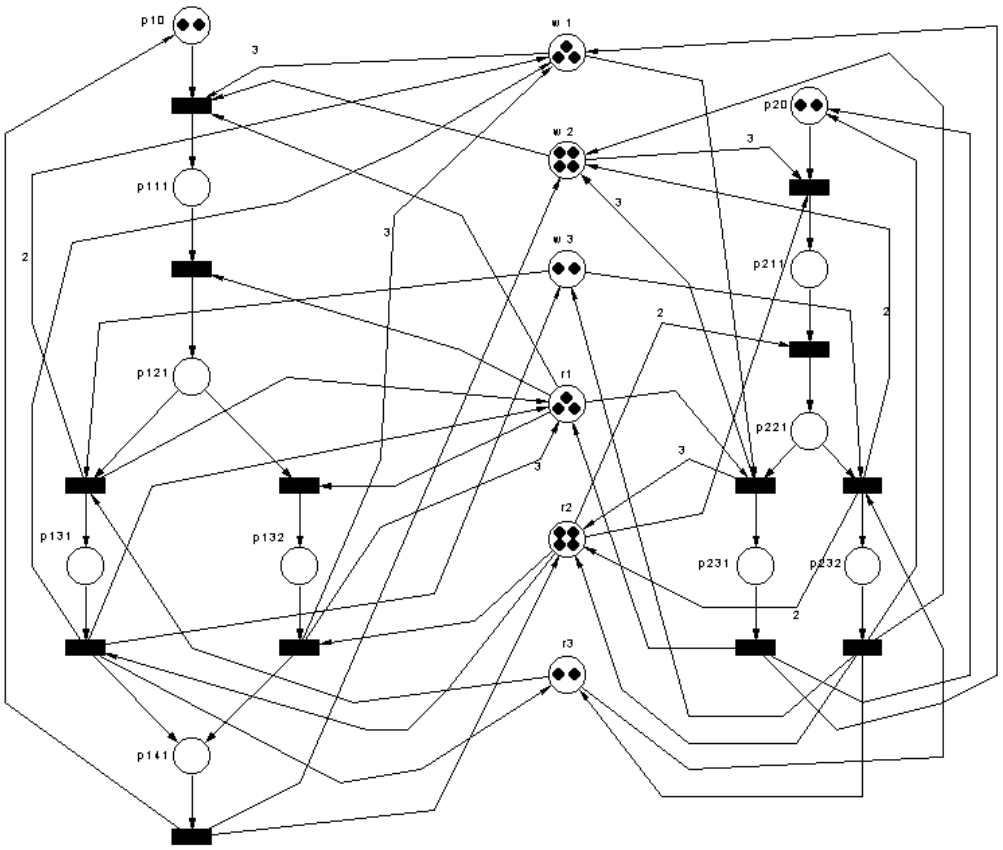

Figure 1: The CS^2PGR^2 net implementing the C/D-RUN DAP for the example RAS configuration

ference on Systems, Man, & Cybernetics, pages 3750–3755. IEEE, 1997.

[4] F. Chu and X-L. Xie. Deadlock analysis of petri nets using siphons and mathematical programming. *IEEE Transactions on Robotics & Automation*, 13(6):793–804, 1997.

[5] J. Ezpeleta, J. M. Colom, and J. Martinez. A petri net based deadlock prevention policy for flexible manufacturing systems. *IEEE Transactions on Robotics & Automation*, 11:173–184, 1995.

[6] M. P. Fanti, B. Maione, S. Mascolo, and B. Turchiano. Event-based feedback control for deadlock avoidance in flexible production systems. *IEEE Transactions on Robotics & Automation*, 13:347–363, 1997.

[7] M. R. Garey and D. S. Johnson. *Computers and Intractability : A Guide to the Theory of NP-Completeness*. W. H. Freeman, New York, 1979.

[8] D. C. Kozen. *Automata and Computability*. Springer Verlag, 1997.

[9] M. Lawley and S. Reveliotis. Deadlock avoidance for sequential resource allocation systems : Hard and easy cases. Technical Report Res. Memo 99-17, School of Industrial Engineering, Purdue University, 1999.

[10] M. Lawley, S. Reveliotis, and P. Ferreira. A correct and scalable deadlock avoidance policy for flexible manufacturing systems. *IEEE Transactions on Robotics & Automation*, 14(5):796–809, 1998.

[11] J. Park and S. A. Reveliotis. Algebraic synthesis of efficient deadlock avoidance policies for sequential resource allocation systems. *IEEE Transactions on Robotics & Automation*, 16(2):190–196, 2000.

[12] J. Park and S. A. Reveliotis. Deadlock avoidance in sequential resource allocation systems with multiple resource acquisitions and flexible routings. *submitted to IEEE Transactions on Automatic Control*, pages –, 2000.

[13] S. A. Reveliotis and P. M. Ferreira. Deadlock avoidance policies for automated manufacturing cells. *IEEE Transactions on Robotics & Automation*, 12(6):845–857, 1996.

[14] S. A. Reveliotis, M. A. Lawley, and P. M. Ferreira. Polynomial complexity deadlock avoidance policies for sequential resource allocation systems. *IEEE Transactions on Automatic Control*, 42(10):1344–1357, 1997.

[15] F. Tricas, F. García-Vallés, J. M. Colom, and J. Ezpeleta. A structural approach to the problem of deadlock prevention in processes with resources. In *Proceedings of the 4th Workshop on Discrete Event Systems*, pages 273–278. IEE, 1998.