Performance optimization for timed weighted marked graphs under infinite server semantics

Zhou He, Zhiwu Li, Alessandro Giua

Abstract

This paper deals with the performance optimization of resource allocation systems with the aim of maximizing the system's throughput under a given budget for acquiring resources. Resources are assumed to be renewable, i.e., they are not consumed by the operations and become available again after they have been released. The systems under consideration are modeled by a subclass of timed Petri nets called deterministic timed weighted marked graphs. In addition, we take into account infinite server semantics, i.e., the degree of self-concurrency of each transition is infinite. We propose an approach which provides an optimal solution, but has a high computational cost. For this reason, we also present two different approaches which can find suboptimal solutions with a reduced computational cost. The performances of the proposed approaches are compared by means of numerical simulations.

Published as:

Zhou He, Zhiwu Li, and Alessandro Giua. "Performance optimization for timed weighted marked graphs under infinite server semantics," *IEEE Trans. on Automatic Control*, vol. 63, no. 8, pp. 2573–2580, 2018. DOI: 10.1109/TAC.2017.2766202

Z. He is with the College of Mechanical and Electrical Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China, email: hezhou@sust.edu.cn.

Z. W. Li is with the Institute of Systems Engineering, Macau University of Science and Technology, Taipa, Macau and also with School of Electro-Mechanical Engineering, Xidian University, Xi'an 710071, China, e-mail: zhwli@xidian.edu.cn.

A. Giua is with DIEE, University of Cagliari, Cagliari 09124, Italy, email: giua@diee.unica.it.

I. INTRODUCTION

Performance optimization plays an important role in *resource allocation systems* [1] which are typical encountered in the domain of automated manufacturing, software engineering, logistics, etc. Timed Petri nets (PNs) have found their extensive applications for modeling and analysis of this class of systems.

Cyclic manufacturing systems such as assembly lines, kanban systems, and transfer lines can be modeled by *timed marked graphs* (TMGs) a class of conflict-free nets that have been studied since the early 90's. Campos *et al*. [2] addressed the performance analysis of TMGs and computed throughput bounds. The *marking optimization problem* for TMGs was investigated in [3] and a heuristic algorithm was developed to find a proper schedule which minimizes the cost of resources under the constraint that the system's throughput should not smaller than a given value. The *cycle time optimization problem* of TMGs was considered in [4] and different approaches were developed to find a proper schedule which maximizes the system's throughput under the constraint that the cost of resources should not exceed a given bound.

However, TMGs cannot model important features that may be present in resource allocation systems such as the use of different quantities of resources in different operations or the processing in batches whose size may change during different processing steps. For this reason, a more general model called *timed weighted marked graph* (TWMG) was studied in the literature. This model, characterized by weighted arcs, can conveniently describe systems with bulk services and arrivals that operate in a cyclic mode, which are commonly encountered in manufacturing, transportation, logistics, and so on.

Several analytical methods were proposed to compute the cycle time of a TWMG — converting it into an equivalent TMG — both under *single server semantics* [5] and under *infinite server semantics* [6]. From a physical point of view, the server semantics can be interpreted as the number of times that an operation can be executed concurrently. Under single server semantics, the same operation can only be executed once at a time, while the same operation can be executed as many times as the number of available servers under infinite server semantics. Note that infinite server semantics is more general than single server (or in general k-server) semantics.

Algebraic approaches based on (max,+) algebra have also been used for the analysis of TWMGs [11]. They can usually be applied to restricted subclasses because the corresponding model is non-linear due the presence of the weights. A polynomial algorithm to check the existence of periodic schedules for a TWMG was presented in [7].

The marking optimization problem of TWMGs under single server semantics was studied in [8], [9]: in these works some efficient heuristic approaches were provided to find a near optimal solution. The same problem under infinite server semantics was addressed in [10] exploiting heuristics based on greedy allocations scheme.

Among other recent works based on net models, we mention a heuristic strategy for resources optimization in process PNs [12], a stochastic approximation algorithm for the performance optimization of stochastic PNs [13] and, finally, the optimization of a new model — called *batch deterministic and stochastic PNs* — introduced to model batch features in supply chains [14].

Based on mixed integer linear programming problem (MILPP) technique, the cycle time optimization problem for TWMGs under single server semantics was investigated in [15]. In this paper we study the same problem under the more general infinite server semantics.

This paper is based on the results initially presented in [16], where we showed that a TWMG under infinite server semantics can be transformed into a finite family of equivalent TMGs, each one valid for a class belonging to a finite partition of the initial markings set. We also presented an MILPP to solve the cycle time minimization problem under a constraint on the cost of available resources. The approach finds an optimal allocation for the equivalent TMG under the constraint that the initial marking belongs to a suitable partition class. However, this procedure has a high computational complexity due to the fact that the number of classes can increase exponentially with the number of places.

In this paper, we better formalize the optimal approach proposed in [16] providing complete proofs for the presented results. In addition, we explore in detail two more efficient suboptimal approaches that can obtain a good approximation of the optimal solution with a significantly reduced computational cost. The first suboptimal approach consists in an MILPP that ensures finding a live marking with an upper bound on the corresponding cycle time: detailed proofs of the main result are here originally presented. The second suboptimal approach, consists in finding a suitable subset of places to which resources are allocated — instead of taking all places into consideration — to reduce the number of partition classes to explore. We extend this approach considering three different objective functions: 1) minimal number of places; 2) minimal cost of resources; 3) minimal number of partition classes. In a series of original numerical experiments, we compare all approaches presented in this paper.

This paper is organized as follows. Basic concepts and the main properties of TWMGs are presented in Section II. In Section III the problem under consideration is formulated. Section IV recalls the approaches proposed in [16] and proposes a place subset allocation approach. Section V investigates the complexity of the three approaches. Applications of the proposed approaches and numerical studies are investigated in Section VI. Conclusions are finally reached in Section VII.

II. Background

A. Preliminaries

In this subsection, we recall the basic concepts and definitions used in this paper. A *place/transition* net (P/T net) is a structure $N = (P, T, \boldsymbol{Pre}, \boldsymbol{Post})$ with a set of n places P; a set of m transitions T; the pre and post incidence matrix $\boldsymbol{Pre}$, $\boldsymbol{Post} \in \mathbb{N}^{n \times m}$, where $\mathbb{N}$ denotes the set of non-negative integers.

A vector $\boldsymbol{x} \in \mathbb{N}^{|T|}$ (resp., $\boldsymbol{y} \in \mathbb{N}^{|P|}$) is a *T-semiflow* (resp., *P-semiflow*) such that $\boldsymbol{C} \cdot \boldsymbol{x} = \boldsymbol{0}$ (resp., $\boldsymbol{y}^T \cdot \boldsymbol{C} = \boldsymbol{0}$), where $\mathbb{N}_+$ denotes the set of positive integers and $\boldsymbol{C} = \boldsymbol{Post} - \boldsymbol{Pre}$ is the incidence matrix. The supports of a T-semiflow and a P-semiflow are defined by $\|\boldsymbol{x}\|$=$\{t_i \in T | x_i > 0\}$ and $\|\boldsymbol{y}\|$=$\{p_i \in P | y_i > 0\}$, respectively. A P-semiflow (resp., T-semiflow) is called minimal if $\|\boldsymbol{y}\|$ (resp., $\|\boldsymbol{x}\|$) is not a superset of the support of any other P-semiflow (T-semiflow), and its components are coprime.

A marking $\boldsymbol{M} : P \to \mathbb{N}$ of a P/T net is a mapping that assigns a non-negative integer of tokens to each place; $M(p)$ denotes the marking of place p. A P/T system or marked net $\langle N, \boldsymbol{M_0} \rangle$ is a net N with an initial marking $\boldsymbol{M_0}$.

A P/T net is *ordinary* if all arcs have unitary weight. A *marked graph* is an ordinary P/T net such that each place has exactly one output and exactly one input transition. A *weighted marked graph* is a net that also meets this structural condition but may not be ordinary.

A P/T net is *strongly connected* if there exists an oriented path from each node to any node. An *elementary circuit* γ of a net is an oriented path that goes from one node back to the same node while all the other nodes are different. We denote the set of elementary circuits by Γ.

Consider a place $p_i \in P$ of a WMG which connects transition $t_{in(p_i)}$ to transition $t_{out(p_i)}$) as shown in Fig. 1. Let $I(p_i)$ and $O(p_i)$ be the weights of its input arc and output arc, respectively. We denote the greatest common divisor of $I(p_i)$ and $O(p_i)$ by gcd_{p_i}.

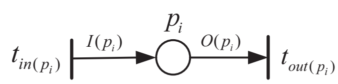

Fig. 1. A place p_i which connects transitions $t_{in(p_i)}$ to transition $t_{out(p_i)}$.

Definition 1: A WMG is said to be *neutral* if for each elementary circuit it holds that $\prod_{p\in\gamma} \frac{I(p)}{O(p)} = 1$. □

A strongly connected and neutral WMG is *conservative*, i.e., there exists a P-semiflow whose components are all positive. In the rest of this paper, we assume that the considered WMG is neutral and strongly connected.

B. Dynamic behavior

There exist two ways of introducing the timing structure in PN models, i.e., associating the timing structure with either transitions or places. A *deterministic transition timed PN* is a pair $N^\delta = (N, \delta)$, where $N = (P, T, \boldsymbol{Pre}, \boldsymbol{Post})$ is a standard PN, and $\delta : T \to \mathbb{N}$, called delay time, assigns a non-negative integer fixed duration to each transition. In the rest of this paper, we will consider *weighted marked graphs* that are *deterministic transition timed* and call them TWMGs.

A transition t is enabled at $\boldsymbol{M}$ if $\boldsymbol{M} \geq \boldsymbol{Pre}(\cdot, t)$ and is denoted by $\boldsymbol{M}[t\rangle$. In a timed net a transition that becomes enabled is assigned a *timer* initially set to $\delta(t)$ and whose value will decrease with time. An enabled transition t whose timer has reached a zero value will fire reaching a new marking $\boldsymbol{M}'$ with

$$\boldsymbol{M}' = \boldsymbol{M} + \boldsymbol{C}(\cdot, t_i). \tag{1}$$

Here $\boldsymbol{Pre}(\cdot, t)$ (resp., $\boldsymbol{C}(\cdot, t)$) denotes the column of $\boldsymbol{Pre}$ (resp., $\boldsymbol{C}$) corresponding to transition t.

The enabling degree of transition t_i enabled at a marking $\boldsymbol{M}_j$, denoted by $\alpha_i(j)$, is the biggest integer number φ such that

$$\boldsymbol{M}_j \geq \varphi \cdot \boldsymbol{Pre}(\cdot, t_i). \tag{2}$$

The server semantics specifies how many clocks are associated with an enabled transition:

- Single server: one clock.
- Infinite server: as many clocks as its enabling degree

- k-server: a number of clocks equal to min $(\varphi,\ \alpha_i(j))$.

Note that single server is a special case of k-server with $k = 1$. Under infinite server semantics, at each time instant the number of clocks associated with a transition t_i is equal to its current enabling degree; this number changes with the enabling degree. If a transition has $k > 1$ clocks which have reached the ZERO value, it will fire k times simultaneously.

C. *Cycle time of a TWMG*

The *cycle time* $\chi(\boldsymbol{M})$ of a TWMG system $\langle N, \boldsymbol{M}\rangle$ is the average time to fire once the minimal T-semiflow. We denote the cycle time of an elementary circuit γ by $\chi_\gamma(\boldsymbol{M})$.

The execution of a live and strongly connected TWMG with integer delays is ultimately repetitive with period Ψ. The number of firings of transition t_i within the period is f_i and the cycle time of the TWMG $\langle N, \boldsymbol{M}\rangle$ is

$$\chi(\boldsymbol{M}) = x_i \cdot \frac{\Psi}{f_i}, \tag{3}$$

where $\boldsymbol{x}$ is the minimal T-semiflow.

The cycle time can be computed by simulation. In addition, a TWMG whose initial marking is known can be transformed into an equivalent TMG whose cycle time can be rather easily computed by several methods. A transformation algorithm for nets under infinite server semantics was presented in [6].

III. Problem statement

In this paper, we deal with the *cycle time optimization problem* for a TWMG under infinite server semantics whose complexity has been proved to be NP-complete [17]. This problem consists in finding an initial marking $\boldsymbol{M}$ such that the weighted sum of tokens in places is less than or equal to a given value and the cycle time is minimized.

$$\begin{aligned} \min \quad & \chi(\boldsymbol{M}) \\ \text{s.t.} \quad & \boldsymbol{y}^T \cdot \boldsymbol{M} \leq R \end{aligned} \tag{4}$$

where $\chi(\boldsymbol{M})$ is the cycle time of the TWMG system $\langle N, \boldsymbol{M}\rangle$, $\boldsymbol{y}^T = (y_1, \ldots, y_n)$ is a non-negative weight vector that represents the cost of the resources, and R is a given positive real number, representing the upper bound on the cost of resources.

In this paper, we assume that R is a large enough number to ensure that there exists a live initial marking $\boldsymbol{M}$. The type of resources considered in this paper are renewable, i.e., the resources are not consumed by the operations and become available again after they have been released, such as machines, tools, and equipments. In addition, we do not necessary consider homogeneous resources. In terms of manufacturing systems, the cost of the resources will remains constant as the production process proceeds.

As a result, we choose the weight vector $\boldsymbol{y}$ as a P-semiflow since the value of $\boldsymbol{y}^T \cdot \boldsymbol{M}$ for every reachable marking is an invariant. In particular, we choose the P-semiflow $\boldsymbol{y}$ that is equal to the weighted sum of all minimal P-semiflows, i.e., $\boldsymbol{y} = \sum_{\gamma\in\Gamma} \lambda_\gamma \cdot \boldsymbol{y}_\gamma$, where $\boldsymbol{y}_\gamma$ represents the minimal P-semiflow corresponding to circuit γ and λ_γ represents the cost of the resources modeled by tokens in the support of $\boldsymbol{y}_\gamma$.

IV. The Proposed Approaches

In this section, we first recall the optimal approach and the Throughput Upper Bound (TUB) approach proposed in [16] and give detailed proofs of the main results. Then, a place subset allocation (PSA) approach is presented.

A. Optimal approach

Nakamura and Silva [6] proved that a TWMG system $\langle N, \boldsymbol{M}\rangle$ with n places and m transitions under infinite server semantics can be transformed into an *equivalent place timed* marked graph (PTMG) system $\langle \hat{N}, \hat{\boldsymbol{M}}\rangle$ with $\hat{n}$ places and $\hat{m}$ transitions. The method consists in transforming the transitions and the places of the TWMG system $\langle N, \boldsymbol{M}\rangle$.

Each transition $t_i \in T$ of the TWMG system $\langle N, \boldsymbol{M}\rangle$ is replaced by a strongly connected circuit with x_i transitions t_i^1, ..., $t_i^{x_i}$ and x_i places q_i^1, ..., $q_i^{x_i}$ (recall that $\boldsymbol{x}$ is the minimal T-semiflow). Each place $p_i \in P$ of the TWMG system $\langle N, \boldsymbol{M}\rangle$ is replaced by a set of $\hat{n}_i$ equivalent timed places $p_i^1, \ldots, p_i^{\hat{n}_i}$ whose number, input and output arcs, and marking depend on both the minimal T-semiflow $\boldsymbol{x}$ and the initial marking $M(p_i)$ (for more details, see [6] and [16]).

Example 1: Let us consider a simple TWMG model N in Fig. 2 whose minimal T-semiflow is $\boldsymbol{x}$=(2, 3)^T. Fig. 3 shows the equivalent PTMG systems $\langle \hat{N}, \hat{\boldsymbol{M}}\rangle$ corresponding to different initial markings. Transitions t_1 and t_2 are replaced by circuits $t_1^1q_1^1t_1^2q_1^2$ and $t_2^1q_2^1t_2^2q_2^2t_2^3q_2^3$, respectively. Place p_1 (resp., p_2) is replaced by $\hat{n}_1$=2 (resp., $\hat{n}_2$=2) places p_1^1 and p_1^2 (resp., p_2^1 and p_2^2). For different initial markings, the structures of equivalent transitions (gray blocks) are always the same, while the structures and markings of equivalent places (blue blocks) may change.

Considering the equivalent PTMG system $\langle \hat{N}_2, \hat{\boldsymbol{M}}_2\rangle$ in Fig. 3(b), we denote the initial markings of equivalent places by $\mu(p_1^1) = \hat{M}(p_1^1) = 0$, $\mu(p_1^2) = \hat{M}(p_1^2) = 1$, $\mu(p_2^1) = \hat{M}(p_2^1) = 0$, and $\mu(p_2^2) = \hat{M}(p_2^2) = 0$. $\diamond$

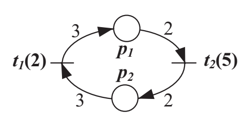

Fig. 2. The TWMG net N for Example 1.

The cycle time optimization problem for a PTMG net $\hat{N}$ can be formulated as follows:

$$\begin{aligned} &\min\ \chi(\hat{\boldsymbol{M}}) \\ \text{s.t.}\quad &\hat{\boldsymbol{y}}^T \cdot \hat{\boldsymbol{M}} \leq R \end{aligned} \tag{5}$$

Next result shows how the cycle time optimization problem for a PTMG can be solved using MILPP.

Proposition 1: Let $(\hat{\boldsymbol{M}}^*, \beta^*, \boldsymbol{\alpha}^*)$ be an optimal solution of the MILPP:

$$\begin{aligned} &\max\ \beta \\ \text{s.t.}\quad &\begin{cases} \hat{\boldsymbol{C}} \cdot \boldsymbol{\alpha} + \hat{\boldsymbol{M}} \geq \boldsymbol{D}_p \cdot \hat{\boldsymbol{Post}} \cdot \boldsymbol{v} \cdot \beta, \\ \hat{\boldsymbol{y}}^T \cdot \hat{\boldsymbol{M}} \leq R, \\ \hat{\boldsymbol{M}} \in \mathbb{N}^{\hat{n}},\ \boldsymbol{\alpha} \in \mathbb{R}^{\hat{m}},\ \beta \in \mathbb{R}^+, \end{cases} \end{aligned} \tag{6}$$

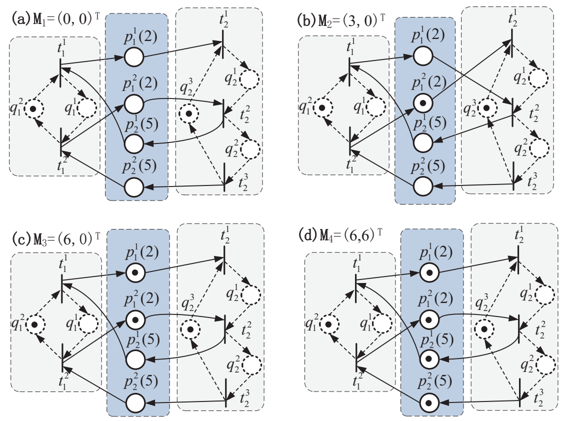

Fig. 3. Equivalent PTMG systems corresponding to different initial markings.

where $\boldsymbol{v}$ is the visit ratio vector which is equal to $\vec{\mathbf{1}}_{\hat{m}\times 1}$ and $\boldsymbol{D}_p$ is a $\hat{n}\times\hat{n}$ matrix such that $D_p(i,j)=\delta(p_i)$, when $i=j$ and otherwise $D_p(i,j)=0$.

Then $\hat{\boldsymbol{M}}^*$ is an optimal solution of problem (5) with an optimal cycle time $\chi(\hat{\boldsymbol{M}}^*)=1/\beta^*$ for a PTMG under infinite server semantics.

Proof: In [6] it was shown that the cycle time of a PTMG system $\langle\hat{N},\hat{\boldsymbol{M}}\rangle$ under infinite server semantics can be directly obtained by solving the following LPP:

$$\begin{aligned} &\max\ \boldsymbol{\sigma}^T\cdot\boldsymbol{D}_p\cdot\hat{\boldsymbol{Post}}\cdot\boldsymbol{v}\\ &\text{s.t.}\ \begin{cases}\boldsymbol{\sigma}^T\cdot\hat{\boldsymbol{C}}=\boldsymbol{0},\\ \boldsymbol{\sigma}^T\cdot\hat{\boldsymbol{M}}=1,\\ \boldsymbol{\sigma}\geq\boldsymbol{0},\end{cases}\end{aligned}\tag{7}$$

The dual problem of LPP (7) is

$$\begin{aligned} &\min\ \chi(\hat{\boldsymbol{M}})\\ &\text{s.t.}\ \ \hat{\boldsymbol{C}}\cdot\boldsymbol{z}+\chi(\hat{\boldsymbol{M}})\cdot\hat{\boldsymbol{M}}\geq\boldsymbol{D}_p\cdot\hat{\boldsymbol{Post}}\cdot\boldsymbol{v}\end{aligned}\tag{8}$$

where the decision variables are $\chi(\hat{\boldsymbol{M}})\in\mathbb{R}^+$ and $\boldsymbol{z}\in\mathbb{R}^{\hat{m}}$. Now let us consider problem (8). This problem can be easily converted into the problem of determining the maximal throughput (i.e., the inverse of cycle time) of the TMG system, given the initial marking. To this end, we only need to replace $\chi(\hat{\boldsymbol{M}})$ with its inverse $\beta=1/\chi(\hat{\boldsymbol{M}})$ and obtain the following LPP:

$$\begin{aligned} &\max\ \beta\\ &\text{s.t.}\ \ \hat{\boldsymbol{C}}\cdot(\beta\boldsymbol{z})+\hat{\boldsymbol{M}}\geq\boldsymbol{D}_p\cdot\hat{\boldsymbol{Post}}\cdot\boldsymbol{v}\cdot\beta,\end{aligned}\tag{9}$$

where $\beta\in\mathbb{R}^+$, and $\beta\boldsymbol{z}\in\mathbb{R}^{\hat{m}}$, or equivalently

$$\begin{aligned} &\max\ \beta\\ &\text{s.t.}\ \ \hat{\boldsymbol{C}}\cdot\boldsymbol{\alpha}+\hat{\boldsymbol{M}}\geq\boldsymbol{D}_p\cdot\hat{\boldsymbol{Post}}\cdot\boldsymbol{v}\cdot\beta,\end{aligned}\tag{10}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^{\hat{m}}$ and $\beta \in \mathbb{R}^{+}$ are the new decision variables. Finally, assuming that $\boldsymbol{M}$ is unknown but under a given constraint on the cost of resources ($\hat{\boldsymbol{y}}^T \cdot \hat{\boldsymbol{M}} \leq R$), we have Eq. (6) where $\hat{\boldsymbol{M}} \in \mathbb{N}^{\hat{n}}$ is a non-negative integer vector. □

To solve the cycle time optimization for a TWMG, we propose to transform it into an equivalent PTMG and use the previous result to compute an optimal initial marking. The main obstacle, however, is the fact that the transformation depends on the (unknown) initial marking. For this reason we need to explore all possible equivalent PTMG structures.

In [16] we showed that the equivalent net structure $\hat{N}$ corresponding to place $p_i \in P$ is periodic with respect to the initial marking $M(p_i)$ and the period is

$$\phi_i = O(p_i) \cdot x_{out(p_i)}. \tag{11}$$

In addition, when looking for an optimal solution we may only consider markings such that for each place p_i the number of assigned tokens is a multiple of gcd_{p_i} to rule out the presence of useless tokens that do not contribute to the cycle time. Thus, we can partition the marking space of a TWMG into Φ subsets, where

$$\Phi = \prod_{p_i \in P} \frac{\phi_i}{gcd_{p_i}}. \tag{12}$$

This allow us to transform a TWMG into a finite family of equivalent PTMGs, each one valid for a partition class of set of initial markings

$$\begin{gathered} \mathcal{M}_j = \bar{\mathcal{M}}_{p_1}^{k_{j,1}} \times \bar{\mathcal{M}}_{p_2}^{k_{j,2}} \times \ldots \times \bar{\mathcal{M}}_{p_n}^{k_{j,n}}, \\ \bar{\mathcal{M}}_{p_i}^{k_i} = \{k_i \cdot gcd_{p_i} + \xi \cdot \phi_i | \xi \in \mathbb{N},\ k_i = 0, \ldots, \tfrac{\phi_i}{gcd_{p_i}} - 1\}. \end{gathered} \tag{13}$$

For instance, in Example 1 it holds: $gcd_{p_1} = gcd_{p_2} = 1$ and $\phi_1 = \phi_2 = 6$. Hence the marking space of the TWMG N can be partitioned into $\Phi = 36$ classes: $\mathcal{M}_1 = (6\xi_{1,1},\ 6\xi_{1,2})^T, \ldots, \mathcal{M}_{36} = (5 + 6\xi_{36,1},\ 5 + 6\xi_{36,2})^T$.

Proposition 2: For each partition class $\mathcal{M}_j$ $(j = 1, \ldots, \Phi)$ in Eq. (13), let $(\beta_j^*, \boldsymbol{M}_j^*, \hat{\boldsymbol{M}}_j^*, \hat{\boldsymbol{\alpha}}_j^*, \boldsymbol{\xi}_j^*)$ be an optimal solution of the MILPP:

$$\begin{aligned} & \max \beta_j \\ & \text{s.t.} \\ & \begin{cases} \hat{\boldsymbol{C}}_j \cdot \boldsymbol{\alpha}_j - \boldsymbol{D}_p \cdot \hat{\boldsymbol{Post}}_j \cdot \boldsymbol{v} \cdot \beta_j + \hat{\boldsymbol{M}}_j \geq 0, & (a) \\ \boldsymbol{y}^T \cdot \boldsymbol{M}_j \leq R, & (b) \\ M_j(p_i) = k_{j,i} \cdot \text{gcd}_{p_i} + \xi_{j,i} \cdot \phi_i,\ \forall p_i \in P, & (c) \\ \hat{M}_j(p_i^s) = \mu_j(p_i^s) + \xi_{j,i}, s = 1, \ldots, \hat{n}_i, & (d) \\ \hat{M}_j(q_i^a) = 0,\ i = 1, \ldots, m,\ a = 1, \ldots, x_i - 1, & (e) \\ \hat{M}_j(q_i^{x_i}) = 1,\ i = 1, \ldots, m, & (f) \\ \beta_j \in \mathbb{R}_{\geq 0}, \xi_{j,i} \in \mathbb{N}, \boldsymbol{M}_j \in \mathbb{N}^n, \hat{\boldsymbol{M}}_j \in \mathbb{N}^{\hat{n}}, \hat{\boldsymbol{\alpha}}_j \in \mathbb{R}^{\hat{m}}. & (g) \end{cases} \end{aligned} \tag{14}$$

where $\mu_j(p_i^s)$ is the marking of equivalent place p_i^s corresponding to $M_j(p_i) = k_{j,i} \cdot \text{gcd}_{p_i}$.

Then $\boldsymbol{M}_j^*$ is an optimal solution of problem (4) restricted to partition class $\mathcal{M}_j$.

Proof: The constraint (a) can provide an optimal solution if $\hat{\boldsymbol{C}_j}$, $\hat{\boldsymbol{Post}_j}$ and $\boldsymbol{D}_p$ are given. The constraint (b) specifies that the weighted sum of tokens in places cannot exceed the upper bound on the cost of resources, and the constraint (c) specifies that feasible markings should be restricted to partition class $\mathcal{M}_j$. The equivalent marking $\hat{\boldsymbol{M}}_j$ is consistent with the marking $\boldsymbol{M}_j$ as ensured by constraints (d), (e) and (f). Thus $\boldsymbol{M}_j^*$ is an optimal solution of problem (4) restricted to partition class $\mathcal{M}_j$. □

Among all the Φ optimal solutions associated with each partition class, we can obtain the maximal throughput and its corresponding marking, i.e., optimal solutions of problem (4).

B. Throughput upper bound

Campos *et al*. [2] proved that an upper bound of the throughput of a TWMG system $\langle N, \boldsymbol{M}\rangle$ under infinite server semantics can be obtained by solving the following LPP:

$$\begin{aligned} &\max\ \beta' \\ \text{s.t.}\ &\begin{cases} \boldsymbol{C}\cdot\boldsymbol{z}+\boldsymbol{M}-\boldsymbol{Pre}\cdot\boldsymbol{\theta}\cdot\beta' \geq \boldsymbol{0}, \\ \beta' \in \mathbb{R}^+, \boldsymbol{z} \in \mathbb{R}^m. \end{cases} \end{aligned} \tag{15}$$

where $\boldsymbol{\theta} = (x_1\cdot\delta(t_1), x_2\cdot\delta(t_2), \ldots, x_m\cdot\delta(t_m))^T$ (recall $\boldsymbol{x}$ is the minimal T-semiflow of the TWMG). The optimal value of β' provides an upper bound of the throughput, i.e.,

$$\beta' \geq \beta. \tag{16}$$

In the following, we first review some conditions concerning the liveness of a TWMG.

Theorem 1: (Teruel et al. [18]) A TWMG system $\langle N, \boldsymbol{M}\rangle$ is live iff each elementary circuit is live. ◇

Proposition 3: (Teruel et al. [18]) If a weighted elementary circuit with an initial marking $\boldsymbol{M}_0$ satisfies $W(\boldsymbol{M}_0) > W(\boldsymbol{M}_D)$, then $\boldsymbol{M}_0$ is a live marking, where $\boldsymbol{M}_D = (O(p_1)-1, O(p_2)-1, \ldots, O(p_n)-1)^T$ and $W(\boldsymbol{M}) = \boldsymbol{y}^T\cdot\boldsymbol{M}$.[1]
◇

Combining these conditions, we present the following proposition.

Proposition 4: Let ($\boldsymbol{M}$, β') be the optimal solution of the MILPP

$$\begin{aligned} &\max\ \beta' \\ \text{s.t.}\ &\begin{cases} \boldsymbol{y}_\gamma^T\cdot\boldsymbol{M} > W(\boldsymbol{M}_D^\gamma), \forall\gamma\in\Gamma, \\ \boldsymbol{C}\cdot\boldsymbol{z}+\boldsymbol{M}-\boldsymbol{Pre}\cdot\boldsymbol{\theta}\cdot\beta' \geq \boldsymbol{0}, \\ M(p_i) \bmod \text{gcd}_{p_i} = 0,\ i=1,\ldots,n, \\ \boldsymbol{y}^T\cdot\boldsymbol{M} \leq R, \\ \beta' \in \mathbb{R}_{\geq 0}, \boldsymbol{M}\in\mathbb{N}^n, \boldsymbol{z}\in\mathbb{R}^m. \end{cases} \end{aligned} \tag{17}$$

[1] Recall that $O(p_i)$ represents the weight of output arc of place p_i and $\boldsymbol{y}$ represents the minimal P-semiflow corresponding to the weighted circuit.

where y_γ denotes the P-semiflow associated with the elementary circuit γ.

Then $\boldsymbol{M}$ is an admissible (possibly non-optimal) live solution for problem (4) and β' is an upper bound of the throughput that it produces.

Proof: According to Theorem 1, a TWMG is live iff each elementary circuit is live. The first constraint is a sufficient condition that ensures the liveness of a weighted elementary circuit according to Proposition 3. Thus, the marking $\boldsymbol{M}$ that we obtain by Eq. (17) will be a live marking. The second condition ensures that marking $\boldsymbol{M}$ is a solution with an upper bound of throughput β'. The number of tokens in place p_i should be a multiple of gcd_{p_i}, which is guaranteed by the third constraint. The fourth constraint is added to limit the cost of resources. □

C. Place Subset Allocation

From a theoretical point of view, it may be interesting to consider a subset of places to which resources are allocated instead of taking all places into consideration, and we believe that in many cases this initial assignment may has a physical meaning that can lead to an optimal solution. As a result, the number of partition classes of marking space can be significantly reduced since we consider $\frac{\phi_i}{gcd_{p_i}}$ partition classes only for a subset of places p_i and one single partition class for other places which are set to be empty. In the following, we will present three different methods to select a subset of places to which resources should be allocated.

According to Theorem 1, a necessary condition to ensure the liveness of a TWMG is that all its elementary circuit are marked. In addition, if the weighted sum of tokens of each elementary circuit is greater than a constant value, then Proposition 3 provides a sufficient condition to ensure the liveness of a TWMG. Combining these two conditions, we select *at least one* place for each elementary circuit to which tokens are allocated. As we discussed in Section III, we assume the upper bound on the weighted sum of tokens is large enough such that the liveness of each elementary circuit can be guaranteed by putting enough tokens into the selected places.

We define a binary vector $\mathcal{I} \in \{0,1\}^n$, i.e., $\mathcal{I} = (\mathcal{I}(1), \cdots, \mathcal{I}(n))^T$. Tokens are initially allocated to the places p_j such that $\mathcal{I}(j) = 1$ and we denote by the set of selected places as $P_r = \{p_j | \mathcal{I}(j) = 1\}$. To fulfill the requirement that each elementary circuit should be marked, we enforce the following constraint:

$$\sum_{p_j \in \gamma} \mathcal{I}(j) \geq 1, \ \forall \gamma \in \Gamma. \tag{18}$$

In addition, in order to reduce the number of partition classes, we present three different approaches to compute the *place subset* P_r based on different objective functions.

PSA1: In this approach, we aim to minimize the number of places to which tokens should be added. Thus, the place subset P_r can be obtained by solving the following problem:

$$\text{PSA1}: \quad \begin{cases} \quad \min \vec{\mathbf{1}}_n^T \cdot \mathcal{I} \\ \text{s.t.} \quad \sum_{p_j \in \gamma} \mathcal{I}(j) \geq 1, \ \forall \gamma \in \Gamma. \end{cases} \tag{19}$$

PSA2: In this approach, we aim to select the places which use the minimal cost of resources. First we define an n-dimensional vector $\boldsymbol{g}_d = (\text{gcd}_{p_1} \cdot y_1, \cdots, \text{gcd}_{p_n} \cdot y_n)^T$, where $\boldsymbol{y}$ is the weight vector used in the criterion that represents the cost of resources. Note that the number of useful tokens in place p_j should be a multiple of gcd_{p_j}.

Therefore, the cost of resources used for place p_j should be a multiple of $\text{gcd}_{p_j} \cdot y_j$. Among all the places, we aim to choose the one whose value of $\text{gcd}_{p_j} \cdot y_j$ is the minimal. As a result, the place subset P_r can be obtained by solving the following problem:

$$\text{PSA2}: \quad \begin{cases} \quad \min \boldsymbol{g}_d^T \cdot \mathcal{I} \\ \text{s.t.} \quad \sum_{p_j \in \gamma} \mathcal{I}(j) \geq 1, \ \forall \gamma \in \Gamma. \end{cases} \tag{20}$$

PSA3: In this approach, we aim to minimize the number of partition classes and reduce the computational cost of optimal approach as much as possible. Thus, the place subset P_r can be obtained by solving the following problem:

$$\text{PSA3}: \quad \begin{cases} \quad \min \prod_{p_i \in P} \phi_i^{\mathcal{I}(i)} \\ \text{s.t.} \quad \sum_{p_j \in \gamma} \mathcal{I}(j) \geq 1, \ \forall \gamma \in \Gamma. \end{cases} \tag{21}$$

Note that the objective function of PSA3 is equivalent to $\min \prod_{p_i \in P_r} \phi_i$. It may be unsolvable when the number of variables is large due to the fact that it is non-linear.

After we obtain the place subset P_r by solving the aforementioned approaches, we can look for possible suboptimal but computationally more efficient solutions of the cycle time optimization problem, as formalized in the following proposition.

Proposition 5: Let ($\boldsymbol{M}$, β) be the optimal solution of the MILPP (14) by replacing constraint (c) with following constraints:

$$\begin{cases} M_j(p_i) = k_{j,i} \cdot \text{gcd}_{p_i} + \xi_{j,i} \cdot \phi_i, \ \forall p_i \in P_r, \quad (c1) \\ M_j(p_i) = 0, \ \forall p_i \notin P_r. \quad (c2) \end{cases} \tag{22}$$

where P_r is the place subset computed by any of the PSA approaches proposed above, i.e., PSA1, PSA2, or PSA3. If $\beta > 0$, then $\boldsymbol{M}$ is a (possibly suboptimal) live solution for problem (4).

Proof: Constraint $c1$ is the same as constraint c in Eq. (14) and is only valid for the selected places. Constraint $c2$ ensures that the number of tokens in places that do not belong to the subset P_r should be zero which is used to reduce the number of partition classes.

Thus, by solving Eq. (14) with new constraints $c1$ and $c2$ in Eq. (22), we obtain a live marking $\boldsymbol{M}$ if $\beta > 0$. □

As a result, the number of partition classes for the PSA approaches is reduced to

$$\Phi' = \prod_{p_j \in P_r} \frac{\phi_j}{\text{gcd}_{p_j}}. \tag{23}$$

According to the results in [16], for a place p_i with zero token there may exist several markings that belong to the same partition class. Thus, the solution obtained by MILPP (14) with new constraints in Eq. (22) may be improved by the following proposition.

Proposition 6: In MILPP (14), constraint $c2$ from Eq. (22) may be relaxed in

$$M_j(p_i) = \xi_{j,i} \cdot \phi_i, \ \xi_{j,i} \in \mathbb{N}, \ \forall p_i \notin P_r. \ \ (c2')$$

The relaxed MILPP has a solution β greater than or equal to the original MILPP (14) and the same number of partition classes Φ'.

Proof: Given constraint $c2$ in Eq. (22), only one partition class may correspond to places $p_i \notin P_r$ since they are marked with zero token. On the basis of Proposition 2 in [16], $M_j(p_i) = 0$ and $M_j(p_i) = \xi_{j,i} \cdot \phi_i$ belong to the same partition class, while $M_j(p_i) = 0$ is a special case of $M_j(p_i) = \xi_{j,i} \cdot \phi_i$ when $\xi_{j,i} = 0$. As a consequence, the number of admissible markings is increased and the obtained throughput β of the PSA approaches may be improved by replacing constraint $c2$ in Eq. (22) with the more general constraint $c2'$, while the number of partition classes Φ' remains the same. □

V. Complexity discussion

In this section, we discuss the computational complexity of the three approaches. It is well known that ILPPs are NP-hard and it is common to characterize the computational burden by the number of variables and constraints.

For the optimal approach, the MILPP in Eq. (14) has $2n + \hat{n} + \hat{m} + 1$ (β_j, $\boldsymbol{M}_j$, $\hat{\boldsymbol{M}}_j$, $\hat{\boldsymbol{\alpha}}_j$, and $\xi_{j,i}$) variables and $2\hat{n} + n + \vec{\mathbf{1}} \cdot \boldsymbol{x} + 1$ constraints totally. The optimal approach requires solving Φ MILPPs in Eq. (14) (recall that $\hat{n}$ and $\hat{m}$ are the numbers of equivalent places and transitions of the TMG, respectively).

For the PSA approaches, we first need to compute the place subset P_r by solving ILPPs (19), (20), and non-linear programming problem (21). These problems have n variables and $|\Gamma|$ constraints totally, where $|\Gamma|$ denotes the number of elementary circuits. While $|\Gamma|$ can grow exponentially with respect to the net size, in practice we found that the time to compute the set P_r is a very small fraction of the total computational time. The MILPP in Eq. (14) with new constraints in Eq. (22) has $n + |P_r| + \hat{n} + \hat{m} + 1$ (β_j, $\boldsymbol{M}_j$, $\hat{\boldsymbol{M}}_j$, $\hat{\boldsymbol{\alpha}}_j$, and $\xi_{j,i}$) variables and $2\hat{n} + n + \vec{\mathbf{1}} \cdot \boldsymbol{x} + 1$ constraints totally, where $|P_r|$ denotes the number of selected places. The PSA approaches require solving Φ' MILPPs in Eq. (14). We observe that the number of variables in Eq. (22) is smaller than that of Eq. (14) and the number of partition classes Φ' is also smaller than that of Eq. (12). In the worst case, $|P_r| = n$ and $\Phi' = \Phi$, i.e., the place subset P_r contains all places. Then, the computational burden of the PSA approaches is the same with the optimal approach. However, in practical example, we find that the computational burden of the PSA approach is much smaller than that of the optimal approach.

For the TUB approach, it has $m + n + 1$ ($\boldsymbol{M}$, $\boldsymbol{z}$, and β') variables and $n + m + |\Gamma| + 1$ constraints totally. In contrast to the optimal approach and the PSA approaches, the TUB approach requires to solve the MILPP only once.

VI. Numerical results

In this section, we apply the proposed approaches to the optimization of the operation of some manufacturing systems taken from the literatures.

A. Application to an FMS

Example 2: Let us consider the flexible manufacturing system in Fig. 4 which combines cyclic assembly process, buffers, work in process, and batch operations. It consists of three machines $\mathcal{U}_1, \mathcal{U}_2$ and $\mathcal{U}_3$, and produce two products $\mathcal{PR}_1$ and $\mathcal{PR}_2$. The production ratios are 3/5 and 2/5 for $\mathcal{PR}_1$ and $\mathcal{PR}_2$, respectively. The production processes of these products are: $\mathcal{PR}_1: \ (\mathcal{U}_1,\ \mathcal{U}_2,\ \mathcal{U}_3)$ and $\mathcal{PR}_2: \ (\mathcal{U}_2,\ \mathcal{U}_1)$. We assume that the upper bound on the cost of resource R is equal to 100.

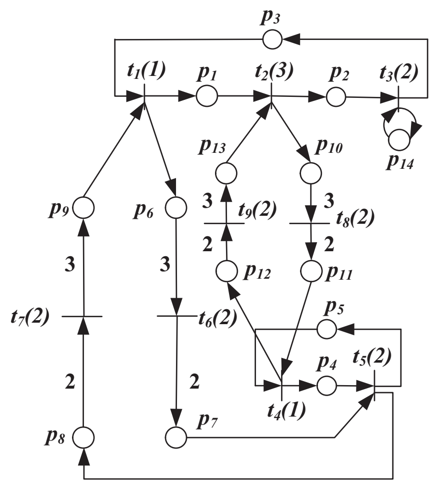

Fig. 4. A TWMG model of a flexible manufacturing system.

TABLE I
SIMULATION RESULTS OF EXAMPLE 2.

Approach X	Nb. of partition classes	Obtained marking $\boldsymbol{M}$	β_{X}	β'	G_X	G'_X	CPU time [s]
Optimal	419904	$(6,0,0,0,2,3,0,4,0,0,0,6,0,2)^T$	0.31		0%	13.9%	156168
TUB	N/A	$(0,0,7,3,0,5,1,0,1,8,0,0,1,1)^T$	0.17	0.36	45.2%	52.8%	4
PSA1	72	$(0,0,6,0,2,0,0,6,0,0,2,4,0,2)^T$	0.31		0%	13.9%	23 (0)
PSA2	162	$(0,6,0,0,2,0,0,6,0,0,2,4,0,2)^T$	0.31		0%	13.9%	48 (0)
PSA3	72	$(0,0,6,0,2,0,0,6,0,0,2,4,0,2)^T$	0.31		0%	13.9%	24 (1)

The Petri net model in Fig. 4 is a strongly connected TWMG which consists of seven elementary circuits: $\gamma_1 = p_1t_2p_2t_3p_3t_1$, $\gamma_2 = p_4t_5p_5t_4$, $\gamma_3 = p_{10}t_8p_{11}t_4p_{12}t_9p_{13}t_2$, $\gamma_4 = p_6t_6p_7t_5p_8t_7p_9t_1$, $\gamma_5 = p_{14}t_3$, $\gamma_6 = p_2t_3p_3t_1p_6t_6p_7t_5p_5t_4p_{12}t_9p_{13}t_2$, and $\gamma_7 = p_{10}t_8p_{11}t_4p_4t_5p_8t_7p_9t_1p_1t_2$. The tokens in circuits γ_3, γ_4, and γ_5 and circuits γ_1 and γ_2 represent the servers and available pallets for products, respectively. Circuits γ_6 and γ_7 are mixed circuits. The minimal T-semiflow is $\boldsymbol{x}$=(3, 3, 3, 2, 2, 1, 1,1, 1)^T. For each circuit we assume that $\lambda_\gamma = 1$ and the weight vector used in the criteria is $y = (3,3,3,4,4,4,6,6,4,4,6,6,4,1)^T$.

For the optimal approach, we have $\phi_1 = 3$, $\phi_2 = 3$, $\phi_3 = 3$, $\phi_4 = 2$, $\phi_5 = 2$, $\phi_6 = 3$, $\phi_7 = 2$, $\phi_8 = 2$, $\phi_9 = 3$, $\phi_{10} = 3$, $\phi_{11} = 2$, $\phi_{12} = 2$, $\phi_{13} = 3$, $\phi_{14} = 3$, and $\text{gcd}_{p_i} = 1$ ($i = 1, \ldots, 14$). Thus, the markings of the TWMG are partitioned into $\Phi = 419904$ subsets.

For the PSA approaches, we solve Eqs. (19), (20), and (21) by using the software Lingo and obtain the place subset P_r and the total number of partition classes of equivalent TMGs as shown in the following.

$$\text{PSA1}: \ P_r = \{p_1, p_4, p_7, p_{11}, p_{14}\}, \ \Phi' = 72,$$
$$\text{PSA2}: \ P_r = \{p_2, p_5, p_9, p_{13}, p_{14}\}, \ \Phi' = 162,$$
$$\text{PSA3}: \ P_r = \{p_2, p_4, p_8, p_{11}, p_{14}\}, \ \Phi' = 72.$$

The experiment results are carried out on a PC with a Pentium Dual-Core CPU 3.0 GHz using MATLAB with YALMIP subroutines. For a given approach X (where $X \in \{Optimal, PSA1, PSA2, PSA3, TUB\}$), we define the *optimality gap*

$$G_X = (\beta_{opt} - \beta_{\mathrm{X}})/\beta_{opt} \tag{24}$$

the difference in $\%$ between the optimal throughput β_{opt} and the throughput computed with approach X, and the *optimality gap upper bound*

$$G'_X = (\beta' - \beta_X)/\beta' \tag{25}$$

the difference in $\%$ between the upper bound on the throughput β' computed with the TUB approach and the throughput computed with approach X.

In Table I, we show the tested approach, the number of partition classes Φ (resp., Φ') that must be considered for the optimal approach (resp., PSA approaches), the obtained marking $\boldsymbol{M}$, the throughput β_X computed with approach X, the throughput upper bound β' obtained with the TUB approach, the optimality gap G_X, the optimality gap upper bound G'_X, and the CPU time for each approach. The number in parenthesis in the last column represents the computational time to determine place subset P_r.

In this example, we solve PSA1, PSA2, and PSA3 approaches by MILPP (14) with new constraints $c1$ in Eq. (22) and $c2'$ in Proposition 6 and the obtained solutions are optimal. Nevertheless, this result does not hold in general, i.e., the PSA approaches cannot always provide an optimal solution. The number of partition classes of PSA1 and PSA3 are smaller than that of PSA2. Due to the reduced number of partition classes, the CPU times required by the PSA approaches are much smaller than that of the optimal approach. The solution obtained by TUB is quite far from the optimal one.

TABLE II
SIMULATION RESULTS FOR DIFFERENT INSTANCES.

	$\lvert P\rvert$	$\lvert T\rvert$	R	Approach X	Nb. of partition classes	β_{X}	β'	G_X	G'_X	CPU time [s]
Case 1	8	5	1000	Optimal	216	0.23	0.23	0%	0%	70
				TUB	N/A	0.21		8.7%	8.7%	3
				PSA1	1	0.23		0%	0%	4 (0)
				PSA2	36	0.23		0%	0%	12 (0)
				PSA3	1	0.23		0%	0%	4 (0)
Case 2	24	12	1000	Optimal	1.00e+11	o.o.t	0.38	N/A	N/A	o.o.t
				TUB	N/A	0.36		N/A	5.3%	7
				PSA1	3456	0.31		N/A	18.4%	1097 (1)
				PSA2	10368	0.27		N/A	28.9%	4982 (1)
				PSA3	3456	0.22		N/A	42.1%	1110 (6)

We mention that the computational cost of the optimal approach and the PSA approach can be influenced by the arcs of the TWMG model tremendously. For example, if we change the production ratios for $\mathcal{PR}_1$ and $\mathcal{PR}_2$ to 2/3 and 1/3, the arcs of the command circuits will be changed accordingly, i.e., $Pre(p_6, t_6) = 2$, $Post(p_7, t_6) = 1$, $Pre(p_8, t_7) = 1$, $Post(p_9, t_7) = 2$, $Pre(p_{10}, t_8) = 2$, $Post(p_{11}, t_8) = 1$, $Pre(p_{12}, t_9) = 1$, and $Post(p_{13}, t_9) = 2$. Thus, the number of partition classes for optimal approach and PSA1/PSA2/PSA3 approaches are 64 and 4/8/4, respectively. Nevertheless, if we change the production ratios for $\mathcal{PR}_1$ and $\mathcal{PR}_2$ to $7/10$ and $3/10$, these numbers will increase to 6.004e+9 and 3087/7203/3087, respectively. ◇

B. More cases study

To better compare the efficiency of the proposed approaches for solutions obtained by the PSA approach and the optimal and TUB approaches proposed in [16], we analyzed some examples taken from literature. Case 1 is an assembly line taken from Fig. 3 in [16]. Case 2 is a jobshop taken from Fig. 6 in [6] that contains four process circuits.

In Table II, we show for each considered instance the number of places and transitions, the upper bound on the cost of resources R, the tested approach, the number of partition classes Φ (resp., Φ') for the optimal approach (resp., PSA approaches), the throughput β_X computed with approach X, the throughput upper bound β' obtained with the TUB approach, the optimality gap G_X as defined in Eq. (24), the optimality gap upper bound G'_X as defined in Eq. (25), and the CPU time for each approach. The number in parenthesis in the last column represents the computational time to determine place subset P_r.

The simulation results show the tradeoff between computational cost and quality of the solution. Note that "o.o.t" (out of time) in Table II means that the solution cannot be found within 48 hours. The computational cost of the optimal solution can grow exponentially as the net size increases. For Case 1, the PSA approaches can provide an optimal throughput which in this case coincides with the upper bound on the throughput. Actually, if we find a solution whose throughput is equal to the upper bound on the throughput by using PSA approaches, we can deduce that this solution is also optimal. For Case 2, the number of partition classes required by the optimal approach is so large that we cannot obtain a solution within a reasonable computational time. The solution obtained by the TUB approach for Case 2 is better than that of the PSA approach. It is not obvious that which approach is the best among the three PSA approaches, i.e., PSA1, PSA2, and PSA3. We observe that the upper bound gap of solutions obtained by the TUB approach for Case 2 are smaller than the one obtained by PSA approaches, which means that these solutions are closer to the optimal one than the solutions obtained by PSA approaches.

VII. Conclusion

In this paper, we deal with the performance optimization for TWMGs under infinite server semantics, which is a more general case than the previous ones [15]. The method proposed in [16] that can provide an optimal solution has a high computational cost. We consider a subset of places to allocate resources instead of taking all places into consideration and develop three practically efficient methods to reduce the computational burden.

The main restriction of TWMGs is the fact that they cannot describe systems with choice, i.e., a condition where several future evolutions are possible but in conflict among them. Future work will aim to extend the considered modelling framework by assuming that choices are possible and must be resolved with a stationary routing, that assigns resources to conflicting processes with a preassigned ratio. The routing parameters will be additional decision variable of our optimization problem.

Acknowledgement

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61374068, 61472295, and 61703321, the Science and Technology Development Fund, MSAR, under Grant No. 078/2015/A3, Shaanxi Province of China Science and Technology Research and Development Project under Grant No. 2014 GY2-04, and the Natural Science Foundation of Shaanxi University of Science and Technology under Grant Nos. 2016BJ-15 and 2017BJ-39

References

[1] S. Reveliotis, "Real-time management of complex resource allocation systems: necessity, achievements and further challenges", *Annual Reviews in Control*, vol. 41, pp.. 147-158, 2016.

[2] J. Campos, G. Chiola, and M. Silva, "Ergodicity and throughput bounds of Petri nets with unique consistent firing count vector," *IEEE Trans. Software Eng.*, vol. 17, no. 2, pp. 117-125, 1991.

[3] S. Lafit, J. M. Proth, and X. L. Xie, "Marking optimization in timed event graphs," *in Proc. Int Conf. Appl. Theory Petri Nets*, Springer, Berlin Heidelberg, pp. 281-300, 1991.

[4] A. Giua, A. Piccaluga, and C. Seatzu, "Firing rate optimization of cyclic timed event graph," *Automatica*, vol. 38, no. 1, pp. 91-103, 2002.

[5] A. Munier, "Régime asymptotique optimal d'un graphe d'événements temporisé généralisé: Application à un problème d'assemblage," *RAIPO-APII*, vol. 27, pp. 487-513, 1992.

[6] M. Nakamura and M. Silva, "Cycle time computation in deterministically timed weighted marked graphs," *in Proc. 7th IEEE Int. Conf. Emerg. Tech. Factory Autom.*, vol. 2, pp. 1037-1046, 1999.

[7] A. Benabid-Najjar, C. Hanen, O. Marchetti, and A. Munier, "Periodic schedules for bounded timed weighted event graphs," *IEEE Trans. Autom. Control*, vol. 57, no. 5, pp. 1222-1232, 2012.

[8] N. Sauer, "Marking optimization of weighted marked graphs," *Discrete Event Dyn. Syst.*, vol. 13, no. 3, pp. 245-262, 2003.

[9] Z. He, Z. W. Li, and A. Giua, "Optimization of deterministic timed weighted marked graphs," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 1084–1095, 2017.

[10] Z. He, Z. W. Li, I. Demongodin, and A. Giua, "Marking optimization of deterministic timed weighted marked graphs under infinite server semantics," *in Proc. 3rd Int. Conf. Control, Decis. Infor. Tech.*, pp. 1-6, Malta, 2016.

[11] B. Cottenceau, L. Hardouin, and J. L. Boimond, "Modeling and control of weight-balanced timed event graphs in dioids," *IEEE Trans. Autom. Control*, vol. 59, no. 5, pp. 1219-1231, 2014.

[12] R. J. Rodriguez, J. Julvez, and J. Merseguer. "On the performance estimation and resource optimization in process Petri nets," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 43, no. 6, pp. 1385-1398, 2013.

[13] R. Li and S. Reveliotis. "Performance optimization for a class of generalized stochastic Petri nets," *Discrete Event Dyn. Syst.*, vol. 25, no. 3, pp. 387-417, 2015.

[14] H. Chen, L. Amodeo, F. Chu, and K. Labadi. "Performance evaluation and optimization of supply chains modelled by batch deterministic and stochastic Petri Nets," *IEEE Trans. Autom. Sci. Eng.*, vol. 2, no. 2, pp. 132-144, 2005.

[15] Z. He, Z. W. Li, and A. Giua, "Cycle time optimization of deterministic timed weighted marked graphs by transformation," *IEEE Trans. Control Syst. Technol.*, vol. 25, no. 4, pp. 1318-1330, 2017.

[16] Z. He, Z. W. Li, and A. Giua, "Cycle time optimization for deterministic timed weighted marked graphs under infinite server semantics," *in Proc. 55th IEEE Conf. Decis. Control*, pp. 3942-3947, Las Vegas, USA, 2016.

[17] O. Marchetti and A. Munier. "Complexity results for weighted timed event graphs," *Discrete Optim.*, vol. 7, no. 3, pp. 166-180, 2010.

[18] E. Teruel, P. Chrzastowski-Wachtel, J. M. Colom, and M. Silva, "On weighted T-Systems," *Appl. Theory Petri Nets*, vol. 616, pp. 348-367, 1992.

[19] Lindo Systems Inc., Lingo User's Guide, 2011. [Online]. Available: http://www.lindo.com