Computers & Industrial Engineering 57 (2009) 660–666

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Simulation-based two-phase genetic algorithm for the capacitated re-entrant line scheduling problem ☆

Jin Young Choi [a,*], Sung-Seok Ko [b,1]

[a] Division of Industrial and Information Systems Engineering, Ajou University, Suwon 443-749, Republic of Korea
[b] Department of Industrial Engineering, Konkuk University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 19 August 2008
Received in revised form 9 December 2008
Accepted 6 January 2009
Available online 14 January 2009

Keywords:
Capacitated re-entrant line
Genetic algorithm
Chromosome
Deadlock avoidance
Banker's algorithm
Scheduling

ABSTRACT

In this paper, we develop a simulation-based two-phase genetic algorithm for the capacitated re-entrant line scheduling problem. The structure of a chromosome consists of two sub-chromosomes for buffer allocation and server allocation, respectively, while considering all possible states of the system in terms of buffer levels of workstations and assigning a preferred job stage to each component of the chromosome. As an implementation of the suggested algorithm, a job priority-based randomized policy is defined, which reflects the job priority and the properness of local non-idling in allocating buffering and processing capacity to available job instances. The algorithm is combined with a polynomial time sub-optimal deadlock avoidance policy, namely, Bankers algorithm, and the fitness of a chromosome was evaluated based on simulation. The performance of the proposed algorithm is evaluated through a numerical experiment, showing that the suggested approach holds considerable promise for providing effective and computationally efficient approximations to the optimal scheduling policy that consistently outperforms the typically employed heuristics.

1. Introduction

As manufacturing systems move into more flexible and automated environment, there are several new issues that must be systematically addressed, which in the past were resolved by the presence of a human operator. One of those issues is a deadlock problem due to the finite buffering capacity, which gives rise to a new set of control problem, referred to as the structural control (Reveliotis, Lawley, & Ferreira, 2001). The structural control seeks to ensure the logical correctness of the automated operations, i.e., deadlock-free operation, by properly managing the allocation of the system resources to the executed jobs, so that all of them are able to run to completion without the need of human intervention. Many research works on the structural control have been performed for resolving deadlock problems (Fanti & Zhou, 2004). Yet, in general, it is very difficult to find a maximally permissive, which is optimal, structural control policy since the number of states of the considered systems grows exponentially as the system configuration becomes complex. Moreover, it should be noted that system performance can be degraded seriously because of the restriction on the system behavior if a sub-optimal structural control policy is applied to the system under consideration. Therefore, we need to deliberately consider how to integrate the performance control and the structural control issues into the system operation.

As a case in point, in this work, we consider a fabrication (fab) scheduling problem in semiconductor manufacturing systems. Fab operations can be characterized by the fact that the production of each unit occurs in several job stages by re-visiting some workstations. This re-entrant nature of the line raises the problem of determining how to allocate the workstation processing capacity to the job stages competing for it, in order to optimize some prespecified performance objectives. This is the re-entrant line scheduling problem, which is the most typical abstraction for the formal modeling and analysis of the fab scheduling problem. The re-entrant line scheduling problem with infinite buffering capacity has been extensively investigated in the last decade, and many of the developed results are analytically strong (Kumar, 1994a, 1994b; Kumar & Kumar, 1996; Lu, Ramaswamy, & Kumar, 1994; Wein, 1998). A representative exposition of these results is provided in the survey paper by Kumar and Kumar (2001).

However, the results derived in past works cannot be immediately transferred to the capacitated re-entrant line (CRL) model, because of the complications arising from the blocking effect taking place in this new environment. Characteristically, Reveliotis (2000) demonstrated through a simple example that these additional material flow dynamics negate in a strong qualitative sense prior analytical results obtained through the study of the basic re-entrant line model with infinite buffering capacity and necessitate the re-examination of the problem in this new operational context,

☆ This manuscript was processed by Area Editor Maged M. Dessouky.

* Corresponding author. Tel.: +82 31 219 2422; fax: +82 31 219 1610.
E-mail addresses: choijy@ajou.ac.kr (J.Y. Choi), ssko@konkuk.ac.kr (S.-S. Ko).
[1] Tel.: +82 2 450 3528.

0360-8352/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cie.2009.01.004

while explicitly considering the logical correctness of the system behavior. Motivated by these remarks, Choi and Reveliotis (2003) developed a formal approach for the analysis and control of capacitated re-entrant lines, based on the modeling framework of Generalized Stochastic Petri Nets (GSPN). As another effort to this end, Choi and Reveliotis (2005) suggested a feature-based parametric approximation method, based on neuro-dynamic programming (NDP), and showed that the proposed approximation framework effectively integrates past results concerning the logical control of these environments by performing a numerical experiment with some prototypical problems concerning the scheduling of capacitated re-entrant lines. However, the practical applicability of these results to "real-world" applications is restricted by the fact that they require the complete enumeration of the underlying state space, and the size of it is a super-polynomial function of the elements defining the considered CRL scheduling problem.

Of particular interest to resolve these issues is the so-called genetic algorithm, which is a computational systematic heuristic method that might provide considerable promise for such scalable and efficient approximation to the optimal solution for the considered CRL scheduling problem. During last decade, there have been a significant number of research works using genetic algorithms in the area of production and operations management (Aytug, Khouja, & Vergara, 2003; Cheng, Gen, & Tsujimura, 1996; Proudlove, Vadera, & Kobbacy, 1998). However, especially for the scheduling problems, most of the research works were focused on the production control problems occurred in job shop and flexible manufacturing systems with infinite buffering capacity. Cheng et al. (1996) provides a tutorial survey of genetic algorithm design methodologies to solve job shop scheduling problems. Even for the scheduling problems in flexible manufacturing systems (FMS) with finite buffering capacity, some special structure of FMSs were considered and the optimal deadlock avoidance policies (DAPs) were applied, resulted in some computational complexities for applying to "real-world" applications.[2] Even though there are a few works for the re-entrant line scheduling problems, they do not explicitly consider the blocking effect taking place in capacitated environment (Chen, Pan, & Lin, 2008; Liu & Wu, 2004).

Motivated by the above remarks, we develop an efficient simulation-based genetic algorithm, which consists of two phases for allocating system resources, while ensuring the logical correctness of the system behavior by an efficient polynomial DAP. More specifically, a chromosome representing a scheduling policy consists of two sub-chromosomes; each of them is for buffer allocation and server allocation, respectively. They are defined to include all the information needed for controlling the system behavior; They consider all possible cases of system states at any time point w.r.t. buffer levels of workstations and assign a prioritized job stage at each possible buffer level for all workstations. A set of chromosomes are randomly selected, forming a population, and the chromosomes in the selected population are subject to an evolution process in order to get more improved solution, in terms of the pre-specified performance objective, specifically the system throughput. The proposed algorithm is evaluated by performing numerical experiments. The results indicate that the suggested approach holds considerable promise for providing effective and computationally efficient approximations to the optimal CRL scheduling policy that consistently outperforms the typically employed heuristics.

The rest of this paper is organized as follows. Section 2 formally defines the capacitated re-entrant line scheduling problem. In Section 3, the suggested genetic algorithm is described including chromosome representation and genetic operations. Section 4 assesses the capability of the suggested method through a numerical experiment. Finally, Section 5 concludes the paper with some directions for future work.

[2] They modeled the system by using petri nets and applied optimal structural control policies, which could be found by characterizing the petri net structure, to control the considered systems.

2. The CRL scheduling problem

In this section, we briefly summarize the formal definition of the CRL scheduling problem with an example, developed in previous work (Choi & Reveliotis, 2005). The basic definition of the re-entrant line (Kumar, 1994a) consists of L workstations, $W_1, W_2, \cdots, W_L$, supporting the production of a single item and each workstation $W_i, i = 1, 2, \cdots, L$, possesses S_i identical servers. The production of each unit occurs in M stages, $J_1, J_2, \cdots, J_M$, and each stage $J_j, j = 1, 2, \cdots, M$, is supported by one of the system workstations, that is to be denoted by $W(J_j)$. The re-entrant nature of the line is expressed by the fact that there exists at least one workstation W_i such that $|j : W(J_j) = W_i| \geqslant 2$, which can be characterized by $M > L$. The capacitated re-entrant line (CRL) (Choi & Reveliotis, 2003), considered in this paper, further assumes that each workstation has B_i buffer slots. Then the controls of the system consist of two phases: (i) buffer allocation and (ii) server allocation; each part visiting the workstation for the execution of some job stage is allocated one unit of buffering capacity, which it holds exclusively during its stay in the station, while blocking other parts coming into the station. Once in the station, the part competes for one of workstation servers for the execution of the requested job stage. Moreover, the part maintains hold of its allocated buffer slot while being processed. After having finished the processing of its current stage at a certain station, the part waits in its allocated buffer for transfer to the next requested station. Due to the finite buffering capacity, this transfer should be authorized by a structural control policy (SCP) (Reveliotis et al., 2001) ensuring that: (1) the destination workstation has available buffering capacity and (2) the transfer is *safe* in that it is still physically possible from the resulting state to process all running jobs to completion.

In the context of this operational framework, the considered scheduling problem for the capacitated re-entrant line (CRL) can be posed as determining how to allocate the workstation processing and buffering capacity to the competing job stages, in order to maximize the long-run system throughput, while maintaining logical correctness of the material flow, i.e., deadlock-free operation. In order to facilitate the subsequent development, without loss of generality, it is further assumed that:

- There exists an infinite amount of raw material waiting for processing at the line's input/output (I/O) station.
- The processing time of stage $j = 1, 2, \cdots, M$, is exponentially distributed with finite nonzero rate μ_j.
- The involved job transfer times are negligible when compared to the processing times.

Example. As an example, let us consider the capacitated re-entrant line depicted in Fig. 1, which consists of two workstations, W_1, W_2, with $S_1 = S_2 = 1$ and $B_1 = 3, B_2 = 2$. The production sequence is $J = \langle J_1, J_2, J_3 \rangle$, with $W(J_1) = W(J_3) = W_1$ and $W(J_2) = W_2$. Stage processing times are exponentially distributed with rate $\mu_j, j = 1, 2, 3$. For this small configuration, it can be easily seen that the system material flow will be deadlock-free as long as

$$|J_1| + |J_2| \leqslant B_1 + B_2 - 1 = 4, \tag{1}$$

where $|J_j|, j = 1, 2, 3$, denotes the number of job instances in $W(J_j)$ executing stage J_j. Under this condition, the CRL scheduling problem is to find an optimal control policy that maximizes the long-run