Journal of Manufacturing Systems 55 (2020) 132–142

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

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

Discrete event-driven model predictive control for real-time work-in-process optimization in serial production systems

Wenchong Chen, Hongwei Liu*, Ershi Qi

College of Management and Economics, Tianjin University, Tianjin, China

ARTICLE INFO

Keywords:
Max-plus algebra
Event-driven model predictive control
Real-time WIP optimization
Disturbing events
Production loss identification

ABSTRACT

Advanced technologies (e.g., distributed sensors, RFID, and auto-identification) can gather processing information (e.g., system status, uncertain machine breakdown, and uncertain job demand) accurately and in real-time. By combining this transparent, detailed, and real-time production information with production system physical properties, an intelligent event-driven feedback control can be designed to reschedule the release plan of jobs in real-time without work-in-process (WIP) explosion. This controller should obtain the operational benefits of pull (e.g., Toyota's Kanban system) and still develop a coherent planning structure (e.g., MRPII). This paper focuses on this purpose by constructing a discrete event-driven model predictive control (e-MPC) for real-time WIP (r-WIP) optimization. The discrete e-MPC addresses three key modelling problems of serial production systems: (1) establish a max-plus linear model to describe dynamic transition behaviors of serial production systems, (2) formulate a model-based event-driven production loss identification method to provide feedback signals for r-WIP optimization, and (3) design a discrete e-MPC to generate the optimal job release plan. Based on a case from an industrial sewing machine production plant, the advantages of the discrete e-MPC are compared with the other two r-WIP control strategies: Kanban and MRPII. The results show that the discrete e-MPC can rapidly and cost-effectively reconfigure production logic. It can decrease the r-WIP without deteriorating system throughput. The proposed e-MPC utilizes the available transparent sensor data to facilitate real-time production decisions. The effort is a step forward in smart manufacturing to achieve improved system responsiveness and efficiency.

1. Introduction

In serial production systems, where a group of producing units are arranged in consecutive order and workpieces move sequentially from one producing unit to the next, throughput is influenced by varied processing times or unexpected disturbing events. To mitigate the effects of these uncertainties, work-in-process (WIP) buffers between any two adjacent producing units can be installed. If an unexpected occurrence (e.g., machine breakdown) causes production to fall behind at a producing unit, WIP buffers can prevent the downstream producing unit from starving. WIP levels fluctuate drastically with random disturbances. The input/output (I/O) control was proposed by [1] to control the real-time WIP (r-WIP). It monitors the WIP level in each process center and reduces/increases/maintains release rate based on current conditions. However, a production system may already be out of control because the I/O control takes measures only when WIP level has become excessive. Pull systems (e.g., Toyota's Kanban system) control WIP directly and coordinate job release time with the current inventory status. They can detect operational problems quickly (e.g., production shortfalls) before WIP level becomes explosive. While a rate-driven system is logistically appealing, it only addresses endogenous uncertainties and could delay the actual job release because of anticipated lack of demand for parts [2]. Considering these issues, if an I/O control can take feedback control actions by quickly detecting endogenous uncertainties in production systems, a coherent planning structure can be developed and WIP explosion can be minimized.

Today, contemporary production enterprises of all sizes rely on a range of professional information technology tools and solutions (e.g. RFID, sensor networks, Enterprise Resource Planning (ERP), and Manufacturing Execution System (MES)) for managing the different levels of their industrial systems. During the production processes, these technologies can perceive large amounts of real-time and multi-source data. For example, an automotive manufacturer in Sweden collects 50 rows of machine data per hour by MES and an average of 500,000 rows of machine data per machine, per year [3]. Using these sensor data, the production system status can be monitored (e.g., the

* Corresponding author.
E-mail address: 1016209007@tju.edu.cn (H. Liu).

https://doi.org/10.1016/j.jmsy.2020.03.002
Received 28 June 2019; Received in revised form 3 March 2020; Accepted 4 March 2020
Available online 19 March 2020

WIP level among workstations and the status of each manufacturing object). This information is significant in supporting the formulation, execution, and modification of a real-time feedback control. Nevertheless, to initiate suitable control actions to improve overall system performance (e.g., WIP level and throughput), a fundamental understanding of real-time production properties is critical. State-driven or data-driven modeling is an efficient approach to identify system's real-time performance and facilitate real-time production control. For instance, Wang et al. [4] studied the transient analysis and real-time control problem in geometric serial production systems, Zou et al. [5] and Zou et al. [6] established a data-driven mathematical model to diagnose and predict system performance. The pitfall of these approaches is the ignorance of the system input. Control theories have been researched for many years, but the research results have not been fully extended to the manufacturing domain [5]. Moreover, due to the stochastic and non-linear properties of production systems, establishing a control policy to enable a fast response to unexpected disturbances is a significant challenge. The model predictive control (MPC) has been applied in many practical process control areas by using receding optimization at every step to generate closed-loop feedback control [7]. It recently has been successfully extended to several discrete event systems, e.g., baggage handling systems and open-station assembly systems [8]. Compared with other design techniques for controllers (e.g., pole placement, LQG, H2, and H_∞), MPC can be easily extended to include additional constraints on the inputs and outputs. Although scholars have implemented various approaches for real-time analysis and control, there are some gaps overlooked in academic research:

(1) Since current state-driven and data-driven modeling approaches are constructed based on flow models to approximate production system performance, the estimation error could be great if job processing time is large, especially in some asynchronous production systems;

(2) Because most of the proposed real-time analysis and control methods ignore the effects of system input, it is difficult to extend these approaches to derive optimal release plan of jobs, and in turn, they cannot be used to provide a coherent planning structure for production systems;

(3) The proposed MPC ignores the integration of available processing sensor data and physical properties of discrete production systems, which will greatly increases the difficulty of production management since it must reconfigure shop-floor operations at every step.

To bridge these gaps, this paper proposes a discrete event-driven MPC (e-MPC) to address optimal feedback control problems in serial production systems. A model-based event-driven diagnostic approach for system performance identification is applied to provide feedback signals for the developed discrete e-MPC. The discrete e-MPC can dynamically adjust original job release plans if permanent production loss exists. The main contributions of this paper are: (1) a max-plus linear dynamic model is developed to describe discrete-event dynamic transition behaviors of serial production systems; (2) a model-based event-driven performance identification method is proposed to provide a feedback signal for the rescheduling of job release plans; and (3) a discrete e-MPC is established to regenerate job release time for r-WIP optimization. The rest of the paper is organized as follows. Section 2 reviews some related work on WIP control in industrial production systems, production system status monitoring for WIP control, and data-driven analysis and real-time control for production systems. Section 3 formulates the r-WIP optimization problem and introduces the assumptions and notations. Section 4 establishes a max-plus linear model for production systems with finite buffer capacity. Section 5 proposes a model-based event-driven approach for permanent production loss identification. Section 6 constructs the discrete e-MPC for r-WIP optimization. Section 7 introduces a case study from an industrial sewing machine production system. Section 8 summarizes conclusions and future research.

2. Related work

2.1. WIP control in industrial production systems

Variabilities (e.g., machine failures, shutdowns due to quality problems, and slowdowns due to product mix changes) exist in all production systems and can have an enormous impact on production performance [9,10]. Considering these variabilities, WIP buffers should be built among producing units to achieve desired throughput and cycle time performance (as shown by Little's Law) [11]. However, WIP buffers do not always mean benefit. Enlarged WIP buffers increase inventory cost and may obstruct the production process for a plant with limited space. The I/O control suggested by [1] is an efficient method to keep WIP buffers under control, especially in make-to-order production environments [12,13]. This controller reduces/increases the job release rate by adjusting the master production schedule. Unfortunately, the I/O control takes actions only when WIP level has become excessive. It tends to suffer from WIP explosion, long cycle time, and poor customer service [14]. Nevertheless, the general nature of I/O control, coupled with widely available commercial software (e.g., MPRII), often makes I/O control seem like the only option. Pull systems (e.g., Kanban, constant WIP, and drum-buffer-rope) do not allow WIP levels to become excessive and quickly detect problems. This may be the main reason that pull systems can work better than push systems [15]. Pull systems maintain flexibility by coordinating release rate with the current inventory status. They reduce average WIP and cycle time [16], reduce variability of cycle times [17], create pressure for quality improvement, and increase flexibility for accommodating changes [18]. Unfortunately, the pull mechanisms may delay the actual work release because of anticipated lack of demand for parts. It provides no inherent mechanism for planning raw material procurement staffing, opportunities for machine maintenance, etc. In contrast, push mechanisms for WIP control are extremely well-suited to solve these problems. The question then is how to obtain the operational benefits of pull and still develop the coherent planning structure of push.

2.2. Production system monitoring for WIP control

To avoid as many WIP explosions as possible, a feedback control should respond to unexpected disturbances in a timely manner. Production status monitoring shows the significance of modeling an efficient feedback control [19,20]. Currently, various distributed sensors are used in production systems to monitor system status in real-time. For instance, inductive proximity sensors can capture material movement and thermal sensors can capture disruption events [6]. RFID and auto-identification techniques can also provide automatic and accurate object-data-capturing [21,22]. Then, advanced technologies (e.g., internet of manufacturing thing and Cyber-physical systems, integrated RFID, sensor networks, artificial intelligence, and virtual simulation with modern management technologies) can realize dynamic perception, intelligent processing, and optimal control of production processes [23,24]. WIP monitoring and management is a key factor in shop-floor manufacturing execution. For instance, Arkan and Van Landeghem [44] and Fang et al. [21] presented RFID-enabled and IoT-enabled management platforms, respectively, to visualize and manage real-time dynamics of shop-floor WIP items. This real-time traceability and visibility of WIP and related information plays a critical role in improving shop-floor performance with better planning, scheduling, and control decisions. Based on these platforms, Cuatrecasas-Arbos et al. [25] provided a data-driven framework for WIP reduction by analyzing the relationship between WIP inventory, manufacturing lead time, and the operational variables they depend on. Unfortunately, the existing research results have only revealed the concept of a WIP management platform or explored an improvement methodology to reduce WIP levels. The management services have not been fully investigated for optimal production control (e.g., real-time production performance analytics, scheduling, and re-scheduling).

2.3. Data-driven analysis and control for production systems

Most traditional methodologies for production system analysis and control are based on steady-state analysis and long-term performance measurements [5,26]. The advantage of detailed sensor data, which is critical for real-time system control, is mostly ignored. Recently, a large number of state-driven and event-driven approaches for real-time production analysis and control have been constructed. For instance, Jia et al. [27] derived a state-based switch on/off feedback control based on analyzing system performance during transients. Pedrielli and Ju [28] constructed a state-based simulation-predictive approach to study optimal policies for complex interacting machines. These research approaches are implemented to plan optimal responses under different arrival patterns of customer orders and various types of real-time events [29]. This kind of controller is not fast enough in response to uncertain real-time production information, as new information is being generated and updated during the operations of a manufacturing system. Event-driven modeling is another part of real-time analysis and control research. Zou et al. [5]Zou et al. [6], and Zou et al. [30] studied the real-time analysis of downtime impacts and the cost of disturbing events in multi-stage production. These analytical methods are extended to generate supervisory control and distributed control for energy efficient production systems [5,31,32]. Since the input of production systems is ignored in these state-driven and event-driven analytical and control models, the optimal feedback control decisions for r-WIP optimization can be generated only by adjusting the parameters of machines. This mechanism is similar to pull systems and would not provide a coherent planning structure for whole production systems.

Control theories have been researched for a long time, but the production domain has not generally been a target for these theories [33]. A serial production system is a discrete-event dynamic system. Max-plus theory can express the non-linearities of such systems with linear equations. Maia et al. [34]Atto et al. [35]da Silva & Maia [36], and Shang et al. [37] explored the control problem of discrete-event dynamic systems. A systematic classification of the relevant studies is demonstrated in a survey by Komenda et al. [38]. However, the proposed methods do not integrate available processing sensor data with the physical properties of practical production systems. Moreover, most of the current control approaches are focused on heuristics rules or require computationally expensive time horizon algorithms, because the optimization of production systems is stochastic and non-linear [5,39]. These approaches are time-consuming and ultimately futile in responding to demand changes or unexpected variations. This characteristic will seriously hinder the effectiveness of the control in an advanced real-time production environment.

3. r-WIP optimization problem

Serial production systems are practically important structures in discrete production system design (e.g., in automotive, electronics, appliances, and aerospace systems). This paper focuses on serial production systems to study the r-WIP optimization problem. Fig. 1 shows the structure of this production system. The rectangles represent *M-1* buffers, and the circles represent *M* machines. The notations used in this paper are listed in Table 1.

Modern production systems are closely connected with information systems (e.g., ERP and MES) supported by RFID systems and various distributed sensors. Particularly, RFID techniques are used to perceive real-time system states (e.g., current input state $u(k)$, output state $y(k)$, production state $x(k)$, processing time $\sigma(k)$, and r-WIP quantity). Typical sensors (e.g., thermal sensors, pressure sensors, and various inspection sensors) are used to capture disturbing events $\vec{e_i}$ (the set of $\vec{e_i}$ is $\boldsymbol{E}$), such as machine breakdown disruptions and quality-related problems. As the real-time state of production systems can be easily perceived, it is technically important to take optimal control actions after analyzing this information. This paper proposes a discrete e-MPC for r-WIP optimization by integrating real-time information and production system physical properties. Fig. 2 shows the event-driven control logic of the discrete e-MPC, which includes three main steps.

Table 1
Notations used in this paper.

Notation	Interpretation
m_i	the i-th machine, $1 \le i \le M$.
B_i	the i-th buffer, $2 \le i \le M$.
$u(k)$	time instant at which the k-th part is fed to the system.
$x_i(k)$	time instant at which machine m_i starts to work on the k-th part.
$y(k)$	time instant at which the k-th part leaves the serial production line.
$\sigma_i(k)$	processing time of the k-th part at i-th machine.
$N_i^-(k)$	the buffer level of B_i after the k-th part entrance into B_i.
$N_i^+(k)$	the buffer level of B_i just after the k-th part leaves B_i.
N_i	the capacity of buffer B_i.
$\vec{e_i} = (j, k_i, d_i)$	a disturbing event last d_i time when machine m_j processes the k_i-th part, $j = 1,2, \ldots, M$, $i = 1,2, \ldots, n$.
N_c	the control horizon of the discrete e-MPC.
N_p	the prediction horizon of the discrete e-MPC, $N_c \le N_p$.
$PL_{\vec{e_i}}$	the permanent production loss attributed to disruption event $\vec{e_i}$.
$r(k)$	the due dates of finished products.
J	the cost criterion of the discrete e-MPC.
J_{out}	the reference tracking error corresponding to job due date $r(k)$.
J_{in}	the control effort to minimize system r-WIP.

Step 1: Discrete e-MPC for initial job release time generation. Design a discrete e-MPC for serial production systems every N_p steps. The objective is to match customer demand with the least WIP cost. Feedback the optimal release time of jobs $u(k+1)$, $u(k+2)$, …, $u(k+N_c-1)$, $u(k+N_c)$, …, $u(k+N_p)$ to MES and production systems, respectively.

Step 2: Production system performance analysis focused on permanent production loss identification. Perceive real-time system state and disturbing event $\vec{e_i}$ with RFID techniques or distributed sensors. Establish a model-based event-driven performance identification approach to determine whether a permanent production loss exists.

Step 3: Discrete e-MPC updates for r-WIP optimization. If a disturbing event causes a permanent production loss, the discrete e-MPC should be updated with new processing parameters. Generate the real-time job release plans with the new MPC model and feedback the results to MES and production systems. Return to Step 1.

To make the three intelligent control steps work efficiently, three technical challenges should be addressed: (1) establish a mathematic model to represent the dynamic behaviors of a serial production system, (2) propose a model-based event-driven production performance identification approach to determine when a permanent production loss exists, and (3) develop an event-driven feedback control – a discrete e-MPC to generate the optimal release time of jobs for r-WIP optimization. Before modeling these technical challenges, several assumptions are defined:

(1) S_{M^*} defines the last slowest machine that is closet to end-of-line,

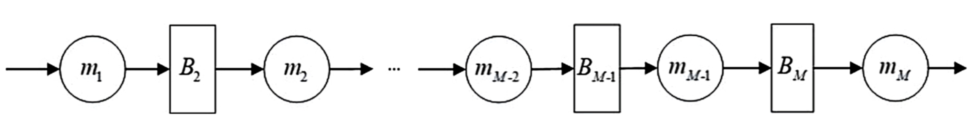

Fig. 1. A serial production line.

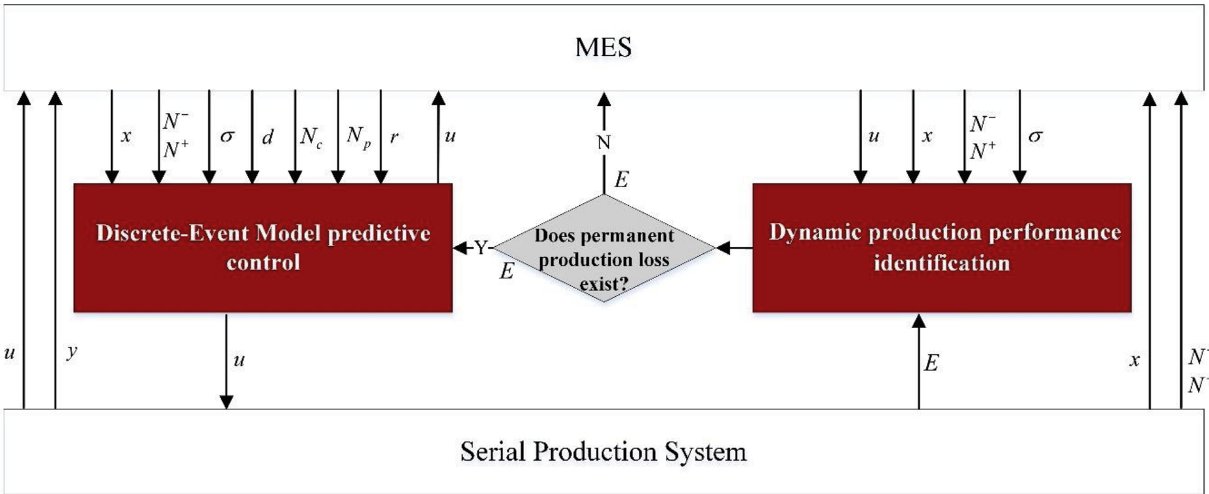

Fig. 2. Event-driven control logic of the discrete e-MPC.

assuming that there might be one or multiple slowest machines in a serial production line;

(2) The processing time of the k_i -th part at the j-th machine is $\sigma_j^{'}(k_i) = \sigma_j(k_i) + d_i$ when a disturbing event occurs;

(3) Each buffer, B_i, $i = 2, ..., M$, has a finite capacity;

(4) The disturbing events are operation dependent and detected instantaneously, e.g., a machine cannot fail when it is starved or blocked;

(5) Customer demand is greater than system production capacity, which means that serial production systems should be operated with maximum throughput;

(6) The transportation time between stations and buffers can be ignored;

(7) The sensor data is appropriately cleaned and filtered, although data cleanup is very important.

4. Modeling of serial production systems with a max-plus framework

The representation of serial production systems aims to model the multi-stage and time-varying transitions of planned and unpredictive discrete events. Such transitions are the foundation for deriving real-time system throughput and generating optimal predictive feedback control actions. Max-plus algebra is an efficient approach to represent the dynamic behaviors of such discrete-event systems [40]. The algebra is based on dioid $\mathbb{R}_{max}$ represented by structure $(\mathbb{R} \cup \{-\infty\}, \bigoplus, \bigotimes)$. $\bigoplus$ and $\bigotimes$ mean maximization and addition, respectively: $a \bigoplus b = max(a, b)$ and $a \bigotimes b = a + b$. ε denotes the zero element (ε$= -\infty$) and e denotes the unit element ($e = 0$). The matrix $\varepsilon_{m\times n}$ is $m \times n$ max-plus algebraic zero matrix: $(\varepsilon_{m\times n})_{ij} = \varepsilon$ for all , j ; E_n is the $n \times n$ max-plus algebraic identity matrix: $(E_n)_{ii} = e$ for all i and $(E_n)_{ij} = \varepsilon$ for all i, j $(i \neq j)$. Define $A, B \in R_{max}^{m\times n}$, $C \in R_{max}^{n\times p}$, then for all i, j: $(A \bigoplus B)_{ij} = a_{ij} \bigoplus b_{ij} = \max(a_{ij}, b_{ij})$, $(A \bigotimes C)_{ij} = \bigoplus_{k=1}^{n} a_{ik} \bigotimes c_{kj}$. Moreover, the max-plus algebraic matrix power of $A \in R_{max}^{n\times n}$ is defined as follows: $A^0 = E_n$ and $A^k = A \bigotimes A^{k-1}$ for $k = 1,2, ...$

4.1. Max-plus linear model for M = two-machine serial production systems

First, a simple two-machine serial production system with finite buffer capacity is used to demonstrate the modeling procedures with a max-plus framework. Timed event graph (TEG) is an effective methodology to iconify and model time varying transitions of discrete-event production systems. A two-machine serial production system is iconified as a TEG in Fig. 3. Transition x_1 represents machine m_1, x_2 represents machine m_2, and place P_2 represents the buffer B_2. P_1 is input place, and its capacity is infinite. The following assumptions are made for this simple model: (1) the buffer B_1 (P_2) capacity is finite, and it is controlled by place $P_2^{'}$; (2) the system starts from an empty condition; (3) machines can start a new cycle only when they finish the previous part and all necessary parts are available (the two machines are controlled by place $P_1^{'}$ and $P_3^{'}$). Then, the time-varying transition rules can be described as:

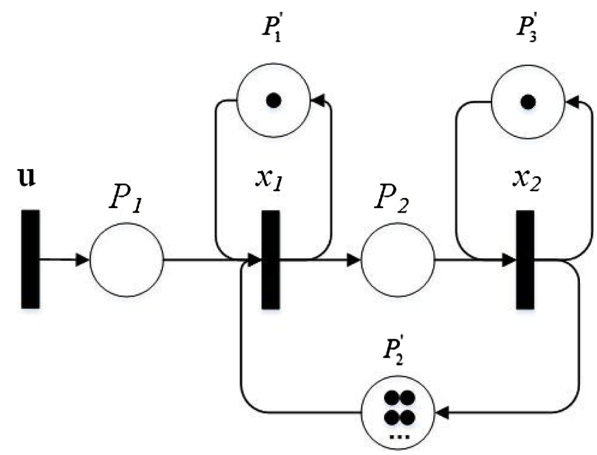

Fig. 3. Timed event graph of a two-machine serial production system.

$$x_1(k) = \sigma_1(k-1) \otimes x_1(k-1) \bigoplus \sigma_2(k-N_2) \otimes x_2(k-N_2) \bigoplus u(k), \quad (1\text{-}1)$$

$$x_2(k) = \sigma_1(k) \otimes x_1(k) \bigoplus \sigma_2(k-1) \otimes x_2(k-1). \quad (1\text{-}2)$$

According to the time-varying transition equations, if the first-input-first-output principle is obtained, the discrete-event state-space model of this serial production system can be represented as:

$$X(k) = \boldsymbol{A}(k-1) \bigotimes X(k-1) \oplus \boldsymbol{B}(k) \bigotimes u(k), \quad (2\text{-}1)$$

$$Y(k) = \boldsymbol{C}(k) \bigotimes X(k), \quad (2\text{-}2)$$

Where $X(k) = [X_1(k), X_1(k-1), ..., X_1(k-N_2+1)]^T$ $(N_2 \geq 2)$ and $X_1(k) = [x_1(k), x_2(k)]^T$. $\boldsymbol{A}(k)$, $\boldsymbol{B}(k)$, and $\boldsymbol{C}(k)$ are $2N_2 \times 2N_2$, $1 \times 2N_2$, and $2N_2 \times 1$ matrixes, respectively. The values of the three matrixes are obtained by Eq. (3).

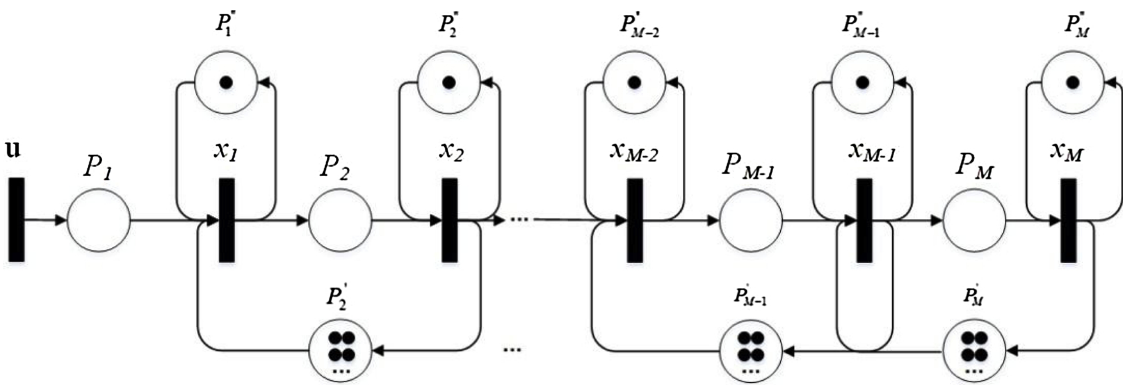

Fig. 4. Timed event graph of a M > two-machine serial production system.

$$A(k-1)=\begin{bmatrix} \sigma_1(k-1) & \varepsilon & \varepsilon & \dots & \varepsilon & \varepsilon & \sigma_2(k-N_2) \\ \sigma_1(k-1)\bigotimes\sigma_1(k) & \sigma_2(k-1) & \varepsilon & \dots & \varepsilon & \varepsilon & \sigma_2(k-N_2)\bigotimes\sigma_1(k) \\ e & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & e & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & e & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \ddots & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & e & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon & e & \varepsilon \end{bmatrix}, \tag{3-1}$$

$$\boldsymbol{B}(k)=[e\sigma_1(k)\varepsilon\dots\varepsilon]^T, \tag{3-2}$$

$$\boldsymbol{C}(k)=[\varepsilon\sigma_2(k)\varepsilon\dots\varepsilon]. \tag{3-3}$$

4.2. Max-plus linear model for M > two-machine serial production systems

Two-machine serial production systems can be extended to M > two-machine tandem systems (as shown in Fig. 4). Transition x_i represents machine m_i, and place P_i represents buffer B_i. Place $P_i^{'}$ ($2 \le i \le M$) controls the buffer capacity in place P_i, and Place $P_i^{''}$ ($1 \le i \le M$) represents that a machine can start a new cycle only when it finishes the previous part. According to the description in Section 4.1, the transition rules of a M > two-machine serial production system can be represented as:

$$\begin{aligned} x_i(k) &= \sigma_i(k-1) \otimes x_i(k-1) \bigoplus \sigma_{i+1}(k-N_{i+1}) \\ &\quad \otimes x_{i+1}(k-N_{i+1}) \bigoplus \sigma_{i-1}(k) \otimes x_{i-1}(k),\ 2 \le i \le M-1, \\ x_1(k) &= \sigma_1(k-1) \otimes x_1(k-1) \bigoplus \sigma_2(k-N_2) \otimes x_2(k-N_2) \bigoplus u(k), \\ x_M(k) &= \sigma_M(k-1) \otimes x_M(k-1) \bigoplus \sigma_{M-1}(k) \otimes x_{M-1}(k),\ k > N_{max}, \end{aligned} \tag{4}$$

Where $N_{max} = \max\{N_2, N_3, \dots, N_M\}$. According to the time-varying Eq. (4), if the first-input-first-output principle is obtained, the discrete-event state-space model of this system can be derived as follows:

$$X(k) = \bigoplus_{s=0}^{N_{max}} \boldsymbol{A_s}(k) \bigotimes X(k-s) \oplus \boldsymbol{B}(k) \bigotimes u(k), \tag{5-1}$$

$$Y(k) = \boldsymbol{C}(k) \bigotimes X(k), \tag{5-2}$$

Where $X(k)$ is a column vector of the machine status with $[x_1(k), x_2(k), \dots, x_M(k)]^T$ and $Y(k)$ is the time instant at which the k-th job leaves the serial production system. $A_0(k) \in R_{max}^{M\times M}$, $A_1(k) \in R_{max}^{M\times M}$, …, $A_{N_{max}}(k) \in R_{max}^{M\times M}$, $B(k) \in R_{max}^{M\times 1}$, and $C(k) \in R_{max}^{1\times M}$ are state transition matrixes with $A_s(k)$, $B(k)$, and $C(k)$, respectively, representing the relationship among machine states, the relationship between job release time and machine states, and the relationship between production throughput and machine states. The element $[A_s(k)]_{ij}$ is the firing time $\sigma_j(k-s)$ of machine m_j if buffer capacity N_j is equal to s and there is a transition path from machine m_j to m_i, as shown in Fig. 4; otherwise $[A_s(k)]_{ij}$ is equal to ε. The element

$[B(k)]_{i1}$ is equal to 0 if there is a system input to machine m_i and $[C(k)]_{1j}$ is equal to the firing time $\sigma_j(k)$ of machine m_j if there is a system output to this machine; otherwise, the two elements are equal to ε.

Sate-space Eq. (5) is not a standard form if we want to develop a real-time controller based on it [41]. However, because $A_0(k)$ is a strictly lower triangular matrix with an appropriate number of transitions, Eq. (5) can be transferred to a standard state-space model as follows:

$$X(k) = \overline{\boldsymbol{A}}(k-1) \bigotimes X(k-1) \bigoplus \overline{\boldsymbol{B}}(k) \bigotimes u(k), \tag{6-1}$$

$$Y(k) = \overline{\boldsymbol{C}}(k) \bigotimes X(k),\ N_{max} - 1 \le k \le K, \tag{6-2}$$

Where $X(k-1) = [X_1^T(k-1), X_1^T(k-1), \dots, X_1^T(k-N_{max})]^T$. The matrix $\overline{\boldsymbol{A}}(k)$, $\overline{\boldsymbol{B}}(k)$, and $\overline{\boldsymbol{C}}(k)$ are represented by Eq. (7). $\overline{\boldsymbol{A}}(k)$, $\overline{\boldsymbol{B}}(k)$, and $\overline{\boldsymbol{C}}(k)$ are $MN_{max} \times MN_{max}$, $MN_{max} \times 1$, and $1 \times MN_{max}$ matrixes, respectively. E is an $M \times M$ -(max, +)-identity matrix. Set $A_0^*(k) = \bigoplus_{s^{'} \ge 0} A_0^{s^{'}}(k)$ ($A_0^*(k)$ is converged because $A_0(k)$ is a strictly lower triangular), the matrix $\overline{A}_1(k)$, $\overline{A}_2(k)$, …, $\overline{A}_{N_{max}}(k)$, and $\overline{B}_0(k)$ in Eq. (7) can be generated by $\overline{A}_s(k) = A_0^*(k) \bigotimes A_s(k)$, $0 < s < N_{max}$, and $\overline{B}_0(k) = A_0^*(k) \bigotimes B(k)$.

$$\overline{\boldsymbol{A}}(k)=\begin{bmatrix} \overline{A}_1(k) & \overline{A}_2(k) & \cdots & \cdots & \overline{A}_{N_{max}}(k) \\ E & \varepsilon & \cdots & \varepsilon & \varepsilon \\ \varepsilon & E & \ddots & \cdots & \varepsilon \\ \vdots & \vdots & \ddots & \dots & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & E & \varepsilon \end{bmatrix}, \tag{7-1}$$

$$\overline{\boldsymbol{B}}(k) = [\overline{B}_0(k), \varepsilon, \dots, \varepsilon]^T, \tag{7-2}$$

$$\overline{\boldsymbol{C}}(k) = [C(k), \varepsilon, \dots, \varepsilon]^T. \tag{7-3}$$

5. r-WIP-based production loss identification approach

The discrete e-MPC adjusts job release time according to production system real-time states. For instance, when a disturbance occurs, the r-WIP in production systems may increase (The r-WIP in this paper contains the parts that are waiting for feeding. This measurement is reasonable as these parts have been transferred to line sides based on job release plans.). If the disturbance does not lead to any permanent production loss, the increased WIP can be consumed over time and the production systems will recover to the initial states. Otherwise, the excessive WIP will be kept in the production systems, and the initial release plan of jobs should be regenerated. Thus, to make the discrete e-MPC work well, a feedback control signal based on permanent production loss identification (the opportunity window $OW(k_i)$, which is elaborated in Definition 1, can be used to identify the existence of a permanent production loss) should be considered. Zou et al. [42] and Zou et al. [6] have proposed an approach for opportunity window calculation based on approximate flow models. As this paper focuses on developing discrete-event state-space models to accurately represent the dynamic behaviors of production systems, a new discrete event-driven approach for opportunity window calculation should be proposed. Before modeling the calculation approach, a definition of predecessors and successors of a place is given in Fig. 4. Based on this definition, the r-WIP $N_i^-(k)$ and $N_i^+(k)$ can be derived by Lemma 1.

Definition 1. Opportunity window $OW(k_i)$: the longest possible disturbing time on machine m_j at transition k_i that does not result in permanent production loss at the end-of-line machine, i.e.,

$$OW(k_i) = sup\{d_i \geq 0: s.\ t.\ \exists\ k^*,\ x_M(k) = x_M(k, \vec{e_i}),\ \forall\ k \geq k^* \geq k_i\}, \quad (8)$$

Where $x_M(k)$ and $x_M(k, \vec{e_i})$ are the k-th transition epoch of the end-of-line machine m_M with and without disturbing event $\vec{e_i} = (j, k_i, d_i)$, respectively. Eq. (8) means that the system can recover to initial state with finite production volume.

Definition 2. Predecessor and successor: If there is a path $x_j \to P_i \to x_l$, then transition x_j is called a predecessor of place P_i and transition x_l is called a successor of place P_i.

Lemma 1. Under the foregoing assumption (1)-(7), the r-WIP $N_i^-(k)$ and $N_i^+(k)$ can be derived as:

$$N_i^-(k) = \sum_{h=1}^{k} 1_{\{x_l(h) > x_j(k)\}}, k = 1,2,\ \ldots, \quad (9\text{-}1)$$

$$N_i^+(k) = \sum_{h=k+1}^{\infty} 1_{\{x_l(k) \leq x_j(h)\}},\ k = 1,2,\ \ldots \quad (9\text{-}2)$$

5.1. Bottleneck-based production loss identification for one-product systems

As shown in Zou et al. [6], the position at which a disturbing event occurs is vital for system diagnosis. Three conditions are considered in this part.

(1) $\vec{e_i} = (j, k_i, d_i)$ and $j > S_{M^*}$: in this condition, machine S_{M^*} stops only when buffers $B_{S_{M^*}+1}, \ldots, B_j$ are full (blockage). When S_{M^*} stops, the system will suffer a permanent production loss. According to Eq. (9), the r-WIP in place P_j at time $x_j(k_i)$ can be calculated by: $N_j^+(k_i) = \sum_{h=k_i+1}^{+\infty} 1_{\{x_j(k_i) \leq x_{j-1}(h)\}}$. Then, the state of machine $j-1$ should be $x_{j-1}(k_i + N_j^+(k_i))$, and the r-WIP in place P_{j-1} at time $x_j(k_i)$ should be: $N_{j-1}^+(k_i + N_j^+(k_i)) = \sum_{h=k_i+N_j^+(k_i)+1}^{+\infty} 1_{\{x_{j-1}(k_i+N_j^+(k_i)) \leq x_{j-2}(h)\}}$. Following these recursive steps, a proposition for r-WIP calculation in each buffer between machine S_{M^*} and machine j can be derived. According to definition 2 and proposition 1, proposition 2 regarding permanent production losses can be given.

Proposition 1. When disturbing event $\vec{e_i}$ occurs, the total WIP in place $P_{S_{M^*}+1}, \ldots,$ and P_j can be formulated by: $WIP_{S_{M^*},j}^{\vec{e_i}} = \sum_{l=S_{M^*}+1}^{j} N_l^+(k_{i,l})$, where $k_{i,j} = k_i$, $k_{i,j-1} = k_i + N_j^+(k_i)$, $k_{i,j-2} = k_i + N_j^+(k_i) + N_{j-1}^+(k_i + N_j^+(k_i))$, ..., $k_{i,S_{M^*}+1} = k_i + N_j^+(k_i) + N_{j-1}^+(k_i + N_j^+(k_i))$. $N_l^+(k_{i,l})$ can be calculated

$+ \ldots + N_{S_{M^*}+2}^+(k_i + N_j^+(k_i) + \ldots)$

by Eq. (9).

Proposition 2. When $\vec{e_i} = (j, k_i, d_i)$ occurs and $j > S_{M^*}$, if $d_i \geq \sigma_{S_{M^*}}(\sum_{l=S_{M^*}+1}^{j} N_l - WIP_{S_{M^*},j}^{\vec{e_i}})$, a permanent production loss exists and can be calculated by:

$$PL_{\vec{e_i}} = \frac{d_i - \sigma_{S_{M^*}}(\sum_{l=S_{M^*}+1}^{j} N_l - WIP_{S_{M^*},j}^{\vec{e_i}})}{\sigma_{S_{M^*}}} \quad (10)$$

(2) $\vec{e_i} = (j, k_i, d_i)$ and $j = S_{M^*}$: it is obvious that any stoppage of the last slowest machine S_{M^*} can contribute to a permanent production loss. In this situation, proposition 3 can be generated.

Proposition 3. When $\vec{e_i} = (j, k_i, d_i)$ occurs and $j = S_{M^*}$, the permanent production loss can be calculated by: $PL_{\vec{e_i}} = d_i/\sigma_{S_{M^*}}$.

(3) $\vec{e_i} = (j, k_i, d_i)$ and $j < S_{M^*}$: in this condition, machine S_{M^*} stops when buffer B_{j+1}, ..., and $B_{S_{M^*}}$ are empty (starvation). Similar to situation (1), the stoppage of machine S_{M^*} will lead to a permanent production loss of whole production systems. According to proposition 1, r-WIP, $N_{j+1}^+(k_{i,j+1})$, ..., and $N_{S_{M^*}}^+(k_{i,S_{M^*}})$, in buffer B_{j+1}, ..., and $B_{S_{M^*}}$ at time $x_j(k_i)$ can be calculated. Thus, proposition 4 can be concluded.

Proposition 4. When $\vec{e_i} = (j, k_i, d_i)$ occurs and $j < S_{M^*}$, if $d_i \geq \sigma_{S_{M^*}} WIP_{j,S_{M^*}}^{\vec{e_i}} - \sum_{l=j}^{S_{M^*}-1} \sigma_l$, a permanent production loss exists and can be calculated by:

$$PL_{\vec{e_i}} = \frac{d_i - (\sigma_{S_{M^*}} WIP_{j,S_{M^*}}^{\vec{e_i}} - \sum_{l=j}^{S_{M^*}-1} \sigma_l)}{\sigma_{S_{M^*}}} \quad (11)$$

5.2. Blockage and starvation-based production loss identification for multi-product systems

If multi-product is produced in a production system, the slowest machine should be difficult to distinguish. In this situation, if the adjacent machine m_{j-1} is blocked or machine m_{j+1} is starved by event $\vec{e_i}$, a permanent production loss may exist. To eliminate the possible negative effects, we develop a new feedback control strategy. According Eq. (9), the buffer size in place P_{j-1} and P_j at time $x_j(k_i)$ can be calculated by: $N_j^-(k_i - 1) = \sum_{h=1}^{k_i} 1_{\{x_{j+1}(h) > x_j(k_i-1)\}}$ and $N_{j-1}^+(k_i) = \sum_{h=k_i+1}^{\infty} 1_{\{x_j(k_i) \leq x_{j-1}(h)\}}$. Then, the state of machine m_{j+1} and m_{j-1} should be $x_{j+1}(k_i - N_j^-(k_i - 1))$ and $x_{j-1}(k_i - N_{j+1}^+(k_i))$, respectively. The opportunity window of a downstream machine and an upstream machine is:

$$OW_{j+1}(k_i) = \sum_{l=k_i-N_j^-(k_i-1)}^{k_i-1} \sigma_{j+1}(l) - \sigma_j(k_i), \quad (12\text{-}1)$$

$$OW_{j-1}(k_i) = \sum_{l=k_i+N_{j-1}^+(k_i)}^{k_i+B_{j-1}} \sigma_{j-1}(l). \quad (12\text{-}2)$$

If $d_i > \max\{\sum_{l=k_i-N_j^-(k_i-1)}^{k_i-1} \sigma_{j+1}(l) - \sigma_m(k_i), \sum_{l=k_i+N_{j-1}^+(k_i)}^{k_i+B_{j-1}} \sigma_{j-1}(l)\}$, a real-time feedback control should be generated.

6. Discrete e-MPC for r-WIP optimization

This discrete e-MPC performs feedback control through real-time event analysis, and in turn, two decision-making stages should be involved: formulation of time-varying discrete-event MPC and event-based MPC switching.

6.1. Time-varying discrete-event MPC

Production system states can be predicted by input sequence of jobs and real-time production data. The job due date provides a control goal to the time-varying discrete-event MPC. This MPC can then be defined as follows: find an optimal series of job release times $u(k+i)$ based on the information available at time step k to minimize the r-WIP (J_{in}) and the deviation (J_{out}) between production output $y(k+i)$ and planned job due date $r(k+i)$ at next N_p steps. According to De Schutter & Van den Boom [41], the objectives J_{in} and J_{out} can be defined as Eq. (13):

$$J_{out} = \sum_{j=1}^{N_p} max\{y(k+jk) - r(k+j), 0\}, \quad (13\text{-}1)$$

$$J_{in} = -\sum_{j=1}^{N^p} u(k+j). \quad (13\text{-}2)$$

Considering a serial production system that can be modeled by Eq. (6), the evaluation of the max-plus linear system from event step k to $k + N_p$ can be represented by Eq. (14):

$$X(k+j|k) = A_{j-1,0}(k)X(k) \bigoplus \oplus_{i=1}^{j}\ A_{j-1,i}(k)B(k+i)u(k+i), 1 \leq j \leq N_p, \quad (14\text{-}1)$$

$$Y(k+j|k) = C(k+j)A_{j-1,0}(k)X(k) \bigoplus \oplus_{i=1}^{j}\ C(k+j)A_{j-1,i}(k)B(k+i)u(k+i), 1 \leq j \leq N_p, \quad ((14)\text{-}2)$$

Where $A_{j-1,i}(k) = A(k+j-1)\ldots A(k+i)$ if $j-1 \geq i$ and $A_{j-1,i}(k) = E$

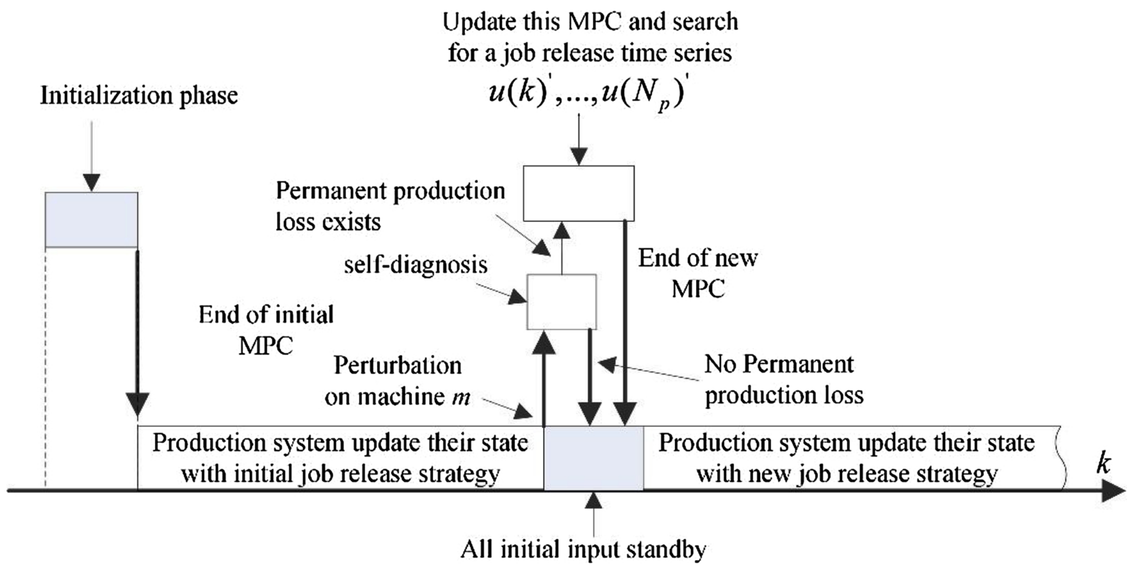

Fig. 5. Implementation structure of discrete e-MPC.

if $j-1<i$. Let $\widetilde{Y}(k)=[Y(k+1k), Y(k+2k),...,Y(k+N_pk)]^T$ and $\widetilde{u}(k)=[u(k+1), u(k+2),...,u(k+N_p)]^T$. The state-space transition equation can be transferred to $\widetilde{Y}(k)=H(k)\bigotimes X(k)\bigoplus L(k)\bigotimes\widetilde{u}(k)$, where $H(k)$ and $L(k)$ are shown in Eq. (15).

$$H(k)=\begin{bmatrix} C(k+1)A_{0,0}(k) & C(k+2)A_{1,0}(k) & \cdots & C(k+N_p)A_{N_p-1,0}(k) \end{bmatrix}^T, \tag{15-1}$$

$$L(k)=\begin{bmatrix} C(k+1)A_{0,1}(k)B(k+1) & \varepsilon & \dots & \varepsilon \\ C(k+2)A_{1,1}(k)B(k+1) & C(k+2)A_{1,2}(k)B(k+2) & \dots & \varepsilon \\ C(k+3)A_{2,1}(k)B(k+1) & C(k+3)A_{2,2}(k)B(k+2) & \dots & \varepsilon \\ \vdots & \vdots & \ddots & \vdots \\ C(k+N_p)A_{N_p-1,1}(k)B(k+1) & C(k+N_p)A_{N_p-1,2}(k)B(k+2) & \dots & C(k+N_p)A_{N_p-1,N_p}(k)B(k+N_p) \end{bmatrix}. \tag{15-2}$$

Combining the evaluation of the max-plus linear system and the two objectives in Eq. (13), the time-varying discrete-event MPC at time step k can be modelled as:

$$minJ=\lambda J_{out}+J_{in}, \tag{16-0}$$

$$\text{s.t. } \widetilde{Y}(k)=H(k)\bigotimes X(k)\bigoplus L(k)\bigotimes\widetilde{u}(k), \tag{16-1}$$

$$\Delta u(k+j)=\Delta u(k+N_c),\ N_c<\text{j}\le N_p, \tag{16-2}$$

$$a(j)<\Delta u(k+j)<b(j),\ 0<\text{j}\le N_c, \tag{16-3}$$

$$u(k+j)>0,\ \text{andisintegerfor} j=1{,}2,\ ...,N_p, \tag{16-4}$$

Where N_c is defined to reduce the number of optimization variables and λ is the weight of function J_{out}. $a(j)$ and $b(j)$ are positive numbers and are used to restrict the value intervals of variables. $\Delta u(k+j)=u(k+j)-u(k+j-1)$, $\forall j\ge 1$. Sometimes, to avoid additional production impacts or order backlog which can be caused by random disturbances, a preventive action with an r-WIP constraint is designed in this MPC, for example, the constant WIP strategy. In a max-plus framework, this constraint can be represented as $E(k)\bigotimes\widetilde{u}(k)+\sum_{i=1}^{M}\sigma_i(k)=\widetilde{y}(k)$ and $E(k)\in\mathbb{R}^{1\times N_p}$. The pre-designed r-WIP capacity can efficiently respond to several kinds of stochastic perturbations.

Eq. (16) is a nonlinear nonconvex optimization problem, as constraint (16–1) is not convex. This part uses an alternative approach based on the extended linear complementarity problem (ELCP) of Eq. (16) [43]. Consider the i-th row of Eq. (16–1) and define $\mathscr{I}_i=\{j|[L(k)]_{ij}\neq\varepsilon\}$. Then, Eq. (16–1) can be transferred into: $\widetilde{y}_i(k)=\max_{j\in\mathscr{I}_i}(l_{ij}+\widetilde{u}_j(k), g_i(k))$ or $\widetilde{y}_i(k)\ge l_{ij}+\widetilde{u}_j(k)$ for $j\in\mathscr{I}_i$ and $\widetilde{y}_i(k)\ge g_i(k)$. The extra condition that at least one inequality should hold with equality is $(\widetilde{y}_i(k)-g_i(k))\prod_{j\in\mathscr{I}_i}(\widetilde{y}_i(k)-l_{ij}-\widetilde{u}_j(k))=0$. After the extension, the time-varying discrete-event MPC can be solved quickly by many optimization algorithms.

6.2. Event-based time-varying MPC switching

Generally, MPC implements the first control sample only, and then is restarted with new system information. Since a frequent update of job release time will greatly increase the difficulty of production management, this paper proposes a discrete e-MPC for r-WIP optimization. The implementation structure of this e-MPC is shown in Fig. 5. This e-MPC can be divided into three main steps.

Step 1: Establish a discrete-event MPC with current production information, and use it to generate an initial predictive series of job release times $u(1)$, $u(2)$, …, $u(N_p)$;

Step 2: The production system updates its states via Equation (6), and perceives the real-time data from physical production systems;

Step 3: If a disturbing event occurs, do self-diagnose to understand how this random event affects system performance. If event $\vec{e_i}$ leads to a permanent production loss, switch parameters of this discrete e-MPC and generate a new series of job release times. Otherwise, update the series of job release times every N_p steps (the mechanism to update the release plan every N_p steps is reasonable if customer demand is deterministic).

If event $\vec{e_i}$ leads to a permanent production loss, set $\sigma_j^{'}(k_i)=\sigma_j(k_i)+d_i$. The matrixes $A_0(k^{'})$, …, $A_{N_{max}}(k^{'})$, $A(k^{'})$, and $B(k^{'})$ should be updated to $A_0^{'}(k^{'})$, …, $A_{N_{max}}^{'}(k^{'})$, $A^{'}(k^{'})$, and $B^{'}(k^{'})$, $k_i+N_{max}\ge k^{'}\ge k_i$. Then, the space-state equation of a serial production system can be transferred to: $X(k^{'}+1)=\boldsymbol{A}^{'}(k^{'})\bigotimes X(k^{'})\bigoplus\boldsymbol{B}^{'}(k^{'}+1)\bigotimes u(k^{'}+1)$ and $Y(k^{'}+1)=\boldsymbol{C}(k^{'}+1)\bigotimes X(k^{'}+1)$, $k^{'}\ge k_i$. Moreover, when a disturbing event $\vec{e_i}$ is perceived, $k^{''}$ parts have been released to the production system ($agr\max_{k^{''}}xu(k^{''}):=\{k^{''}|x_j(k_i)-u(k^{''})>0\}$). Then, the new MPC can update the series of job release times from $k^{''}$ to $k^{''}+N_p$.

7. Case study

To illustrate the effectiveness of the proposed discrete e-MPC, a case

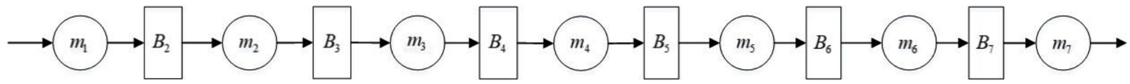

Fig. 6. A real industrial sewing machine production system.

study based on a serial production system consisting of seven machines and six buffers is proposed. The structure of this system is shown in Fig. 6. This is a segment of a real industrial sewing machine production line. To develop an ability to make real-time decisions, an RFID system, a sensor-network, and an image recognition system are applied in this line. The RFID system is constructed to collect real-time materials and jobs data, such as material shortage data, job transition data, and job position data; the sensor-based network is formulated to perceive real-time machine status data, such as machine breakdown event data (as listed in Table 2) and job processing data (as listed in Table 3); the image recognition system is designed to identify worker data (e.g., worker absence data). There is a control center that easily and rapidly conducts real-time data fusion and analysis (we do not list the data processing procedures in detail, because data volumes and how to process real-time raw data is not the research topic of this paper). Because the real-time production data can be quickly perceived and analyzed, an optimal controller for r-WIP optimization can be developed to automatically control this production system. Fig. 7 shows the timed event graph of this serial production system. To identify the parameters of operation, a three-day time study was carried out to collect data during active production hours ($K = 700$). Only one kind of product was processed during this period. The discrete e-MPC is solved by Lingo 11.0. The ELCP model of the case example is a non-linear optimization problem and can be solved by the non-linear algorithm in Lingo in ten minutes. The matrix operation in max-plus algebra is conducted in MATLAB 2016a.

7.1. Dynamic equation and discrete-event MPC model

By numbering the transitions and places from 1 to 7 from left to right (excluding input *u(k)*), the dynamic equation of this serial production line can be written as:

$$X(k+1) = \overline{\boldsymbol{A}} \otimes X(k) \oplus \overline{\boldsymbol{B}} \otimes u(k+1), \tag{17-1}$$

$$Y(k+1) = \overline{\boldsymbol{C}} \otimes X(k+1), \tag{17-2}$$

Where $X(k) = [X_1(k), X_1(k-1), X_1(k-2), X_1(k-3), X_1(k-4), X_1(k-5), X_1(k-6)]^T$, $X_1(k) = [x_1(k), x_2(k), x_3(k), x_4(k), x_5(k), x_6(k), x_7(k)]^T$. Matrix $\overline{\boldsymbol{A}}$, $\overline{\boldsymbol{B}}$, and $\overline{\boldsymbol{C}}$ can be calculated according to Equation (6) and (7). The initial condition is $\boldsymbol{X}_1(-1) = \boldsymbol{X}_1(-2) = \boldsymbol{X}_1(-3) = \boldsymbol{X}_1(-4) = \boldsymbol{X}_1(-5) = [-80 - 95 - 90 - 104 - 80 - 120 - 90]^T$ and $\boldsymbol{X}_1(0) = [080175265369449659]^T$. Let $r(k) = 659 + 120k, 1 \le k \le 700$ (the maximum throughput without any disturbing events), $N_c = 80$, $N_p = 20$, $\lambda = 200$, $a(j) = 0$ and $b(j) = 150$, $\forall$ j$\in [1, N_c]$. These parameters are set based on real production system running. Based on Equation (16), the discrete-event MPC can be written as:

$$\mathrm{Min}J = 200\sum_{j=1}^{100} max\{y(k+1k) - r(k+j), 0\} - \sum_{j=1}^{100} u(k+j), \tag{18-(0)}$$

$$\text{s.t. } \widetilde{Y}(k) = HX(k) \bigoplus L\widetilde{u}(k), \tag{18-1}$$

$$\Delta u(k+j) = \Delta u(k+80),\ 80 < \text{j} \le 100, \tag{18-2}$$

$$0 \le \Delta u(k+j) < 150,\ 0 < \text{j} \le 80, \tag{18-3}$$

$$u(k+j) > 0, \text{andisintegerfor} j = 1{,}2, \ldots, N_p, \tag{18-4}$$

Where $\widetilde{Y}(k) = [Y(k+1|k), Y(k+2|k), \ldots, Y(k+100|k)]$ and $\widetilde{u}(k) = [u(k+1|k), u(k+2|k), \ldots, u(k+100|k)]$. H and L can be calculated by Equation (15). If preventive r-WIP should be designed in this system, a r-WIP constraint, $y_{k+j} - u_{k+j} = 659 + 40j$, $1 \le j \le 18$, should be added in Equation (18).

Table 3
Processing time of machines and buffer capacity of this serial production system.

Parameters	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$
Cycle time (σ_i) (sec)	80	95	90	104	80	120	90
Buffer capacity (N_i)	–	4	2	5	2	6	3

7.2. Performance of the discrete e-MPC

To analyze the efficiency of the proposed discrete e-MPC, this part develops seven cases with different control mechanisms for comparison. These cases are: (1) $\Delta u(k) = 80$, for all $1 \le k \le 700$, (2) $\Delta u(k) = 120$, for all $1 \le k \le 700$, (3) discrete-event MPC which updates job release time every N_p steps (d-MPC), (4) discrete e-MPC without r-WIP constraints (e-MPC), (5) d-MPC with r-WIP constraints (d-MPC-WIP), (6) discrete e-MPC with r-WIP constraints (e-MPC-WIP), and (7) Kanban mechanism.

7.2.1. Permanent production loss identification and real-time control ability

First, the model-based event-driven identification method is used to diagnose the real-time system performance focused on permanent production losses under different control mechanisms. The self-diagnostic results are shown in Table 4. For example, event $\vec{e_1}$ and $\vec{e_2}$ cause permanent production loss in case 2, case 3, and case 4, while no permanent production loss exists in case 1, case 5, or case 6. The fifteen disturbing events cause permanent production losses five times in case 1, nine times in case 2, twelve times in case 3, fifteen times in case 4, eight times in case 5, six times in case 6, and six times in case 7.

Table 2
Sample disruption events list.

Perturbation event $\vec{e_i}$	Perturbation machine No.	Transition k_i	Last time d_i (sec)	Perturbation event $\vec{e_i}$	Perturbation machine No.	Transition k_i	Last time d_i (sec)
$\vec{e_1}$	5	52	300	$\vec{e_9}$	2	478	300
$\vec{e_2}$	3	70	600	$\vec{e_{10}}$	3	504	637
$\vec{e_3}$	7	72	240	$\vec{e_{11}}$	1	544	1800
$\vec{e_4}$	2	95	180	$\vec{e_{12}}$	5	587	1800
$\vec{e_5}$	6	120	600	$\vec{e_{13}}$	7	587	1205
$\vec{e_6}$	1	248	1200	$\vec{e_{14}}$	6	604	930
$\vec{e_7}$	5	302	200	$\vec{e_{15}}$	3	669	472
$\vec{e_8}$	4	429	660	–	–	–	–

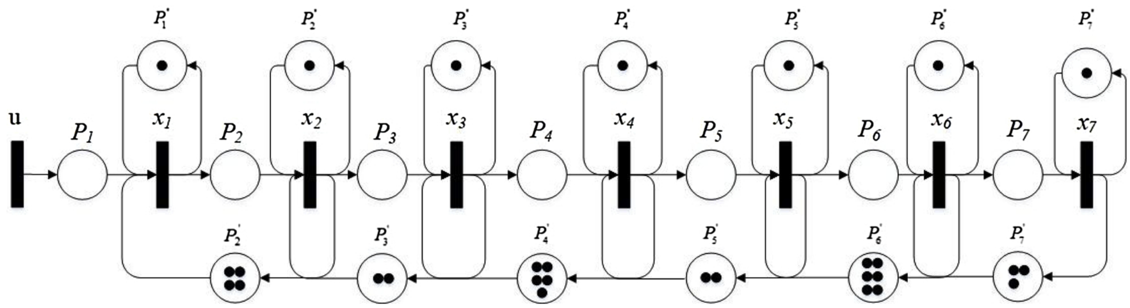

Fig. 7. Timed event graph of the real industrial sewing machine production system.

Table 4
System self-diagnosis results.

Event	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
$\vec{e}_1$	N	Y	Y	Y	N	N	N
$\vec{e}_2$	N	Y	Y	Y	N	N	N
$\vec{e}_3$	N	Y	Y	Y	Y	N	N
$\vec{e}_4$	N	N	N	Y	N	N	N
$\vec{e}_5$	Y	Y	Y	Y	Y	Y	Y
$\vec{e}_6$	Y	Y	Y	Y	Y	Y	Y
$\vec{e}_7$	N	N	Y	Y	Y	N	N
$\vec{e}_8$	N	N	Y	Y	N	N	N
$\vec{e}_9$	N	N	N	Y	N	N	N
$\vec{e}_{10}$	N	N	Y	Y	N	N	N
$\vec{e}_{11}$	N	Y	Y	Y	Y	Y	Y
$\vec{e}_{12}$	Y	Y	Y	Y	Y	Y	Y
$\vec{e}_{13}$	Y	Y	Y	Y	Y	Y	Y
$\vec{e}_{14}$	Y	Y	Y	Y	Y	Y	Y
$\vec{e}_{15}$	N	N	N	Y	N	N	N

According to Table 4, it is known that the control strategy of job release time can significantly affect the system's capacity in resisting disturbances. Because these disturbances are uncertain, a good control mechanism should not only adjust job release plans in real-time, but also maintain enough system flexibility to respond to uncertainties.

Fig. 8 shows the series of job release times corresponding to different control strategies. The x-axis is the product number (the k -th input) and the y-axis is the input gap ($\Delta u(k) = u(k+1) - u(k)$). As shown in Fig. 8, case 1 and case 2 release jobs based on *master production schedule*. They cannot update their release plan by current inventory condition. In contrast, case 3 - case 7 can dynamically adjust their release plan. However, only case 4, case 6, and case 7 can adjust the release plan via real-time system performance. Comparing cases 5–7 with cases 3–4, it is known that d-MPC-WIP, e-MPC-WIP, and Kanban can maintain certain system flexibility in responding to uncertain events. Moreover, e-MPC-WIP and Kanban show the same performance in dynamically adjusting job release time. The difference is that Kanban adjusts the release plan more frequently than e-MPC-WIP, which increases the difficulty of raw material procurement.

7.2.2. System throughput and r-WIP

In addition to the real-time control effort J_{in}, the reference tracking error J_{out} should also be considered in evaluating the performance of control strategies. Fig. 9 shows the tardiness in different cases. $y(700) - r(700)$ is equal to 4,185 s, 4,825 s, 8,085 s, 10,844 s, 4,825 s, 4,885 s, and 4,185 s in case 1-case 7, respectively, as shown in Fig. 9(a). Case 1 and case 7 (Kanban) can lead to maximum throughput when uncertainties exist. However, this performance should be attained by maintaining higher WIP buffering. Fig. 10 shows the r-WIP in the seven cases. The r-WIP at each step is calculated approximately by: $WIP(k) \longleftarrow [y(k) - r(k)]/max(\sigma_i)$. As shown in Fig. 10(a), the r-WIP in case 1 increases as a straight line with the job release plan. Fig. 10(b) is the box diagram of the r-WIP in case 2 - case 7. This figure shows that the r-WIP in case 2 and case 7 is obviously higher than in other cases. The average r-WIP in case 2 and case 7 is approximately 22 units, compared to 12 units, 6 units, 13 units, and 13 units in case 3, case 4, case 5, and case 6, respectively. The performance of d-MPC and e-MPC can be increased if preventive work-in-process is allocated in

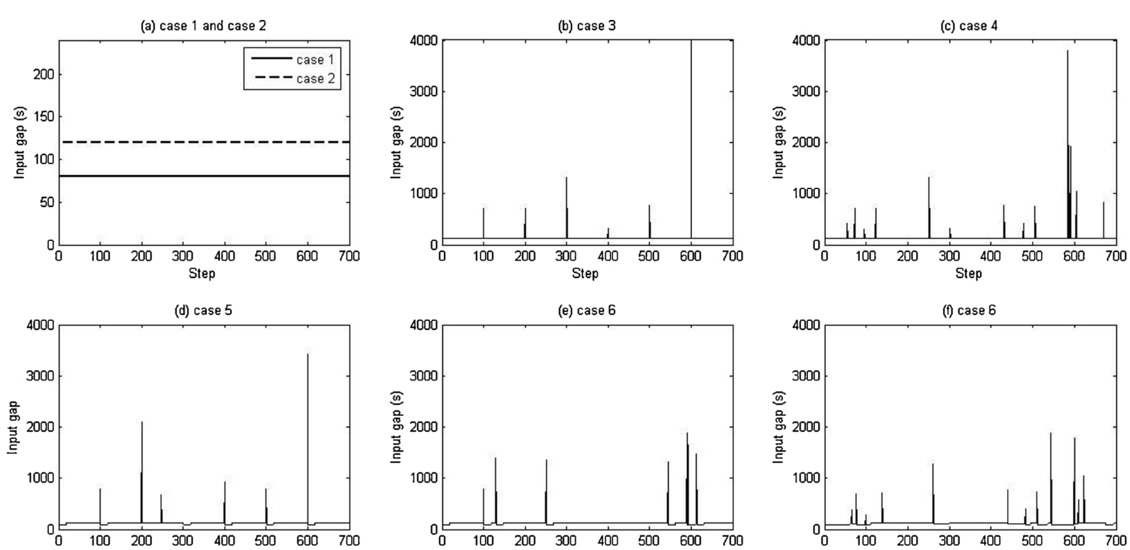

Fig. 8. Real-time system input with different control mechanisms.

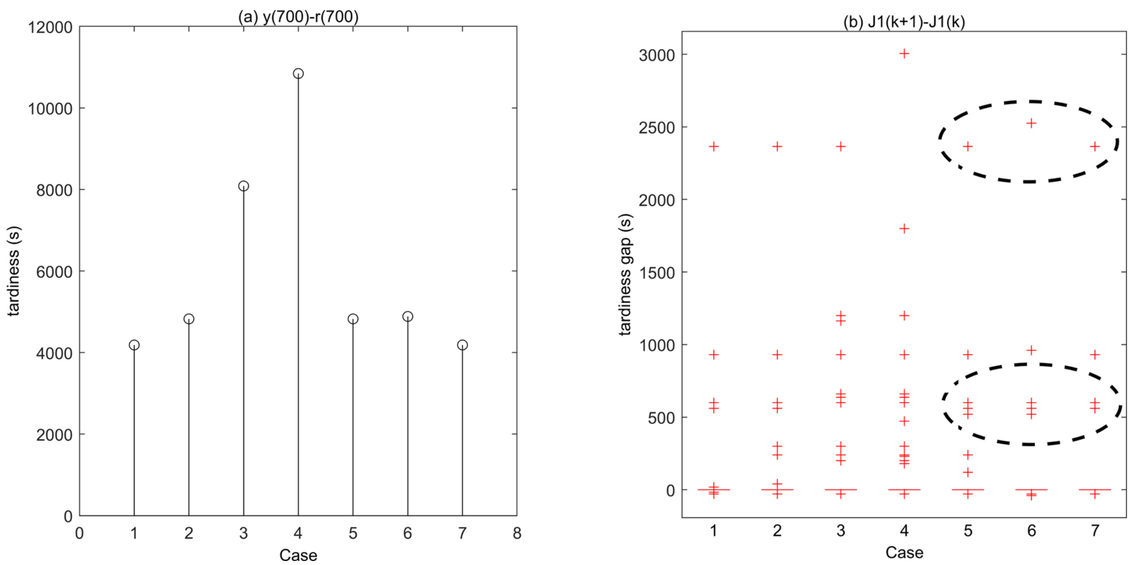

Fig. 9. System throughput with different control mechanisms.

production systems. To observe what the average r-WIP should be for e-MPC to achieve the same throughput as case 1 and case 7, several sensitive experiments on r-WIP constraint are conducted in this section. The results reveal that the e-MPC-WIP can show the same throughput as Kanban when the average r-WIP is 19 units. The average r-WIP is smaller than this index in case 1 and case 7.

Moreover, compared with Kanban, d-MPC-WIP and e-MPC-WIP show insufficient flexibility only when a disturbing event has a long duration. Fig. 9(b) shows the tardiness gap $J_{out}(k+1) - J_{out}(k)$ in different cases. Combining Table 4 with Fig. 9(b), the tardiness gap of d-MPC-WIP and e-MPC-WIP is greater than Kanban when disturbing events $\vec{e}_{11}$, $\vec{e}_{12}$, and $\vec{e}_{13}$ occur, as shown by the circles in Fig. 9(b). This phenomenon means that the throughput of production systems can be increased by d-MPC-WIP and e-MPC-WIP only if extra work-in-process is added in advance of unexpected events. Although d-MPC-WIP shows similar output performance compared with e-MPC-WIP, as shown in Fig. 9(a), its stability is obviously poor. The r-WIP in d-MPC-WIP fluctuates drastically, as shown in Fig. 10(b), and it may increase the difficulty of process management. In contrast, e-MPC-WIP can keep the WIP at a constant level by adjusting job release plans in real-time.

Through case studies, it can be demonstrated that d-MPC, e-MPC, d-MPC-WIP, e-MPC-WIP, and Kanban can efficiently adjust job release time when disturbing events occur unexpectedly. Due to the lack of process flexibility, the tracking error in d-MPC and e-MPC is greater than other cases. Most important, the e-MPC-WIP shows the greatest performance by synthetically considering the reference tracking error J_{out} and control effort J_{in}. It can obtain the operational benefits of Kanban and still develop a coherent planning structure for organization.

8. Conclusion and future work

The smart manufacturing environment drives factories to be based on data-driven production planning and control systems. Recently, with the application of professional information technologies to perceive and collect real-time production data, intelligent and automated planning and control systems are becoming easier to implement. This paper uses an e-MPC to address intelligent feedback control problems by using transparent, detailed and real-time production information. Max-plus linear models are built for serial production systems with finite buffer capacities. This mathematical model can reveal the dynamic relations between system input and system machine status. It can be utilized as a tool to iteratively derive system states at any time. Based on the max-plus linear models, model-based event-driven performance

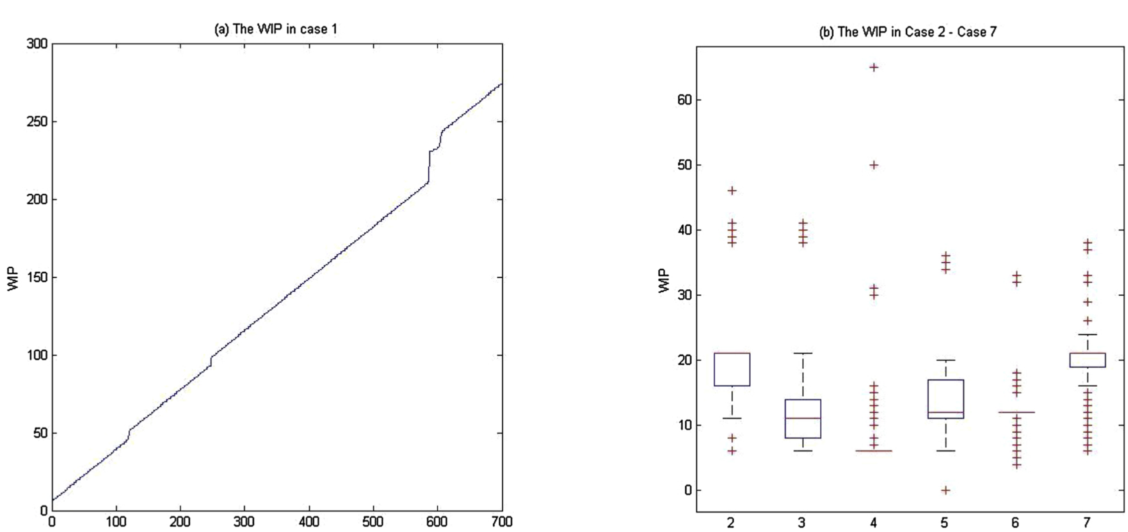

Fig. 10. r-WIP with different control mechanisms.

identification methods are proposed to diagnose the real-time permanent production loss when an unexpected disturbance occurs. Considering the physical properties of production systems, a time-varying event-driven MPC is built to address the real-time feedback control problem. Compared with other control strategies, the results of case studies reveal that the e-MPC can adjust job release plans in real-time without any throughput deterioration. This proposed e-MPC can obtain the operational benefits of pull (e.g., Kanban system) and still develop a coherent planning structure. The e-MPC integrates real-time production information with production system physical properties. It can be efficiently embedded in ERP or MES for system self-diagnosis and real-time feedback control.

This study has several limitations which provide potential opportunities for future study. First, the major challenge in e-MPC is the modeling of state-space equations. This research presents discrete-event state-space equations for serial production systems only, and other complex production structures should be addressed in the future (e.g., assembly/disassembly systems and hybrid flow shops). Second, with the development of smart manufacturing and intelligent algorithms, the status of production resources, such as machine reliability, can be predicted in real-time. This ability drives a robust predictive-reactive controller for production decisions, which could be studied in further detail.

Declaration of Competing Interest

We declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported in part by: Major Program of National Fund of Philosophy and Social Science of China under Grant 15ZDB151; National Fund of Philosophy and Social Science of China under Grant 16BGL001.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jmsy.2020.03.002.

References

[1] Wight O. Input/output control: a real handle on lead time. Pro Invent Manage J 1970;11(3):9–31.
[2] Pergher I, de Almeida AT. A multi-attribute decision model for setting production planning parameters. J Manuf Syst 2017;42:224–32.
[3] Subramaniyan M, Skoogh A, Salomonsson H, et al. A data-driven algorithm to predict throughput bottlenecks in a production system based on active periods of the machines. Comput Ind Eng 2018;125:533–44.
[4] Wang FF, Ju F, Kang NX. Transient analysis and real-time control of geometric serial lines with residence time constraints. IISE Trans 2019;51(7):709–28.
[5] Zou J, Chang Q, Arinez J, et al. Data-driven modeling and real-time distributed control for energy efficient manufacturing systems. Energy 2017;127:247–57.
[6] Zou J, Chang Q, Arinez J, et al. Dynamic production system diagnosis and prognosis using model-based data-driven method. Expert Syst Appl 2017;80:200–9.
[7] Lu JZ. Closing the gap between planning and control: a multiscale MPC cascade approach. Annu Rev Control 2015;40:3–13.
[8] Ruppert T, Dorgo G, Abonyi J. Fuzzy activity time-based model predictive control of open-station assembly lines. J Manuf Syst 2020;54:12–23.
[9] Ju F, Li J, Horst JA. Transient analysis of serial production lines with perishable products: bernoulli reliability model. IEEE Trans Automat Contr 2017;62(2):694–707.
[10] Vinod KT, Prabagaran S, Joseph OA. Dynamic due date assignment method: a simulation study in a job shop with sequence-dependent setups. J Manuf Technol Manag 2019;30(2):421–37.
[11] Perona M, Saccani N, Bonetti S, et al. Stabilisation by means of workload control: an action research and a new method. Prod Plan Control 2016;27(7–8):660–70.
[12] Biller S, Meerkov SM, Yan CB. Raw material release rates to ensure desired production lead time in Bernoulli serial lines. Int J Prod Res 2013;51(4):4349–64.
[13] Thurer M, Stevenson M, Land MJ. On the integration of input and output control: workload control order release. Int J Prod Econ 2016;174:43–53.
[14] Hong P, Leffakis ZM. Managing demand variability and operational effectiveness: case of lean improvement programmes and MRP planning integration. Prod Plan Control 2017;28(13):1066–80.
[15] Roy D, Ravikumaran V. An extensive evaluation of CONWIP-card controlled and scheduled start time based production system designs. J Manuf Syst 2019;50:119–34.
[16] Rafiei H, Rabbani M, Nazaridoust B, et al. Multi-objective cell information problem considering work-in-process minimization. Int J Adv Manuf Technol 2015;76(9–12):1947–55.
[17] Selcuk B. Adaptive lead time quotation in a pull production system with lead time responsive demand. J Manuf Syst 2013;32(1):138–46.
[18] Lander E, Liker JK. The Toyota production system and art: making highly customized and creative products the Toyota way. Int J Prod Res 2007;45(16):3681–98.
[19] Guo ZX, Ngai EWT, Yang C, et al. An RFID-based intelligent decision support system architecture for production monitoring and scheduling in a distributed manufacturing environment. Int J Prod Econ 2015;159:16–28.
[20] Cao YX, Subramaniam V, Chen RF. Performance evaluation and enhancement of multistage manufacturing systems with rework loops. Comput Ind Eng 2012;62:161–76.
[21] Fang J, Huang GQ, Li Z. Event-driven multi-agent ubiquitous manufacturing execution platform for shop floor work-in-progress management. Int J Prod Res 2013;51(4):1168–85.
[22] Yao XF, Zhang JM, Li YX, et al. Towards flexible FRID event-driven integrated manufacturing for make-to-order production. Int J Comput Integr Manuf 2018;31(3):228–42.
[23] Iarovyi S, Mohammed WM, Lobov A, et al. Cyber-physical systems for open-knowledge-driven manufacturing execution. Proc IEEE 2016;104(5):1142–54.
[24] Tao F, Qi QL, Liu A, et al. Data-driven smart manufacturing. J Manuf Syst 2018;48:157–69.
[25] Cuatrecasas-Arbos L, Fortuny-Santos J, Ruiz-de-Arbulo-Lopez P, et al. Monitoring processes through inventory and manufacturing lead time. Ind Manag Data Syst 2015;115(5):951–70.
[26] Li M, Yang F, Uzsoy R, et al. A metamodel-based Monte Carlo simulation approach for responsive production planning of manufacturing systems. J Manuf Syst 2016;38:114–33.
[27] Jia ZY, Zhang L, Arinez J, et al. Performance analysis for serial production lines with Bernoulli machines and real-time WIP-based machine switch-on/off control. Int J Prod Res 2016;54(21):6285–301.
[28] Pedrielli G, Ju F. Simulation-predictive control for manufacturing systems. 2018 IEEE 14th International Conference on Automation Science and Engineering 2018:1310–5.
[29] Arica E, Haskins C, Strandhagen JO. A framework for production rescheduling in sociotechnical manufacturing environments. Prod Plan Control 2016;27(14):1191–205.
[30] Zou J, Chang Q, Arinez J, et al. Production performance prognostics through model-based analytical method and recency-weighted stochastic approximation method. J Manuf Syst 2018;47:107–14.
[31] Chang Q, Ni J, Bandyopadhyay P, et al. Supervisory factory control based on real-time production feedback. J Manuf Sci Engin-Trans ASME 2007;129(3):653–60.
[32] Li Y, Chang Q, Ni J, et al. Event-based supervisory control for energy efficient manufacturing systems. IEEE Trans Autom Sci Eng 2018;15(1):92–103.
[33] San-Millan A, Aphale SS, Feliu V. A fast algebraic estimator for system parameter estimation and online controller tuning – a nanopositioning application. IEEE Trans Ind Electron 2019;66(6):4534–43.
[34] Maia CA, Andrade CR, Hardouin L. On the control of max-plus linear system subject to state restriction. Automatica 2011;47:988–92.
[35] Atto AM, Martinez C, Amari S. Control of discrete event systems with respect to strict duration: supervision of an industrial manufacturing plant. Comput Ind Eng 2011;61:1149–59.
[36] da Silva GG, Maia CA. On just-in-time control of timed event graphs with input constraints: a semimodule approach. Discrete Event Dyn S 2016;26(2):351–66.
[37] Shang Y, Hardouin L, Lhommeau M, et al. An integrated control strategy to solve the disturbance decoupling problem for max-plus linear systems with applications to a high throughput screening system. Automatica 2016;63:338–48.
[38] Komenda J, Lahaye S, Boimond JL, et al. Max-plus algebra in the history of discrete event systems. Annu Rev Control 2018;45:240–9.
[39] Chuang SH, Chan FTSK, Chan HK. A modified genetic algorithm approach for scheduling of perfect maintenance in distributed production scheduling. Eng Appl Artif Intell 2009;22(7):1005–14.
[40] Baccelli F, Schmidt V. Taylor expansions for Poisson driven (max, +)-linear systems. An Appl Probab 1996;6(1):138–85.
[41] De Schutter B, van den Boom T. Model predictive control for max-plus-linear discrete event systems. Automatica 2001;37(7):1049–56.
[42] Zou J, Arinez J, Chang Q, et al. Opportunity window for energy saving and maintenance in stochastic production systems. J Manuf Sci Engin-Trans ASME 2016;138(12):121009. 2016.
[43] De Schutter B, De Moor B. The extended linear complementarity problem. Math Program 1995;71(3):289–325.
[44] Arkan I, Van Landeghem H. Evaluating the performance of a discrete manufacturing process using RFID: a case study. Rob Comput Integr Manuf 2013;29(6):502–12.