Improving the Performance of MMPP/M/C Queue by Convex Optimization—A Real-World Application in Iron and Steel Industry

YANHE JIA¹, ZHE GEORGE ZHANG²,³, AND TE XU⁴,⁵
¹School of Economics and Management, Beijing Information Science & Technology University, Beijing 100192, China
²Beedie School of Business, Simon Fraser University, Burnaby, BC V5A 1S6, Canada
³Department of Decision Sciences, Western Washington University, Bellingham, WA 98225, USA
⁴Liaoning Key Laboratory of Manufacturing System and Logistics, Northeastern University, Shenyang 110819, China
⁵Institute of Industrial and Systems Engineering, Northeastern University, Shenyang 110819, China

Corresponding author: Te Xu (xute@ise.neu.edu.cn)

This work was supported in part by the Major International Joint Research Project of the National Natural Science Foundation of China under Grant 71520107004, in part by the Major Program of National Natural Science Foundation of China under Grant 71790614, in part by the Fund for Innovative Research Groups of the National Natural Science Foundation of China under Grant 71621061, and in part by the 111 Project under Grant B16009.

ABSTRACT

Now the information system of the iron and steel industry is a typical sensor cloud computing system. The system has accumulated a lot of relevant manufacturing data. Using these data can well solve the slab storage problem in the production process for the iron and steel industry. In this paper, we investigate a queueing system where customers arrive according to a Markov Modulated Poisson Process (MMPP). MMPP can describe how the arrival rate changes with the environment, which is more realistic. We develop an MMPP(3)/M/C queueing model to solve the congestion problem in the iron and steel industry. In the actual production process, the slab arrival rates vary with states, therefore MMPP is used to model the arrival process in this paper. Based on explicit performance measures, we develop a nonlinear optimization model of queueing system, and convert the model into a convex optimization problem. Through the convex optimization method, the MMPP(3)/M/C model, resulted from the practical system, can be analyzed by the M/M/C model approximately.

INDEX TERMS

Sensor-cloud systems, Markov modulated Poisson process, convex optimization, slab, crane.

I. INTRODUCTION

The iron and steel industry collects the data of production and transportation through wireless sensors in the machine shop. In addition, the data is uploaded to the cloud database. Therefore, the information system of the iron and steel industry is a typical sensor cloud computing system. Managers can read relevant data from the database through the system and optimize it. This can reduce the company’s operating costs and improve production efficiency [1].

This paper focuses on the optimization problem of slab storage in the iron and steel industry of China. In actual production, the slab arriving rate to the slab yard varies with the rate of upstream production. This situation affects the number of crane which needs to serve the arrival slabs.

A good arrangement for the arrival rate of slabs and the service rate of cranes can improve the work load of queueing system [2]. After continuous casting, slabs need to be stored in the slab yard for the next production stage. Slab storage is completed by cranes. A simplified diagram of the slab production process is shown in Fig.1. Based on the process,
we analyze the storage problem by the queueing theory. Slabs are considered as customers in the queueing system. Cranes are equivalent to servers.

We mainly analyze the optimal configuration between arrival rate and service rate in the queueing system. When the service rate is high, it will increase the efficiency of the slab storage, but some servers will be idle. This will result in low utilization of equipment and waste of resources. When the arrival rate is high, it will produce slab backlog. This will generate waiting costs. So there is a tradeoff between the arrival rate of slabs and the service rate of cranes.

Therefore, it is very important to determine a reasonable optimal arrival rate and service rate in slab storage process. This can make the performance of the queueing system optimal.

In the actual production process, the arrival rate of slab varies with steel grade. MMPP can describe how the arrival rate changes with the environment. Therefore, we use MMPP to describe the arrival process, which makes the model more realistic. And for the optimal model of MMPP/M/C queueing system about service load, the solving process is NP hard. So we discuss the approximate method to solve the problem, which can get the optimal solution easily. After a long period of investigation, the arrival rates of slabs have three levels, which can describe the arrival process very appropriately. We assume that the service time of crane obeys exponential distribution. In the production process of slabs, the service guideline is “first come first service”. During the service process, sometimes there is a situation of “two slabs by one crane”, i.e., one crane can pick up two slabs at the same time, but this situation is rare in actual production. This article analyzes the general situation, i.e., “one slab by one crane”.

II. RELATED WORK AND MAIN FRAME

About the research of slab storage, Tang et al. [3] modelled and optimized slab storage in slab warehouses, and focused on the slab stack shuffling problem in the warehouses and solved the problem by a genetic algorithm, which reduced the number of slab shuffles. Jia et al. [4] discussed the MMPP/M/C queue with thresholds, which used the matrix geometry method to optimize the problem of slab yard. They also used the matrix geometry method to optimize other problem of the iron and steel industry [5]. In this paper, we also optimize the slab storage in the warehouse, but focus on the service load of the queueing system. According to queueing models, the nonlinear optimization model is established for the queueing system. Then the MMPP(3)/M/C model can be optimized approximately by the results of M/M/C. Dohn and Clausen [6] studied the issues involving slab warehouse plans and crane scheduling, and a two-stage heuristic algorithm is used to optimize the calculation. We study the optimization problem of slab stocking through a new method. According to the queueing theory, the queueing model of slab storage is established, and the corresponding optimal service load is solved by convex optimization method.

There are many studies on the MMPP model. Baiocchi and Blefari-Melazzi [7] analysed the MMPP/G/1/K queueing model and designed a corresponding algorithm. Ching et al. [8] discussed the application of the MMPP queueing system, and proposed a hedging strategy for production control. They obtained the optimal hedging point by solving the model with a numerical algorithm to obtain the steady state distribution of the system. Ching [9] investigated the inventory system of multi-location and built a new model for the inventory system of consumable goods. The inventory system of each warehouse and main warehouse were modelled by Markov queueing network. The transshipments were described by MMPP. Shah-Heydari and Le-Ngoc [10] presented a study about Markov modulated Poisson processes, which characterized the multimedia traffic with short-term and long-term correlation. Alain et al. showed how the analysis of Markov modulated rate processes can be used to address the problem of computing the distribution of W [11]. Wang et al. [12] proposed various buffer management and congestion control mechanisms to support differentiated service-of-quality(QoS) requirements. They developed an analytical model using the non-bursty Poisson process and the bursty MMPP for a finite buffer queueing system. They extensive simulation experiments are employed to validate the accuracy of the analytical model in their work. Lu [13] discussed a MMPP whose arrival time was associated with state-dependent marks, and addressed parameter estimation through an EM algorithm and got the optimal policy. Nasr and Maddah [14] studied a MMPP system, which was a continuous inventory replenishment system. The demands varied with the environment. They proposed a heuristic algorithm to optimize the queueing system. Romano et al [15] gave an approximate method to calculate the MMPP/M/1 queueing model, which also considered the accuracy of the approximate calculation and the calculation cost. The accuracy and calculation cost are evaluated by the super exponential approximation in their work. For a comprehensive analysis of MMPP, see Fischer and Meier-Hellstern [16]. All of the above studies create MMPP models for practical problems and developed the corresponding algorithms to optimize the solutions. In this paper, we aim at the problem of slab storage in the iron and steel industry. First, an MMPP(3)/M/C queueing system is investigated. Since the special structure of MMPP(3)/M/C that the queue length and waiting time of it can just be calculated by matrix geometric method. According to the queueing theory, a nonlinear optimization model about the configuration of queueing system is established. Then we design a transformation method to convert the model into convex optimization model, which makes it easy to solve the optimization problem of multi-server queueing system.

There are few studies on the convex optimization methods to optimize queueing systems. Chiang et al [17] first provided a suite of generalized weighted fair queueing formulations for output link scheduling, where the weights can be dynamically optimized under QoS constraints using the tool of
geometric programming. Chiang et al. [18] proposed an effective nonlinear optimization system. They used convex optimization calculation tools and fast polynomial time algorithms to obtain the global optimality. Several convex structures of the queueing system were shown and followed by numerical examples for a single server queue. Neely and Modiano [19] studied the convexity of the G/G/1 queueing system. Bertsimas and Natarajan [20] used a semi-positive optimization method to analyze the steady state of the queueing system. Li and Neely [21] studied the convex optimization problem in a multi-class M/G/1 queue with controllable service rate. Chiang et al. [17] first provided a suite of generalized weighted fair queueing formulations for output link scheduling, where the weights can be dynamically optimized under QoS constraints using the tool of geometric programming. Marques et al. [22] relied on stochastic convex optimization to develop optimal algorithms that used instantaneous fading and queue length information to allocate resources at the transport (flow-control), link, and physical layers. Lardjane and Messaci [23] discussed a new numerical method. It is based on linear programming and convex optimization, is performed for the computation of the steady state solution of the queueing system. Egan and Collings [24] achieved queue stability by employing a stochastic scheduling technique, where the realizations of tuned random variables determine whether each BS transmits, in addition to the minimum SINR target. Ganesh and Anantharam [25] used large deviations estimates of the probabilities of these paths, and solved a constrained convex optimization problem to find the most likely path leading to a large queue size. Casmir and Effanga [26] discussed the use of response surface methodology to search for the optimal conditions for improving grinding process in case of convex situations in paper producing industries. Ziegler [27] considered the problem of minimizing a special convex function subject to one linear constraint, which is applied on production planning. All of the above studies are combined with convex optimization methods to analyze the queueing systems with single server. In this paper, we also use the convex optimization method to optimize the queueing system. However, the queueing system with several servers is discussed, which is more complex and realistic.

The main contributions of this paper are as follows:

1. An MMPP(3)/M/C queueing model is developed according to the slab storage in the iron and steel industry.
2. According to the queueing theory, a nonlinear optimization model is established, and a transformation method is designed to transform the model into a convex optimization model.
3. The convex optimization model established in this paper can be applied to the warehousing in the slab logistics system. By optimizing the arrival rate of slab and service rate of crane, the performance of the queueing system can be optimized.
4. The convex optimization model established in this paper can be applied to the warehousing in the slab logistics system. By optimizing the arrival rate of slab and service rate of crane, the performance of the queueing system can be optimized.

This paper is organized as follows: section 2 introduces the related work and main work. Section 3 sets up an MMPP(3)/M/C model for the slab storage in the iron and steel industry. Section 4 presents the comparison of MMPP(3)/M/C model with M/M/C and does a significance test between them. Section 5 describes the nonlinear optimization model of queueing system and the model is transformed into convex optimization model. Section 6 optimizes the problems of slab stocking in the Iron and Steel Complex. By convex optimization model, we give the policy to optimize the queueing system of slab storage. Finally, section 7 concludes.

III. THE MODEL OF SLAB STORAGE

In the iron and steel industry, after slabs being produced from the upstream production, the slabs need to be stored in warehouses and placed in a designated location by cranes. Different slabs have different positions, so the service time for them varies. In the queueing system, Slabs are considered as customers and cranes are equivalent to servers. The arriving slabs are the input in the queueing model. After the slabs being stored in the warehouse by cranes, the stored slabs are output. We assume that the service time obeys exponential distribution. The flow chart of slab production can be seen from Fig.2.

The number of cranes is $C$ in the slab yard. Because the slabs are produced at different rates from the upstream production, the slab arrival rate varies with the states. We use MMPP to describe the arrival process. Based on the investigation from the industry, and divide the slab arrival rates into the following three categories: low arrival rate $\lambda_0$, medium arrival rate $\lambda_1$, and high arrival rate $\lambda_2$. Each arrival state corresponds to a Poisson process, which has different arrival rate. These Poisson processes can transfer with each other. The transition matrix is

$$
\begin{pmatrix}
- (\alpha_1 + \alpha_3) & \alpha_1 & \alpha_3 \\
\beta_1 & - (\beta_1 + \alpha_2) & \alpha_2 \\
\beta_3 & \beta_2 & - (\beta_3 + \beta_2)
\end{pmatrix}
$$

Fig.3 shows the state transition of the MMPP(3)/M/C queueing system.
Through the diagram of state transition, the corresponding transition matrix $Q$ can be get, which is

$$Q = \begin{bmatrix}
q_0 & q_1 & \cdots & q_6 & A \\
q_1 & \ddots & & & \\
\vdots & & \ddots & & \\
A & & & \ddots & \end{bmatrix}$$

The sub-matrices are $q_0, q_1, q_6, A, A_0, A_1,$ and $A_2,$ as shown at the bottom of the next page.

When the number of customers in the system exceeds $C,$ the corresponding elements of $Q$ matrix are the same, which are $A_0, A_1$ and $A_2$ respectively.

$$\pi_i Q = 0 \quad (1)$$

$\pi_i$ is the steady state distribution of $i$ customers in the system. There is an $R$ matrix that makes $\pi_i = \pi C R^{i-C}, \ i = C + 1, C + 2, \ldots;$ therefore, we can get:

$$A_0 + RA_1 + R^2 A_2 = 0 \quad (2)$$

The normalized condition is $\sum_{i=0}^{\infty} \pi_i = 1 \quad (3).$ The steady state distribution of the MMPP(3)/M/C model can be obtained from formulas (1), (2) and (3). The average queue length of the model can be obtained by:

$$E_q = \sum_{j=0}^{C} j \pi_j e + \sum_{i=C+1}^{\infty} i \pi C R^{i-C} e \quad (3)$$

In this paper, MMPP(3)/M/C model with three states is considered, which are high, medium and low arrival rates. These three states can describe the process of slab storage more realistic. So we can get the balance equation:

$$\begin{align*}
\alpha_1 r_0 + \alpha_3 r_0 &= \beta_1 r_1 + \beta_3 r_2 \\
\beta_1 r_1 + \alpha_2 r_1 &= \alpha_1 r_0 + \beta_2 r_2 \\
\beta_2 r_2 + \beta_3 r_2 &= \alpha_3 r_0 + \alpha_2 r_1 \\
r_0 + r_1 + r_2 &= 1
\end{align*} \quad (4)$$

$r_0, r_1, r_2$ are probabilities which are respectively for customers in the high, medium and low arrival rate. By balance equation (4), we can get $r_0, r_1, r_2 \quad (5),$ as shown at the bottom of the next page.

Then we can obtain the average arrival rate:

$$\lambda = r_0 \lambda_0 + r_1 \lambda_1 + r_2 \lambda_2 \quad (6)$$

Since the average queue length of the MMPP(3)/M/C is formula (3), the waiting time is:

$$W = \frac{E_q}{\lambda} \quad (7)$$

When the waiting time of the queueing system is taken, we can use the convex optimal method to optimize the performance of the queueing system. But for the complexity of the formula (3) and (7), the convex optimal process is NP hard. Next we find a method to simplified convex optimization model.
IV. ANALYSIS AND COMPARISON BETWEEN MMPP(3)/M/C AND M/M/C

Due to the special structure of MMPP/M/C, there is no analytical expression to describe the queue length and waiting time for it. Therefore, this phenomenon motivates us to do an analysis between MMPP/M/C and M/M/C model which has specific analytical expression. Next the analysis and comparison are displayed in detail.

A. APPROXIMATION BETWEEN MMPP(3)/M/C AND M/M/C

By the above equation (3) and (7), we can get the numerical solutions of the average queue length and waiting time of the MMPP/M/C model, which are calculated by the matrix geometry method.

Fig. 4 shows the average number length for the two models with the same parameters. The number of servers C is 6; the service rate of each server is 0.1; the average arrival rate of MMPP(3)/M/C and M/M/C are same. It can be seen from the graph that the average queue length of the models are almost same.

Finally, we can also get the waiting time of MMPP/M/C and M/M/C. Fig. 5 shows the waiting times for the two models with the same parameters. It can be seen from the graph that the two models also have almost the same waiting time.

Besides doing the calculation of MMPP(3)/M/C and M/M/C, we also calculate the MMPP(3)/M/C model, while the MMPP has four states. From Fig. 6 and Fig. 7, we can see that the trend of MMPP(3)/M/C is different from M/M/C. And the value of the queue length and waiting time of the

\[
q_0 = \begin{bmatrix}
-(\lambda_0 + \alpha_1 + \alpha_3) & \alpha_1 & \alpha_3 & \lambda_0 \\
\beta_1 & -(\lambda_1 + \alpha_2 + \beta_1) & \alpha_2 & \lambda_1 \\
\beta_3 & \beta_2 & -(\lambda_2 + \alpha_3 + \beta_3) & \lambda_2
\end{bmatrix}
\]

\[q_1 = \begin{bmatrix}
\mu & -(\mu + \lambda_0 + \alpha_1 + \alpha_3) & \alpha_1 & \alpha_3 & \lambda_0 \\
\mu & \beta_1 & -(\mu + \lambda_1 + \alpha_2 + \beta_1) & \alpha_2 & \lambda_1 \\
\beta_3 & \beta_2 & -(\mu + \lambda_2 + \beta_2 + \beta_3) & \lambda_2
\end{bmatrix}
\]

\[q_6 = \begin{bmatrix}
6\mu & -(6\mu + \lambda_0 + \alpha_1 + \alpha_3) & \alpha_1 & \alpha_3 & \lambda_0 \\
6\mu & \beta_1 & -(6\mu + \lambda_1 + \alpha_2 + \beta_1) & \alpha_2 & \lambda_1 \\
6\mu & \beta_3 & -(6\mu + \lambda_2 + \beta_2 + \beta_3) & \lambda_2
\end{bmatrix}
\]

\[A_0 = \begin{bmatrix}
A_0 & A_1 & A_2
\end{bmatrix}
\]

\[A_0 = \begin{bmatrix}
6\mu \\
6\mu \\
6\mu
\end{bmatrix}
\]

\[A_1 = \begin{bmatrix}
-(6\mu + \lambda_0 + \alpha_1 + \alpha_3) & \alpha_1 & \alpha_3 \\
\beta_1 & -(6\mu + \lambda_1 + \alpha_2 + \beta_1) & \alpha_2 \\
\beta_3 & \beta_2 & -(6\mu + \lambda_2 + \beta_2 + \beta_3)
\end{bmatrix}
\]

\[A_2 = \begin{bmatrix}
\lambda_0 & \lambda_1 & \lambda_2
\end{bmatrix}
\]

\[r_0 = \frac{\alpha_2 \beta_3 + \beta_1 \beta_2 + \beta_1 \beta_3}{\alpha_1 \alpha_2 + \alpha_1 \beta_2 + \alpha_1 \beta_3 + \alpha_2 \alpha_3 + \alpha_2 \beta_3 + \alpha_3 \beta_1 + \alpha_3 \beta_2 + \beta_1 \beta_2 + \beta_1 \beta_3}
\]

\[r_1 = \frac{\alpha_1 \alpha_2 + \alpha_1 \beta_2 + \alpha_1 \beta_3 + \alpha_2 \alpha_3 + \alpha_2 \beta_3 + \alpha_3 \beta_1 + \alpha_3 \beta_2 + \beta_1 \beta_2 + \beta_1 \beta_3}{\alpha_1 \alpha_2 + \alpha_2 \alpha_3 + \alpha_3 \beta_1}
\]

\[r_2 = \frac{\alpha_1 \alpha_2 + \alpha_1 \beta_2 + \alpha_1 \beta_3 + \alpha_2 \alpha_3 + \alpha_2 \beta_3 + \alpha_3 \beta_1 + \alpha_3 \beta_2 + \beta_1 \beta_2 + \beta_1 \beta_3}{\alpha_1 \alpha_2 + \alpha_1 \beta_3 + \alpha_2 \alpha_3 + \alpha_2 \beta_3 + \alpha_3 \beta_1 + \alpha_3 \beta_2 + \beta_1 \beta_2 + \beta_1 \beta_3}
\]
B. SIGNIFICANCE TEST BETWEEN MMPP(3)/M/C AND M/M/C MODEL

Since the above results are obtained when the average arrival rate and the service rate are same, in this section, we will use the F-test to analyze the significance between the different models (MMPP(3)/M/C and M/M/C).

First, we use SPSS to analyze the queue length of MMPP(3)/M/C and M/M/C model, we can get the detail information from Table 1 and Table 2.

Before doing the F-test, we need to check the homogeneity of variance for the data sets. In Table 1, we find that $P = 0.982 > 0.05$, which means the variance has the homogeneity. Then we can do the F-test in the next step.

From Table 2, we see that $P = 0.99 > 0.05$, which means there is no significant difference for the queue length between the MMPP(3)/M/C model and M/M/C model.

Similarly, we use F-test to analyze the waiting time of MMPP(3)/M/C model and M/M/C model. From Table 4, we find that $P = 0.988 > 0.05$. It means there is no significant difference for the waiting time between the MMPP(3)/M/C model and M/M/C model.

In the M/M/C queueing system, the customer arrival process is Poisson process. When $n$ of binomial distribution is large and $p$ is small, the binomial distribution can be approximated by a Poisson distribution. The Poisson distribution is suitable for describing the random events occurring in a unit of time (or space). The Poisson distribution is one of the most important discrete distributions, and many arrival processes can be approximated by the Poisson process.

In the MMPP(3)/M/C queueing system, the customer’s arrival process obeys the Markov modulated Poisson process. The M/M/C model and MMPP(3)/M/C model are multi-servers queueing systems which the distribution of service time is exponential distribution. The Markov modulated Poisson process in the paper has only three states, which may not

### TABLE 1. Homogeneity of variance test for queue length.

| Homogeneity of variance test | Levine statistics | Degree of freedom 1 | Degree of freedom 2 | Statistical significance |
|------------------------------|-------------------|---------------------|---------------------|--------------------------|
| Queue length                 |                   |                     |                     |                          |
|                               | 0.001             | 1                   | 108                 | 0.982                    |
change the properties of Poisson distribution greatly. Therefore, through analysis and comparison, we can conclude: in general, the M/M/C queueing model can be used to approximate the MMPP(3)/M/C queueing model.

V. CONVEX OPTIMIZATION METHOD TO OPTIMIZE THE QUEUEING SYSTEM

From above analysis, we acknowledge that, the MMPP(3)/M/C model can be effectively approximated by M/M/C. Next, we establish a queueing system optimization model about the service load based on the M/M/C model, and then we can get the optimal policy of MMPP(3)/M/C model approximately.

In the iron and steel industry, the number of cranes in the slab yard is represented as \( C \); \( \lambda \) is the arrival rate; the service rate of the crane is \( \mu \); the time that slabs stay in the queueing system is \( W \); the time that slabs wait for storage is \( W^q \); the queue length is \( L^q \). Based on M/M/C queueing system, an optimal model of queueing system about service load is developed.

\[
\min_{\lambda, \mu} \frac{C\mu}{\lambda} \\
\text{s.t.} \begin{cases} 
W \leq W_{\text{max}} \\
W^q \leq W^q_{\text{max}} \\
L^q \leq L^q_{\text{max}} \\
\lambda \geq \lambda_{\text{min}} \\
\mu \leq \mu_{\text{max}} 
\end{cases} 
\]

The specific expressions in the model are as follows:

\[
W = \frac{P_0 \left( \frac{\lambda}{\mu} \right)^{C+1}}{C! \left( 1 - \frac{\lambda}{\mu} \right)^2 \lambda} \\
W^q = \frac{P_0 \left( \frac{\lambda}{\mu} \right)^{C+1}}{C! \left( 1 - \frac{\lambda}{\mu} \right)^2 \lambda} + \frac{1}{\mu} \\
L^q = \frac{P_0 \left( \frac{\lambda}{\mu} \right)^{C+1}}{C! \left( 1 - \frac{\lambda}{\mu} \right)^2} \\
P_0 = \frac{r^C}{C! (1 - \rho) + \sum_{n=0}^{C-1} \frac{r^n}{n!}}^{-1}, \frac{r}{C} = \rho < 1, \rho = \frac{\lambda}{C\mu}
\]

It can be seen from the model (8) that the objective function and the constraints of it are all non-linear functions with respect to \( \lambda \) and \( \mu \). Even the simplest queueing system (M/M/1) is hard-pressed to optimize the performance. Fortunately, we can use the convexity of the queueing system to solve this problem, and the calculation time is polynomial time. Subsequently, we will do some conversion, and get the convex optimization model. First, an intermediate variable \( t_1 \) is introduced; the first inequality of model (8) can be
transformed to the followings:

\[
\begin{align*}
  P_0 \left( \frac{\lambda}{\mu} \right)^{C+1} &\leq W_{\text{max}} \\
  C! \left( 1 - \frac{\lambda}{\mu} \right)^2 \lambda \\
  P_0 \left( \frac{\lambda}{\mu} \right)^{C+1} &\leq \left( 1 - \frac{\lambda}{\mu} \right)^2 \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! W_{\text{max}}}{C! W_{\text{max}}} &\leq t_1 \leq \left( 1 - \frac{\lambda}{\mu} \right)^2 \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! W_{\text{max}}}{C! W_{\text{max}}} &\leq \lambda \mu^{-1} + t_1 \leq 1
\end{align*}
\]

Let \( t_1 \) be a constant number.

\[
\frac{P_0 \lambda^C \mu^{-(C+1)}}{C! W_{\text{max}}} \leq t_1 \leq \left( 1 - \frac{\lambda}{\mu} \right)^2
\]

Then we can get

\[
\begin{align*}
  \frac{P_0 \lambda^C \mu^{-(C+1)}}{C! W_{\text{max}}} &\leq t_1 \\
  t_1 &\leq \left( 1 - \frac{\lambda}{\mu} \right)^2 \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! (1 - \frac{\lambda}{\mu})^2}{C! W_{\text{max}}} &\leq W_{\text{max}} \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! (1 - \frac{\lambda}{\mu})^2 \mu^{-1}}{C! W_{\text{max}}} &\leq \left( 1 - \frac{\lambda}{\mu} \right)^2 \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! (1 - \frac{\lambda}{\mu})^2 \mu^{-1}}{C! W_{\text{max}}} &\leq \left( 1 - \frac{\lambda}{\mu} \right)^2 \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! (1 - \frac{\lambda}{\mu})^2 \mu^{-1}}{C! W_{\text{max}}} &\leq t_2 \leq \left( 1 - \frac{\lambda}{\mu} \right)^2 \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! (1 - \frac{\lambda}{\mu})^2 \mu^{-1}}{C! W_{\text{max}}} &\leq t_2 \\
  \lambda \mu^{-1} + \frac{1}{\mu} &\leq \lambda \mu^{-1} + t_2 \leq 1
\end{align*}
\]

Next, we change the second inequality in the constraints of the model (8). The difference from the first inequality is that the left side of this inequality is a polynomial, which needs to be divided before the transformation. In addition to the introduction of variable \( t_2 \), it is also necessary to introduce a variable \( t_3 \) to transform the model. Both \( t_2 \) and \( t_3 \) are positive real numbers. The specific inequality transformation process is as follows:

\[
\begin{align*}
  \frac{P_0 \left( \frac{\lambda}{\mu} \right)^{C+1}}{C! \left( 1 - \frac{\lambda}{\mu} \right)^2 \lambda} &\leq \lambda \mu^{-1} \leq W_{\text{max}}^{\lambda} \\
  P_0 \lambda^C \mu^{-(C+1)} &\leq \left( 1 - \frac{\lambda}{\mu} \right)^2 \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! (1 - \frac{\lambda}{\mu})^2}{C! W_{\text{max}}} &\leq \left( 1 - \frac{\lambda}{\mu} \right)^2 \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! (1 - \frac{\lambda}{\mu})^2 \mu^{-1}}{C! W_{\text{max}}} &\leq \left( 1 - \frac{\lambda}{\mu} \right)^2 \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! (1 - \frac{\lambda}{\mu})^2 \mu^{-1}}{C! W_{\text{max}}} &\leq \left( 1 - \frac{\lambda}{\mu} \right)^2 \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! (1 - \frac{\lambda}{\mu})^2 \mu^{-1}}{C! W_{\text{max}}} &\leq t_2 \leq \left( 1 - \frac{\lambda}{\mu} \right)^2 \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! (1 - \frac{\lambda}{\mu})^2 \mu^{-1}}{C! W_{\text{max}}} &\leq t_2 \\
  \lambda \mu^{-1} + \frac{1}{\mu} &\leq \lambda \mu^{-1} + \frac{1}{\mu} \leq 1
\end{align*}
\]

Finally, a group of inequalities can be obtained:

\[
\begin{align*}
  \frac{P_0 \left( \frac{\lambda}{\mu} \right)^{C+1}}{C! \left( 1 - \frac{\lambda}{\mu} \right)^2 \lambda} &\leq \lambda \mu^{-1} \leq W_{\text{max}}^{\lambda} \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! \left( 1 - \frac{\lambda}{\mu} \right)^2}{C! W_{\text{max}}} &\leq \left( 1 - \frac{\lambda}{\mu} \right)^2 \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! \left( 1 - \frac{\lambda}{\mu} \right)^2 \mu^{-1}}{C! W_{\text{max}}} &\leq \left( 1 - \frac{\lambda}{\mu} \right)^2 \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! \left( 1 - \frac{\lambda}{\mu} \right)^2 \mu^{-1}}{C! W_{\text{max}}} &\leq \left( 1 - \frac{\lambda}{\mu} \right)^2 \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! \left( 1 - \frac{\lambda}{\mu} \right)^2 \mu^{-1}}{C! W_{\text{max}}} &\leq t_2 \leq \left( 1 - \frac{\lambda}{\mu} \right)^2 \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! \left( 1 - \frac{\lambda}{\mu} \right)^2 \mu^{-1}}{C! W_{\text{max}}} &\leq t_2 \\
  \lambda \mu^{-1} + \frac{1}{\mu} &\leq \lambda \mu^{-1} + \frac{1}{\mu} \leq 1
\end{align*}
\]

While transforming the constraints above, according to the characteristics of the convex optimization objective function, we transform the objective function into \( \min \frac{C \mu \lambda^{-1}}{\lambda} \).

The similar transformation is done for the other inequalities. The model (8) can be transformed into:

\[
\begin{align*}
  \min \lambda \mu^{-1} \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! \left( 1 - \frac{\lambda}{\mu} \right)^2}{C! W_{\text{max}}} &\leq \lambda \mu^{-1} \\
  \frac{P_0 \lambda^C \mu^{-(C+1)} + C! \left( 1 - \frac{\lambda}{\mu} \right)^2 \mu^{-1}}{C! W_{\text{max}}} &\leq \lambda \mu^{-1} + t_2 \leq 1
\end{align*}
\]

The model (10) obtained by transformation is a geometric programming model. At this time, the MATLAB toolkit can
be used to optimize the model. In order to more intuitively reflect the characteristics of the convex model from the model, we let $\lambda = e^{\lambda}$, $\mu = e^{\mu}$, $t_1 = e^{t_1}$, $t_2 = e^{t_2}$, $t_3 = e^{t_3}$, $t_4 = e^{t_4}$. Then the final convex optimization model can be obtained:

$$
\begin{align*}
\min_{\tilde{\lambda}, \tilde{\mu}, \tilde{t}_1, \tilde{t}_2, \tilde{t}_3, \tilde{t}_4} & C e^{\tilde{\lambda} - (C + 1) \tilde{t}_1 - \tilde{t}_1} \leq 1 \\
& \frac{P_0}{C W_{\text{max}}} e^{\tilde{\lambda} - (C + 1) \tilde{t}_2 + \frac{1}{W_{\text{max}}}} e^{2\tilde{t}_2 - \tilde{t}_2} \leq 1 \\
& \frac{P_0}{C W_{\text{max}}} e^{\tilde{\lambda} - (C + 1) \tilde{t}_3 - \tilde{t}_3} \leq 1 \\
\text{s.t.} & \tilde{\lambda} \geq \ln \lambda_{\text{min}} \\
& \tilde{\mu} \leq \ln \mu_{\text{max}} \\
\end{align*}
$$

Obviously, the model (10) is one kind of convex optimization model, which can be directly solved by various convex optimization packages. Although the number of variables and constraints of the convex optimization model increase, the convex optimization calculation is still fast. Therefore, the model can still be solved quickly.

VI. ANALYSIS OF NUMERICAL EXAMPLE

The data is from the sensor cloud computing system of the iron and steel industry, which is used to analyze. The production process of slabs is iron making, refining, continuous casting, and slab storage. In this paper, we focus on the process of slab storage. As mentioned before, the arrival rate of slabs is determined by the upstream production. Different steel grades lead to different arrival rates of slabs. We find that the arrival rate varies with the state and the number of states is three. So the MMPP(3) is the best choice to describe this situation. Based on the calculations of previous sections and analysis, we conclude that M/M/C model can approximate MMPP(3)/M/C model. Therefore, we use the M/M/C queueing system to build a nonlinear optimization model. The optimization model can be converted into convex optimization model. Through the optimal solution of M/M/C, we can get the approximate optimal policy of MMPP(3)/M/C.

According to the data of actual production, the total number of cranes in the production line C is 5, which is servers’ number of the queueing system. If the number of arrival slabs is too large, the number of cranes may be not enough. Then the slabs need to wait, and there will be waiting costs. If the number of arrival slabs is too small, cranes may not work at full capacity, which will cause the waste of resources. In the model, the unit time is one day. The minimum value of arrival rate ($\lambda_{\text{min}}$) is 20; the maximum service rate of each crane ($\mu_{\text{max}}$) is 99; the maximum time that slabs stay in the queueing system ($W_{\text{max}}$) is 1.8; the maximum time that slabs waited for storage ($W_q$) is 1.5; the maximum queue length ($L_q$) is 36.

Finally, we can get the optimal arrival rate and service rate calculated by MATLAB. The arrival rate is 39.25 and the service rate of each crane is 56.63. This policy optimizes system configuration. The optimal service load is 19.05. In this situation, the performance of the queueing system in the slab storage process is optimal. Through the results of M/M/C queueing model, we can get the approximate policy to optimize MMPP(3)/M/C queue. Then we can use equation (5) and (6) to assign the arrival rate for the three different levels, while the average arrival rate is 39.25. That is to say, this policy also can make the system configuration of MMPP(3)/M/C queueing system optimal approximately.

VII. CONCLUSION

This paper develops the MMPP(3)/M/C model based on the slab storage problem. And then the comparison between MMPP(3)/M/C model and M/M/C model is given by using QBD and matrix geometry method. Results show that MMPP(3)/M/C model can be approximated well by M/M/C model. Then the optimization model of service load is built according to the M/M/C queueing system. Through some transformations, the model can be converted into a convex optimization model. After doing the transformation, the nonlinear optimization model can be solved easily, which is intractable before. By optimizing the M/M/C model, we obtain the approximate optimal solution of the MMPP(3)/M/C model.

We applied the proposed convex optimization model to the actual slab storage process of the iron and steel enterprise in China to evaluate its performance. Based on the actual data, we calculate the optimal arrival rate and service rate in slab storage process to optimize the performance of the queueing system.

ACKNOWLEDGMENT

The authors would like to thank the referees and editors, whose comments significantly helped the presentation and analysis in this article.

REFERENCES

[1] L. Tang, F. Li, and Z.-L. Chen, “Integrated scheduling of production and two-stage delivery of Make-to-Order products: Offline and online algorithms,” INFORMS J. Comput., vol. 31, no. 3, pp. 493–514, Jul. 2019.
[2] D. Sun, L. Tang, and R. Baldacci, “A benders decomposition-based framework for solving quay crane scheduling problems,” Eur. J. Oper. Res., vol. 273, no. 2, pp. 504–515, Mar. 2019.
[3] L. Tang, J. Liu, A. Rong, and Z. Yang, “Modelling and a genetic algorithm solution for the slab stack shuffling problem when implementing steel rolling schedules,” Int. J. Prod. Res., vol. 40, no. 7, pp. 1583–1595, Jan. 2002.
[4] Y.-H. Jia, L.-X. Tang, Z. G. Zhang, and X.-F. Chen, “MMPP/MC queue with congestion-based staffing policy and applications in operations of steel industry,” J. Iron Steel Res. Int., vol. 26, no. 7, pp. 659–668, Jul. 2019.
[5] Y. Jia, Z. G. Zhang, and L. Tang, “M/M/C queue under a congestion-based staffing policy with applications in steel industry operations,” Int. J. Prod. Res., pp. 1–12, Mar. 2020, doi: 10.1080/00207543.2020.1735656.
[6] A. Dohn and J. Clausen, “Optimising the slabs yard planning and crane scheduling problem using a two-stage heuristic,” Int. J. Prod. Res., vol. 48, no. 15, pp. 4585–4608, Aug. 2010.
YANHE JIA received the B.Eng. degree in mathematics and applied mathematics from the Shanxi University of Finance and Economics, China, in 2011, the M.Eng. degree in industrial engineering from Southwest Jiaotong University, China, in 2013, and the Ph.D. degree in systems engineering from Northeastern University, China, in 2018. From 2015 to 2017, she was a Visiting Student with Simon Fraser University, Canada. She is currently a Lecturer with the School of Economics and Management, Beijing Information Science & Technology University, China. Her research interests include industrial big data science, queuing systems, random process, data analytics and machine learning, dynamic optimization, convex optimization, logistics planning, production and logistics scheduling and engineering applications in manufacturing, resources industry, logistics systems, and service industry.

ZHE GEORGE ZHANG received the B.Sc. degree in computer science and the M.A. degree in economics from Nankai University, China, the M.B.A. degree from the Schulich School of Business, York University, and the Ph.D. degree in operations research from the University of Waterloo.

His research interests include queueing theory and applications, stochastic dynamic programming, probability models in reliability, and supply chain management issues in manufacturing and service organizations. His studies include both theoretical work and a wide range of applications in business, engineering, economics, and applied mathematics. The main theme of his research is to bridge the gap between theory and application, obtaining unobservable and sometimes counter-intuitive but significant/practical insights via modeling and quantitative analysis. He is one of Editor-in-Chiefs of Journal of the Operational Research Society (JORS). As one of the premier Operation Research (OR) journals, JORS is the first OR journal in the world. He is also one of the founding Editors-in-Chief of Queueing Models and Service Management. He was an Associate Editor of Information Systems and Operational Research (INFOR). He is on the editorial board of several international journals.

TE XU received the B.Eng. degree in computer science from Northeastern University, China, in 2002, and the M.Eng. and Ph.D. degrees in industrial engineering & systems from Changwon National University, South Korea, in 2006 and 2010, respectively. He is currently a Lecturer with the Key Laboratory of Data Analytics and Optimization for Smart Industry, Northeast University. His research interests include industrial big data science, data analytics and machine learning, reinforcement learning and dynamic optimization, computational intelligent optimization, plant-wide production and logistics planning, production and logistics batch scheduling and engineering applications in manufacturing (steel, petroleum-chemical, and nonferrous), energy, resources industry, and logistics systems.

* * *

[7] A. Baiocchi and N. Blefari-Melazzi, “Steady-state analysis of the MMPP/GI/1 queue,” IEEE Trans. Commun., vol. 41, no. 4, pp. 531–534, Apr. 1993.

[8] W. K. Ching, R. H. Chan, and X. Y. Zhou, “Circulant preconditioners for Markov-modulated Poisson processes and their applications to manufacturing systems,” SIAM J. Matrix Anal. Appl., vol. 18, no. 2, pp. 464–481, Apr. 1997.

[9] W. K. Ching, “Markov-modulated Poisson processes for multi-location inventory problems,” Int. J. Prod. Econ., vol. 53, no. 2, pp. 217–223, Nov. 1997.

[10] S. Shah-Heydari and T. Le-Ngoc, “MMPP models for multimedia traffic,” Telecommun. Syst., vol. 15, nos. 3–4, 273–293, 2000.

[11] A. Jean-Marie, Z. Liu, P. Nain, and D. Towsley, “Computational aspects of the workload distribution in the MMPP/GI/1 queue,” IEEE J. Sel. Areas Commun., vol. 16, no. 5, pp. 640–652, Jun. 1998.

[12] L. Wang, G. Min, and I. Awan, “Performance analysis of buffer allocation schemes under MMPP and Poisson traffic with individual thresholds,” Cluster Comput., vol. 10, no. 1, pp. 17–31, Apr. 2007.

[13] S. Lu, “Markov modulated Poisson process associated with state-dependent marks and its applications to the deep earthquakes,” Ann. Inst. Stat. Math., vol. 64, pp. 87–106, Feb. 2012.

[14] W. W. Nasr and B. Maddah, “Continuous (s, S) policy with MMPP correlated demand,” Eur. J. Oper. Res., vol. 246, no. 3, pp. 874–885, Nov. 2015.

[15] P. Romano, B. Ciciani, A. Santoro, and F. Quaglia, “Accuracy versus efficiency of hyper-exponential approximations of the response time distribution of MMPP/MAI queues,” Int. J. Parallel, Emergent Distrib. Syst., vol. 24, no. 2, pp. 107–125, 2016.

[16] W. Fischer and K. Meier-Hellsten, “The Markov-modulated Poisson process (MMPP) cookbook,” Perform. Eval., vol. 18, no. 2, pp. 149–171, Sep. 1993.

[17] M. Chiang, B. L. F. Chan, and S. Boyd, “Convex optimization of output link scheduling and active queue management in QoS constrained packet switches,” in Proc. IEEE Int. Conf. Commun. Conf. (ICC), Apr. 2002, pp. 2126–2130.

[18] M. Chiang, A. Sutivong, and S. Boyd, “Efficient nonlinear optimizations of queuing systems,” in Proc. IEEE Global Telecommun. Conf., Taipei, Taiwan, Nov. 2002, pp. 17–21.

[19] M. J. Neely and E. Modiano, “Convexity in queues with general inputs,” IEEE Trans. Inf. Theory, vol. 51, no. 2, pp. 706–714, Feb. 2005.

[20] D. Bertsimas and K. Natarajan, “A semidefinite optimization approach to the steady-state analysis of queueing systems,” Queueing Syst., vol. 56, no. 1, pp. 27–39, May 2007.

[21] C.-P. Li and M. J. Neely, “Solving convex optimization with side constraints in a multi-class queue by adaptive $c_\mu$ rule,” Queueing Syst., vol. 77, no. 3, pp. 331–372, Jul. 2014.

[22] A. G. Marques, L. M. Lopez-Ramos, G. B. Giannakis, J. Ramos, and A. J. Caamaño, “Optimal cross-layer resource allocation in cellular networks using channel-and-queue state information,” IEEE Trans. Veh. Technol., vol. 61, no. 6, pp. 2789–2807, Jul. 2012.

[23] L. Tayeb and R. Messaci, “On a new numerical analysis for the symmetric shortest queue problem,” Int. J. Math. Comput. Sci., vol. 6, no. 8, pp. 1185–1193, 2012.

[24] M. Egan and I. B. Collings, “Base station cooperation for queue stability in wireless heterogeneous cellular networks,” in Proc. IEEE 24th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC), Sep. 2013, pp. 3344–3348.

[25] A. Ganesh and V. Anantharam, “Stationary tail probabilities in exponential metric shortest queue problem,” Int. J. Theor. Appl. Math., vol. 2, no. 1, pp. 13–23, 2016.

[26] H. Ziegler, “Solving certain singly constrained convex optimization problems in production planning,” Oper. Res. Lett., vol. 1, no. 6, pp. 246–252, Dec. 1982.