Computers & Industrial Engineering 113 (2017) 46–63

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Discrete time model of a two-station one-buffer serial system with inventory level-dependent operation

Elisa Gebennini *, Andrea Grassi, Cesare Fantuzzi, Bianca Rimini

Dipartimento di Scienze e Metodi dell'Ingegneria, Universitá degli Studi di Modena e Reggio Emilia, Via Amendola 2, 42122 Reggio Emilia, Italy

ARTICLE INFO

Article history:
Received 7 June 2016
Received in revised form 21 July 2017
Accepted 6 September 2017
Available online 8 September 2017

Keywords:
Transfer line
Discrete time model
Control policy
Restart level
Performance analysis

ABSTRACT

The paper deals with the discrete-time discrete-state modeling of two-station one-buffer serial systems in which the condition of the first station (operative/idle) is controlled according to the inventory level of the intermediate buffer. Briefly, the first station is forced to remain idle each time the buffer fills up until it empties to a predefined inventory level (referred to as *restarting-inventory level*). Previous works have proved that this control policy, called *restart policy*, is effective when outage costs (e.g., waste production) are generated each time the first station restarts production after an interruption.

While the works currently available in the literature assume that the buffer has to become completely empty before allowing the first station to resume production, the proposed paper develops a new analytical Markov model in which the restarting-inventory level can be greater than zero and arbitrarily set in the range $[2, N-2]$, where N is the buffer size. The proposed model is solved in closed form by means of a partitioning procedure and a solution technique described in detail. Then, the most important performance measures are obtained as a function of the restarting-inventory level, the buffer size and the reliability parameters of both stations.

Finally, some numerical results are discussed in order to validate the model and draw some concluding remarks about the values of the restarting-inventory level which maximize the effective efficiency of the system.

1. Introduction

A serial production system, or production line, is a manufacturing process in which the product passes through the same sequence of operations, and the machines and other devices (e.g., product transfer conveyors) are laid-out in the order they are used. More specifically, the focus of this paper is on *transfer* lines, that are serial production systems in which parts are moved between machines (or stations) in a synchronous manner, so that all the stations have the same available time to perform their tasks.

The performance of such production systems depends not only on the performance of the single stations, but also on their cross-interaction. In other words, flow interruptions (e.g., failures) occurring on a specific station not only affect its own performance, but can propagate their effects on stations located in the upstream and in the downstream. Thus, an operational station is said to be "starved" if it becomes idle because it does not receive products to process from the upstream, and "blocked" if it becomes idle because it is not allowed to release products in the downstream. Since starvation and blocking generate idle times, they result in a loss of performance for the production system as a whole. Hence, the need to mitigate the interaction between stations arises with the aim of reducing the probability that a failure in any station can generate idle times in the operational ones. An approach commonly adopted is to introduce buffers between stations. Buffers act as decoupling points (if they are neither empty nor full), thus avoiding starvation and blocking situations by accumulating and releasing material.

To be able to predict the performance of a serial system formed by stations and buffers, specific analytical approaches, able to take stochastic phenomena into consideration, have been developed in years by the scientific community. The first contributions were from Buzacott (1967) and Buzacott and Hanifin (1978), who modeled the basic transfer line formed by two machines separated by one finite buffer as a Markov process. A more accurate model was proposed by Schick and Gershwin (1978). Then, asynchronous and non-homogeneous lines were modeled by means of continuous-time mixed-state Markov processes, as reported Gershwin and Schick (1980). Those authors started a scientific track devoted to the modeling of the so-called *building block*, that is the basic line formed by two Markovian stations (or machines) decoupled with one finite buffer. The reader may refer to works such as Dallery and Gershwin (1992), Papadopolous, Heavey, and

* Corresponding author.
E-mail address: elisa.gebennini@unimore.it (E. Gebennini).

http://dx.doi.org/10.1016/j.cie.2017.09.007
0360-8352/

Browne (1993), Papadopoulos and Heavey (1996), Gershwin (2002), and Li and Meerkov (2008) for extensive reviews.

Some of the most recent advances in building block modeling are briefly discussed in the following. Tolio, Matta, and Gershwin (2002) presented an analytical method where each machine can fail according to different failure modes, while Colledani and Tolio (2011) added the capability to model general repair time by means of phase-type distributions. Tan and Gershwin (2009, 2011) and Colledani and Gershwin (in press) developed models able to consider general Markovian machines. Then, lines operating in multi-product environments were addressed by Colledani, Matta, and Tolio (2005, 2008). Also quality related issues were investigated by scientists. Kim and Gershwin (2005) investigated the relationship between productivity and quality and proposed a model with both quality and operational failures; the work was further extended in Kim and Gershwin (2008). Liberopoulos, Kozanidis, and Tsarouhas (2007) proposed a model in which the quality of the material trapped in the stopped stations deteriorates with time. Gebennini and Gershwin (2013) developed a model for the exact computation of the waste produced as a consequence of outages in the first machine. Other contribution were devoted to the modeling of buffer level related control policies. Gebennini, Grassi, Fantuzzi, Gershwin, and Schick (2013) and Gebennini, Grassi, and Fantuzzi (2015) developed building blocks for the discrete-time and continuous-time cases (respectively) in which the first station can be put in a controlled idle state each time the buffer gets full until it empties again. Tolio and Ratti (2013) proposed a two-machine one-buffer continuous-time model in which the system changes its behavior each time the buffer level goes above or below certain thresholds.

For what concerns the analysis of longer production systems, no exact analytical models are generally available for reasons of mathematical tractability. Hence, approximate techniques have been developed that can be classified into two main categories: *decomposition techniques* which decompose the system into as many building blocks as the number of buffers in the original line (see, e.g. Gershwin & Burman, 2000; Colledani & Tolio, 2005; Maggio, Matta, Gershwin, & Tolio, 2009), and *aggregation techniques* which replace any two-machine one-buffer sub-system by one single equivalent machine (see, e.g. Koster, 1987; Li, Meerkov, & Zhang, 2010). The interesting point is that both techniques make use of the exact models developed for the simpler building block. Consequently, building block models, such as the model proposed in this paper, are not only important for the analysis of two-station one-buffer systems, but also for the analysis of longer and more complex production lines, since they are the basic elements of the aforementioned techniques for the analysis of long systems.

Then, it is important to emphasize that evaluation models for the performance analysis of production systems are also the basic elements for optimization techniques. Such techniques aim to define the optimal design of the system (e.g., optimal buffer space allocation, throughput maximization, work-in-process minimization, etc.). To this extent, the scientific literature proposes various heuristics (Gershwin & Schor, 2000; Papadopoulos & Vidalis, 2001a, 2001b), controller design techniques (Denardo & Lee, 1996; Veatch & Wein, 1994) and methods based on artificial intelligence techniques (Lutz, Davis, & Sun, 1998; Spinellis & Papadopoulos, 2000).

Finally, it can be noted that there is a recent growing interest in modeling building blocks with Markovian stations not only in the manufacturing process, but also in the inventory control problem for multi-echelon supply chains.

To sum up, the literature about the analysis and design of serial production systems shows that great interest exists to model two-station one-buffer building blocks and to include more and more sophisticated features and operative conditions. The original contribution of this paper, which falls into the *building block* modeling track, is related to the modeling of a particular control policy, called "restart policy", that prevents the first station from processing parts each time the buffer becomes full and until it reaches a predefined inventory level. This restart policy is an important matter of discussion in production control since certain types of machines exhibit an outage cost at each production stoppage, either an internal stoppage (i.e., failure) or a stoppage caused by the interaction with the rest of the system, so that it is important to keep their blocking frequency as low as possible. Such outage cost is due to the particular technological processes those stations execute. For example, the first machine of an aseptic packaging line in the food industry, called "filling machine", fills packages with liquid dairy products (or other beverages) by means of a continuous processes which makes use of hydrogen peroxide. If the process is interrupted for any reason (e.g. a failure or a blocking event) a portion of the packaging material remains in contact with the hydrogen peroxide for too long. As a consequence, the first packages produced when the stoppage is removed must be rejected as waste. More in general, different types of stations performing a continuous process are strongly affected by outage costs as a consequence of production waste occurring at each restart.

As mentioned above, the first modeling of a restart policy in two-station one-buffer transfer lines was due to Gebennini et al. (2013). However, this paper provides a significant step ahead in modeling the effects of a restart policy by introducing the possibility to set an arbitrary restarting-inventory level $L \in [2, N-2]$, being N the buffer size. Hence, the "controlled idle" state on the first station is removed as soon as the buffer level reaches L. The possibility to tune the restarting-inventory level is important in practice since two opposite effects can be distinguished. On one hand, forcing the first station to stay idle when the buffer is emptying produces a reduction of its blocking probability and then a reduction of the expected outage costs (waste production); on the other hand, that "controlled idle" state reduces the available production time of the station and, consequently the overall throughput of the production system. Hence, it is critical to identify the proper restarting-inventory level L that makes it possible to reduce the blocking frequency of the first station (and, consequently, the production of waste) while keeping an efficient use of the buffer with respect to the productivity of the whole system.

The remainder of the paper is organized as follows. Section 2 formalizes the model and describes the partitioning procedure applied to the Markov chain representing the system of interest. Section 3 provides the model equations. Section 4 discusses the main performance measures. Section 5 derives the analytical solution. Finally, Section 6 provides some numerical results and comparisons, and Section 7 offers some concluding remarks.

2. Building block model with restarting-inventory level

The system under study, also referred as *building block*, is a transfer line made up of two stations and a finite buffer of size N in between. Both stations are unreliable and times to failure and times to repair are assumed to be distributed according to a geometric distribution. As described in Section 1 and shown in Fig. 1, an arbitrary restarting-inventory level L is here introduced, so that the first station is forced to remain idle each time the buffer fills up until it reaches that level L.

Hence, the two-station one-buffer transfer line of interest can be described by means of two Markovian behaviors, i.e., the so-called *standard operation* behavior and the *buffer drainage* behavior. When the system operates according to the *standard operation* behavior, the generic station i, for $i = 1, 2$, can fail and be repaired according to the corresponding failure probability p_i and repair probability r_i. When the first station is operational (i.e., not under repair) it processes parts continuously except when it is "blocked" because the downstream buffer is full. Similarly, the second station