CUSTOMERS’ JOINING BEHAVIOR IN AN UNOBSERVABLE GI/Geo/m QUEUE

VEENA GOSWAMI
School of Computer Applications
Kalinga Institute of Industrial Technology
Bhubaneswar-751024, India

GOPINATH PANDA
Department of Electrical and Computer Engineering
University of Central Florida, USA

(Communicated by Gerhard-Wilhelm Weber)

Abstract. This paper studies the equilibrium balking strategies of impatient customers in a discrete-time multi-server renewal input queue with identical servers. Arriving customers are unaware of the number of customers in the queue before making a decision whether to join or balk the queue. We model the decision-making process as a non-cooperative symmetric game and derive the Nash equilibrium mixed strategy and optimal social strategies. The stationary system-length distributions at different observation epochs under the equilibrium structure are obtained using the roots method. Finally, some numerical examples are presented to show the effect of the information level together with system parameters on the equilibrium and social behavior of impatient customers.

1. Introduction. Discrete-time queueing systems with strategic customers’ have received substantial attention because of their application in real-time communication, cloud computing, call centers, inventory systems, etc. Upon arrival, strategic customers decide whether to join the queue and wait for his service completion or to not join the queue at all. Their decisions are made on the basis of available system information and a linear cost-reward structure. Each customer try to maximize his/her net benefit in opponent to other customers having the same objective, which may be viewed as a non-cooperative symmetric game among them. The game-theoretic point of view in an M/M/1 queue was presented in Naor [29]. Naor assumed that customers could observe the queue length upon arriving; relying on that information, they decide whether to join or balk from the queue. Edelson and Hildebrand [6] considered the strategic behavior in the corresponding unobservable counterpart. Recently, Goswami and Panda [11] studied the strategic behavior of the customer in discrete-time Geo/Geo/1 queue for both observable and unobservable cases.

The game-theoretic analysis of single server queues is further extended to the multi-server queueing systems. However, there are few works on the multi-server

2010 Mathematics Subject Classification. Primary: 60K25, 68M20; Secondary: 90B22.

Key words and phrases. Strategic customers, Nash equilibrium, Discrete-time, Multi-server, Social benefit, Renewal input.
queues due to their analytical complexity. The earlier work reported on discrete-time multi-server queues is Chan and Maa [2]. They considered an infinite-buffer renewal input queue with an early arrival system. Further, Chaudhry et al. [3] discussed the relations among the distributions at different epochs for the same queueing model in detail. A queueing system with balking incorporates the characteristics of the customer’s impatience policy. From the perspective of an arriving customer, the number of servers and the waiting time are significant measures. A multi-server queueing system with balking in discrete-time time have been carried out in Goswami [9]. Guha et al. [13] analyzed an unobservable GI/M/c queue with an equilibrium balking strategy and reneging. To the best of our knowledge, the customers’ strategic behavior in a discrete-time multi-server with renewal inputs is not available in the literature. In the present study, we try to fill this gap.

This paper considers the economic analysis of discrete-time multi server queueing system with identical servers. This model is of unobservable type, that is, arriving customers do not have the queue-length information, but they have information regarding the service rate, waiting cost and service completion reward. Based on a linear cost-reward framework that comprises their wish for service and also their reluctance to wait, customers choose whether to join or balk the queue. We denote this queueing system as GI/Geo/m unobservable model with early arrival system (EAS) or arrival first (AF) policy. More details on the EAS and other classifications are discussed in the next section. The main contributions of our work are:

- steady-state analysis of a discrete-time multi-server renewal input queueing model of unobservable type.
- explore the equilibrium balking strategy and obtain the closed-form results for the system-length distribution at pre-arrival epochs.
- discuss the social benefit of the investigated system to make it economically viable.
- present study gives helpful insights for planning an optimal management policy of discrete-time queues applicable in call centre modelling, inventory systems, computer networks with packet loss, etc.

The rest of this paper is organized as follows. Section 2 presents the relevant literature survey related to game-theoretic analysis in a queueing system. Section 3 illustrates an interesting application of our model in the optimal design of an asynchronous transfer mode (ATM) switch. Section 4 specifies the model dynamics and the reward-cost structure. Section 5 analyzes the stationary probabilities at pre-arrival epochs and examines the equilibrium balking strategies as well as socially optimal strategies of the non-cooperative game. Further, the inefficiency of the selfish behavior is measured in terms of the price of anarchy (PoA). Section 6 illustrates extensive numerical results for equilibrium balking behavior of customers and society. Finally, Section 7 concludes and provides then future scope of the paper.

2. Related works. This section presents the relevant literature survey related to game-theoretic analysis in a queueing system. Naor [29] was the first to study strategic behavior in queues. After successful service completion, the customers receive the fixed reward and incur the fixed waiting cost per unit. Naor presented a comparison of individual and socially optimized behavior in a single-server Markovian queue, where arriving customers have the queue-length information before deciding whether to join or to balk (observable case). Edelson and Hildebrand [6] examined the strategic behavior in the corresponding unobservable M/M/1 model,
where the customers do not observe the queue length upon arrival. One may refer to comprehensive literature and references on the strategic customers in Hassin and Haviv [16] and Hassin [15]. Naor [29] did not investigate the Price of Anarchy (PoA). Gilboa-Freedman et al. [8] were the first to discuss PoA in a single-server Markovian queue. Goswami and Panda [10] studied optimal information revelation policies in discrete-time Geo/Geo/1 queue. They considered a partial information disclosure policy, where a service provider depending upon the number of waiting customers, reveals information to some customers and hides it from others. Recently, Goswami and Panda [11] presented strategic customer behavior in discrete-time single server queues in both observable and unobservable cases. There are various literature available on strategic customer behavior in discrete-time single server queues with different variations such as server vacations (Gao and Wang [7], Ma et al. [26]), interruption (Yu and Alfa [35]), batch service (Panda and Goswami [30]), pricing (Ma and Liu [27]), breakdown and repairs (Yang et al. [33]).

Table 1. Survey on queueing models related to game-theoretic analysis

| Reference | Model       | Buffer size | Findings                                      |
|-----------|-------------|-------------|-----------------------------------------------|
| [29]      | M/M/1       | Finite      | Individual and social optimal behavior        |
| [6]       | M/M/1       | Infinite    | Individual and social optimal behavior        |
| [8]       | M/M/1       | Finite      | Price of Anarchy                             |
| [11]      | Geo/Geo/1   | Finite & infinite | Individual & social optimal behavior and PoA |
| [34]      | GI/M/s      | finite & infinite | Self & social optimization                   |
| [19]      | M/M/s       | Infinite    | Individual & social optimization behavior of customers under a non-linear holding cost |
| [23]      | M/M/s       | Infinite    | Individual & social optimization behavior of customers under a linear holding cost |
| [1]       | M/G/s       | Infinite    | Individual & social optimization with holding cost |
| [14]      | GI/M/c      | Infinite    | Equilibrium balking strategy with reneging    |
| Present study | GI/Geo/m   | Infinite    | Individual & social optimal behavior and PoA |

Table 1 presents the works that are available related to the game-theoretic analysis in single and multi-server queueing systems. Yechiali [34] extended Naor’s work to a multi server setting. He studied the self and social optimization in observable single channel GI/M/s queue with both finite and infinite buffer. Knudsen [19] discussed the individual and social optimization behavior of customers in a single channel M/M/s queue under a non-linear holding cost. Lippman and Stidham Jr [23] studied the same M/M/s model with linear holding cost and shown that individual optimization leads to over utilization of the service facility. The individual as well as social optimization in an unobservable M/G/s multi-channel system with
heterogeneous service and holding cost is studied in Bell and Stidham Jr [1]. Martin and Pankoff [28] discussed optimal balking, reneging and jockeying decisions for the multi-channel GI/M/c queue with heterogeneous servers. Guha et al. [13] discussed an equilibrium balking strategy with reneging when the customers take up a server in an unobservable GI/M/c queue. Further, the same authors Guha et al. [14] presented a finite buffer multi-server renewal input state-dependent balking and reneging queue with or without vacations. Recently, Tang et al. [32] studied a single channel Poisson queue with two servers and two categories of customers.

3. Application of the suggested model. ATM is a telecommunications standard for digital transmission of voice, data, video, and image signals in a network without utilizing any overlay networks, Jeffrey [18]; De Prycker [5]; Ra’ed and M-ouftah [31]. It can handle the conventional high-throughput data traffic as well as low-latency, live streaming content like voice and video. ATM includes a utility that uses characteristics of both packet and circuit switching networks. It encodes data into small, fixed-sized network packets by employing asynchronous time-division multiplexing and it allows traffic over a high-speed single access circuit, Cosmas et al. [4]. It divides the information into equal sized cells of 53 bytes before transmission, of which 48 bytes contain information, and 5 bytes constitute control information in the header. The ATM switch has input ports and output ports for data flow. Fig 1 shows a queueing representation of an ATM switch. The basic functions of an ATM switch are routing, queueing and header translation, Le Boudec [21]. Its task is to transports cells from the input port to output ports with the help of routing information and information stored at switching nodes employing a connection set-up procedure, called virtual circuit (VC). It may be of two types: (i) a permanent VC (PVC), created at the endpoints administratively, (ii) switched VC (SVC), created by users on the requirement. Call admission is then performed by the network to verify the availability of requested resources, and a connection route; otherwise, the packet will be blocked or lost. In ATM networks, it is better to keep the loss of packets at a meagre rate, generally less than $10^{-8}$, Kuehn [20]. One may apply the presented model in the renewable energy and sustainable supply chain network problem [24, 25].

The output port can be conveniently modelled as a multi-server discrete-time queueing system. The output port is made up of two parts: the buffer and $m$ servers. The buffer has infinite waiting capacity to store packets from input ports.
Table 2. Notations and model parameters

| Operational parameters | Economic parameters |
|------------------------|----------------------|
| $1/\lambda$            | $R$                  |
| $1/\mu$                | customers gets a reward after completion of service |
| $m$                    | $C$                  |
| $d \in [0, 1]$         | waiting cost per time unit in the system |
|                        | Performance measures |
|                        | $W_s$                |
|                        | mean system-length   |
|                        | $L_s$                |
|                        | net benefit of the tagged customer |
|                        | $\Delta_s(d)$       |
|                        | social benefit per time unit |
|                        | $\Delta_e(d)$       |
|                        | price of anarchy     |
|                        | $d_e$                |
|                        | equilibrium joining probability |
|                        | $d^*$                |
|                        | socially optimal joining probability |

Each input/output port can carry exactly one packet during each slot. The arriving packets wait in the buffer if $m$ channels (servers) are busy. In every slot, a bunch of packets from the input ports, arrives at the output port. These packets wait in the buffer until they are served at the output port. The generally independent arrival distribution is suitable to model the packet arrivals to the output queue, whereas the multi-server queue will help in the efficient management of the buffer space. If call admission or connection for a packet fails, the packet is considered lost, which is similar to balking in our model. The control plane of the ATM switch behaves like a central decision-maker that manages the packets for better quality of service (QoS) and minimal loss.

4. **Unobservable GI/Geo/m queue under EAS policy.** Discrete-time queueing systems has two different but associated cases based on the simultaneity of arrivals and departures at slot boundaries: (i) early arrival system (EAS) and (ii) late arrival system with delayed access (LAS-DA). For more aspects on this subject, one may refer to Hunter [17] and Gravey and Hébuterne [12]. In the LAS-DA model, a probable arrival happens in the slot $(t^-, t)$ and a probable departure occurs in the slot $(t, t^+)$, where $t^-$ and $t^+$ are time instants just before and just after $t$, respectively. Even if a customer arrives in the interval $(t^-, t)$ and finds the server to be idle, his service will not start until the time interval $(t, t^+)$ is reached. In the EAS model, a probable departure happens in the slot $(t^-, t)$ and a probable arrival occurs in the slot $(t, t^+)$. In EAS model, an arrival may receive service in the same slot. The various time epochs at which events take place in the EAS model are described in Fig.2. The state of the system changes only around the slot boundaries, that is, both arrivals and departures are feasible only at slot boundaries. Further, we presume that departure in $(0^-, 0)$ is not possible. The associated model parameters and notations are summarized in Table 2 for convenience.

We consider a discrete-time renewal input multi-server queueing system with s-
strategic customers under the unobservable case. Suppose that customers arrive at time epochs \(0 = t_0, t_1, t_2, \ldots\), with \(t_i > t_j\) for \(i > j\) and the random variables \(T_k = t_{k+1} - t_k\), \(k = 0, 1, \ldots\), are the inter-arrival times. Let \(T_k\)'s are independent and identically distributed random variables with probability mass function (pmf)

\[ a_n = P(T = n), \quad n \geq 1, \quad a_0 = 0, \]

probability generating function (pgf)

\[ A(z) = \sum_{n=0}^{\infty} a_n z^n, \quad (|z| \leq 1) \]

and mean \(1/\lambda\), where \(T\) is the generic of inter-arrival times. There are \(m\) independent homogeneous servers with a single queue. The service times \(S\) of each server is independent and geometrically distributed with pmf

\[ b_n = P(S = n) = (1 - \mu)^{n-1} \mu, \quad n \geq 1, \quad 0 < \mu < 1 \]

and the mean service time of each server is \(E(S) = 1/\mu\). The customers are served in a first-come first-served (FCFS) order. The inter-arrival times and service times are assumed to be mutually independent. The traffic intensity of the system is \(\rho = \lambda/m\mu\).

At time \(t_k\), let \(N(t_k)\) be the state of the system, that is, the number of customers in the system at the arrival instant \(t_k\). The evolution of the system state process \(\{N(t_k)\}\) at two consecutive arrival epochs \(t_k, t_{k+1}\) depends on the number of departures during that inter-arrival time \(T_k\). To compute the number of departures in an interval, define \(b(j|i)\) as the probability that \(j\) customers accomplish service in the next interval out of \(i > j\) customers in the system such that it pursues the binomial distribution

\[ b(j|i) = \begin{cases} \binom{i}{j} \mu^j (1 - \mu)^{i-j}, & 1 \leq i < m, \quad 0 \leq j \leq i \\ \binom{m}{j} \mu^j (1 - \mu)^{m-j}, & i \geq m, \quad 0 \leq j \leq m \end{cases} \]

with \(b(0|0) = 1\) and \(\binom{u}{v} = 0\) for \(u < v\) or \(u < 0\).

In a theoretical framework, the game theory optimizes independent and participating rational brokers. A non-cooperative game is a contest between individual players without making any association or connection with other players. The players are indistinguishable in a symmetric game; that is, the corresponding payoffs and strategy sets are the same for all players in the game. We aim to deduce the (Nash) equilibrium strategies that are best responses opposed to themselves; that is, no player has a reward for varying from this strategy unilaterally.

We assume that upon arrival to the unobservable queue, each customer acquires some system information like, service rate, reward value after service completion,
waiting cost and decides whether to balk or join the system based on this information. Customers get a reward of \( R \) units after completion of service and incur a waiting cost of \( C \) units per unit of time waiting in the system. All customers are assumed to be risk neutral and want to maximize their expected utility. Their decisions are unalterable in the sense that neither balking customers are permitted to retry nor joining customers are granted to leave the system. We assume that a balking customer’s net utility is zero.

5. **Equilibrium balking strategy under EAS policy.** We consider the unobservable GI/Geo/m queue under EAS policy. As the arriving customers do not observe the number in system at their arrival instants, they follow a mixed strategy of joining, that is, they choose to join or balk with certain probability. Since they observe the number in system at their arrival instants, they follow a mixed strategy in the individual and social optimization cases. For deriving the equilibrium strategies, we need to compute the stationary system-length distributions at different time instants in the following sections.

5.1. **System-length distribution at pre-arrival epoch.** Consider the system just before the arrival instants, \( t_k, k \geq 0 \) which are considered as embedded points. Let \( t_k^+ \) denote the pre-arrival epochs, that is, the time epochs just before the arrival instants \( t_k \). The state of the system at these pre-arrival epochs is set as \( \{ N(t_k^+) \} \), where \( \{ N(t_k^+) \} \) is the number of customers in the system at the pre-arrival epoch \( t_k^+ \). In steady-state, let us assume \( Q_k^- = \lim_{j \to \infty} P (N(t_j^-) = k), k \geq 0 \), where \( Q_k^- \) refers the probability that just prior to an arrival epoch there are \( k \) customers in the system. The sequence of random variables \( \{ N(t_k^+) \} \) forms a Markov chain with stationary probabilities \( Q_k^- \). The state of the system at two consecutive pre-arrival epochs can be derived once we obtain the following transition probabilities.

Let \( f_j (j \geq 0) \) describes the probability that during an inter-arrival time, \( j \) customers have been served when all the \( m \) servers are busy. Let \( g_{i,j} \) be the probability that an arriving customer notices \( i \) customers in the system whereas the next arriving customer gets \( j \) customers in the system. Consequently, \( (i-j) \) customers have been served during an inter-arrival time. Thus, we have

\[
f_j = \sum_{n=1}^{\infty} a_n \binom{mn}{j} \mu^j (1 - \mu)^{mn-j}, \quad j \geq 0,
\]

\[
g_{i,j} = \left\{ \begin{array}{ll}
\sum_{n=1}^{\infty} a_n \sum_{y=0}^{m} \frac{m!}{y!(m-y)!} (1 - \mu)^{m-y} (1 - (1-\mu)^{m-y})^{i-y} \mu^{i-y-s}(1 - \mu)^{m(i-y-s)} (y\mu)^{y-s} \\
\times (1 - \mu)^{m-y} (1 - \mu)^{m-n-y} (1 - (1-\mu)^{m-n-y})^{i-y-s-j}, \quad i > m, 0 \leq j \leq m - 1.
\end{array} \right.
\]

One may refer to Chan and Ma \[2\] for a detailed derivation of \( f_j \) and \( g_{i,j} \). The probability generating function of \( f_j \) is given by

\[
F(z) = \sum_{j=0}^{\infty} f_j z^j = \sum_{j=0}^{\infty} z^j \sum_{n=1}^{\infty} a_n \binom{mn}{j} \mu^j (1 - \mu)^{mn-j}
\]
\[\sum_{n=1}^{\infty} a_n(\mu + \mu z)^m = A^*(\mu + \mu z)^m.\]

Relating the state of the system at two consecutive pre-arrival epochs, using definitions and probabilities discussed before, at steady-state, we have

\[Q_0 = d \sum_{k=0}^{\infty} Q_k g_{k+1,0} + (1 - d) \sum_{k=0}^{\infty} Q_k g_{k,0},\]  
(1)

\[Q_n = d \sum_{k=n-1}^{\infty} Q_k g_{k+1,n} + (1 - d) \sum_{k=n}^{\infty} Q_k g_{k,n}, 1 \leq n \leq m - 1,\]  
(2)

\[Q_n = d \sum_{k=n-1}^{\infty} Q_k f_{k+1-n} + (1 - d) \sum_{k=n}^{\infty} Q_k f_{k-n}, n \geq m.\]  
(3)

The system-length distribution at pre-arrival epochs \(Q_n, n \geq 0\) can be computed by solving the above system. To solve the difference equation (3), we take the displacement operator \(E\) defined by

\[E_j Q_n = Q_{n+j},\]

and rewrite (3) as

\[E - d (f_0 + f_1 E + f_2 E^2 + \ldots) - \bar{d} (f_0 E + f_1 E^2 + \ldots) Q_n = 0, n \geq m.\]  
(4)

After simplification, the characteristic equation related with (4) reduces to

\[k(z) \equiv z - (d + \bar{d} z) F(z) = 0.\]

Using Rouché’s theorem, it can be shown that \(z - (d + \bar{d} z) F(z)\) has a unique real root, say \(\xi\) inside the region \(|z| < 1\) iff \(\rho = \lambda/m\mu < 1\). So, the solution of the homogeneous difference equation (3) can be obtained as

\[Q_n = K \xi^{n-m}, n \geq m - 1,\]  
(5)

where \(K\) is an unknown to be evaluated. Applying (5) in (2) and then working out recursively, we have

\[Q_n = K \Psi_n, n = m - 2, m - 3, \ldots, 1, 0,\]  
(6)

where \(\Psi_n\) is given by

\[
\Psi_{m-2} = \frac{1}{d} \xi g_{m-1,m-1} \left[ 1 - \bar{d} g_{m-1,m-1} - (d + \bar{d} \xi) \sum_{k=m}^{\infty} \xi^{k-m} g_{k,m-1} \right],
\]

\[
\Psi_n = \frac{1}{d} \xi g_{n+1,n+1} \left[ \xi \Psi_{n+1} - \xi \sum_{k=n+1}^{m-2} (dg_{k+1,n+1} + \bar{d}g_{k,n+1}) - \bar{d}g_{m-1,n+1}
\right.
\]
\[-(d + \bar{d} \xi) \sum_{k=m}^{\infty} \xi^{k-m} g_{k,n+1} \right], n = m - 3, \ldots, 1, 0.
\]

Now the only unknown \(K\), can be obtained using the normalization condition as

\[K = \frac{\xi (1 - \xi)}{1 + \xi (1 - \xi) \sum_{j=0}^{m-2} \Psi_j}.\]  
(7)

Thus, we computed the pre-arrival epoch probabilities \(Q_j^-, j \geq 0\) under the mixed joining strategy \(d\).
5.2. **Sojourn time analysis.** We find the sojourn time distribution of a tagged customer in accordance with the FCFS discipline. Let the random variable $T_q$ represent the queueing time of the tagged customer who decides to enter the system, and has pmf $w_j = P(T_q = j), \ j \geq 0$. On arrival, the tagged customer may find the system in one of the following two instances.

**Case 1.** $w_0 = P(T_q = 0)$. At an arrival epoch, if there are $i$ ($0 \leq i \leq m - 1$) customers in the system, the service of the tagged arrival begins immediately. The probability that the tagged customer does not wait is given by

$$w_0 = P(T_q = 0) = \sum_{i=0}^{m-1} Q^{-}. \tag{5.2}$$

**Case 2.** $w_j = P(T_q = j), \ j \geq 1$. At an arrival epoch, if there are $i$ ($i \geq m$) customers in the system and all the servers busy, that is, $m$ in service and $n = i - m$, ($n \geq 0$), customers waiting in the queue. The tagged customer’s service will start after the departure of $(n + 1)$ customers. The tagged customer will wait up to $j$ slots, if there will be at most $n$ service completions during the $j$ slots duration. Therefore, the probability that the tagged customer waits for greater than $j$ slots is

$$P(T_q > j) = \sum_{i=m}^{\infty} Q_{i}^{-} \sum_{r=0}^{i-m} \binom{m}{r} \mu^r (1 - \mu)^{m-r}, \ j \geq 1$$

$$= K \sum_{i=m}^{\infty} \xi^{i-m} \sum_{r=0}^{i-m} \binom{m}{r} \mu^r (1 - \mu)^{m-r} = \frac{K (\bar{\mu} + \mu \xi)^{m}}{1 - \xi}. \tag{5.3}$$

The probability of waiting exactly $j$ slots by the tagged customer is

$$w_j = P(T_q = j) = P(T_q > j - 1) - P(T_q > j), \ j \geq 1,$$

$$= \frac{K}{1 - \xi} (\bar{\mu} + \mu \xi)^{m(j-1)} [1 - (\bar{\mu} + \mu \xi)]^{m}. \tag{5.4}$$

The average waiting-time in the queue ($W_q$) is given by

$$W_q = \sum_{j=1}^{\infty} j w_j = \frac{K}{(1 - \xi) [1 - (\bar{\mu} + \mu \xi)]^{m}}. \tag{5.5}$$

The average sojourn time in the system is given by

$$W_s = W_q + \frac{1}{\mu} = \frac{K}{(1 - \xi) [1 - (\bar{\mu} + \mu \xi)]^{m}} + \frac{1}{\mu}. \tag{5.6}$$

We can obtain the mean system-length ($L_s$) using Little’s law as

$$L_s = \lambda d \left( \frac{K}{(1 - \xi) [1 - (\bar{\mu} + \mu \xi)]^{m}} + \frac{1}{\mu} \right). \tag{5.7}$$

5.3. **Equilibrium balking strategies.** Consider a tagged customer who decides to join the system with probability $d$ upon arrival to the unobservable system. Let $\Delta_c(d)$ denote the net benefit of the tagged customer, which is a function of the joining probability and the strategy followed by other customers. When all arriving
customers follow the same joining strategy \( d \), the net benefit of the tagged customer is

\[
\Delta_e(d) = R - C \cdot W_s(d) = R - C \left( \frac{K}{(1 - \xi) [1 - (\bar{\mu} + \mu \xi)^m]} + \frac{1}{\mu} \right). \quad (11)
\]

The tagged customer wishes to join if \( \Delta_e(d) > 0 \), it wishes to balk if \( \Delta_e(d) < 0 \), and is unconcerned between joining and balking if \( \Delta_e(d) = 0 \). Let us denote the equilibrium joining probability as \( d_e \). Solving \( \Delta_e(d) = 0 \), we get the positive and feasible root \( d_e^* \). A closed form expression for \( d_e^* \) is intractable due to the complex expression of the root as a function of the joining probability. However, we have computed the equilibrium balking strategy numerically for different system parameters. We numerically observe that \( W_s(d) \) is increasing with \( \lambda (\lambda < m\mu) \). It can be seen that the customer’s equilibrium joining strategy \( d_e \) is unique and \( d_e = \min \{d_e^*, 1\} \).

5.4. Socially optimal balking strategies. A central planner is interested in the net benefit of the whole society, known as the social benefit, which is the sum of the net benefits of the individual customers who joined the system. If all customers follow the joining strategy \( d \), then the effective arrival rate is \( \lambda d \). The social benefit per time unit, which can be expressed as

\[
\Delta_s(d) = \lambda d (R - CW_s(d)) = \lambda d R - C \lambda d \left( \frac{K}{(1 - \xi) [1 - (\bar{\mu} + \mu \xi)^m]} + \frac{1}{\mu} \right). \quad (12)
\]

We observe that \( \Delta_s(d) \) is differentiable with respect to \( d \) in \([0, 1]\). The first two derivatives are given by

\[
\Delta_s'(d) = \lambda (R - C \cdot W_s(d)) - \lambda C \cdot W_s'(d), \quad (13)
\]

\[
\Delta_s''(d) = -\lambda C \cdot W_s'(d) - \lambda C \cdot W_s''(d) = -2\lambda C \cdot W_s'(d). \quad (14)
\]

By solving \( \Delta_s'(d) = 0 \), one may find the socially optimal joining probability \( d^* \). As \( \rho < 1 \) and \( W_s'(d) \geq 0 \), \( \Delta_s''(d) < 0 \) for any \( d \in [0, 1] \). The function \( \Delta_s(d) \) is concave in \([0, 1] \cap \{d \mid W_s''(d) \geq 0\} \), and it reaches a unique maximum at the point \( d = d^* \), that is, \( \Delta_s(d^*) = \max \{\Delta_s(d), d \in [0, 1]\} \). The social benefit per time unit, when all customers follow the equilibrium joining strategy \( d_e \) is \( \lambda d_e R - C \lambda d_e \left( \frac{K}{(1 - \xi) [1 - (\bar{\mu} + \mu \xi)^m]} + \frac{1}{\mu} \right) \). A detailed numerical computation of the social benefit is presented in the next section.

5.5. Price of anarchy. The price of anarchy often measures the inefficiency of selfish behavior. It is the ratio of optimal and equilibrium value of social benefit. It measures the extent to which non-cooperation approximates cooperation. We can evaluate PoA using the formula

\[
PoA = \frac{\Delta_s(d^*)}{\Delta_s(d_e)}. \quad (15)
\]

The social benefit under equilibrium is \( \Delta_s(d_e) = \lambda d_e (R - CW_s(d_e)) = 0 \) when \( d_e \neq 1 \), and \( \Delta_s(1) = \lambda R - C W_s(1) \). The optimal social benefit \( \Delta_s(d^*) = \lambda d^*(R - CW_s(d^*)) \) when \( d^* \neq 1 \). As \( R\mu > C \), \( \Delta_s(d^*) > 0 \) for \( 0 \leq d^* < 1 \). Since, the social benefit under equilibrium is zero for \( 0 \leq d_e < 1 \), the PoA is infinity. When the joining strategy \( d_e = 1 = d^* \), the equilibrium and optimal social benefits are equal, so PoA is 1. A PoA value close to 1 suggests that the equilibrium is close to socially optimal, and hence the effects of selfish behavior are relatively beneficial.
6. **Numerical illustrations.** In this section, we illustrate the behavior of equilibrium joining and socially optimal strategies as well as equilibrium and social benefits under the variation of several system parameters. We present the influence of several parameters on the performance characteristics and discuss some qualitative aspects of the unobservable system under consideration. Figs. 3 - 5 show the influence of system parameters $\lambda$, $R$ and $C$ on the strategic joining behavior of customers. Fig. 3 depicts the impact of $\lambda$ on equilibrium joining probabilities $d_e$ and social optimal joining probability $d^*$. We observe that with the increase of $\lambda$, both equilibrium and social optimal joining probabilities decrease. One may note that the equilibrium probability $d_e$ is always higher than the social optimal joining probability $d^*$. It is because the customers who maximize their benefit and disregard the negative externalities that were led to affect the customers who join later. In Fig. 4, we notice that with an increase of $R$, the equilibrium joining probability $d_e$ increases. Intuitively, the customers are more likely to join when the reward rate $R$ is higher. The reason is that when customers get more rewards, they will prefer to join. There is a discontinuity in the social optimal joining probability $d^*$; that is, the social benefit function is bimodal. The change of the optimal rate from higher to the lower local optimum occurs in the continuous segment. From Fig. 5, we note that the equilibrium and socially optimal joining probabilities decrease monotonically with the increase of cost $C$. The reason is that a customer prefers to balk as the expected waiting time cost increases, and the expected net benefit is negative. We note that $d^* < d_e$, which indicates that the social optimization differs from the individual optimization.

Figs. 6 and 7 present the impact of $\lambda$ and waiting cost $C$ on behavior of equilibrium and social benefit. The equilibrium and social benefit decrease and reaches an optimum value with an increase of $\lambda$. There will be congestion in the system with an increase of $\lambda$, resulting in more waiting costs, which negatively affects the equilibrium benefit, and decreases. The equilibrium and social benefit become zero for sufficiently large values of $\lambda$ and $C$. Fig. 8 illustrates the impact of service rate $\mu$ on the expected waiting time for various number of servers. The expected waiting time decreases monotonically with the increase of service rate $\mu$.

Fig. 9 describes the impact of reward rate $R$ on PoA. We notice that the optimal social benefit decreases as the reward $R$ increases. It is intuitive since by increasing
the reward rate, customers are more probable to join and the effectiveness of the entire system drops. Thus, PoA decreases with reward rate $R$. In Fig. 10, we depict the impact of $\lambda$ on PoA. We observe that the PoA moderately increases and reaches an optimum value, then decreases and again increases with further increase.
in $\lambda$. For some values of $\lambda$, the social benefit are distributed in two separate intervals. When the arrival probability is higher, arriving customers are less inclined to join as they expect that the system is overloaded. Thus, the joining probability decreases with the increase of $\lambda$. When the proportion of the time that the server works increases, the delay of the joining customer is reduced; as a result, the socially optimal mixed strategy increases. Further, as $\lambda$ increases, the additional load imposed on the system by the joining customers exceeds the social benefit due to the higher arrival rate. The joining probability starts decreasing again. There is a discontinuity in PoA due to the bimodality of the social benefit function.

7. Conclusions. In this paper, we studied the equilibrium and socially optimal strategies in a multi-server discrete-time renewal input queue under an arrival first policy. A closed-form solution of the steady-state probabilities at pre-arrival epochs are derived, using the embedded Markov chain approach and the displacement operator method. Various numerical results are discussed to show the dependence of system parameters on the performance measures. Several comparison results for social and individual optimization are presented. Interesting features like bimodality of the PoA function against customer arrival rate are observed. Such exciting results will help the managers to look into pricing concerns and enable customers to have optimal strategies. The outcomes of this paper are as follows:

- We study the equilibrium balking strategy and social benefit of the presented system to make it economically practicable.
- From Fig. 3, we observe that both equilibrium and social optimal joining probabilities decrease with the increase of $\lambda$. Also, the equilibrium probability $d_e$ is always higher than the social optimal joining probability $d^\ast$.
- From Fig. 4, we note that with an increase of $R$, the equilibrium joining probability $d_e$ increases because the customers are more likely to join when the reward rate is higher.
- We illustrate the inefficiency of social optimization via the price of anarchy measures.
- The impact of reward rate $R$ on PoA is shown in Fig. 9. We notice that the PoA decreases as the reward $R$ increases. Because by increasing the reward rate, customers are more likely to join, and the effectiveness of the entire system drops.

Further enhancement of this study is perhaps to examine the equilibrium and socially optimal behavior under other information situations. Another challenging work may be to consider a multi-server queue with customer abandonments and different server vacation policies.

REFERENCES

[1] C. E. Bell and S. Stidham Jr, Individual versus social optimization in the allocation of customers to alternative servers, Management Science, 29 (1983), 831–839.
[2] W. Chan and D. Maa, The GI/Geom/N queue in discrete time, INFOR: Information Systems and Operational Research, 16 (1978), 232–252.
[3] M. L. Chaudhry, U. C. Gupta and V. Goswami, Relations among the distributions at different epochs for discrete-time GI/Geom/m and continuous-time GI/M/m queues, International Journal of Information and Management Sciences, 12 (2001), 71–82.
[4] J. P. Cosmas, G. H. Petit, R. Lehnert, C. Blondia, K. Kontovassilis, O. Casals and T. Theimer, A review of voice, data and video traffic models for atm, European Transactions on Telecommunications, 5 (1994), 139–154.
[5] M. De Prycker, *Asynchronous Transfer Mode solution for broadband ISDN*, Prentice Hall International (UK) Ltd., 1995.
[6] N. M. Edelson and D. K. Hilderbrand, *Congestion tolls for poisson queuing processes*, *Econometrica*, 43 (1975), 81–92.
[7] S. Gao and J. Wang, *Equilibrium balking strategies in the observable Geo/Geo/1 queue with delayed multiple vacations*, *RAIRO-Operations Research*, 50 (2016), 119–129.
[8] G. Gilboa-Freedman, R. Hassin and Y. Kerner, *The price of anarchy in the Markovian single server queue*, *IEEE Transactions on Automatic Control*, 59 (2013), 455–459.
[9] V. Goswami, *Analysis of discrete-time multi-server queue with balking*, *International Journal of Management Science and Engineering Management*, 9 (2014), 21–32.
[10] V. Goswami and G. Panda, *Optimal information policy in discrete-time queues with strategic customers*, *Journal of Industrial & Management Optimization*, 15 (2019), 689–703.
[11] V. Goswami and G. Panda, *Optimal customer behavior in observable and unobservable discrete-time queues*, *Journal of Industrial & Management Optimization*, 17 (2021), 299–316.
[12] A. Gravey and G. Hébuterne, *Simultaneity in discrete-time single server queues with bernoulli inputs*, *Performance Evaluation*, 14 (1992), 123–131.
[13] D. Guha, A. D. Banik, V. Goswami and S. Ghosh, *Equilibrium balking strategy in an unobservable GI/M/c queue with customers' impatience*, in *Distributed Computing and Internet Technology*, Springer, (2014), 188–199.
[14] D. Guha, V. Goswami and A. Banik, *Algorithmic computation of steady-state probabilities in an almost observable GI/M/c queue with or without vacations under state dependent balking and reneging*, *Applied Mathematical Modelling*, 40 (2016), 4199–4219.
[15] R. Hassin, *Rational Queueing*, CRC press, 2016.
[16] R. Hassin and M. Haviv, *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems*, Springer Science & Business Media, 2003.
[17] J. J. Hunter, *Mathematical Techniques of Applied Probability: Discrete Time Models: Basic Theory*, vol. 1, Academic Press, 1983.
[18] M. Jeffrey, *Asynchronous transfer mode: the ultimate broadband solution*, *Electronics & Communication Engineering Journal*, 6 (1994), 143–151.
[19] N. C. Knudsen, *Individual and social optimization in a multiserver queue with a general cost-benefit structure*, *Econometrica: Journal of the Econometric Society*, 40 (1972), 515–528.
[20] P. J. Kuehn, Reminder on queueing theory for atm networks, *Telecommunication Systems*, 5 (1996), 1–24.
[21] J.-Y. Le Boudec, *The asynchronous transfer mode: a tutorial*, *Computer Networks and ISDN Systems*, 24 (1992), 279–309.
[22] D. H. Lee, *A note on the optimal pricing strategy in the discrete-time Geo/Geo/1 queuing system with sojourn time-dependent reward*, *Operations Research Perspectives*, 4 (2017), 113–117.
[23] S. A. Lippman and S. Stidham Jr, *Individual versus social optimization in exponential congestion systems*, *Operations Research*, 25 (1977), 233–247.
[24] R. Lotfi, N. Mardani and G. W. Weber, Robust bilevel programming for renewable energy location, *International Journal of Energy Research*, 45 (2021), 7521–7534.
[25] R. Lotfi, B. Kargar, S. H. Hoseini, S. Nazari, S. Safavi and G. W. Weber, Resilience and sustainable supply chain network design by considering renewable energy, *International Journal of Energy Research*.
[26] Y. Ma, W.-q. Liu and J.-h. Li, *Equilibrium balking behavior in the Geo/Geo/1 queueing system with multiple vacations*, *Applied Mathematical Modelling*, 37 (2013), 3861–3878.
[27] Y. Ma and Z. Liu, *Pricing analysis in Geo/Geo/1 queueing system with multiple vacations*, *Mathematical Problems in Engineering*, 2015, Article ID 181653.
[28] G. Martin and L. Pankoff, *Optimal customer decisions in a G/M/c queue*, *Mathematical and Computer Modelling*, 10 (1988), 251–256.
[29] P. Naor, The regulation of queue size by levying tolls, *Econometrica*, 37 (1969), 15–24.
[30] G. Panda and V. Goswami, *Effect of information on the strategic behavior of customers in a discrete-time bulk service queue*, *Journal of Industrial & Management Optimization*, 16 (2020), 1369–1388.
[31] Y. A. Ra'ed and H. T. Mouftah, Survey of ATM switch architectures, *Computer Networks and ISDN Systems*, 27 (1995), 1567–1613.
[32] Y. Tang, P. Guo and Y. Wang, Equilibrium queueing strategies of two types of customers in a two-server queue, *Operations Research Letters*, 46 (2018), 99–102.

[33] T. Yang, J. Wang and F. Zhang, Equilibrium balking strategies in the Geo/Geo/1 queues with server breakdowns and repairs, *Quality Technology & Quantitative Management*, 11 (2014), 231–243.

[34] U. Yechiali, Customers’ optimal joining rules for the GI/M/s queue, *Management Science*, 18 (1972), 434–443.

[35] M. Yu and A. S. Alfa, Strategic queueing behavior for individual and social optimization in managing discrete time working vacation queue with bernoulli interruption schedule, *Computers & Operations Research*, 73 (2016), 43–55.

Received June 2021; 1st and 2nd revision September 2021; Final revision October 2021, Early access November 2021.

*E-mail address: veena.goswami@yahoo.com*

*E-mail address: gopinath.panda@gmail.com*