# Job pricing with online platforms

Kaiyue Yang<sup>1</sup>, Benjamin Legros<sup>2</sup>, Oualid Jouini<sup>1</sup>

<sup>1</sup> Université Paris-Saclay, CentraleSupélec, Laboratoire Genie Industriel, 91190 Gif-sur-Yvette,

France

**Mots-clés** : Double-ended Queue ;Markov Decision Process ; Queueing Games ;Blended Workforce ;Pricing Strategy.

# 1 Introduction

The emerging sharing economy significantly changed how employees work on a fixed timetable in the traditional service system. In contrast, now employees have more flexibility to decide when and if they want to work. Some employees prefer a flexible working schedule. However, the flexible agents are viewed as independent contractors who are not eligible for social welfare like unemployment insurance, which harms labor's benefit[2, 3]. Companies prefer flexible agents because they can avoid paying for their idle periods, reducing overall labor costs. To address this challenge and pursue a mutually beneficial solution, they proposed a blended workforce model with a portfolio of permanent and flexible agents[2, 7]. The introduction of flexible agents in staffing management adds complexity to balancing supply and demand because of the unpredictable availability. Some scholars consider the pricing strategy to control the number of customers to match supply and demand [1, 10]. However, they ignored the strategic behavior of flexible agents. The research conducted by Shi et al.[9] and Nguyen et al. [6] extended the research and further consider the strategic behavior of flexible agents. However, they did not consider permanent agents.

Our paper introduces a queueing model incorporating both permanent and flexible agents. Customers' and agents' demand is endogenous and a function of the difference between a reward and a waiting cost. Here, we summarize the main contributions of this paper.

1. We model the system as a Markov chain and obtain the stationary probabilities using the Matrix geometric approach of Neuts.

2. In the case with priority for flexible agents, we obtain the stationary probabilities in closed form.

3. We prove that there is a unique equilibrium for the arrival of flexible agents and customers. 4. We then can optimize the price that customers pay to the platform and the reward that the platform should pay to the agents in order to maximize revenue of the platform.

5. We optimize price and reward and discuss the effect of the system parameters, such as the level of priority given to flexible agents and the number of permanent agents that are hired.

6. We find that revenue increases with the number of permanent agents. Contrary to the common practice, revenue reduces with the level of priority for permanent agents.

# 2 Model Description

We consider an on-demand platform that plays the role of intermedia to allow some customers to access some workers. The platform has two kinds of workforce, including permanent agents and flexible agents. Assuming they have the same processing speeds, however, they differ in their required working periods and unit operating costs. Permanent agents are present and available at all time when they are not serving a customer, while flexible agents join the plateform only if the difference between their reward and the time wasted before finding a customer is sufficiently high.

We consider a market size of customers that can generate up to  $\Lambda$  arrivals per time unit. Customers decide to join the platform depending on a utility function  $U_c$  that depends on the reward for service R, the price paid to the platform, p, and the cost of waiting c, for an expected waiting time,  $EW_c$ . Therefore, we have  $U_c = R - p - cEW_c$  as in [4]. We assume that customers are not informed about the state of the system, leading to a randomizing joining decision, which keeps the Poisson property of the arrival process. Given that the utility of balking is zero, the arrival rate of customers  $\lambda$  is the solution in  $\lambda$  of  $U_c = 0$ , with  $0 \le \lambda \le \Lambda$ .

Assuming that there is also a market size of flexible agents that can generate up to M flexible agents per time unit. Flexible agents decide to join the platform depending on the reward rthat they might get (the payment) with  $0 \le r \le p$  and the cost of waiting d for an expected wait  $EW_f$ . Therefore, their utility is  $U_f = r - dEW_f$ . Again, we assume that the utility of balking is zero so that agents decide to work for the platform only if  $U_f \ge 0$ . The arrival process of agents (assuming that they are uninformed) is then also Poisson with rate  $\mu$  that is the solution in  $\mu$  of  $U_f = 0$ , with  $0 \le \mu \le M$ . We assume the agents won't come back to the system once they match with customers.

Assuming there are N permanent agents hired by the on-demand platform. These agents will always come back to the queue when completing the service. We assume the service rate of permanent agents is distributed exponentially with rate  $\theta$ . When both a flexible agent and a permanent agent are available, a customer is routed to a permanent agent with probability  $\alpha$ . So, with  $\alpha = 1$ , it means giving priority to permanent agents.

The objective for the platform is to find the optimal values for p and r, such that the revenue per time unit,  $TR = \lambda p - \lambda_1 r - wN$ , is maximized.  $\lambda_1$  means the ratio of the customers served by flexible agents.

As a first step of the analysis, we assume customers(agents) will immediately leave the system when matching with agents(customers). Therefore we cannot find some customers and agents waiting at the same time. The utility function is the same for each individual customer, and likewise for each flexible agent. After they decide to join the queue, reneging is not allowed. In this model, a customer can be taken by one flexible or permanent agent. The customers and agents are arranged based on a first-come-first-served discipline(FCFS).

At a given instant t, we denote by  $\{n_1(t), t \ge 0\}$ ,  $\{(n_2(t), n_3(t)), t \ge 0\}$  the process of the number of customers in the queue, the number of permanent agents and flexible agents in the queue, respectively. Figure (1) illustrates the transition of the system state.

# **3** Performance Analysis

#### 3.1 Stationary probabilities

Before obtaining the stationary probabilities of the system state, we establish the stability condition as in Theory 1 for the Markov chain through the Matrix Geometric method.

**Théorème 1** The system is stable if and only if

$$\frac{\mu}{\alpha\lambda x_0 + (1-\alpha)} < 1, \frac{\lambda}{\mu + N\theta} < 1$$

where

$$x_0 = \left(\sum_{j=0}^{N} \frac{N! \theta^j}{(N-j)! (\alpha \lambda)^j}\right)^{-1}.$$



FIG. 1 – Markov chain for the number of agents and customers in the queue

Let us consider the process  $\{n_1(t), t \ge 0\}$ . The stationary probabilities of n customers in the queue are denoted by  $p_n$ . For ease of exposition, note  $p_0 = p_{0,0}$ , then  $p_n$  is given by

$$p_n = \left(\frac{\lambda}{\mu + N\theta}\right)^n p_0, n \ge 1.$$
(1)

For the process  $\{(n_2(t), n_3(t)), t \ge 0\}$ , the stationary probabilities of x permanent agents and y flexible agents are denoted by  $p_{x,y}$ . Based on Equation 1, the summation of  $p_{x,y}$  is given by

$$\sum_{x=0}^{N}\sum_{y=0}^{\infty}p_{x,y} = 1 - \sum_{n=1}^{\infty}p_n = 1 - \frac{\lambda}{\mu + N\theta}\sum_{n=0}^{\infty}(\frac{\lambda}{\mu + N\theta})^n p_0 = 1 - \frac{\lambda}{\mu + N\theta - \lambda}p_0.$$
 (2)

Computing the stationary probabilities of  $p_{x,y}$  is a complicated task. We obtained the explicit expression  $p_{x,y}$  when  $\alpha = 0$  by difference equations. When  $\alpha = 0$ , the stationary probabilities of the system state can be given by

$$p_{0,y} = X_0^y p_0, (3)$$

$$p_{x,y} = a_{x,0}X_0^y + a_{x,1}X_1^y + a_{x,2}X_2^y + \dots + a_{x,x}X_x^y = \sum_{i=0}^x a_{x,i}X_i^y, 1 \le x \le N.$$
(4)

Here,

$$X_{i} = \frac{\lambda + \mu + (N - i)\theta - \sqrt{(\lambda + \mu + (N - i)\theta)^{2} - 4\lambda\mu}}{2\lambda}$$
$$a_{0,0} = 1, a_{x,0} = \frac{(-1)^{x}}{x!} \prod_{i=0}^{x-1} (N - i), x \ge 1,$$

$$a_{x,i} = (-1)^{x-i} \prod_{k=i}^{x-1} \frac{(N-k)}{(x-i)!} a_{i,i}, 0 < i < x,$$
$$a_{x,x} = \left\{ \frac{1}{x!} + \sum_{j=0}^{x-1} \prod_{i=j}^{x-1} (\frac{\theta}{\lambda})^{x-j} \frac{1}{j!(1-X_i)} \right\} \prod_{i=0}^{x-1} (N-i), x \ge 1,$$
$$p_0 = \left\{ \sum_{x=0}^{N} \sum_{i=0}^{x} \frac{a_{x,i}}{1-X_i} + \frac{\lambda}{\mu + N\theta - \lambda} \right\}^{-1}.$$

When  $\alpha \neq 0$ ,  $p_{x,y}$  can be derived by the Matrix Geometric method because the transition matrix in the two-dimensional part of the Markov chain has a block-tridiagonal structure which indicates that is the quasi-birth-death processes(QBD) structure. For further details, please refer to [5]. We define  $u = (u_0, u_1, u_2, \ldots, u_y, \ldots)$  and  $u_y = (p_{0,y}, p_{1,y}, p_{2,y}, \ldots, p_{N,y})$ , according to the theoretical foundations of the method, a constant matrix R should exist such that  $u_y = u_{y-1}R$  and successive substitution can be used to solve matrix R. After solving the matrix R, all vectors  $u_y$  could be derived by  $u_y = u_0 R^y$ . The only remaining problem is deriving the value of  $u_0$ , which could be derived by Equation 5 and Equation 6. Here  $B_0$  and  $A_0$  is transition matrix block.

$$u_0(B_0 + RA_0) = 0 (5)$$

$$1 - \frac{\lambda}{\mu + N\theta - \lambda} p_0 = ue = u_0 e + \sum_{y=1}^{\infty} u_y e = u_0 \sum_{y=0}^{\infty} R^y e \tag{6}$$

#### 3.2 Waiting time of customers

The expected waiting time of customers is given by

$$EW_c = \frac{1}{\mu + N\theta} p_0 + \sum_{n=1}^{\infty} \frac{n+1}{\mu + N\theta} p_n = \sum_{n=0}^{\infty} \frac{(n+1)\lambda^n p_0}{(\mu + N\theta)^{n+1}} = \frac{(\mu + N\theta)p_0}{(\lambda - \mu - N\theta)^2}, n \ge 1.$$
(7)

#### 3.3 Waiting time of flexible agents

The waiting time of flexible agents is related to  $\alpha$ . When  $\alpha = 0$ , the waiting time for a newly arriving flexible agent is solely related to the number of flexible agents already in the queue ahead due to the priority. When  $\alpha = 1$ , flexible agents must wait for all permanent agents that are already in the queue and those expected in the future to leave the system. Hence, the waiting time of flexible agents is influenced by all future permanent agent's arrivals. When  $\alpha \in (0, 1)$ , flexible agents could be matched directly with probability  $1 - \alpha$ . Alternatively, they need to wait for the service of permanent agents, including those already in the queue and future arrivals. Therefore, the expected waiting time of flexible agents is determined through the following scenarios.

**Priority to flexible agents** Denote  $P_1(y) = \sum_{x=0}^{N} p_{x,y}$ , the expected waiting time of flexible agents is given by Equation 8.

$$EW_f = \sum_{y=0}^{\infty} \frac{y+1}{\lambda} P_1(y) \tag{8}$$

**Priority to permanent agents.** We denote random variable  $V_x$  as the first passage time from state x to state x - 1. Based on strong Markov property, we could get the expected value of  $V_x$  shown Equation 9.

$$V_x = \frac{1}{\lambda} + \frac{1}{\lambda} \sum_{j=0}^{N-2-x} \prod_{k=x}^{x+j} \frac{(N-k)\theta}{\lambda} + \prod_{k=x}^{N-1} \frac{(N-k)\theta}{\lambda} V_N, 0 \le x < N$$
(9)

Then, the expected waiting time for flexible agents is given by

$$EW_f = \sum_{x=0}^{N} \sum_{y=0}^{\infty} (\sum_{k=0}^{x} V_k + yV_0) p_{x,y}$$
(10)

**Random allocation with**  $\alpha$  We consider the first passage time in the one-dimensional birthand-death process with N+1 absorbing states. The state transition can be conceptualized as a phase-type distribution, which captures the time taken for absorption in a finite-time Markov chain. Utilizing the first step analysis[8], we can derive the probability of absorbing in state k,  $r_{xk}$ , and the time it takes  $v_k$ . Then, the expected waiting time is derived.

$$EW_f = \sum_{x=0}^{N} \sum_{y=0}^{\infty} (v_x + \sum_{k=0}^{N} \prod_{n=1}^{y} r_{xk}^n v_k) p_{x,y}$$

The results are consistent with Little's theorem, so formula  $EW_f = \frac{EQ_f}{\mu}$  and  $EW_c = \frac{EQ_c}{\lambda}$  could also be used to infer the expected waiting time of customers and flexible agents.

## 4 Equilibrium behavior

Suppose the market size  $\Lambda$  and M is infinity. In this context, we can analyze the Nash equilibrium by examining the equations  $U_c(\lambda^*, \mu^*) = R - p - cEW_c(\lambda^*, \mu^*) = 0$  and  $U_f(\lambda^*, \mu^*) = r - dEW_f(\lambda^*, \mu^*) = 0$ . Consequently, the equilibrium arrival rates  $\lambda^*, \mu^*$  represent the roots of two utility functions. However, solving this problem analytically is challenging due to the complexity of the function set. Given that the expected waiting time  $EW_f(EW_c)$  is strictly increasing with  $\mu(\lambda)$ , and the expected waiting time  $EW_f(EW_c)$  is strictly decreasing with  $\lambda(\mu)$ , we obtain the equilibrium rates via the following algorithm.

Given initial value  $\lambda_0$ , we start to solve  $\mu_1^*$  by  $U_f(\lambda_0, \mu_1^*) = 0$  using the bisection method. Substituting  $\mu_1^*$  into  $U_c(\lambda_1^*, \mu_1^*) = 0$  yields  $\lambda_1^*$ . By comparing  $\lambda_0$  and  $\lambda_1^*$ , we update the value accordingly : if the  $\lambda_1^*$  is greater, we set it to  $2 * \lambda_1^*$ ; otherwise, we keep it as  $\frac{\lambda_1^*}{2}$ . This iterative process aids in identifying the range of roots. Subsequently, we employ the bisection method until the optimal value is attained. Additionally, we assert the uniqueness of the solution.

### 5 Optimization

We consider now the online platform that sets a optimal price p and reward r, such that the revenue per time unit,  $\lambda p - \lambda_1 r - wN$  is maximized. Assuming a monopoly scenario where the platform operates without market competition, they does not leave a positive customer surplus and agent surplus since, in such a situation, the platform will improve the price and decrease the payment to agents as much as possible and don't reduce the arrival rate of customers and agents[4]. Therefore, the optimal value for  $p_m$  and  $r_m$  will make the utility function of customers and agents equal 0. In other words,  $p_m = R - cEW_c(\lambda, \mu)$ ,  $r_m = dEW_f(\lambda, \mu)$ . Therefore, the monopoly's problem is to maximize

$$\max_{\lambda,\mu} TR = \lambda \left[ R - cEW_c(\lambda,\mu) \right] - \lambda_1 dEW_f(\lambda,\mu) - wN$$

under the constraints  $\lambda \ge 0$  and  $\mu \ge 0$ , and find the optimal value  $\lambda_m$  and  $\mu_m$ . The optimization problem seems to be complicated, and it is hard to solve the analytical solutions. We mainly use the Particle Swarm Optimization(PSO) algorithm to derive the optimal results.

## 6 Results and Discussion

We mainly analyze how  $N, \alpha, \theta, R, c$ , and d affect the equilibrium behavior and optimal pricing strategy in this section. As the number of permanent agents (N) increases, we observe a simultaneous rise in the Nash equilibrium arrival rates  $(\lambda \text{ and } \mu)$ , optimal price, and optimal platform revenue. However, the optimal reward experiences a decrease. This intuitive phenomenon can be attributed to the diminishing waiting time for customers as N increases, leading to a higher inclination for customers to join the queue. Consequently, the entry probabilities for flexible agents increase, allowing the platform to set a higher price and a lower reward, optimizing overall revenue.

The escalation of R and  $\theta$  exerts a positive influence on the equilibrium arrival rates  $\lambda$  and  $\mu$ . This conclusion aligns with intuition. An increase in service rate or customer reward for service increases customers' inclination to join the queue, consequently increasing the arrival rate of flexible agents. Besides, An increase in R contributes to an enhancement in both optimal pricing and optimal profit. Conversely, the service rate  $\theta$  plays a dual role by lowering optimal pricing and simultaneously boosting optimal profit.

The waiting costs(c) for customers and for flexible agents(d) exert a detrimental impact on equilibrium arrival rates, prompting a higher optimal price and reward but simultaneously decreasing platform revenue.

The concept of prioritizing permanent agents with parameter ( $\alpha$ ) is intriguing. As  $\alpha$  increases, the arrival rate of flexible agents tends to decrease, influencing entry probabilities of customers. Simultaneously, the platform's optimal strategy involves a substantial increase in both price and reward. Remarkably, when  $\alpha$  equals zero, both customer and flexible agent arrival rates tend towards infinity, and the platform's revenue tends towards infinity.

## Références

- Jiaru Bai, Kut C So, Christopher S Tang, Xiqun Chen, and Hai Wang. Coordinating supply and demand on an on-demand service platform with impatient customers. *Manu*facturing & Service Operations Management, 21(3):556–570, 2019.
- [2] Amiya K Chakravarty. Blending capacity on a rideshare platform : Independent and dedicated drivers. Production and Operations Management, 30(8) :2522-2546, 2021.
- [3] Sarah A Donovan, David H Bradley, and Jon O Shimabukuru. What does the gig economy mean for workers? 2016.
- [4] Refael Hassin and Moshe Haviv. To queue or not to queue : Equilibrium behavior in queueing systems, volume 59. Springer Science & Business Media, 2003.
- [5] Marcel F Neuts. *Matrix-geometric solutions in stochastic models : an algorithmic approach.* Courier Corporation, 1994.
- [6] Hung Q Nguyen and Tuan Phung-Duc. Supply-demand equilibria and multivariate optimization of social welfare in double-ended queueing systems. Computers & Industrial Engineering, 170 :108306, 2022.
- [7] Edieal J Pinker and Richard C Larson. Optimizing the use of contingent labor when demand is uncertain. *European Journal of Operational Research*, 144(1):39–55, 2003.
- [8] Mark Pinsky and Samuel Karlin. An introduction to stochastic modeling. Academic press, 2010.
- [9] Ying Shi and Zhaotong Lian. Equilibrium strategies and optimal control for a doubleended queue. Asia-Pacific Journal of Operational Research, 33(03) :1650022, 2016.
- [10] Terry A Taylor. On-demand service platforms. Manufacturing & Service Operations Management, 20(4):704–720, 2018.