Delay Information Sharing in Two-Sided Queues

Mehmet Berat Aydemir
Walmart Global Tech, gsberat@gmail.com

Mohammad Delasay
College of Business, Stony Brook University, Stony Brook, NY 11794, mohammad.delasay@stonybrook.edu

Siddharth Prakash Singh
UCL School of Management, London E14 5AA siddharth.singh@ucl.ac.uk

Mustafa Akan
Tepper School of Business, Carnegie Mellon University, Pittsburgh PA 15217, akan@andrew.cmu.edu

We study delay information disclosure policies in on-demand platforms, modeled as two-sided queues, with two user classes—customers and providers—who seek matches to each other using the platform. The primary objective is maximizing the rate at which these matches occur by adopting one of three information regimes: the occupancy regime (disclosing the current system occupancy to both user categories) or two distinct asymmetric regimes (sharing no information with one user class while sharing occupancy information with the other). The users of each class are strategic and decide whether to join based on the available delay information. We use continuous-time Markov chains to model the system as a two-sided queue and employ equilibrium analysis to characterize the users' joining behavior and the platform's match rate under each information regime. The optimal policy reveals a complex dependence on system parameters and is strongly influenced by the users' patience profiles. We demonstrate analytically that it is strategically advantageous to withhold delay information from one user class, especially when it consists of a substantial number of relatively delay-insensitive users. The optimality of the asymmetric information-sharing regimes becomes more prevalent as the discrepancy in the patience profiles or the market sizes of the two user classes increases. However, our extensive numerical analyses find that the occupancy regime proves to be optimal in many other settings. In cases where it falls short of optimality, the sub-optimality gap is usually minimal (on average, $\sim 5\%$). Our findings hold crucial implications for platform managers, indicating that in such two-sided systems, the occupancy regime is a safe choice unless the two user classes exhibit a large patience profile discrepancy or a large market size imbalance. In such cases, opting for the occupancy information regime could adversely impact the platform's match rate. In such situations, carefully evaluating the chosen information-sharing strategy becomes imperative.

Key words: On-demand platforms; Two-sided queues; Strategic queues; Delay announcement

1. Introduction

Service firms often communicate delay information to influence customers' patronage behavior. Extensive research has been conducted on the choice of delay information disclosure policies in traditional one-sided markets (see Ibrahim 2018b, for a comprehensive review). However, two-sided on-demand platforms, which facilitate matching customers needing services with providers offering them, have introduced additional complexities to whether to share delay information with users. Managers of such platforms now face the challenge of determining what delay information to disclose to *each* side of the market and how it will impact users' behavior (UberPeople.net 2017). This decision carries significant implications for various on-demand platforms, including ride-sharing services (such as Uber and Lyft), on-demand food and grocery delivery (such as InstaCart and Postmates), and labor marketplaces (such as Handy and TaskEasy). We design and analyze strategic queueing models to investigate the effects of different delay disclosure policies on the match rate of on-demand platforms.

The primary technical challenge in choosing a delay disclosure policy for a two-sided platform is the interdependence between the participation decisions of providers and customers, which depend on the delay disclosure policy. Participation is endogenously determined – providers are free to determine their work schedules, and providers and customers typically lack strong allegiance to a particular platform. To tackle the challenges arising from "at-will participation", platforms employ diverse incentive structures to influence users' engagement and maximize overall profitability. Surge pricing mechanisms, for instance, empower platforms to adjust service prices to balance supply and demand (Ridester 2023, Handy 2023). Our research centers on a different mechanism to influences users' decisions to participate: The nature of delay information shared by the platform with users on each side of the market.

In the past, firms have exhibited uncertainty regarding the most effective approach for leveraging delay information sharing. As an illustration, Uber temporarily removed its drivers' access to information regarding the number of other drivers in their vicinity (UberPeople.net 2017), prompting opposition from drivers due to the resulting reduction in their ability to estimate waiting times for ride requests. Furthermore, there exist various types of information that can be disclosed. For instance, some ride-sharing platforms provide riders with current wait time estimates (see "Ride With Lyft" in Lyft 2023) while providing drivers with surge pricing maps, signifying varying anticipated delays (see "Busy Zones" in Lyft 2019).

We explore the effectiveness of various delay information disclosure policies for an on-demand platform represented as two-sided queueing models. Users are strategic in their joining behavior and belong to two *classes* – customers and providers. In the base model, users from both classes are either impatient (unwilling to wait for a match) or patient (willing to wait a finite amount of time for a match). Matches occur instantly and based on the first-come-first-served (FCFS) discipline. We aim to identify the policy that maximizes the *match rate* among those examined in this paper (which will be further elaborated upon). The match rate is a proxy for the platform's revenue and profitability, reflecting the average rate at which customers and providers are successfully paired. We analyze three policies: the *occupancy information regime* (Regime $\mathbb{O}$), wherein *both* customers and providers receive updates regarding the platform's current occupancy level (or equivalently the expected delay conditioned on the current number of users within the platform), and two *asymmetric* information regimes: one where customers (Regime $\mathbb{C}$) receive occupancy information while providers do not, and another where providers (Regime $\mathbb{P}$) receive occupancy information while customers do not. Building upon established models from one-sided delay announcement literature, we consider that users receive a reward upon receiving the service (being matched) while incurring a cost associated with the delay. Incoming users formulate an expectation of the delay cost based on the delay information available from the platform and decide about joining the platform or balking, aiming to maximize their overall expected utility.

Prior literature on delay announcement in *one-sided* queueing systems (e.g., Hassin 1986, Chen and Frank 2004) suggests that disclosing queue information is advisable when the system load is sufficiently high while concealing such information is preferred otherwise. However, it is not immediately clear how to translate these findings into a two-sided setting. One can plausibly apply these findings to compare the effectiveness of Regime $\mathbb{O}$ (where both providers and customers receive information) and Regime $\mathbb{P}$ (where only providers receive information) as a function of the volume of customers: Implementing Regime $\mathbb{O}$ may be appropriate when there is a sufficiently high volume of customers, while Regime $\mathbb{P}$ could be preferred otherwise. However, existing findings do not offer any guidance on the performance of Regime $\mathbb{C}$ in this relative comparison. This motivates the need for additional modeling and analysis in two-sided platforms where customers' and providers' joining decisions mutually influence each other.

We contribute by illustrating how the insights from one-sided systems can be translated into two-sided systems. Our work advances the modeling and analysis of two-sided systems (see Diamant and Baron 2019, for a recent example), specifically focusing on on-demand platforms that involve strategic users with varying degrees of sensitivity to delays. We formulate and examine queueing models for these two-sided systems, where arrival rates of users are endogenously determined. By conducting equilibrium analyses, we uncover users' decisions regarding joining or balking. This investigation allows us to determine the match rates under different information regimes, facilitating a comprehensive comparison. Our analytical findings reveal a parallel with the results from one-sided setups: As with one-sided settings, it is not universally optimal in two-sided settings to disclose delay information to both user classes. Notably, we find it optimal to hide delay information from the user class that has patient users with:

- Sufficiently low delay sensitivity and an ample number compared to patient users of the other class (Proposition 1).
- High delay sensitivity and an ample number compared to patient users of the other class, while the users of the other class are significantly delay-insensitive (Proposition 2).

- Moderate delay sensitivity when its users considerably outnumber those of the other class (Proposition 3).

Despite the scenarios mentioned above where it is beneficial to conceal delay information from one user class (i.e., Regime $\mathbb{P}$ or $\mathbb{C}$ may be optimal), we find from a comprehensive numerical investigation that it is generally optimal to share delay information with both user classes (Regime $\mathbb{O}$ is optimal in about 67% of our experiments). Further, our numerical investigation reveals that even when Regime $\mathbb{O}$ is sub-optimal, it does not typically result in a significant match rate loss compared to the optimal regime (on average, about 5%) unless there is a large patience profile discrepancy between the two user classes or a significant imbalance in the user classes' market sizes.

In addition, our analysis reveals that the optimal delay information regime for the platform may not always align with the users' preferences: Users may find that a different regime, other than the one maximizing the platform's match rate, maximizes their expected welfare. Specifically, when the platform finds Regime $\mathbb{P}$ or $\mathbb{C}$ optimal, either providers or customers prefer a different regime. However, the platform and both user classes may align on the choice of Regime $\mathbb{O}$. Overall, it is generally safe for the platform to implement Regime $\mathbb{O}$, as this generates minimal match rate regret and often aligns with the users' preferences.

Our main findings and insights are robust to two model extensions. First, users are matched according to the random service order instead of FCFS. Second, users are allowed to abandon after joining the queue. In a third extension, we demonstrate that in systems characterized by greater levels of users' patience heterogeneity (number of distinct patience sensitivity levels), the platform finds Regime $\mathbb{O}$ optimal over a broader range of parameters. This finding complements the results documented in Dobson and Pinker (2006) and Guo and Zipkin (2007) for one-sided systems.

2. Literature Review

We contribute to the growing literature on on-demand platforms. Wang and Yang (2019) have conducted a comprehensive literature survey covering various aspects of on-demand platforms, among which the Operations Management field has emphasized on the following: (i) Pricing (e.g., Cachon et al. 2017, Taylor 2018, Hu and Zhou 2019, Afeche et al. 2020, Afche and Akan 2016, 2023), (ii) matching (e.g., Dickerson et al. 2018, Lyu et al. 2019, Özkan and Ward 2020), (iii) relocation and dispatching of agents in ride-sharing systems (and sometimes combined with other decisions such as pricing) (e.g., Afeche et al. 2018, Braverman et al. 2019, Ata et al. 2020, Hosseini et al. 2021, Alwan et al. 2023), and (iv) sharing different types of information with the users of such platforms. In light of these research areas, we position our study within the broader context of the existing literature on on-demand platforms and contribute to understanding information sharing in these platforms.

The existing literature about information sharing in on-demand platforms primarily revolves around two key areas. Firstly, it explores the disclosure of *fare* or *destination* information to drivers within ride-sharing platforms (e.g., Rosenblat and Stark 2016, Chu et al. 2018). Secondly, it investigates the sharing of customers' attributes with providers (e.g., Romanyuk 2017, Romanyuk and Smolin 2019). These studies collectively suggest that, within specific contextual conditions and problems, complete disclosure of information may adversely affect the performance of platforms.

Our focus in this paper diverges from the aforementioned information-sharing studies as we focus on *delay information*, which refers to the time required for an agent to be matched with another agent from the opposing side of the queue. Furthermore, we explore the consequences of sharing such delay information with both classes of users, namely customers and providers. Additionally, we consider that users have heterogeneous levels of patience. While the disclosure of delay information has received considerable attention within traditional one-sided queues (as we review below), it remains relatively unexplored within the two-sided setting. Thus, our research fills this gap by examining the implications of delay information sharing in the context of two-sided queueing models of on-demand platforms.

Hassin (2016) offers a comprehensive review of the literature on one-sided strategic queueing systems that consider strategic customers making joining decisions based on available delay information. Ibrahim (2018b) provides a comprehensive review of literature specifically focusing on delay announcement in strategic queues, including the accuracy of different delay information structures (Armony et al. 2009, Ibrahim et al. 2017), models of the impact of anticipated delays on system performance (Jouini et al. 2011, Singh et al. 2023), and empirical investigations into the effects of delay information on user behavior (Akşin et al. 2013, Batt and Terwiesch 2015, Yu et al. 2017).

Our research is particularly relevant to the literature on delay announcements, which examines how different delay announcement structures can enhance service systems' performance. For instance, Hu et al. (2018) investigate the effect of informing only a fraction of customers about the real-time delay in a single-server queue and discover that some level of information heterogeneity can increase throughput and social welfare. Dimitrakopoulos et al. (2021) study a single-server system with alternating 'observable' and 'unobservable' queue periods and find that optimizing the duration of these periods generally improves equilibrium throughput and social welfare compared to a solely observable or unobservable queue. Another notable study by Lingenbrink and Iyer (2019) explores the optimal information disclosure policy in a single-server queue where strategic customers estimate their expected delay using Bayesian updating. They ascertain that the delay signaling mechanism that maximizes throughput is of a threshold nature, indicating that the platform's delay signal must encourage customer participation up to a specific occupancy level.

The two closely related papers Dobson and Pinker (2006) and Guo and Zipkin (2007) analyze and compare different delay information richness levels, albeit within one-sided settings. However, it should be noted that the findings of these papers regarding the optimal announcement of delay information cannot be directly extrapolated to a two-sided setting. Dobson and Pinker (2006) examine the sharing of lead-time information in a supply chain context and demonstrate that, with greater heterogeneity in user patience levels, sharing more detailed lead-time information improves throughput. Similarly, Guo and Zipkin (2007) employ a general model of customer patience heterogeneity, focusing on a single-server queue, and find that disclosing richer information leads to higher throughput when customers exhibit sufficient heterogeneity in their patience levels. We extend these insights to the two-sided setting. By comparing the match rate (equivalent to throughput in our case) under three specific patience heterogeneity distributions, we establish that sharing richer delay information benefits the platform when users are more heterogeneous. Our findings align with the results of Guo and Zipkin (2007) for a one-sided queue, demonstrating that, in the context of a two-sided platform, the optimal information structure determined by the platform *can* align with the users' best interests.

In a very recent working paper, Zhu et al. (2023) examine information disclosure for on-demand platforms with exogenous arrival rates using single and two-sided queues to model such platforms. In an extension, they consider a reduced form model wherein users' decisions to join are dictated by a probability equal to the likelihood of being matched. In contrast, our model takes a more intricate approach, wherein users' join/bulk decisions are determined endogenously by utility specifications that resonate with the rational user behavior modeling approach in Naor (1969) and the bulk of literature on strategic queueing, as surveyed in Hassin (2016) and Hassin and Haviv (2003). This utility-based join/balk behavior results in users employing a mixed strategy in the absence of delay information and joining according to an equilibrium arrival rate determined by their patience level. Our modeling approach requires solving for this equilibrium arrival in line with Edelson and Hilderbrand (1975) and Guo and Zipkin (2007). Furthermore, Additionally, our model accommodates heterogeneity in users' patience levels. These modeling differences also lead to different conclusions because, in our model, the platform takes into account the equilibrium implications of its disclosure policies. While Zhu et al. (2023) find that the choice of information disclosure policy is primarily driven by system load and exogenous arrival rates, we find that the choice of optimal disclosure policy is contingent upon the composition of patient and impatient users, the degree of patience exhibited by these users, and their potential arrival rates.

3. Model

We consider a two-sided queuing system (the platform) with two *user classes*—customers and providers—where customers arrive (notionally) at one side of the system, and providers arrive at the other. Customers use the services offered on the platform while providers supply them. Users of both classes are delay-sensitive and seek to maximize their utility (we define their utility functions in §3.2). A match occurs if a customer (resp., provider) arrives while a provider (resp., customer) is waiting for a match. In our base model, users are served on a first-come-first-served (FCFS) basis, ensuring fairness in clearing the waiting users. (In a model extension in §7.1, we allow the matching process to occur based on the random discipline and find that this does not significantly impact our insights). Once a user pair is successfully matched, they exit the system instantly; we discuss how the zero post-match delay assumption may be relaxed in §4. As a result, there are never simultaneous queues of customers and providers waiting to be matched at any given time.

The platform may decide to relay delay information about time-to-match to one or both user classes, based on which arriving users form an expectation of their system delay and accordingly make join/balk decisions. We do not consider users' abandonment after joining the system in the base model; however, we relax this assumption in a model extension and discuss its implications in §7.2. The types of relayed delay information form an information regime I, which induces a mapping from the current *system state* (the number and class of users waiting to be matched) to *delay signals* provided to each user class. In §3.1, we describe the information regimes considered in this paper. Subsequently, §3.2 introduces our model for the users' join/balk decisions.

3.1. Information Regimes

The platform's manager chooses an information regime that maximizes the *match rate* (i.e., the rate users are matched and leave the system). We use the match rate as a proxy for the revenue of the on-demand platform. Building on the "No Information" and "Partial Information" paradigms studied for one-sided queuing systems in Guo and Zipkin (2007), we construct and study the following information regimes for the two-sided setting, all of which result in stable systems:

- **Occupancy information to both user classes (Regime $\mathbb{O}$):** Under Regime $\mathbb{O}$, the platform signals the current system state to the arriving users of both classes. This is equivalent to signaling the current queue length-based-expected delay (Ibrahim and Whitt 2009) under the FCFS discipline.
- **Occupancy information to providers only (Regime $\mathbb{P}$):** Under Regime $\mathbb{P}$, the platform relays the current system state to providers and relays no information to customers.
- **Occupancy information to customers only (Regime $\mathbb{C}$):** Under Regime $\mathbb{C}$, the platform relays the current system state to customers and relays no information to providers.

Our analysis does not consider the fourth possible information regime, where neither user class receives any information. In this regime, both user classes arrive at a fixed rate independent of the system state. However, in a two-sided queue, the arrival rate of one user class is the service rate of the other. Therefore, this information regime leads to an unstable system.

3.2. Users' Joining Decisions

We consider two types of users denoted by superscript $u \in \{c, p\}$, where c and p represent customers and providers, respectively. Both types of users arrive at their respective sides of the platform at independent Poisson rates Λ^c and Λ^p. Users obtain no utility from balking and will join if it yields a positive expected utility. Once a user decides to join, their decision is irreversible, and they wait until they are matched, i.e., no abandonment is allowed in the base model; however, we include it in a model extension in §7.2. A user's utility is determined by their valuation $R^u \geq 0$ for a match and their expected delay cost, which depends on the following factors:

- The user's delay sensitivity θ^u, *individually drawn* from the probability distribution function $f^u(\theta^u), 0 \leq \theta^u \leq 1$, which captures the heterogeneous delay sensitivity of users of the same class.
- The expectation $\mathbb{E}[c^u(W_I^u(s))]$ of a cost function $c^u(\cdot)$ of the delay W_I^u the user will experience (identical for all users of the same class). This expectation is based on the delay signal s the user observes under the delay information Regime $I \in \{\mathbb{O}, \mathbb{P}, \mathbb{C}\}$. For tractability, we consider a linear cost function, i.e., $\mathbb{E}[c^u(W_I^u(s))] = a^u \, \mathbb{E}[W_I^u(s)] + b^u$ with parameters a^u and b^u.

Therefore, the expected utility of user u with delay sensitivity θ^u who receives signal s under delay information Regime I follows:

$$U_I^u(s) = R^u - \theta^u \left(a^u \, \mathbb{E}[W_I^u(s)] + b^u\right), \quad u \in \{c, p\}, \tag{1}$$

where we set $b^u = R^u$ to ensure: (i) $U_I^u(s) \geq 0$ when $\mathbb{E}[W_I^u(s)] = 0$ and (ii) $U_I^u(s) < 0$ when $\mathbb{E}[W_I^u(s)] > 0$ and $\theta^u = 1$. Therefore, (i) those users who would balk even when the signal indicates no delay are excluded, and (ii) the most delay-sensitive users balk when the signal indicates a non-zero delay. For convenience and without loss of generality, we scale time by a multiplicative factor of a^c/R^c, and then, a^c and a^p by a factor of R^c/a^c to leave the utility function unchanged. The resulting utility functions follow:

$$\begin{aligned} U_I^c(s) &= R^c \left(1 - \theta^c \left(1 + \mathbb{E}[W_I^c(s)]\right)\right), \\ U_I^p(s) &= R^p \left(1 - \theta^p \left(1 + K^p \, \mathbb{E}[W_I^p(s)]\right)\right), \end{aligned} \tag{2}$$

where $K^p = (a^p/a^c)(R^p/R^c)$ represents the relative value of time for the providers compared to the customers; i.e., for the same R^u and θ^u values, if $K^p > 1$, providers value time more than customers do, and therefore, they are willing to wait less (compared to customers) to be matched.

Users employ the provided delay signals to form beliefs about their expected delay (when it is not directly provided). This resembles the typical assumption in the strategic queueing literature with unobservable queues (for example, check Hassin 2016). These beliefs may depend on the equilibrium joining behavior of users of both classes, which is affected by the signals they observe.

We model this behavior formally by defining an equilibrium joining strategy as a set of probabilities, $\{J_I(\theta^u|s); \forall \theta^u, s\}$, each specifying the probability that a user of class u with delay sensitivity θ^u joins upon observing signal s under information Regime I. For these joining probabilities to form an equilibrium, it must be that every focal user has no incentive to deviate from the equilibrium joining probability $J_I(\theta^u|s)$, keeping all other equilibrium joining probabilities fixed.

To account for the heterogeneity in users' delay sensitivities and to provide tractability and clarity of insights, our base model considers two user *types* within each class depending on their delay sensitivity: *Patient* and *impatient*. Patient users are willing to wait for a match, whereas impatient users are not. Formally,

$$f^u(\theta^u) = \begin{cases} \delta^u & \text{for } \theta^u = t^u < 1 \\ 1-\delta^u & \text{for } \theta^u = 1 \end{cases}, \qquad u \in \{c, p\}. \tag{3}$$

Under the above two-point user heterogeneity model, *patient* users (for whom $\theta^u = t^u < 1$) arrive at rate $\Lambda^u \delta^u$, and *impatient* users (for whom $\theta^u = 1$) arrive at rate $\Lambda^u(1-\delta^u)$. In equilibrium, impatient users join with positive probability only if they expect a zero delay. However, patient users may join even if they expect some delay (based on their utility given in Eq. (2)). For convenience, we define the respective willingness-to-wait of patient customers and providers as

$$\omega^c = \frac{1-t^c}{t^c}, \tag{4}$$

$$\omega^p = \frac{1-t^p}{K^p t^p}, \tag{5}$$

where customers (resp., providers) obtain a zero expected utility if they expect to wait an amount of time ω^c (resp., ω^p) for a match. The two-point distribution captures the impact of user heterogeneity on the platform's choice of information regime in a tractable manner. In §7.3, we consider more granular delay sensitivity distributions for the users.

4. Analyzing the Delay Information Regimes

The delay information regime influences users' equilibrium joining strategies and, consequently, the platform's match rate. To obtain the match rates, we first construct a general underlying continuous-time Markov chain (CTMC) that we use later to represent and analyze the system's dynamics under each delay information regime. As the matches occur instantaneously, there are either no users or only providers or customers (but never both) in the system at any time. Therefore, the CTMC under any of the information regimes has a one-dimensional state space with the state variable $n \in \mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ representing that there are currently $|n|$ customers (resp. providers) in the system if $n < 0$ (resp. $n > 0$) and no customers or providers when $n = 0$. For a specific state n, we denote the *delay signal* that the platform broadcasts to user class $u \in \{c, p\}$ under information Regime I as $s_I^u(n)$. We denote the set of delay signals as $S_I^u = \cup_n s_I^u(n)$.

The users' instantaneous equilibrium arrival rates $\lambda_I^u(s_I^u(n))$ at state n depend on the delay signal $s_I^u(n)$ they receive. Accordingly, the long-run average arrival rate of users (customers or providers) is the weighted average of the instantaneous arrival rates $\lambda_I^u(s_I^u(n))$. The weights are the steady-state probabilities $\pi_I(n),\ n \in \mathbb{Z}$. In turn, these probabilities depend on the delay information regime I through the instantaneous arrival rates $\lambda_I^u(s_I^u(n))$ the delay information regime I induces. Since any user who joins will be matched eventually (in the absence of abandonment), the long-run average arrival rate equals the match rate. Therefore, we can write the match rate under delay information regime I as:

$$M_I = \sum_{n\in\mathbb{Z}} \pi_I(n)\lambda_I^c(s_I^c(n)) = \sum_{n\in\mathbb{Z}} \pi_I(n)\lambda_I^p(s_I^p(n)), \quad I \in \{\mathbb{O}, \mathbb{C}, \mathbb{P}\}. \tag{6}$$

In §§4.1-4.2, we obtain the match rates for all the delay information regimes. However, it is worth noting that users may choose not to join the system in equilibrium regardless of the signal they receive, resulting in a match rate of zero. Hence, we disregard this trivial equilibrium joining strategy for the rest of the paper unless it is the unique one: assuming that the system can have a non-zero match rate in equilibrium, we consider that users will join the system in a way that leads to this non-zero match rate.

4.1. Occupancy Information (Regime $\mathbb{O}$)

Under Regime $\mathbb{O}$, the platform provides the current system state to the arriving users; i.e., at state n, the signal that the users receive is $s_{\mathbb{O}}^c(n) = s_{\mathbb{O}}^p(n) = n$, leading to the signal sets $S_{\mathbb{O}}^c = S_{\mathbb{O}}^p = \mathbb{Z}$. A positive (resp., negative) signal indicates that providers (resp., customers) are waiting to be matched. An alternative signaling that results in an identical joining behavior is to signal to providers (resp., customers) the number of providers (resp., customers) when there are no customers (resp., providers) and to signal "no delay" when there are customers (resp., providers). This roughly mimics the queue position information ride-hailing platforms provide drivers in airports (Paul 2015) and the expected delay information these platforms provide riders.

Under Regime $\mathbb{O}$, the expected delay of an arriving user from class u is entirely determined by the signal s the user observes, and it does not depend on the joining behavior of other users of the same class receiving signal s. Accordingly, each arriving user either balks or joins with probability one (i.e., no mixed strategy equilibrium). More explicitly, each arriving *customer* who observes $s_{\mathbb{O}}^c(n) = n > 0$ faces a zero expected delay; therefore, they will join with probability one regardless of their type (i.e., whether they are patient with delay sensitivity $\theta^c = t^c$ or they are impatient with $\theta^c = 1$), which leads to an instantaneous joining rate $\lambda_{\mathbb{O}}^c(s_{\mathbb{O}}^c(n)) = \Lambda^c$ for $n > 0$. Following a similar argument for *providers*, we have $\lambda_{\mathbb{O}}^p(s_{\mathbb{O}}^p(n)) = \Lambda^p, n < 0$.

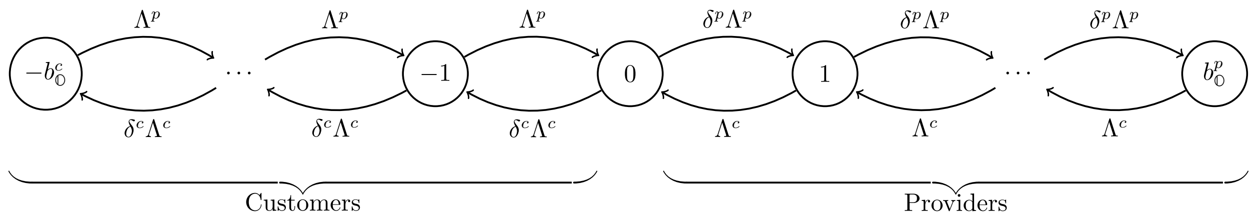


Figure 1 **Regime $\mathbb{O}$ CTMC**

Users join differently when their signal indicates a delay. An arriving customer who observes $s^c_{\mathbb{O}}(n) = n \le 0$ faces an expected delay of $\mathbb{E}[W^c_{\mathbb{O}}(n)] = \dfrac{|n|+1}{\Lambda^p}$ (each customer ahead of them takes an exponentially distributed time with mean $\dfrac{1}{\Lambda^p}$ to be matched). If the matched users do not leave the system instantly (i.e., the post-match delay is non-zero), we can add the expected post-match delay to $\mathbb{E}[W^c_{\mathbb{O}}(n)]$ (assuming that the delay signal is updated the moment the match occurs); for ease of exposition, we do not consider this possibility. If the customer is impatient, she will not join; otherwise, she will join with probability one when $|n|$ is sufficiently small. Accordingly, $\lambda^c_{\mathbb{O}}(s^c_{\mathbb{O}}(n)) = \delta^c\Lambda^c$ when $|n|$ is sufficiently small, and $\lambda^c_{\mathbb{O}}(s^c_{\mathbb{O}}(n)) = 0$ otherwise. Formally, patient customers join as long as they gain a positive expected utility (i.e., $U^c_{\mathbb{O}}(n) \ge 0$), which yields the condition in Eq. (7). By an analogous argument, when $n \ge 0$, impatient providers balk, and patient providers join with probability one if and only if the condition in Eq. (8) holds.

$$0 \le |n| \le \Lambda^p\omega^c - 1. \tag{7}$$

$$0 \le n \le \Lambda^c\omega^p - 1. \tag{8}$$

According to Eqs. (7) and (8), patient users only join up to a threshold system state on their side of the platform (i.e., when the signal indicates a delay). Therefore, the resulting CTMC will be bounded on both sides at the *bounding states* $b^c_{\mathbb{O}}$ and $b^p_{\mathbb{O}}$, as illustrated in Fig. 1. From Eqs. (7) and (8), those bounding states follow:

$$b^u_{\mathbb{O}} = \left\lfloor \Lambda^{u'}\omega^u \right\rfloor, \qquad u, u' \in \{c, p\}, u \ne u'. \tag{9}$$

Note that patient customers (resp., providers) balk even at state 0 when the bounding state $b^c_{\mathbb{O}} = 0$ (resp., $b^p_{\mathbb{O}} = 0$); i.e., they balk given any signal that indicates a non-zero expected delay. We characterize the match rate under Regime $\mathbb{O}$ in Lemma 1.

Lemma 1. *Regime $\mathbb{O}$'s match rate $M_{\mathbb{O}}$ follows:*

$$M_{\mathbb{O}} = \Lambda^p \frac{\rho^c(1-\rho^p)\left(1-(\rho^c)^{b^c_{\mathbb{O}}}\right) + \delta^p(1-\rho^c)\left(1-(\rho^p)^{b^p_{\mathbb{O}}}\right)}{\rho^c(1-\rho^p)\left(1-(\rho^c)^{b^c_{\mathbb{O}}}\right) + (1-\rho^c)\left(1-(\rho^p)^{b^p_{\mathbb{O}}+1}\right)}, \tag{10}$$

where $\rho^u = \dfrac{\delta^u\Lambda^u}{\Lambda^{u'}}, u \in \{c, p\}, u \ne u'$.

The proof of this and all subsequent Lemmas and Propositions are in the Appendix.

4.2. Asymmetric Information Regimes (Regimes $\mathbb{P}$ and $\mathbb{C}$)

In asymmetric regimes, the platform shares the current system state with one user class and shares no information with the other. For brevity, we present our analysis for the regime where the current system state information is shared with providers (i.e., Regime $\mathbb{P}$); we present the results for Regime $\mathbb{C}$ at the end of this section. Under Regime $\mathbb{P}$, when the system state is n, an arriving provider receives signal $s^p_{\mathbb{P}}(n) = n$, but an arriving customer receives no information (which we denote by signal $s^c_{\mathbb{P}}(n) = \emptyset$, $\forall n \in \mathbb{Z}$). This leads to signal sets $S^p_{\mathbb{P}} = \mathbb{Z}$ and $S^c_{\mathbb{P}} = \emptyset$.

Since customers receive no information under Regime $\mathbb{P}$, their joining behavior is independent of the system state, i.e., $\lambda^c_{\mathbb{P}}(s^c_{\mathbb{P}}(n)) = \lambda^c_{\mathbb{P}}(\emptyset), \forall n \in \mathbb{Z}$ (we explain how to obtain the customers' equilibrium arrival rate $\lambda^c_{\mathbb{P}}(\emptyset)$ later). However, the providers' joining behavior is state-dependent. Similar to Regime $\mathbb{O}$, when the state indicates no delay for providers under Regime $\mathbb{P}$, i.e., $s^p_{\mathbb{P}}(n) = n < 0$, an arriving provider joins with probability one (independent of being patient or impatient); hence, $\lambda^p_{\mathbb{P}}(s^p_{\mathbb{P}}(n)) = \Lambda^p$ for $n < 0$. On the other hand, if an arriving provider receives a signal $s^p_{\mathbb{P}}(n) = n \geq 0$, her expected delay depends on the equilibrium arrival rate $\lambda^c_{\mathbb{P}}(\emptyset)$ of customers, and we have $\mathbb{E}[W^p_{\mathbb{P}}(n)] = \dfrac{n+1}{\lambda^c_{\mathbb{P}}(\emptyset)}$ (each provider already in the system and ahead of the arriving provider takes an exponentially distributed amount of time with mean $\dfrac{1}{\lambda^c_{\mathbb{P}}(\emptyset)}$ to be matched; as with Regime $\mathbb{O}$, if matches do not occur instantly, we can add the expected post-match delay to this expectation). If the provider is patient, she joins with probability one if n is sufficiently small; if she is impatient, she balks. Accordingly, $\lambda^p_{\mathbb{P}}(n) = \delta^p \Lambda^p$ if n is sufficiently small, and $\lambda^p_{\mathbb{P}}(n) = 0$, otherwise. Formally, patient providers join with probability one if and only if their utility from joining $U^p_{\mathbb{P}}(n) \geq 0$, which yields the following condition:

$$0 \leq n \leq \lambda^c_{\mathbb{P}}(\emptyset)\omega^p - 1. \tag{11}$$

Accordingly, the bounding state $b^p_{\mathbb{P}}$ up to which patient providers join with probability one under Regime $\mathbb{P}$ is given by:

$$b^p_{\mathbb{P}} = \lfloor \lambda^c_{\mathbb{P}}(\emptyset)\omega^p \rfloor. \tag{12}$$

Fig. 2 shows the resulting CTMC under Regime $\mathbb{P}$. The bounding state $b^p_{\mathbb{P}}$ depends on the equilibrium arrival rate $\lambda^c_{\mathbb{P}}(\emptyset)$ of customers given that they receive no information. We now explain how to obtain the equilibrium arrival rate $\lambda^c_{\mathbb{P}}(\emptyset)$. Any equilibrium with $\lambda^c_{\mathbb{P}}(\emptyset) > 0$ involves customers joining at state 0 (otherwise, customers would never join, resulting in a trivial equilibrium with zero match rate). Customers experience a positive delay if they join at states $n \leq 0$, and they experience no delay if they join at states $n > 0$. Therefore, their expected delay under Regime $\mathbb{P}$ will be positive, i.e., $\mathbb{E}[W^c_{\mathbb{P}}(\emptyset)] > 0$. As a result, *impatient* customers do not join as their utility $U^c_{\mathbb{P}}(\emptyset) = R^c\left(1 - 1\left(1 + \mathbb{E}[W^c_{\mathbb{P}}(\emptyset)]\right)\right) < 0$. On the other hand, *patient* customers either join with a

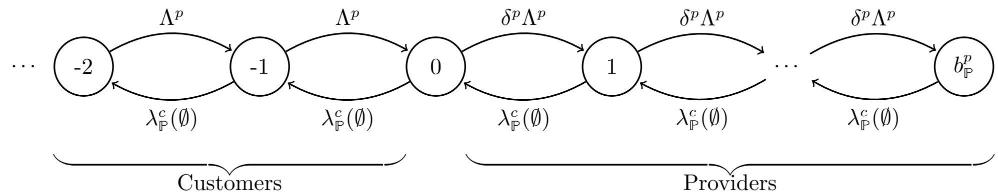


Figure 2 **Regime $\mathbb{P}$ CTMC**

positive probability (i.e., $J_{\mathbb{P}}(t^c|\emptyset)>0$) or balk (i.e., $J_{\mathbb{P}}(t^c|\emptyset)=0$), depending on the impact of their equilibrium joining behavior on their expected delay. The expected delay of patient customers, given the equilibrium joining probability $J_{\mathbb{P}}(t^c|\emptyset)$, when they receive no information, follows:

$$\mathbb{E}[W^c_{\mathbb{P}}(\emptyset)|J_{\mathbb{P}}(t^c|\emptyset)] = \sum_{n=-\infty}^{0} \frac{|n|+1}{\Lambda^p}\pi_{\mathbb{P}}(n), \tag{13}$$

where the stationary probabilities $\pi_{\mathbb{P}}(i)$ can be obtained in terms of $b^p_{\mathbb{P}}$ and $\lambda^c_{\mathbb{P}}(\emptyset)$ which depend on the joining probability $J_{\mathbb{P}}(t^c|\emptyset)$.

The values $b^p_{\mathbb{P}}$ and $\lambda^c_{\mathbb{P}}(\emptyset)$ need to be found to be consistent with each other. For this, we examine different cases depending on the customers' joining probability ($J_{\mathbb{P}}(t^c|\emptyset)=0,\ 0<J_{\mathbb{P}}(t^c|\emptyset)<1$, or $J_{\mathbb{P}}(t^c|\emptyset)=1$) and the bounding state for providers ($b^p_{\mathbb{P}}=0$ or $b^p_{\mathbb{P}}>0$). For each case that leads to a non-zero match rate, we characterize this match rate in Lemma 2 using Eq. (6), which simplifies to $M_{\mathbb{P}}=\lambda^c_{\mathbb{P}}(\emptyset)$ since customers receive the same signal in all states.

Lemma 2. *The match rate under Regime $\mathbb{U}\in\{\mathbb{C},\mathbb{P}\}$ could be $M_{\mathbb{U}}=\delta^{u'}\Lambda^{u'}$, $M_{\mathbb{U}}=\Lambda^u-\frac{1}{\omega^{u'}}$, or the $M_{\mathbb{U}}$ that solves (14):*

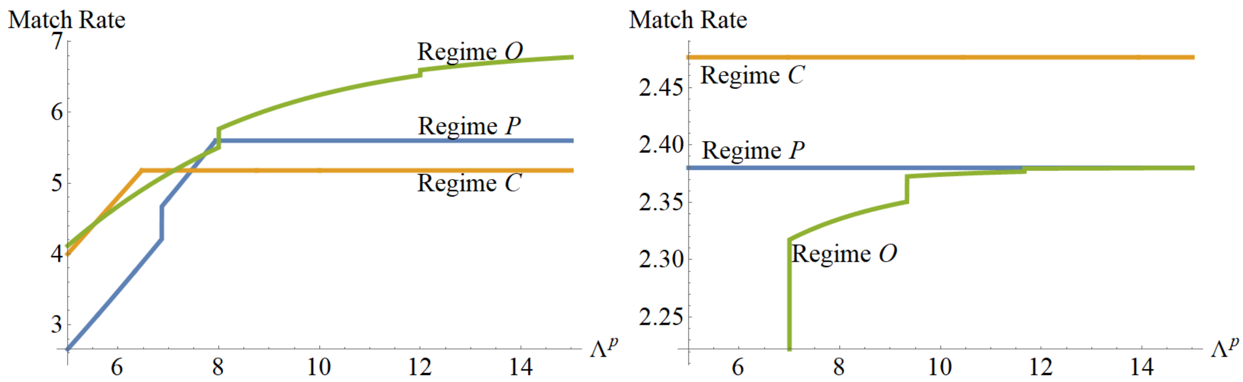
$$\frac{\Lambda^u - M_{\mathbb{U}}}{\delta^u\Lambda^u - M_{\mathbb{U}}}\left(M_{\mathbb{U}}\left(1-\delta^u\right) - \delta^u\left(\Lambda^u - M_{\mathbb{U}}\right)\left(\frac{\delta^u\Lambda^u}{M_{\mathbb{U}}}\right)^{\lfloor M_{\mathbb{U}}\omega^u\rfloor}\right) = \frac{1}{\omega^{u'}}, \tag{14}$$

depending on the specific conditions on the parameters presented in closed form in Appendix B.1.

If none of the conditions in Lemma 2 are satisfied, users do not join the system, leading to a trivial match rate of zero. Under Regimes $\mathbb{P}$ and $\mathbb{C}$, multiple equilibria may arise for a given parameter setting because either more than one of the cases in Lemma 2 hold or Eq. (14) may yield multiple solutions. In such cases, we choose the equilibrium with a higher match rate. We elaborate on the multiple equilibria issue in Appendix C.

5. Platform's Optimal Information Regime

In this section, we utilize the match rate expressions derived in §4 to ascertain the conditions under which each information regime maximizes the match rate of the platform. The characterization of these conditions is intricate due to the case-specific match rates for Regimes $\mathbb{P}$ and $\mathbb{C}$, as outlined in Lemma 2, and the analytical complexity of Eq. (14). Consequently, the optimal regime



(a) $\{t^c, t^p, \delta^c, \delta^p\} = \{0.8, 0.7, 0.8, 0.8\}$ (b) $\{t^c, t^p, \delta^c, \delta^p\} = \{0.7, 0.88, 0.34, 0.67\}$

Figure 3 Complicated structure of the optimal information regime; illustration for $\Lambda^c = 7$, $K^p = 1$

is highly sensitive to parameter changes and does not exhibit straightforward comparative statics. For instance, Fig. 3a demonstrates multiple shifts in the regime that maximizes the match rate as Λ^p increases. Moreover, in contrast to Fig. 3a, where Regime $\mathbb{O}$ outperforms other regimes for most values of Λ^p, Regime $\mathbb{O}$ is consistently sub-optimal in the scenario depicted in Fig. 3b.

While a complete characterization of the optimal regime remains challenging, we present Proposition 1, which provides conditions for each asymmetric regime (Regime $\mathbb{P}$ or $\mathbb{C}$) to be optimal. These conditions are significant as they indicate instances where the platform's match rate is negatively affected by providing information to both user classes (Regime $\mathbb{O}$).

Proposition 1. *Providing delay information to both user classes may hurt the match rate of two-sided platforms. Specifically, hiding information from user class $u \in \{c, p\}$ (i.e., Regime $\mathbb{U}'$) is optimal when patient users of class u (a) have sufficiently low delay sensitivity such that they always join under Regime $\mathbb{U}'$ (the conditions for which are specified in the proof), (b) arrive at a higher rate (i.e., $\delta^u \Lambda^u > \delta^{u'} \Lambda^{u'}$), and (c) their proportion is higher than a threshold such that:*

$$\delta^u > \left(1 + \frac{1-\rho^{u'}}{\rho^{u'}} \times \frac{(\rho^u)^{b^u_{\mathbb{O}}}}{1-(\rho^{u'})^{b^{u'}_{\mathbb{O}}}}\right)^{-1}. \tag{15}$$

We explain the intuition for Proposition 1 when $u = c$ and $u' = p$ (i.e., the optimality of Regime $\mathbb{P}$); the intuition for when $u = p$ and $u' = c$ (i.e., the optimality of Regime $\mathbb{C}$) follows similarly. For clarity of discussions, we first restate Proposition 1 when $u = c$ and $u' = p$: *Regime $\mathbb{P}$ is optimal when patient customers (a) have sufficiently low delay sensitivity such that they always join under Regime $\mathbb{P}$, (b) arrive at a higher rate ($\delta^c \Lambda^c > \delta^p \Lambda^p$), and (c) their proportion is higher than a threshold such that:*

$$\delta^c > \left(1 + \frac{1-\rho^p}{\rho^p} \times \frac{(\rho^c)^{b^c_{\mathbb{O}}}}{1-(\rho^p)^{b^p_{\mathbb{O}}}}\right)^{-1}. \tag{16}$$

Proposition 1 indicates that Regime $\mathbb{P}$ tends to be optimal when the resulting loss of impatient customers (who *always* balk when they receive no information) is relatively small. When patient customers consistently join under Regime $\mathbb{P}$ (as stated in condition (a) of the proposition), the match rate $M_{\mathbb{P}}$ for Regime $\mathbb{P}$ is equal to $\delta^c\Lambda^c$, as per Lemma 2(a). If $\delta^c\Lambda^c > \delta^p\Lambda^p$ (condition (b) in the proposition), Regime $\mathbb{P}$ outperforms the best possible match rate $\delta^p\Lambda^p$ of Regime $\mathbb{C}$.

Next, let us compare Regime $\mathbb{P}$ to Regime $\mathbb{O}$. Switching from Regime $\mathbb{P}$ to Regime $\mathbb{O}$ brings about two significant changes that affect the match rate. First, impatient customers join when they observe zero delays, leading providers to be willing to join even when the queue is longer compared to Regime $\mathbb{P}$. Second, there will be a maximum queue length on the customers' side, beyond which patient customers no longer join, resulting in a finite customers' bounding state. While the first change favors Regime $\mathbb{O}$, the second favors Regime $\mathbb{P}$. When the impact of the first change is negligible (i.e., when there are only a few impatient customers), the combined effect of both changes favors Regime $\mathbb{P}$. Consequently, there exists a threshold on δ^c, above which Regime $\mathbb{P}$ is preferable to Regime $\mathbb{O}$ (condition (c) in the proposition).[1]

In cases where the proportions of patient customers and providers are not adequately high, and impatient users are abundant on both sides of the market (i.e., condition (c) of Proposition 1 does not hold for either $u = p$ or $u = c$), Regime $\mathbb{O}$ is optimal. Therefore, we can make the following remark:

REMARK 1. Regime $\mathbb{O}$ maximizes the match rate (among the regimes in our study) when condition (15) fails to hold for $u = p$ and $u = c$.

The general intuition from the one-sided literature suggests that it is optimal to reveal delay information in high-congestion systems. Interestingly, disclosing delay information to the user class experiencing relatively higher congestion in our two-sided setting is not always the best strategy due to the interdependence between the two user classes. To illustrate this, we provide two numerical examples in Appendix E in which the parameters satisfy Proposition 1's conditions for the optimality of Regime $\mathbb{P}$. The first example involves higher congestion on the providers' side, while the second involves higher congestion on the customers' side.

Proposition 1 explains how providing delay information to both user classes can detrimentally impact the platform's match rate. To enhance our understanding of these circumstances, we examine two extreme scenarios in §§5.1-5.2. In §5.1, we construct a scenario with a significant discrepancy in the delay sensitivity parameters, resulting in one side of the platform being considerably more patient than the other. This is achieved by allowing one of the parameters, t^c or t^p, to approach its limiting value of 0, thereby rendering the users of that class insensitive to delays. In §5.2, we

[1] In Eq. (15), the right hand side also includes expressions with δ^u. However, the right-hand side is decreasing in δ^u, which results in a threshold on δ^u beyond which Eq. (15) holds.

Table 1 Experiment settings

$\Lambda^c = \{1, 5, 10, 15, 20\} + 0.001$	$\Lambda^p = \{1, 5, 10, 15, 20\} - 0.001$
$t^c = \{0.05, 0.1, 0.15, 0.2, 0.3, ..., 0.95\} + 0.001$	$t^p = \{0.05, 0.1, 0.15, 0.2, 0.3, ..., 0.95\} - 0.001$
$\delta^c = \{0.15, 0.35, 0.55, 0.75, 0.95\} + 0.001$	$\delta^p = \{0.15, 0.35, 0.55, 0.75, 0.95\} - 0.001$

create an imbalanced situation by pushing the market size of one user class to its limit. This is accomplished by either increasing the market size of customers ($\Lambda^c \to \infty$) or providers ($\Lambda^p \to \infty$), creating disparities in the potential congestion levels between the two sides of the platform.

While our primary emphasis in §§5.1-5.2 is on analytical findings and insights, we also include brief numerical experiments to complement the analytical results. Subsequently, we conduct extensive numerical experiments in Section §6 to obtain more comprehensive insights regarding the optimal information regime. To provide context for the numerical experiments, we present the parameter values used in our numerical settings in Table 1.

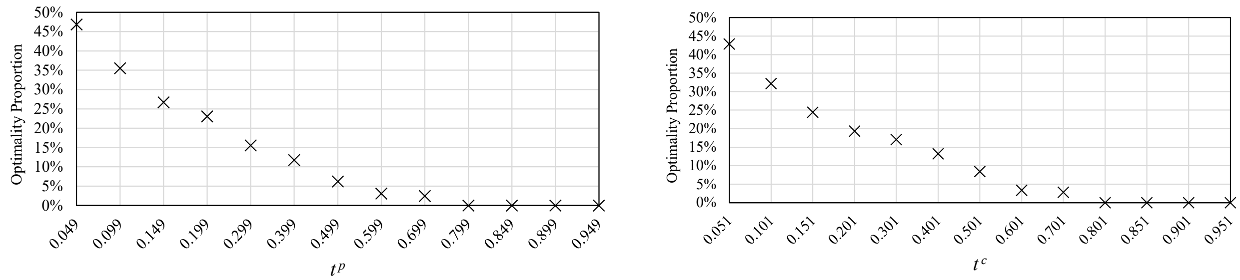
5.1. High Discrepancy in User's Patience Profiles

The level of delay sensitivity among users significantly influences their decision to join the platform, and a large discrepancy in the patience profiles of the two user classes could result in one side of the platform being more suitable for information disclosure/concealment than the other. This section investigates how the patience profile disparity between providers and customers determines whether disclosing information to both user classes (Regime $\mathbb{O}$) harms the platform's match rate.

In Proposition 2, we explore an extreme scenario where patient users of one class exhibit insensitivity to delays while patient users of the other class are highly sensitive to delays. This significant contrast in patience profiles enables us to identify the conditions under which concealing delay information from one user class proves more effective than adopting Regime $\mathbb{O}$.

PROPOSITION 2. *It may be optimal to hide information from one user class when there is a large discrepancy between the patience profiles of the two user classes. Specifically, it is favorable to conceal information solely from user class $u \in \{c, p\}$ (i.e., Regime $\mathbb{U}'$) when its patient users arrive at a higher rate (i.e., $\delta^u \Lambda^u > \delta^{u'} \Lambda^{u'}, u' \neq u$) and are sufficiently delay-sensitive (i.e., $t^c > T_1^c := \dfrac{\Lambda^p}{1 + \Lambda^p}$ when $u = c$ and $t^p > T_1^p := \dfrac{\Lambda^c}{K^p + \Lambda^c}$ when $u = p$), while the other class's patient users are delay-insensitive (i.e., $t^{u'} \to 0$).*

We provide an intuitive explanation for Proposition 2 specifically when $u = c$ and $u' = p$, demonstrating the optimality of Regime $\mathbb{P}$. The rationale for when $u = p$ and $u' = c$, leading to the optimality of Regime $\mathbb{C}$, follows a similar line of reasoning. To ensure clarity in our discussions, we begin by restating Proposition 2 when $u = c$ and $u' = p$: *Regime $\mathbb{P}$ is optimal when patient customers arrive at a higher rate (i.e., $\delta^c \Lambda^c > \delta^p \Lambda^p$) and are sufficiently delay-sensitive (i.e., $t^c > T_1^c$), while patient providers are delay-insensitive (i.e., $t^p \to 0$).*



(a) $u=c$ in Proposition 2 (Regime $\mathbb{P}$ optimality) (b) $u=p$ in Proposition 2 (Regime $\mathbb{C}$ optimality)

Figure 4 **Asymmetric regimes become optimal more frequently as t^u decreases (among experiments where Proposition 2 conditions for patient users of class u' hold).**

When patient customers exhibit high sensitivity to delays (i.e., $t^c > T_1^c$), they tend to avoid joining an empty system (state 0) when provided with expected delay information under Regimes $\mathbb{C}$ and $\mathbb{O}$. As a result, impatient providers are discouraged from joining the system, leading to a maximum achievable match rate of $\delta^p\Lambda^p$ under Regimes $\mathbb{C}$ and $\mathbb{O}$. However, under Regime $\mathbb{P}$, the platform can incentivize impatient providers to join when there is an excess of customers. This is possible because customers, whose joining rate is independent of the system state under Regime $\mathbb{P}$, can be encouraged to join at a higher rate than $\delta^p\Lambda^p$ if there are enough patient customers available. This is facilitated by the fact that patient providers, who are almost insensitive to delay ($t^p \to 0$), are willing to join the system up to a significantly higher bounding state, thereby reducing the expected delay experienced by customers towards zero.

Proposition 2 offers analytical insights into the optimality of concealing delay information from one user class in the case of an extreme difference in patience profiles ($t^c \to 0$ or $t^p \to 0$). However, our numerical experiments, conducted using the parameter settings outlined in Table 1, reveal that the asymmetric regimes (Regimes $\mathbb{P}$ and $\mathbb{C}$) can frequently emerge as the optimal choice even when the discrepancy in users' patience profiles is not as pronounced as in Proposition 2 (given the other conditions in the proposition hold). Specifically, the results indicate that the asymmetric regimes are more commonly optimal as the patient users of one class become less sensitive to delays (i.e., more delay-insensitive), while the conditions outlined in Proposition 2 hold for patient users of the other class. This is evident in Fig. 4a (for customers) and Fig. 4b (for providers), where we focus solely on experiments where the conditions specified in Proposition 2 apply to the respective user class. In these cases, the proportion of experiments for which Regime $\mathbb{P}$ (in Fig. 4a) and Regime $\mathbb{C}$ (in Fig. 4b) are optimal increases as the respective delay sensitivities of providers and customers decrease.

5.2. Market Size Imbalance

In this section, we construct a situation where one side of the platform is more congested than the other by creating an imbalance between the market sizes of the two sides of the platform.

Specifically, we investigate how the difference in population sizes (and, therefore, arrival rates) affects the effectiveness of disclosing delay information to both classes. Proposition 3 considers the extreme scenario where the arrival rate of one user class is *unbounded* while the arrival rate of the other is bounded. It establishes conditions under which disclosing delay information solely to the user class with the *bounded* arrival rate is advantageous, rendering Regime $\mathbb{O}$ suboptimal.

PROPOSITION 3. *When the arrival rate of user class $u \in \{c,p\}$ is unbounded (i.e., $\Lambda^u \to \infty$):*

(a) It is always optimal to share delay information with user class $u' \neq u$ (i.e., Regime $\mathbb{O}$ or $\mathbb{U}'$ are optimal).

(b) It is optimal to hide information from user class u (i.e., Regime $\mathbb{U}'$ is optimal) when their delay sensitivity is intermediate, i.e., the range $t^u \in [T_1^u, T_2^u)$ is defined (i.e., $T_1^u < T_2^u$) where

$$T_1^u := \frac{\Lambda^{u'}}{i + \Lambda^{u'}} \quad \text{and} \quad T_2^u := \left(1 + \frac{i}{\Lambda^{u'}(1-\delta^{u'})}\left(1 + (1-\delta^{u'})\left\lfloor \delta^{u'}\Lambda^{u'}\omega^{u'}\right\rfloor\right)^{-1}\right)^{-1}, \tag{17}$$

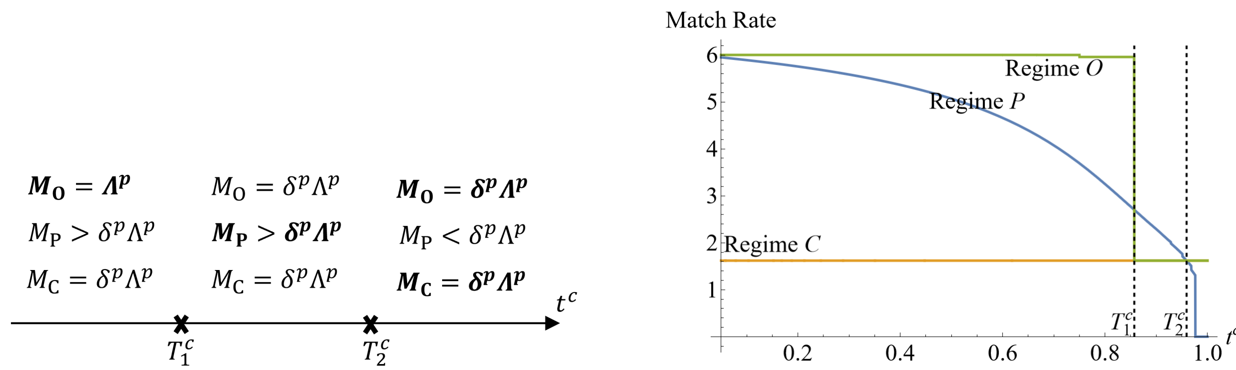
where $i = 1$ (resp., $i = K^p$) when $u = c$ (resp., $u = p$).

For clarity of our discussions, we first reiterate Proposition 3 when $u = c$ and $u' = p$: *When $\Lambda^c \to \infty$), (a) it is always optimal to share delay information with providers (i.e., optimality of Regime $\mathbb{O}$ or $\mathbb{P}$), and (b) it is optimal to hide information from customers (i.e., optimality of Regime $\mathbb{P}$) when their delay sensitivity is intermediate, i.e., the range $t^c \in [T_1^c, T_2^c)$.*

To provide intuition for Proposition 3, we explain below each regime's system dynamics and match rates when $\Lambda^c \to \infty$; check the proof for a more detailed discussion:

- Under Regime $\mathbb{O}$, the patient customers' joining behavior to an empty system determines the match rate. If they are sufficiently patient to join when the system is empty, the system almost always will have an excess of customers, leading to $M_{\mathbb{O}} = \Lambda^p$. Otherwise, the system will almost always be empty, leading to $M_{\mathbb{O}} = \delta^p\Lambda^p$.
- Under Regime $\mathbb{P}$, customers must employ a mixed strategy to avoid crowding, leading to a match rate between 0 and Λ^p depending on the patient customers' delay sensitivity: When $t^c > T_2^c$ (resp., $t^c < T_2^c$), $M_{\mathbb{P}} < \delta^p\Lambda^p$ (resp., $M_{\mathbb{P}} > \delta^p\Lambda^p$).
- Under Regime $\mathbb{C}$, patient providers do not receive any delay information upon arrival and join at rate $\delta^p\Lambda^p$ (regardless of their delay sensitivity) as they expect to be matched almost instantly, leading to $M_{\mathbb{C}} = \delta^p\Lambda^p$.

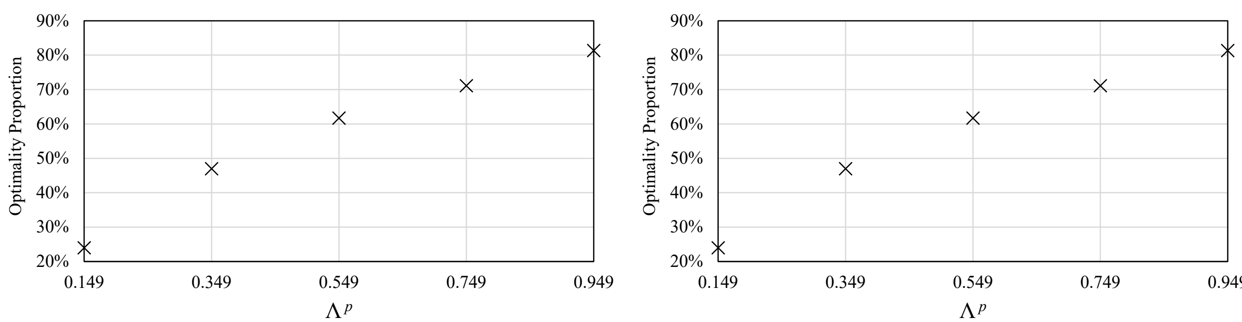
Based on the above discussion, Fig. 5a visualizes the comparison between the three information regimes for different regions of the customers' delay sensitivity t^c when $\Lambda^c \to \infty$, and Fig. 5b illustrates the comparison in a specific numerical example where Λ^c is much larger than Λ^p (i.e., $\Lambda^c = 1000$ and $\Lambda^p = 10$).



(a) $\Lambda^c \to \infty$; the optimal regime is highlighted in each region.

(b) Numerical example; $\Lambda^c = 1000 \gg \Lambda^p = 10, \delta^c = 0.6, \delta^p = 0.27, t^p = 0.38, K^p = 0.41$.

Figure 5 Match rates under market size imbalance when $T_{\mathbb{O}} < T_{\mathbb{P}}$



(a) $u = p$ in Proposition 3 (Regime $\mathbb{P}$ optimality) (b) $u = c$ in Proposition 3 (Regime $\mathbb{C}$ optimality)

Figure 6 Asymmetric regimes become optimal more frequently as Λ^u increases (among experiments where $t^u \in [T_1^u, T_2^u)$).

When the condition $T_1^u < T_2^u$ in Proposition 3(b) does not hold, we can make the following remark, supported by the match rate characterizations and illustrated by a numerical example presented in Appendix H:

Remark 2. When $\Lambda^u \to \infty$ and $T_1^u > T_2^u$ (for either $u = c$ or $u = p$), sharing delay information with both user classes (i.e., Regime $\mathbb{O}$) is weakly optimal regardless of t^u.

Proposition 3 provides analytical insights about when hiding delay information from one user class is optimal for the extreme market sizes imbalance (i.e., $\Lambda^c \to \infty$ or $\Lambda^p \to \infty$). The results of our numerical experiments (based on the parameter settings in Table 1) show that the asymmetric regimes (i.e., Regimes $\mathbb{P}$ and $\mathbb{C}$) can frequently be the optimal regime even when the market size imbalance is not as extreme as in Proposition 3. Specifically, as Figs. 6a-6b show, the asymmetric Regime $\mathbb{U}$ is optimal more frequently among experiments where $t^u \in [T_1^u, T_2^u)$ as $\Lambda^{u'}$ increases.

To summarize, when one user class is significantly more abundant than the other, the platform should share delay information with the less abundant class to inform them about the abundance of the more abundant class. However, whether to share delay information with the abundant

Table 2 Summary of the results of the numerical experiments

	Regime $\mathbb{O}$	Regime $\mathbb{P}$	Regime $\mathbb{C}$
Optimality proportion (%)	66.72	12.10	12.47
Avg. % optimality (sub-optimality) magnitude	63.35 (5.08)	7.48 (38.38)	6.43 (38.18)
Max. % optimality (sub-optimality) magnitude	18634.53 (55.92)	126.84 (100)	120.44 (100)

class depends on their sensitivity to delays. When the abundant type has an intermediate delay sensitivity, it may be optimal to withhold delay information from them.

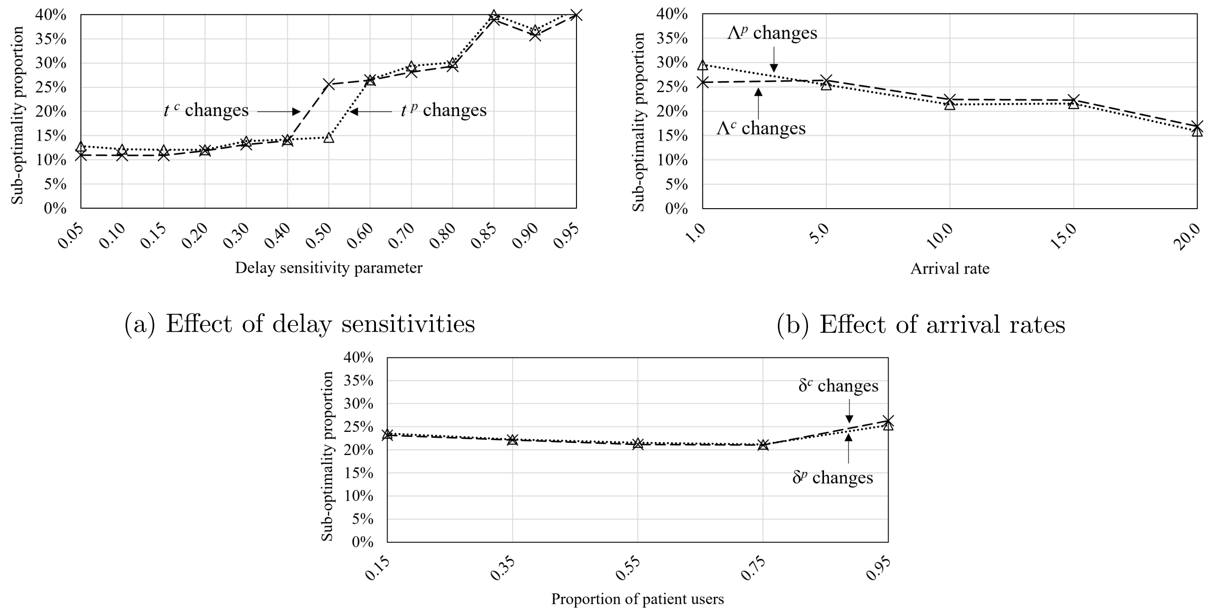
6. Numerical Analysis

Our analytical characterizations in §5 specify scenarios in which disclosing delay information to both parties of the platform may adversely affect the platform's match rates. To explore the generality of these findings and uncover other insights, we design an extensive experiment set by setting $K^p = 1$ and varying the remaining parameters as outlined in Table 1. This amounts to a total of $105,625$ experiments, of which we omit the $3,300$ that yield a match rate of zero under all regimes, leaving us with $102,325$ experiments. In each experiment, we document the match rate for each information regime and identify the regime yielding the highest match rate. In 8.71% of the experiments, we observe that two regimes exhibit an equal maximum match rate, resulting in a tie; in all of those experiments, the tie is between either Regimes $\mathbb{C}$ and $\mathbb{O}$ or Regimes $\mathbb{P}$ and $\mathbb{O}$.

Table 2 provides a summary of the results, indicating that Regime $\mathbb{O}$ emerges as the optimal choice in a considerable proportion (66.72%) of the experiments, yielding an average of 63.35% (and a maximum of 18634.53%) higher match rate than the second-best regime. On the other hand, when Regime $\mathbb{O}$ falls short of achieving the maximum match rate, its average sub-optimality magnitude compared to the optimal regime is merely 5.08%, although it can reach as high as 55.92%. In approximately 24% of the experiments, the asymmetric regimes $\mathbb{P}$ and $\mathbb{C}$ are optimal, offering respective average match rate advantages of 6.43% and 7.48% over Regime $\mathbb{O}$.

Based on the plots in Fig. 7, the occurrence of Regime $\mathbb{O}$'s sub-optimality increases with the users' delay sensitivities t^c and t^p (Fig. 7a) and decreases with their arrival rates Λ^c and Λ^p (Fig. 7b). In Fig. 8, we plot the 99% confidence interval of Regime $\mathbb{O}$'s sub-optimality magnitude in response to changes in the customers' parameters. Notably, the sub-optimality magnitude decreases with δ_c (Fig. 8c). However, it tends to be higher for more extreme values of t^c (Fig. 8a) but more intermediate values of Λ_c (Fig. 8b). We obtain similar insights when analyzing the plots corresponding to the changes in the provider parameters, as presented in Appendix I.

The findings illustrated in these plots indicate that even when platform managers opt to implement Regime $\mathbb{O}$, despite it not being the optimal choice, it is unlikely to result in a significant detriment to the match rate. This observation is further supported by the second row of Table 2,

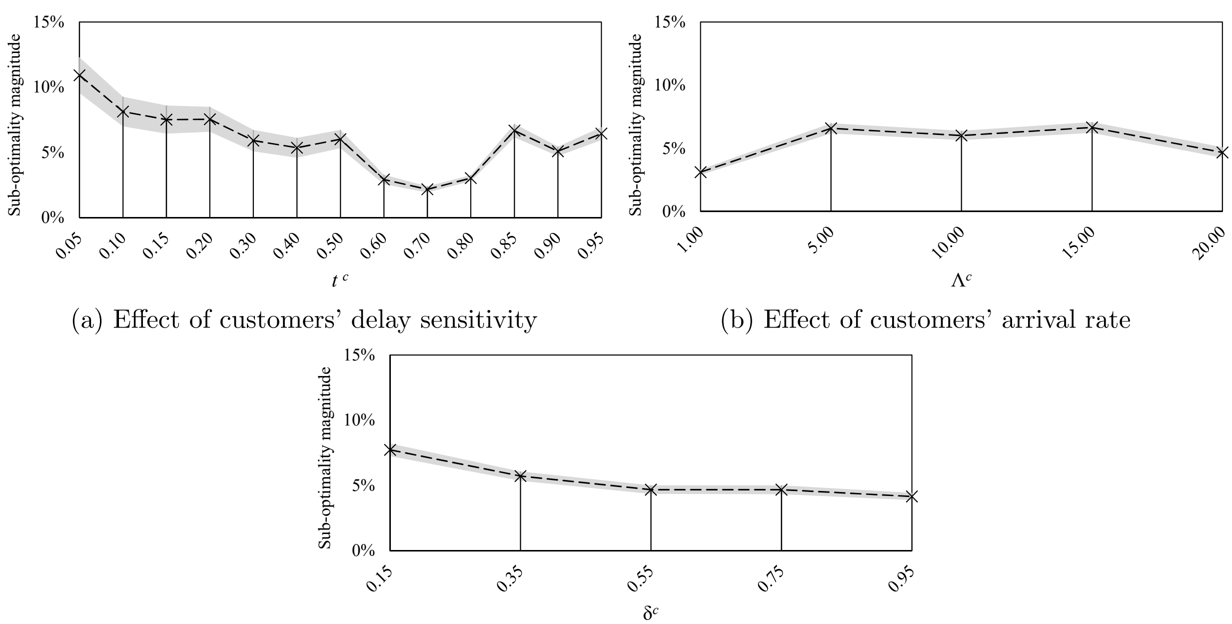


(a) Effect of delay sensitivities

(b) Effect of arrival rates

(c) Effect of patient users proportions

Figure 7 The marginal effect of parameters on Regime $\mathbb{O}$'s sub-optimality proportion



(a) Effect of customers' delay sensitivity

(b) Effect of customers' arrival rate

(c) Effect of patient customers' proportion

Figure 8 The marginal effect of parameters on Regime $\mathbb{O}$'s sub-optimality magnitude (99% confidence interval)

where it is evident that the average sub-optimality of Regime $\mathbb{O}$ amounts to a mere 5.08%. Therefore, the implementation of Regime $\mathbb{O}$, even when not the optimal regime, generally has a limited negative impact on the match rate.

Table 3 Optimality magnitude of Regimes $\mathbb{P}$ and $\mathbb{C}$ over $\mathbb{O}$ under relaxed versions of Propositions 2 and 3

(a) Regime $\mathbb{P}$

		Prop. 3 Not hold	Prop. 3 Hold
Prop. 2	Not hold	0.43%	21.74%
	Hold	8.75%	30.63%

(b) Regime $\mathbb{C}$

		Prop. 3 Not hold	Prop. 3 Hold
Prop. 2	Not hold	0.45%	22.09%
	Hold	7.26%	29.91%

Nevertheless, it is important to note that the asymmetric regimes, namely Regimes $\mathbb{P}$ and $\mathbb{C}$, exhibit substantially greater match rate enhancements in specific experiments, with improvements reaching as high as over 126% compared to Regime $\mathbb{O}$ (as indicated in the third row of Table 2). By focusing on the specific experiments in which Regimes $\mathbb{P}$ and $\mathbb{C}$ emerge as the optimal choices, we can leverage Propositions 2 and 3 to effectively characterize the parameter configurations that lead to significant optimality magnitudes for these regimes:

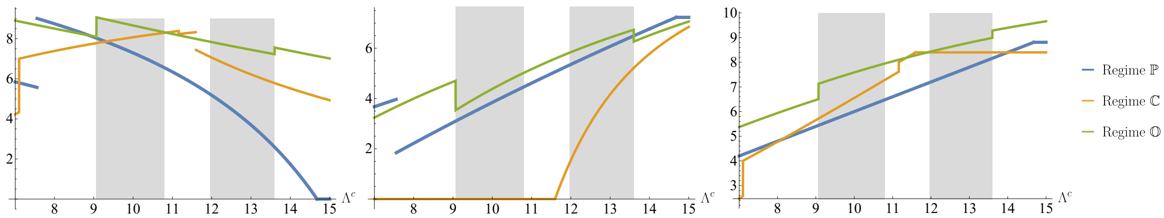
- According to Proposition 2, Regime $\mathbb{P}$ (resp., $\mathbb{C}$) is optimal when $\delta^c\Lambda^c > \delta^p\Lambda^p$, $t^c > T_1^c$, and $t^p \to 0$ (resp., $\delta^p\Lambda^p > \delta^c\Lambda^c$, $t^p > T_1^p$, and $t^c \to 0$). Now, consider relaxed versions of Proposition 2 in our experiments by setting $t^p \le 0.199$ and $t^c \le 0.201$ for the optimality of Regimes $\mathbb{P}$ and $\mathbb{C}$, respectively.
- According to Proposition 3, Regime $\mathbb{P}$ (resp., $\mathbb{C}$) is optimal when $t^c \in [T_1^c, T_2^c)$ (resp., $t^p \in [T_1^p, T_2^p)$) and $\Lambda^c \to \infty$ (resp., $\Lambda^c \to \infty$). Now, consider relaxed versions of Proposition 3 in our experiments by setting $\Lambda^c = 20.001$ and $\Lambda^p = 19.999$ (their highest values) for the optimality of Regimes $\mathbb{P}$ and $\mathbb{C}$, respectively.

Table 3 reports the magnitude of the optimality of Regimes $\mathbb{P}$ and $\mathbb{C}$ over Regime $\mathbb{O}$ depending on whether the above-relaxed versions of the conditions in Propositions 2 and 3 hold. In cases where neither of the relaxed conditions from the propositions is met, the asymmetric regimes demonstrate a minimal improvement, averaging around 0.43%-0.45% over Regime $\mathbb{O}$. However, when either set of the relaxed conditions is satisfied, the asymmetric regimes outperform Regime $\mathbb{O}$ with an average improvement ranging from 7.26%-22.09%. The most substantial improvement, with an average improvement of 29.9%-30.63%, is observed when both sets of the relaxed conditions are met.

The above analysis provides valuable guidance on when the sub-optimality of Regime $\mathbb{O}$ may be considerable. Specifically, when a platform manager faces a situation with a large patience profile discrepancy between the user classes and/or imbalances in market sizes between the two classes, an asymmetric regime may offer a significantly better match rate.

6.1. Alignment Between Users' Preferences and Platform's Optimal Regime

Lastly, we examine the effect of the platform's regime choice on the users' average utility under the three information regimes. Specifically, we investigate whether the platform's optimal regime also maximizes the utility of both user classes, which we refer to as "full alignment."



(a) Customers' Avg. Utility (b) Providers' Avg. Utility (c) Platform's Match Rate

Figure 9 **Users' utility and platform's match rate** **(**$\Lambda^p = 12, R^c = R^p = 100, K^p = 0.8, t^c = 0.8, t^p = 0.85, \delta^c = 0.6, \delta^p = 0.7$**)**

The average utility of user class u under information Regime I can be calculated as $\mathbb{E}[U_I^u] = \sum_{s\in S_I^u} \Pr(s)\mathbb{E}[U_I^u(s)]$, where $\Pr(s)$ is the probability that user class $u \in \{c,p\}$ receives signal s under information Regime $I \in \{\mathbb{O}, \mathbb{P}, \mathbb{C}\}$.and evaluate them numerically to assess the possibility of alignment between the platform and the users.[2]

Our results indicate that in certain cases, a user class might achieve the highest utility when information is concealed from the other user class. For example, in Fig. 9a, customers benefit when information is hidden from providers when Λ^c is approximately between 10.8 and 11.6 (the region where Regime $\mathbb{C}$'s utility for customers is the highest of all three regimes). More interestingly, a user class may be better off when information is hidden from *their own class*. For customers, in Fig. 9a, this occurs when Λ^c is approximately between 7.5 and 9.1 (the region where Regime $\mathbb{C}$'s utility for customers is the lowest of all three regimes). We explain this phenomenon in Appendix J. Our finding that users might prefer less information in a two-sided setting aligns with a similar effect found in one-sided settings as noted in Guo and Zipkin (2007).

Although a user class's welfare may be maximized under a regime where one class (as explained, possibly themselves) receives no information, we observe in our experiments that it is always the case that either the other user class or the platform finds that regime suboptimal. As a result, full alignment can only be achieved when Regime $\mathbb{O}$ maximizes the platform's match rate. The shaded regions in the plots of Fig. 9, which specify parameter values where this alignment is achieved, illustrate this general observation. Note, however, that full alignment under Regime $\mathbb{O}$ is not guaranteed. We observe full alignment in 53.4% of the experiments where Regime $\mathbb{O}$ is optimal for the platform.

7. Alternative Modeling Choices

This section discusses other modeling choices and demonstrates their impact on the results. In §7.1, we consider the random order matching process instead of FCFS and find no significant impact on

[2] The average utility expressions as a function of the model primitives and the match rate are available as a Mathematica notebook at https://tinyurl.com/twosideduserutility.

our results. In §7.2, we consider the possibility of users abandoning the platform and discover that while Regime $\mathbb{O}$ remains the most frequently optimal, Regimes $\mathbb{P}$'s and $\mathbb{C}$'s optimality proportions increase. In §7.3, we examine how the choice of users' patience distributions affects our results, revealing that greater variation in patience levels results in Regime $\mathbb{O}$ being optimal more frequently.

7.1. Random Service Order Matching Discipline

When using the random service order (RSO) matching discipline, a user arriving on the empty side of the platform is paired randomly with one of the users waiting on the opposite side, if there are any. Under RSO, computing the match rates is relatively straightforward when the bounding states are known, as the structures of the resulting CTMCs closely resemble those of the FCFS discipline. Therefore, the primary challenge lies in determining the bounding states up to which users are willing to join the system. The bounding states can be obtained using the recursive equations in Proposition 4:

PROPOSITION 4. *Consider a user from class $u \in \{c,p\}$ arriving at state n. Given the bounding state b_I^u when the information regime is I, the solution to the following recursive equations yields the user's expected waiting time $\mathbb{E}(W_I^u(n))$:*

$$\mathbb{E}[W_I^u(n)] = \frac{1}{\delta^u\Lambda^u + a} + \frac{\delta^u\Lambda^u}{\delta^u\Lambda^u + a}\,\mathbb{E}[W_I^u(n+i)] + \left(\frac{a}{\delta^u\Lambda^u + a}\right)\left(\frac{|n|}{|n|+1}\right)\mathbb{E}[W_I^u(n-i)], \tag{18}$$

$$\mathbb{E}[W_I^u(0)] = \frac{1}{\delta^u\Lambda^u + a} + \frac{\delta^u\Lambda^u}{\delta^u\Lambda^u + a}\,\mathbb{E}[W_I^u(i)], \tag{19}$$

$$\mathbb{E}[W_I^u(ib_I^u - i)] = \frac{1}{a} + \frac{b_I^u - 1}{b_I^u}\,\mathbb{E}[W_I^u(ib_I^u - 2i)], \tag{20}$$

where $n \in \{1,...,b_I^u - 2\}$ and $i = 1$ when $u = p$, and $n \in \{-1,...,-b_I^u + 2\}$ and $i = -1$ when $u = c$. Additionally, $a = \Lambda^{u'}$, $\lambda_{\mathbb{P}}^c(\emptyset)$, and $\lambda_{\mathbb{C}}^p(\emptyset)$ when $I = \mathbb{O}$, $\mathbb{P}$, and $\mathbb{C}$, respectively. Finally, when $I = \mathbb{P}$ (resp., $I = \mathbb{C}$), b_I^u is only defined for $u = p$ (resp., $u = c$).

For Regime $\mathbb{O}$, we use the procedure outlined in Proposition 4 iteratively to arrive at the appropriate bounding states consistent with t^c and t^p and then use these bounding states to compute the match rate according to Eq. (6). For Regimes $\mathbb{P}$ (resp., $\mathbb{C}$), we jointly determine $\lambda_{\mathbb{P}}^c(\emptyset)$ and $b_{\mathbb{P}}^p$ (resp., $\lambda_{\mathbb{C}}^p(\emptyset)$ and $b_{\mathbb{C}}^c$) that are consistent with each other, using the procedure in Proposition 4 with an artificial truncation limit $b_{\mathbb{P}}^c$ (resp. $b_{\mathbb{C}}^p$). Then, we compute the match rates for the truncated chains according to Eq. (6). We present the details in Appendix M.

Using the RSO discipline, we replicate the experiments described in Table 1. Under RSO, compared to FCFS, users who join the platform when a few (resp., many) users of their class are already in the system could anticipate longer (resp., shorter) waiting times. Therefore, the impact of the service discipline change on the platform's match rate is uncertain. We find that the average

Table 4 Summary of the results of the numerical experiments under RSO

	Regime $\mathbb{O}$	Regime $\mathbb{P}$	Regime $\mathbb{C}$
Optimality proportion (%)	66.59	9.66	9.72
Avg. % optimality (sub-optimality) magnitude	62.79 (6.28)	9.65 (38.22)	8.51 (38.08)
Max. % optimality (sub-optimality) magnitude	18484.91 (55.99)	127.24 (100)	120.44 (100)

increase in match rate resulting from transitioning to the RSO discipline is merely 0.18%, with a 90% confidence interval of [0%, 1.93%].

Table 4 summarizes the outcomes of the experiments. Upon comparing the values reported in Tables 2 and 4, it becomes evident that transitioning from FCFS to RSO has minimal impact on the optimality proportions and the optimality and sub-optimality gaps. The only noticeable change is a reduction in the optimality proportions of Regimes $\mathbb{P}$ and $\mathbb{C}$, which can be attributed to increased cases where there is a tie between Regime $\mathbb{O}$'s match rate and the match rate of either Regime $\mathbb{P}$ or $\mathbb{C}$. Therefore, our key findings concerning regime optimality closely align with those observed under the FCFS discipline.

Upon a closer examination of the experiments, we observe that among the experiments where Regime $\mathbb{O}$ is optimal under the FCFS discipline, it remains optimal in 99.6% under RSO (i.e., only in 0.4% of such experiments, the optimal regime changes to $\mathbb{P}$ or $\mathbb{C}$ after switching to RSO). Furthermore, among the experiments where Regime $\mathbb{P}$ (resp., Regime $\mathbb{C}$) is optimal under FCFS, Regime $\mathbb{P}$ (resp., Regime $\mathbb{C}$) retains optimality in 81.6% (resp., 82.8%) under RSO. Overall, the choice of service discipline has a negligible effect on our main results.

7.2. Allowing for Users' Abandonment

One assumption in our base model is that once users join the platform, they do not abandon it. In this section, we consider that users may abandon their queues; we consider exponential abandonment times with rates $\alpha^u, u \in \{c,p\}$. We depict the CTMCs of the abandonment models in Appendix L. Considering the updated transition rates, we can find the new bounding states and match rates by excluding users who join but subsequently abandon the system.[3]

We replicated the experiments described in Table 1 for the abandonment model by setting the abandonment rates α^c and α^p according to the users' willingness to wait ω^u: Recognizing that users are generally more willing to wait after committing to join compared to their initial willingness (attributed to an effect similar to the "sunk cost fallacy" documented in Kuzu (2015) and Liu et al. (2022) and modeled in Ibrahim (2018a)), we set $\alpha^u = (2\omega^u)^{-1}, u \in \{c,p\}$.

Including users' abandonment in the model leads to two opposing forces on the match rate. First, users are more inclined to join as they anticipate that some users ahead of them will abandon. More

[3] By allowing abandonment, a fourth information regime becomes relevant under which neither side receives occupancy information, as the resulting model will be stable. To maintain consistent comparisons with our base model, we exclude the fourth regime from our analysis and continue to examine only Regimes $\mathbb{O}$, $\mathbb{P}$, and $\mathbb{C}$.

Table 5 Summary of the results of the numerical experiments when abandonment is allowed

	Regime $\mathbb{O}$	Regime $\mathbb{P}$	Regime $\mathbb{C}$
Optimality proportion (%)	43.75	27.61	25.73
Avg. % optimality (sub-optimality) magnitude	81.35 (12.75)	18.18 (44.04)	14.67 (44.03)
Max. % optimality (sub-optimality) magnitude	17869.39 (66.42)	146.57 (100)	195.95 (100)

users joining the system impacts the match rate positively. However, the actual act of abandoning negatively impacts the match rate. We find in our experiments that the net effect of these two forces results in an average increase of 8.6% in the match rate under the abandonment model, with a 90% confidence interval of $[-6.24\%, \ 46.62\%]$. This average increase could be due to adjusting the abandonment rates based on the users' willingness to wait, as we explained above.

Table 5 summarizes the outcomes of the experiments under the abandonment models. Comparing Tables 2 and 5 reveals that while Regime $\mathbb{O}$ remains the most frequently optimal regime, its optimality proportion is lower compared to the base model without abandonment. Specifically, the experiments that identified Regimes $\mathbb{P}$ or $\mathbb{C}$ as optimal in the base model still find these regimes optimal when abandonment is considered. However, approximately 17% (resp., 15%) of experiments that identified Regime $\mathbb{O}$ as optimal in the base model find Regime $\mathbb{P}$ (resp., $\mathbb{C}$) optimal when abandonment is considered. Moreover, the optimality and sub-optimality magnitudes increase under all regimes relative to the base model, indicating that the regime choice has a more significant impact when abandonment is considered. This finding underscores the importance of carefully evaluating the regime choice, particularly for platforms with users who tend to abandon.

7.3. Higher Degrees of Heterogeneity in Users' Patience Levels

Our base model considers a two-point distribution for the users' delay sensitivities, presented in Eq. (3), in which we consider two user types: patient and impatient. In this section, we explore the impact of the users' delay sensitivity distribution on the optimal information regime. Specifically, if one thinks of different values that users' delay sensitivities could take as the degree of heterogeneity in patience levels (similar to Guo and Zipkin 2007), we investigate the effect of increased user heterogeneity by considering the following three settings, all of which use uniform distributions with the same means for a fair comparison:

- Two-point uniform: Both user classes' delay sensitivities exhibit two-point uniform distributions (following Eq. (21)). This is a special case of the delay sensitivity distributions in Eq. (3) where $\delta^c = \delta^p = 0.5$.
- Hybrid: Providers' delay sensitivity exhibits a two-point uniform distribution (following Eq. (21)), while customers' delay sensitivity exhibits a continuous uniform distribution (following Eq. (22)). For the sake of brevity, we omit this setting's symmetric analog.
- Continuous uniform: Both user classes' delay sensitivities exhibit continuous uniform distributions (following Eq. (22)).

Table 6 **Optimality proportions for different delay sensitivity distributions**

	Regime $\mathbb{O}$	Regime $\mathbb{P}$	Regime $\mathbb{C}$
Two-point uniform	81.38%	9.61%	8.99%
Hybrid	90.3%	6.44%	3.15%
Continuous uniform	100%	0%	0%

$$f_2^u(\theta^u) = 0.5 \quad \text{for} \quad \theta^u \in \{t^u, 1\}, \tag{21}$$

$$f_{uni}^u(\theta^u) = \frac{1}{1-t^u} \quad \text{for} \quad \theta^u \in [t^u, 1]\,. \tag{22}$$

We conduct our experiments for the above settings using Table 1's parameters, except we fix $\delta^c = 0.5 = \delta^p = 0.5$, resulting in 4225 experiments. Table 6 summarizes the results, indicating that as we transition from the two-point uniform setting to hybrid and subsequently to continuous uniform (increasing the users' patience levels heterogeneity), Regime $\mathbb{O}$ becomes optimal for a broader range of parameters, while the optimality of Regimes $\mathbb{P}$ and $\mathbb{C}$ decrease. Regime $\mathbb{O}$ consistently outperforms the other regimes across all experiments in the continuous uniform setting.

The primary insight gained from these experiments is that the platform generally prefers to share occupancy information as the heterogeneity among users within a class (i.e., the number of distinct user types) increases.[4] This finding aligns with the conclusions drawn by Dobson and Pinker (2006) and Guo and Zipkin (2007) in one-sided settings. Dobson and Pinker (2006) find that a firm benefits from sharing more detailed lead time information when customers' tolerances for waiting vary significantly. Similarly, Guo and Zipkin (2007) conclude that sharing occupancy information (referred to as "partial information" in their study) outperforms the absence of information sharing when "the cost-scale distribution is spread out, so customers are heterogeneous."

8. Concluding Remarks

We study on-demand platforms' delay information disclosure policies when the platform matches two user classes to maximize the match rate. We find it optimal for the platform to withhold delay information from one user class (Regime $\mathbb{P}$ or $\mathbb{C}$) under certain settings. Specifically, concealing information from a user class is considered ideal when its patient users exhibit high levels of delay insensitivity, have a higher arrival rate compared to their patient counterparts in the other class, and constitute a significant proportion of their respective user class (Proposition 1). The prevalence of an asymmetric regime's optimality increases as the discrepancy in the patience profiles (Proposition 2) or the market sizes (Proposition 3) of the two user classes increases. These findings do not necessarily align with the one-sided literature findings documenting that it is optimal to hide delay information in low-congestion systems (see Ibrahim 2018b, and the references therein).

[4] Note that this is not true for every individual parameter setting. See Aydemir (2021), p. 51, for specific parameter settings for which Regime $\mathbb{O}$ is optimal in the two-point setting, whereas Regime $\mathbb{P}$ or $\mathbb{C}$ is optimal in the hybrid setting.

In two-sided systems, hiding delay information from the relatively *more* congested user class could be the best strategy (Appendix E).

Although Regimes $\mathbb{P}$ or $\mathbb{C}$ may be optimal, Regime $\mathbb{O}$ is optimal in most settings (66.72% of the experiments). Our extensive numerical analysis shows that even when Regime $\mathbb{P}$ or $\mathbb{C}$ is optimal, Regime $\mathbb{O}$ typically results in an average match rate loss of only 5.08%, suggesting that the presence of a second side in the market erodes the possible advantage of hiding delay information. Furthermore, alignment between the users' and the platform's preferred regimes only occurs under Regime $\mathbb{O}$. Therefore, Regime $\mathbb{O}$ is also less likely to invite opposition from the platform's users. Given all these considerations, Regime $\mathbb{O}$, under which the platform discloses delay information to both user classes, is almost always a safe choice.

We consider several extensions. When users are matched according to the random service discipline instead of FCFS, our main results remain unchanged (§7.1). Similarly, when we allow for users' abandonment, Regime $\mathbb{O}$ continues to be the dominant regime, albeit to a smaller extent (§7.2). Finally, as users of the same class become more heterogeneous (having more granular patience sensitivity levels), the platform finds it optimal to disclose occupancy information to both its user classes for a broader range of parameters (§7.3).

Our work opens up several avenues for future research in this domain. As we find in §7.3, the choice of the distribution governing users' delay sensitivities has a significant impact on our results. Although we consider more granular distributions in §7.3, a more in-depth analysis is required to find a general link between the users' delay sensitivity distribution and the optimal regime—a problem that is only solved partially even for one-sided queuing systems (Guo and Zipkin 2007). Another avenue to expand our models is to consider a more general class of information regimes, for example, by considering different structures for the announced delay information depending on the system state. Furthermore, studying non-linear delay cost functions could be a useful but analytically challenging extension. This would supplement the work of Guo and Zipkin (2007), who study the effect of customer delay cost functions with general forms on information disclosure in a one-sided system. A more drastic extension to the models studied in our paper is considering multiple queues on each side of the platform (with each queue, for example, representing a region or a skill) with a network structure specifying feasible matches. It would be interesting to study different, possibly state-dependent delay information regimes in this setting.

All the URLs below were last accessed on August 3, 2023.

References

Afeche P, Liu Z, Maglaras C (2018) Ride-hailing networks with strategic drivers: The impact of platform control capabilities on performance. *Rotman School of Management Working Paper* (3120544):18–19.

Afeche P, Liu Z, Maglaras C (2020) Surge pricing and dynamic matching for hotspot demand shock in ride-hailing networks. Technical report, Working paper.

Afche P, Akan M (2016) Pricing in a two-sided market with time-sensitive customers and suppliers. *INFORMS Annual Meeting, Nashville, TN* .

Afche P, Akan M (2023) Pricing in a two-sided market with time-sensitive customers and suppliers. *Working Paper* .

Akşin Z, Ata B, Emadi SM, Su CL (2013) Structural estimation of callers' delay sensitivity in call centers. *Management Science* 59(12):2727–2746.

Alwan AA, Ata B, Zhou Y (2023) A queueing model of dynamic pricing and dispatch control for ride-hailing systems incorporating travel times. *arXiv preprint arXiv:2302.02265* .

Armony M, Shimkin N, Whitt W (2009) The impact of delay announcements in many-server queues with abandonment. *Operations Research* 57(1):66–81.

Ata B, Barjesteh N, Kumar S (2020) Dynamic dispatch and centralized relocation of cars in ride-hailing platforms. *Available at SSRN 3675888* .

Aydemir M (2021) *Time Sensitive Operations*. Ph.D. thesis, Carnegie Mellon University.

Batt RJ, Terwiesch C (2015) Waiting patiently: An empirical study of queue abandonment in an emergency department. *Management Science* 61(1):39–59.

Braverman A, Dai JG, Liu X, Ying L (2019) Empty-car routing in ridesharing systems. *Operations Research* 67(5):1437–1452.

Cachon GP, Daniels KM, Lobel R (2017) The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing & Service Operations Management* 19(3):368–384.

Chen H, Frank M (2004) Monopoly pricing when customers queue. *IIE Transactions* 36(6):569–581.

Chu LY, Wan Z, Zhan D (2018) Harnessing the double-edged sword via routing: Information provision on ride-hailing platforms. *Available at SSRN 3266250* .

Diamant A, Baron O (2019) Double-sided matching queues: Priority and impatient customers. *Operations Research Letters* 47(3):219–224.

Dickerson JP, Sankararaman KA, Srinivasan A, Xu P (2018) Allocation problems in ride-sharing platforms: Online matching with offline reusable resources. *Thirty-Second AAAI Conference on Artificial Intelligence*.

Dimitrakopoulos Y, Economou A, Leonardos S (2021) Strategic customer behavior in a queueing system with alternating information structure. *European Journal of Operational Research* 291(3):1024–1040.

Dobson G, Pinker EJ (2006) The value of sharing lead time information. *IIE Transactions* 38(3):171–183.

Edelson NM, Hilderbrand DK (1975) Congestion tolls for poisson queuing processes. *Econometrica: Journal of the Econometric Society* 81–92.

Guo P, Zipkin P (2007) Analysis and comparison of queues with different levels of delay information. *Management Science* 53(6):962–970.

Handy (2023) What is peak pricing? https://help.handy.com/hc/en-us/articles/219851127-What-is-Peak-Pricing.

Hassin R (1986) Consumer information in markets with random product quality: The case of queues and balking. *Econometrica: Journal of the Econometric Society* 1185–1195.

Hassin R (2016) *Rational queueing* (Chapman and Hall/CRC).

Hassin R, Haviv M (2003) *To Queue or not to Queue: Equilibrium Behavior in Queueing Systems* (USA: Kluwer Academic Publishers).

Hosseini M, Milner J, Romero G (2021) Dynamic relocations in car-sharing networks. *Rotman School of Management Working Paper* (3774324).

Hu M, Li Y, Wang J (2018) Efficient ignorance: Information heterogeneity in a queue. *Management Science* 64(6):2650–2671.

Hu M, Zhou Y (2019) Price, wage and fixed commission in on-demand matching. *Available at SSRN 2949513* .

Ibrahim R (2018a) Managing queueing systems where capacity is random and customers are impatient. *Production and Operations Management* 27(2):234–250.

Ibrahim R (2018b) Sharing delay information in service systems: A literature survey. *Queueing Systems* 89(1-2):49–79.

Ibrahim R, Armony M, Bassamboo A (2017) Does the past predict the future? The case of delay announcements in service systems. *Management Science* 63(6):1762–1780.

Ibrahim R, Whitt W (2009) Real-time delay estimation based on delay history. *Manufacturing & Service Operations Management* 11(3):397–415.

Jouini O, Akşin Z, Dallery Y (2011) Call centers with delay information: Models and insights. *Manufacturing & Service Operations Management* 13(4):534–548.

Kuzu K (2015) Comparisons of perceptions and behavior in ticket queues and physical queues. *Service Science* 7(4):294–314.

Lingenbrink D, Iyer K (2019) Optimal signaling mechanisms in unobservable queues. *Operations Research* 67(5):1397–1416.

Liu Y, QI J, Yang H, Tang Y (2022) The passenger's willingness to wait with sunk waiting time: An empirical study in ride-sourcing market. *Available at SSRN 4285203* .

Lyft (2019) Explore the Lyft driver app. https://www.lyft.com/hub/posts/explore-the-lyft-driver-app.

Lyft (2023) Lyft. https://www.lyft.com/.

Lyu G, Cheung WC, Teo CP, Wang H (2019) Multi-objective online ride-matching. *Available at SSRN 3356823* .

Naor P (1969) The regulation of queue size by levying tolls. *Econometrica: journal of the Econometric Society* 15–24.

Özkan E, Ward AR (2020) Dynamic matching for real-time ride sharing. *Stochastic Systems* 10(1):29–70.

Paul K (2015) The new system Uber is implementing at airports has some drivers worried. https://www.vice.com/en/article/gvy357/the-new-system-uber-is-implementing-at-airports-has-some-drivers-worried.

Ridester (2023) How Instacart peak boost can help shoppers earn more money. https://www.ridester.com/instacart-peak-boost/.

Romanyuk G (2017) Ignorance is strength: Improving the performance of matching markets by limiting information. *Harvard Univeristy Cambridge, MA Working Paper.*

Romanyuk G, Smolin A (2019) Cream skimming and information design in matching markets. *American Economic Journal: Microeconomics* 11(2):250–76.

Rosenblat A, Stark L (2016) Algorithmic labor and information asymmetries: A case study of Uber's drivers. *International Journal of Communication* 10:3758–3784.

Singh SP, Delasay M, Scheller-Wolf A (2023) Real-time delay announcement under competition. *Production and Operations Management* 32(3):863–881.

Smith DR, Whitt W (1981) Resource sharing for efficiency in traffic systems. *Bell System Technical Journal* 60(1):39–55.

Taylor TA (2018) On-demand service platforms. *Manufacturing & Service Operations Management* 20(4):704–720.

UberPeoplenet (2017) Passenger app no longer shows cars in area. https://uberpeople.net/threads/passenger-app-no-longer-shows-cars-in-area.136892/.

Wang H, Yang H (2019) Ridesourcing systems: A framework and review. *Transportation Research Part B: Methodological* 129:122–155.

Yu Q, Allon G, Bassamboo A (2017) How do delay announcements shape customer behavior? An empirical study. *Management Science* 63(1):1–20.

Zhu D, Minner S, Bichler M (2023) Information design for on-demand service platforms: A queueing-theoretic approach. *Working Paper* .

Appendix

A. Proof of Lemma 1

Providers join at the instantaneous rate $\delta^p \Lambda^p$ at states $n \in \{0, 1, \dots, b^p_{\mathbb{O}} - 1\}$ (as impatient providers do not join due to the positive expected delay) and at rate Λ^p at states $n \leq -1$ (since their expected delay is zero). Analogously, customers join at the instantaneous rate $\delta^c \Lambda^c$ at states $n \in \{-b^c_{\mathbb{O}} + 1, \dots, -1, 0\}$ and at rate Λ^c at states $n \geq 1$.

Eq. (6) gives the general expression for the match rate under all regimes. Under Regime $\mathbb{O}$, we have a truncated state space, i.e., some terms in Eq. (6) are zero. We can express $M_{\mathbb{O}}$ under Regime $\mathbb{O}$ as follows:

$$M_{\mathbb{O}} = \sum_{n=-b^c_{\mathbb{O}}}^{-1} \pi_{\mathbb{O}}(n) \Lambda^p + \sum_{n=0}^{b^p_{\mathbb{O}}-1} \pi_{\mathbb{O}}(n) \delta^p \Lambda^p. \tag{23}$$

We can find the steady-state probabilities $\pi_{\mathbb{O}}(n)$ by solving the following set of balance equations and the normalization equation (check the CTMC in Fig. 1):

$$\begin{cases} \pi_{\mathbb{O}}(n) = \pi_{\mathbb{O}}(0) \left(\rho^p\right)^n, & \forall n \in \{0, 1, \dots, b^p_{\mathbb{O}}\}, \\ \pi_{\mathbb{O}}(n) = \pi_{\mathbb{O}}(0) \left(\rho^c\right)^{|n|}, & \forall n \in \{-b^c_{\mathbb{O}}, \dots, -1, 0\}, \\ \sum_{n=-b^c_{\mathbb{O}}}^{b^p_{\mathbb{O}}} \pi_{\mathbb{O}}(n) = 1, \end{cases} \tag{24}$$

which results in the probability $\pi_{\mathbb{O}}(0)$ that the system is empty as follows:

$$\pi_{\mathbb{O}}(0) = \left(\frac{1 - \left(\rho^p\right)^{b^p_{\mathbb{O}}+1}}{1 - \rho^p} + \frac{\rho^c - \left(\rho^c\right)^{b^c_{\mathbb{O}}+1}}{1 - \rho^c} \right)^{-1}. \tag{25}$$

Substituting Eq. (25) in Eq. (24) results in the expressions for the steady state probabilities under Regime $\mathbb{O}$. Substituting the resulting steady-state probabilities in Eq. (23) results in Regime $\mathbb{O}$'s match rate as provided in Lemma 1 (Eq. (10)).

B. Proof and expressions for Lemma 2

B.1. Expressions for Lemma 2

Regime $\mathbb{P}$ conditions.

$$(1)\text{Eq. } (36) \wedge (2)\text{Eq. } (37) \wedge (3)\text{Eq. } (38). \tag{26}$$

$$(1)\text{Eq. } (36) \wedge (2)\text{Eq. } (39) \wedge (3)\text{Eq. } (40). \tag{27}$$

$$(1)\text{Eq. } (41) \wedge \Big((2)\text{Eq. } (45) \vee (3)\text{Eq. } (43) \Big) \wedge (4)\text{Eq. } (44). \tag{28}$$

$$\Big((1)\text{Eq. } (45) \vee (2)\text{Eq. } (46) \Big) \wedge (3)\text{Eq. } (50). \tag{29}$$

Regime $\mathbb{C}$ conditions.

$$(1)\text{Eq. }(52) \wedge (2)\text{Eq. }(53) \wedge (3)\text{Eq. }(54). \tag{30}$$

$$(1)\text{Eq. }(52) \wedge (2)\text{Eq. }(55) \wedge (3)\text{Eq. }(56). \tag{31}$$

$$(1)\text{Eq. }(57) \wedge \Big((2)\text{Eq. }(58) \vee (3)\text{Eq. }(59)\Big) \wedge (4)\text{Eq. }(60). \tag{32}$$

$$\Big((1)\text{Eq. }(58) \vee (2)\text{Eq. }(61)\Big) \wedge (3)\text{Eq. }(62). \tag{33}$$

B.2. Proof of Lemma 2

We present the proof for the optimality of Regime $\mathbb{P}$; the proof for Regime $\mathbb{C}$ follows analogously. Under Regime $\mathbb{P}$, customers do not receive any delay information. We begin by describing the two possible joining strategies that patient customers can employ that result in a non-zero match rate:

- All arriving patient customers join with probability one, resulting in the instantaneous arrival rate $\delta^c\Lambda^c$. This occurs when their utility is positive even if they all join, i.e.,

$$U^c_{\mathbb{P}}(\emptyset) = R^c\left(1 - t^c\left(1 + \mathbb{E}[W^c_{\mathbb{P}}(\emptyset)|J_{\mathbb{P}}(t^c|\emptyset) = 1]\right)\right) > 0. \tag{34}$$

- All arriving patient customers join with probability $j \in (0,1)$, resulting in the instantaneous arrival rate $j\delta^c\Lambda^c$, where j is chosen such that they are indifferent between joining and not joining, i.e.,

$$U^c_{\mathbb{P}}(\emptyset) = R^c\left(1 - t^c\left(1 + \mathbb{E}[W^c_{\mathbb{P}}(\emptyset)|J_{\mathbb{P}}(t^c|\emptyset) = j]\right)\right) = 0. \tag{35}$$

The above strategies and the providers' joining behavior result in four cases with non-zero match rates:

(a1) $J_{\mathbb{P}}(t^c|\emptyset) = 1$ and $b^p_{\mathbb{P}} > 0$,

(a2) $J_{\mathbb{P}}(t^c|\emptyset) = 1$ and $b^p_{\mathbb{P}} = 0$,

(b) $J_{\mathbb{P}}(t^c|\emptyset) = j \in (0,1)$ and $b^p_{\mathbb{P}} = 0$,

(c) $J_{\mathbb{P}}(t^c|\emptyset) = j \in (0,1)$ and $b^p_{\mathbb{P}} > 0$.

In presenting Lemma 2, we merge Cases (a1) and (a2) into Case (a) as they result in the same match rate.

Case (a1) ($J_{\mathbb{P}}(t^c|\emptyset) = 1$ and $b^p_{\mathbb{P}} > 0$). In this case, customers' instantaneous arrival rate $\lambda^c_{\mathbb{P}}(\emptyset) = \delta^c\Lambda^c$ (as $J_{\mathbb{P}}(t^c|\emptyset) = 1$). Fig. 10 represents the CTMC under Regime $\mathbb{P}$ where following Eq. (12), the bounding state can be presented as $b^p_{\mathbb{P}} = \lfloor \delta^c\Lambda^c\omega^p \rfloor$.

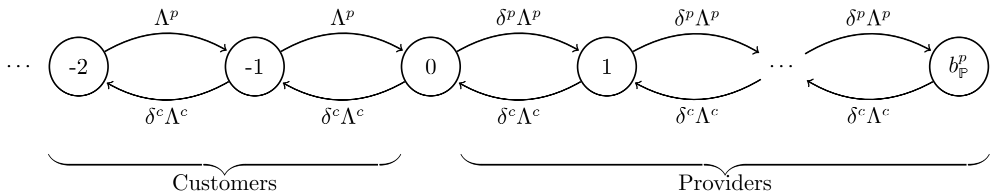


Figure 10 **Regime $\mathbb{P}$ CTMC when $\lambda^c_{\mathbb{P}}(\emptyset) = \delta^c\Lambda^c$ and $b^p_{\mathbb{P}} > 0$ (Case (a1))**

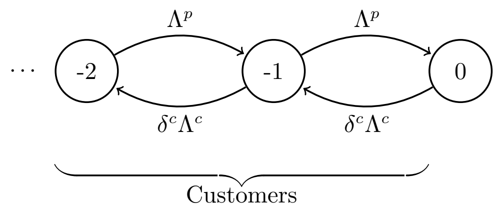


Figure 11 **Regime $\mathbb{P}$ CTMC when $\lambda^c_{\mathbb{P}}(\emptyset) = \delta^c\Lambda^c$ and $b^p_{\mathbb{P}} = 0$ (Case (a2))**

As all patient customers join in this case, we first need to ensure that the resulting system is stable, i.e., we need to have the following:

$$\rho^c = \frac{\delta^c\Lambda^c}{\Lambda^p} < 1. \tag{36}$$

Next, we find the stationary probabilities $\pi_{\mathbb{P}}(i)$ for the CTMC in Fig. 10, which we use to derive $\mathbb{E}[W^c_{\mathbb{P}}(\emptyset)|J_{\mathbb{P}}(t^c|\emptyset) = 1]$ using Eq. (13). We then substitute $\mathbb{E}[W^c_{\mathbb{P}}(\emptyset)|J_{\mathbb{P}}(t^c|\emptyset) = 1]$ into Eq. (34) and obtain the following condition on t^c:

$$t^c \leq \left(1 + \frac{\delta^c\Lambda^c - \delta^p\Lambda^p}{(\Lambda^p - \delta^c\Lambda^c)\left(\delta^c\Lambda^c(1-\delta^p) - \delta^p(\Lambda^p - \delta^c\Lambda^c)\left(\frac{\delta^p\Lambda^p}{\delta^c\Lambda^c}\right)^{b^p_{\mathbb{P}}}\right)}\right)^{-1}. \tag{37}$$

We then substitute $\lambda^c_{\mathbb{P}}(\emptyset) = \delta^c\Lambda^c$ in Eq. (11), and re-arrange it to obtain the required condition on t^p (corresponding to third condition in (26)):

$$\frac{\delta^c\Lambda^c\,(1-t^p)}{K^p t^p} - 1 \geq 0 \implies t^p \leq \frac{\delta^c\Lambda^c}{K^p + \delta^c\Lambda^c}. \tag{38}$$

The resulting match rate in Case (a1) is equal to $\lambda^c_{\mathbb{P}}(\emptyset) = \delta^c\Lambda^c$.

Case (a2) ($J_{\mathbb{P}}(t^c|\emptyset) = 1$ and $b^p_{\mathbb{P}} = 0$). In this case, customers' instantaneous arrival rate $\lambda^c_{\mathbb{P}}(\emptyset) = \delta^c\Lambda^c$, and providers' instantaneous arrival rate $\lambda^p_{\mathbb{P}}(s^p_{\mathbb{P}}(n) = n) = \Lambda^p$, $\forall n < 0$.

As the bounding state $b^p_{\mathbb{P}} = 0$ in this case, the system resembles an $M/M/1$ queue (see Fig. 11) with the respective arrival and service rates $\delta^c\Lambda^c$ (as $J_{\mathbb{P}}(t^c|\emptyset) = 1$) and Λ^p. Therefore, to ensure system stability, similar to Case (a1), we need to have $\rho^c = \dfrac{\delta^c\lambda^c}{\Lambda^p} < 1$.

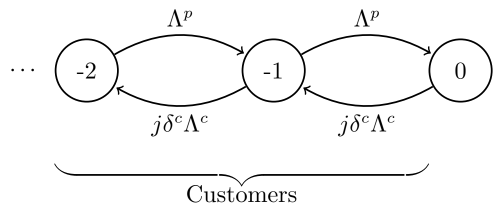


Figure 12 **Regime $\mathbb{P}$ CTMC when $J_{\mathbb{P}}(t^c|\emptyset)=j\in(0,1)$ and $b^p_{\mathbb{P}}=0$ (Case (b))**

Following the $M/M/1$ expected delay expressions, the customers' expected delay in this case follows $\mathbb{E}[W^c_{\mathbb{P}}(\emptyset)|J_{\mathbb{P}}(t^c|\emptyset)=1]=\dfrac{1}{\Lambda^p-\delta^c\Lambda^c}$, which we use in Eq. (34) to obtain the following condition that ensures customers join at rate $\delta^c\Lambda^c$ in equilibrium:

$$t^c \le \frac{\Lambda^p-\delta^c\Lambda^c}{1+\Lambda^p-\delta^c\Lambda^c}. \tag{39}$$

We then re-arrange Eq. (11) to obtain the condition that ensures the providers' bounding state $b^p_{\mathbb{P}}=0$ (corresponding to the third condition in Eq. (27)):

$$\frac{\delta^c\Lambda^c\,(1-t^p)}{K^p t^p}-1<0 \implies t^p>\frac{\delta^c\Lambda^c}{K^p+\delta^c\Lambda^c}. \tag{40}$$

The match rate in Case (a2) (like Case (a1)) is $\lambda^c_{\mathbb{P}}(\emptyset)=\delta^c\Lambda^c$.

Case (b) ($J_{\mathbb{P}}(t^c|\emptyset)=j\in(0,1)$ and $b^p_{\mathbb{P}}=0$). In this case, customers' expected delay $\mathbb{E}[W^c_{\mathbb{P}}(\emptyset)|J_{\mathbb{P}}(t^c|\emptyset)=j]$ depends only on their own instantaneous arrival rate $\lambda^c_{\mathbb{P}}(\emptyset)=j\delta^c\Lambda^c$. To find the customers' equilibrium joining strategy, we must find $J_{\mathbb{P}}(t^c|\emptyset)$ that solves Eq. (35).

Since $b^p_{\mathbb{P}}=0$, the system simplifies to an $M/M/1$ queue with respective arrival and service rates $\lambda^c_{\mathbb{P}}(\emptyset)$ and Λ^p. The CTMC is shown in Fig. 12. Hence, the expected delay in terms of the customers' instantaneous arrival rate $j\delta^c\Lambda^c$ is $\mathbb{E}[W^c_{\mathbb{P}}(\emptyset)|J_{\mathbb{P}}(t^c|\emptyset)=j]=\dfrac{1}{\Lambda^p-j\delta^c\Lambda^c}$. Then, using Eq. (35) we solve for the customers' equilibrium joining probability j and obtain the customers' arrival rate as $\lambda^c_{\mathbb{P}}(\emptyset)=\Lambda^p-\dfrac{t^c}{1-t^c}$. For Case (b) to be valid, providers must be unwilling to wait in the system. Using the derived customers' arrival rate, we derive this condition using Eq. (11) to obtain the following:

$$\frac{\left(\Lambda^p-\dfrac{t^c}{1-t^c}\right)(1-t^p)}{K^p t^p}<1 \implies t^p>\frac{\Lambda^p\omega^c-1}{K^p\omega^c+\Lambda^p\omega^c-1}. \tag{41}$$

Next, we must ensure that customers' joining probability is between 0 and 1, i.e.,

$$0<\lambda^c_{\mathbb{P}}(\emptyset)=\Lambda^p-\frac{t^c}{1-t^c}<\delta^c\Lambda^c. \tag{42}$$

We distinguish between two sub-cases:

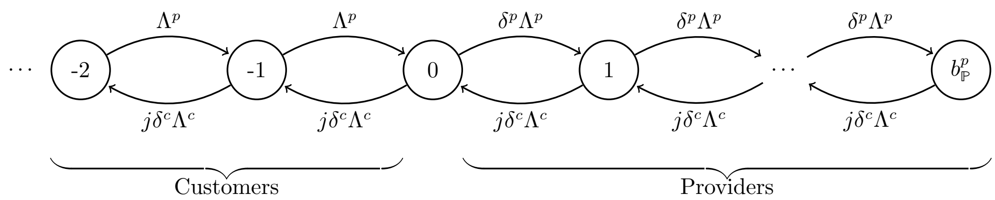


Figure 13 **Regime $\mathbb{P}$ CTMC when $J_{\mathbb{P}}(t^c|\emptyset) = j \in (0,1)$ and $b_{\mathbb{P}}^p > 0$ (Case (c))**

Case (b)(i): $\rho^c < 1$. In this case, we need to ensure that both sides of Eq. (42) hold, which results in the following two conditions:

$$\frac{\Lambda^p - \delta^c\Lambda^c}{1 + \Lambda^p - \delta^c\Lambda^c} < t^c, \tag{43}$$

$$t^c < \frac{\Lambda^p}{1 + \Lambda^p}. \tag{44}$$

Case (b)(ii): $\rho^c \geq 1$. In this case, we have:

$$\rho^c = \frac{\delta^c\Lambda^c}{\Lambda^p} \geq 1, \tag{45}$$

under which the condition $\lambda_{\mathbb{P}}^c(\emptyset) < \delta^c\Lambda^c$ in Eq. (42) holds trivially. Therefore, we only need to ensure that $\lambda_{\mathbb{P}}^c(\emptyset) > 0$, which simplifies to Eq. (44).

The resulting match rate in Case (b) is $\lambda_{\mathbb{P}}^c(\emptyset) = \Lambda^p - \dfrac{t^c}{1 - t^c}$.

Case (c) ($J_{\mathbb{P}}(t^c|\emptyset) = j \in (0,1)$ and $b_{\mathbb{P}}^p > 0$). In this case, customers' instantaneous arrival rate $\lambda_{\mathbb{P}}^c(\emptyset) = j\delta^c\Lambda^c, \forall n$, and provider's instantaneous arrival rate is $\lambda_{\mathbb{P}}^p(s_{\mathbb{P}}^p(n) = n) = \delta^p\Lambda^p, 0 \leq n < b_{\mathbb{P}}^p$, and $\lambda_{\mathbb{P}}^p(s_{\mathbb{P}}^p(n) = n) = \Lambda^p, \forall n < 0$. The CTMC is shown in Fig. 13.

Since the joining decisions of customers and providers are intertwined, we derive conditions for Case (c) by considering the behavior of customers and providers together. For Case (c) to occur, customers must join the system with a probability of less than one. This occurs when (i) customers joining with probability one results in an unstable system (i.e., $\rho^c \geq 1$) or (ii) customers' patience level is too small for them to join the system with probability one, which happens when t^c is larger than the upper bound on t^c presented for Case (a1). The former case corresponds to the first condition in Eq. (29). In the latter case, the upper bound on t^c presented for Case (a1) forms a lower bound on t^c, corresponding to the second condition in Eq. (29):

$$t^c > \left(1 + \frac{\delta^c\Lambda^c - \delta^p\Lambda^p}{(\Lambda^p - \delta^c\Lambda^c)\left(\delta^c\Lambda^c(1 - \delta^p) - \delta^p(\Lambda^p - \delta^c\Lambda^c)\left(\delta^p\Lambda^p/\delta^c\Lambda^c\right)^{b_{\mathbb{P}}^p}\right)}\right)^{-1}. \tag{46}$$

As customers become more impatient (as t^c increases), customers join with a lower rate $\lambda_{\mathbb{P}}^c(\emptyset)$, leading to a smaller providers' bounding state $b_{\mathbb{P}}^p$. Therefore, there must be an upper bound on t^c corresponding to the lowest induced value of $b^p_{\mathbb{P}}$ in Case (c), i.e., $b^p_{\mathbb{P}}=1$. When $b^p_{\mathbb{P}}=1$, we can derive in closed form the expected delay $\mathbb{E}[W^c_1(\emptyset)|\lambda^c_{\mathbb{P}}(\emptyset)]$ and the eqilibrium arrival rate $\lambda^c_{\mathbb{P}}(\emptyset)$ that are consistent with each other by solving Eq. (47) for the arrival rate l:

$$1-t^c\left(1+\mathbb{E}[W^c_1(\emptyset)|\lambda^c_{\mathbb{P}}(\emptyset)=l]\right)=0, \tag{47}$$

where $\mathbb{E}[W^c_1(\emptyset)|\lambda^c_{\mathbb{P}}(\emptyset)=l]=\dfrac{l}{(\Lambda^p-l)(l(1-\delta^p)+\Lambda^p\delta^p)}$, which results in:

$$l=\frac{\sqrt{\left(\Lambda^p\omega^c\right)^2-2\Lambda^p\omega^c(1-2\delta^p)+1}+\Lambda^p\omega^c(1-2\delta^p)-1}{2\omega^c(1-\delta^p)}. \tag{48}$$

Then, we ensure that providers are patient enough to join at state 0 when customers join at rate l. (If they are, the resulting l is strictly positive.) From Eq. (11), we obtain this condition as:

$$t^p\leq\frac{l}{K^p+l}, \tag{49}$$

where l is given in Eq. (48). Rearranging the terms in Eq. (49), we obtain the required upper bound on t^c:

$$t^c\leq\frac{\left(\Lambda^p\omega^p-1\right)\left(\delta^p\Lambda^p\omega^p+1-\delta^p\right)}{\delta^p\left(\Lambda^p\omega^p\right)^2+\omega^p\left(\Lambda^p\left(2\delta^p-1\right)-1\right)+\left(1-\delta^p\right)}. \tag{50}$$

Although we can obtain the conditions under which Case (c) holds in closed form, we cannot find the match rate in closed form. To do so, we must find the value of $J_{\mathbb{P}}(t^c|\emptyset)$ that solves Eq. (35). This equation cannot, in general, be solved in closed-form because the expression for the expected delay $\mathbb{E}[W^c_{\mathbb{P}}(\emptyset)|J_{\mathbb{P}}(t^c|\emptyset)]$ contains the providers' bounding state $b^p_{\mathbb{P}}$ in the exponent, in which the unknown $J_{\mathbb{P}}(t^c|\emptyset)$ appears inside a floor function as well as outside the exponent. We present the relevant equation in terms of the primitive parameters below:

$$\frac{\Lambda^p-M_{\mathbb{P}}}{\delta^p\Lambda^p-M_{\mathbb{P}}}\left(M_{\mathbb{P}}\left(1-\delta^p\right)-\delta^p\left(\Lambda^p-M_{\mathbb{P}}\right)\left(\frac{\delta^p\Lambda^p}{M_{\mathbb{P}}}\right)^{\lfloor M_{\mathbb{P}}\omega^p\rfloor}\right)=\frac{1}{\omega^c}. \tag{51}$$

In Cases (a1)-(c), we presented conditions under which users are willing to join with positive probability. If none of these cases applies, we conclude that users are not willing to join the system, and hence the system is empty, leading to a match rate of 0.

Expressions for Regime $\mathbb{C}$: Following an analogous process to that for Regime $\mathbb{P}$, we derive the following expressions for Regime $\mathbb{C}$ for use in Lemma 2:

$$\rho^p=\frac{\delta^p\Lambda^p}{\Lambda^c}<1, \tag{52}$$

$$t^p \le \left(1 + \frac{K^p(\delta^p\Lambda^p - \delta^c\Lambda^c)}{(\Lambda^c - \delta^p\Lambda^p)\left(\delta^p\Lambda^p(1-\delta^c) - \delta^c(\Lambda^c - \delta^p\Lambda^p)\left(\delta^c\Lambda^c/\delta^p\Lambda^p\right)^{b^c_{\mathbb{C}}}\right)}\right)^{-1}, \tag{53}$$

$$t^c \le \frac{\delta^p\Lambda^p}{1+\delta^p\Lambda^p}, \tag{54}$$

$$t^p \le \frac{\Lambda^c - \delta^p\Lambda^p}{K^p + \Lambda^c - \delta^p\Lambda^p}, \tag{55}$$

$$t^p > \frac{\delta^p\Lambda^p}{1+\delta^p\Lambda^p}. \tag{56}$$

$$t^c > \frac{\Lambda^c\omega^p - 1}{\omega^p + \Lambda^c\omega^p - 1}, \tag{57}$$

$$\rho^p = \frac{\delta^p\Lambda^p}{\Lambda^c} \ge 1, \tag{58}$$

$$t^p > \frac{\Lambda^c - \delta^p\Lambda^p}{K^p + \Lambda^c - \delta^p\Lambda^p}, \tag{59}$$

$$t^p < \frac{\Lambda^c}{K^p + \Lambda^c}, \tag{60}$$

$$t^p > \left(1 + \frac{K^p(\delta^p\Lambda^p - \delta^c\Lambda^c)}{(\Lambda^c - \delta^p\Lambda^p)\left(\delta^p\Lambda^p(1-\delta^c) - \delta^c(\Lambda^c - \delta^p\Lambda^p)\left(\delta^c\Lambda^c/\delta^p\Lambda^p\right)^{b^c_{\mathbb{C}}}\right)}\right)^{-1}, \tag{61}$$

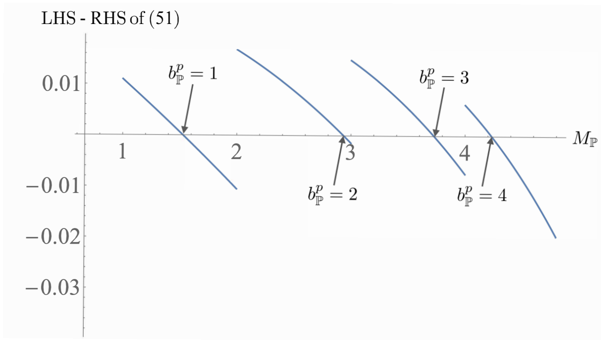
$$t^p \le \left(\frac{1}{K^p}\right)^2 \frac{(\Lambda^c\omega^c - 1)((1-\delta^c) + \delta^c\Lambda^c\omega^c)}{((1-\delta^c) + \omega^c(\Lambda^c(2\delta^c - 1) - K^p) + \delta^c(\Lambda^c\omega^c)^2}. \tag{62}$$

The equation for the match rate under Case (c) is:

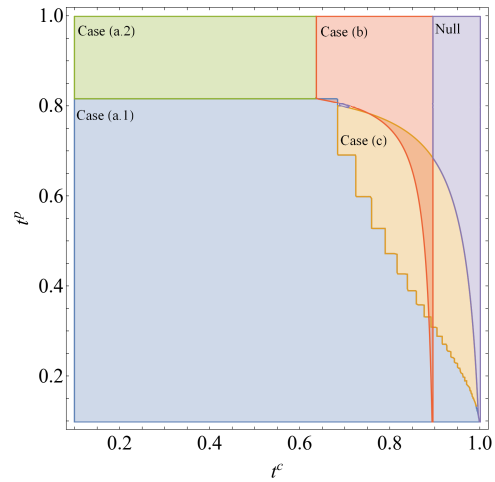
$$\frac{\Lambda^c - M_{\mathbb{C}}}{\delta^c\Lambda^c - M_{\mathbb{C}}}\left(M_{\mathbb{C}}\left(1-\delta^c\right) - \delta^c\left(\Lambda^c - M_{\mathbb{C}}\right)\left(\frac{\delta^c\Lambda^c}{M_{\mathbb{C}}}\right)^{\lfloor M_{\mathbb{C}}\omega^c\rfloor}\right) = \frac{1}{\omega^p}. \tag{63}$$

C. Existence of Multiple Equilibria

For brevity, we illustrate the case of multiple equilibria for Regime $\mathbb{P}$. There are two instances of multiple equilibria in Lemma 2. First, the match rate associated with Case (c) in Lemma 2, which is the solution to Eq. (51), is not necessarily unique; i.e., Case (c) may yield multiple equilibria (multiple combinations of $b^p_{\mathbb{P}}$ and $\lambda^c_{\mathbb{P}}(\emptyset)$ that are consistent with each other). To illustrate, we plot the left-hand side minus the right-hand side of Eq. (51) against $M_{\mathbb{P}}$ for a particular parameter setting in Fig. 14a, in which $M_{\mathbb{P}}$ crosses zero four times at $M_{\mathbb{P}} \in \{1.53, 2.92, 3.75, 4.23\}$ corresponding to providers' bounding states $b^p_{\mathbb{P}} \in \{1, 2, 3, 4\}$. Second, the cases of Lemma 2 are not mutually exclusive, i.e., there are parameter settings under which more than one case holds, which again leads to multiple equilibria. We illustrate this in Fig. 14b, highlighting the regions corresponding to the different cases as a function of t^c and t^p. For ease of exposition, we use the cases as introduced in the proof of Lemma 2 (we consider two separate sub-cases for case (a)). In Proposition C.1, we list the possible overlaps between the cases of Lemma 2 and identify the case that results in a higher match rate.



(a) $\{\Lambda^c, \Lambda^p, K^p, t^c, t^p, \delta^c, \delta^p\} = \{8, 9, 1.5, 0.97, 0.4, 0.6, 0.68\}$



(b) $\{\Lambda^c, \Lambda^p, K^p, \delta^c, \delta^p\} = \{9, 8.5, 1.5, 0.75, 0.9\}$

Figure 14 Cases of Multiple Equilibria in Regime $\mathbb{P}$

PROPOSITION C.1. *Based on the cases (a1), (a2), (b), and (c) used in the proof of Lemma 2, there are only two possible overlaps between the cases of Lemma 2: (i) The conditions for cases (a1) and (b) could hold together; if so, the match rate under Case (a1) is higher. (ii) The conditions for cases (b) and (c) could hold together; if so, the match rate under Case (c) is higher.*

Proof of Proposition C.1. We first show that the only two possible case overlaps are between Cases (a1) and (b) and between Cases (b) and (c). To do so, we first consider every possible pair of cases below and discuss whether they can overlap:

Cases (a1) and (c). Note that the conditions for t^c are negations of each other; hence there doesn't exist any t^c that satisfies both of the conditions.

Cases (a1) and (a2). Note that the conditions for t^p are negations of each other; hence there does not exist any t^p that satisfies both of the conditions.

Cases (a2) and (c). Note that the RHS of the condition for t^p under Case (c) is increasing in l, i.e., $\dfrac{l}{K^p + l}$ is increasing in l. Also note that l is bounded above by $\delta^c\Lambda^c$, i.e., $l \le \delta^c\Lambda^c$. Accordingly, the RHS of the t^p condition for Case (c) is smaller than the RHS of the t^p condition for Case (a2). Hence, there does not exist any t^p value that satisfies both conditions.

Cases (a2) and (b). Note that the upper bound for t^c for Case (a2) is the lower bound for t^c for Case (b). Hence there does not exist any t^c value that satisfies both of the conditions.

Therefore, the possible candidates for overlap are Cases (a1) and (b) and Cases (b) and (c). Both these overlaps are possible, as shown in Fig. 14, which uses parameters $\Lambda^c = 9, \Lambda^p = 8.5, K^p = 1.5\delta^c = 0.75$ and $\delta^p = 0.9$.

Now, we compare the effective match rate between the overlapping cases. First, note that under Case (b), patient customers join with probability $J_{\mathbb{P}}(t^c|\emptyset) = j \in (0,1)$ while under Case (a1) they join with probability $J_{\mathbb{P}}(t^c|\emptyset) = 1$. Therefore, it is immediate that the match rate under Case (a1) is higher than that of Case (b).

Next, we show that the match rate (equivalently, the customers' effective arrival rate) under Case (c) is higher than that under Case (b). In particular, consider a Case (c) equilibrium in which providers join up to state $B-1$ such that the bounding state is $B \geq 1$. We will show that:

(i) For a fixed arrival rate λ^c, the customers' utility is weakly increasing in bounding state B.

(ii) For a fixed bounding state, the customers' utility is weakly decreasing in the arrival rate λ^c.

Part (i) implies that fixing the arrival rate at the equilibrium arrival rate corresponding to Case (b) ($B = 0$) and increasing the bounding state to $B > 0$ leaves customers with excess utility, implying that customers' equilibrium joining rate is different from λ^c. Part (ii) then implies that to find the equilibrium arrival rate under Case (c), λ^c needs to increase to a higher value than the Case (b) equilibrium arrival rate.

Proof of Part (i). Showing Part (i) is equivalent to showing that for a fixed λ^c the customers' expected delay $\mathbb{E}[W^c_{\mathbb{P}}(\emptyset)|\lambda^c]$ decreases in the bounding state B. In order to do so, we take the first derivative of the delay under Case (c) with respect to bounding state B and show that it is non-positive. With some abuse of notation, we have:

$$\mathbb{E}[W^c_{\mathbb{P}}(\emptyset)|\lambda^c] = \frac{\Lambda^p\delta^p - \lambda^c}{(\lambda^c - \Lambda^p)\left(\delta^p(\lambda^c - \Lambda^p)\left(\frac{\Lambda^p\delta^p}{\lambda^c}\right)^B + \lambda^c - \lambda^c\delta^p\right)}, \tag{64}$$

$$\frac{\partial\,\mathbb{E}[W^c_{\mathbb{P}}(\emptyset)|\lambda^c]}{\partial B} = -\frac{\delta^p\left(\frac{\delta^p\Lambda^p}{\lambda^c}\right)^B \overbrace{(\Lambda^p\delta^p - \lambda^c)\log\left(\frac{\delta^p\Lambda^p}{\lambda^c}\right)}^{>0}}{\left(\delta^p(\lambda^c - \Lambda^p)\left(\frac{\Lambda^p\delta^p}{\lambda^c}\right)^B + \lambda^c - \lambda^c\delta^p\right)^2} \leq 0. \tag{65}$$

Proof of Part (ii). We equivalently show that for a fixed bounding state B, the customers' expected delay $\mathbb{E}[W^c_{\mathbb{P}}(\emptyset)|\lambda^c]$ increases in their arrival rate λ^c. Observe that the CTMC in Fig. 2 is a birth-death process. As a direct consequence of Theorem 5 in Smith and Whitt (1981) we have that:

$$\Pr(i \leq j) \text{ is increasing in } \lambda^c,\ \forall j, \tag{66}$$

where i and j are state indices corresponding to the CTMC in Fig. 2. (In particular, the result in Smith and Whitt (1981) implies a likelihood ratio ordering between the Markov chain with a lower λ^c and that with a higher λ^c; this, in turn, implies the first-order stochastic dominance in Eq. (66).) From Eq. (66), we have:

$$\sum_{j=-\infty}^{0} \Pr(i \leq j) \text{ is increasing in } \lambda^c \Leftrightarrow \sum_{j=-\infty}^{0} (i+1)\Pr(i=j) \text{ is increasing in } \lambda^c$$
$$\Leftrightarrow \sum_{j=-\infty}^{0} \frac{i+1}{\Lambda^p}\Pr(i=j) \text{ is increasing in } \lambda^c \Leftrightarrow \mathbb{E}[W^c_{\mathbb{P}}(\emptyset)|\lambda^c] \text{ is increasing in } \lambda^c.$$

This completes the proof.

D. Proof of Proposition 1

We provide the proof for the optimality of Regime $\mathbb{P}$. The proof for the optimality of Regime $\mathbb{C}$ follows a similar approach. Proposition 1 for the optimality of Regime $\mathbb{P}$ reads as: *Regime $\mathbb{P}$ is optimal when patient customers (a) always join under Regime $\mathbb{P}$ (i.e., conditions (26) or (27)), (b) arrive at a higher rate ($\delta^c\Lambda^c > \delta^p\Lambda^p$), and (c) their proportion is higher than a threshold such that:*

$$\delta^c > \left(1 + \frac{1-\rho^p}{\rho^p}\frac{(\rho^c)^{b^c_{\mathbb{O}}}}{1-(\rho^p)^{b^p_{\mathbb{O}}}}\right)^{-1}. \tag{67}$$

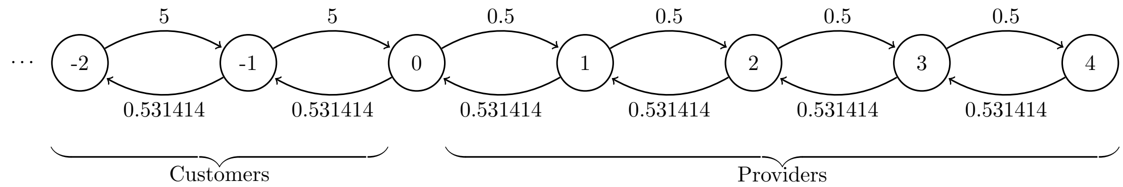
When patient customers always join under Regime $\mathbb{P}$ (condition (a) of the proposition), Regime $\mathbb{P}$'s match rate $M_{\mathbb{P}} = \delta^c\Lambda^c$ and Regime $\mathbb{C}$'s match rate $M_{\mathbb{C}}$ is bounded above by $\delta^p\Lambda^p$ (i.e., $M_{\mathbb{C}} < \delta^p\Lambda^p$). Accordingly, for Regime $\mathbb{P}$ to yield a higher match rate than Regime $\mathbb{C}$, it is sufficient that $\delta^c\Lambda^c > \delta^p\Lambda^p$ (condition (b) of the proposition).

Now, we derive conditions under which Regime $\mathbb{O}$'s match rate is lower than Regime $\mathbb{P}$'s match rate (i.e., $M_{\mathbb{O}} < \delta^c\Lambda^c$). We write Regime $\mathbb{O}$'s match rate in general form:

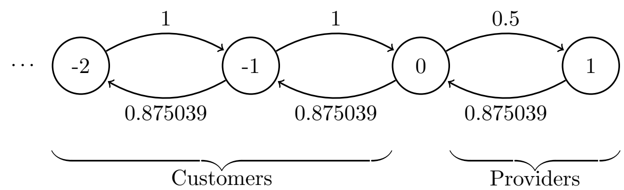
$$M_{\mathbb{O}} = \sum_{n=-b^c_{\mathbb{O}}+1}^{0} \delta^c\Lambda^c\pi_{\mathbb{O}}(n) + \sum_{n=1}^{b^p_{\mathbb{O}}} \Lambda^c\pi_{\mathbb{O}}(n), \tag{68}$$

where the first and second terms vanish if $b^c_{\mathbb{O}} = 0$ and $b^p_{\mathbb{O}} = 0$, respectively. Now we use the normalization condition $\pi_{\mathbb{O}}(-b^c_{\mathbb{O}}) + \sum_{n=-b^c_{\mathbb{O}}+1}^{0}\pi_{\mathbb{O}}(n) + \sum_{n=1}^{b^p_{\mathbb{O}}}\pi_{\mathbb{O}}(n) = 1$. Multiplying both sides of the normalization equation by $\delta^c\Lambda^c$ and some algebraic manipulations yields:

$$\delta^c\Lambda^c\pi_{\mathbb{O}}(-b^c_{\mathbb{O}}) + \delta^c\Lambda^c\sum_{n=-b^c_{\mathbb{O}}+1}^{0}\pi_{\mathbb{O}}(n) + \delta^c\Lambda^c\sum_{n=1}^{b^p_{\mathbb{O}}}\pi_{\mathbb{O}}(n) = \delta^c\Lambda^c \Rightarrow$$
$$\delta^c\Lambda^c\pi_{\mathbb{O}}(-b^c_{\mathbb{O}}) + \underbrace{\delta^c\Lambda^c\sum_{n=-b^c_{\mathbb{O}}+1}^{0}\pi_{\mathbb{O}}(n) + \Lambda^c\sum_{n=1}^{b^p_{\mathbb{O}}}\pi_{\mathbb{O}}(n)}_{M_{\mathbb{O}}(\text{Eq. (68)})} - (1-\delta^c)\,\Lambda^c\sum_{n=1}^{b^p_{\mathbb{O}}}\pi_{\mathbb{O}}(n) = \underbrace{\delta^c\Lambda^c}_{M_{\mathbb{P}}}. \tag{69}$$



(a) Example 1



(b) Example 2

Figure 15 The CTMC for Regime $\mathbb{P}$ under the two examples

In Eq. (69) if Ineq. (70) holds, we will have $M_{\mathbb{P}} > M_{\mathbb{O}}$:

$$\delta^c \Lambda^c \pi_{\mathbb{O}}(-b^c_{\mathbb{O}}) > (1-\delta^c)\Lambda^c \sum_{n=1}^{b^p_{\mathbb{O}}} \pi_{\mathbb{O}}(n). \tag{70}$$

By substituting the expressions for the stationary probabilities and the bounding states in Ineq. (70), we derive:

$$\delta^c \left(\rho^c\right)^{b^c_{\mathbb{O}}} > \frac{\rho^p \left(1-\delta^c\right)\left(1-\left(\rho^p\right)^{b^p_{\mathbb{O}}}\right)}{1-\rho^p}. \tag{71}$$

E. Illustrative examples of cases when Regime $\mathbb{P}$ is optimal

This section provides two examples illustrating that disclosing delay information in our two-sided setting to the user class experiencing relatively higher congestion is not always the best strategy. We select parameters for these two examples such that the conditions for the optimality of Regime $\mathbb{P}$ in Proposition 1 are met. The first example involves relatively higher congestion on the providers' side, while the second involves relatively higher congestion on the customers' side.

Example 1. $\Lambda^c = 2$, $\Lambda^p = 5$, $K^p = 1$, $\delta^c = 0.265707$, $\delta^p = 0.1$, $t^c = 0.942665$, and $t^p = 0.1$. In this case, $M_{\mathbb{P}} = 0.531414 > M_{\mathbb{O}} = M_{\mathbb{C}} = 0.5$. Fig. 15a shows the resulting CTMC under Regime $\mathbb{P}$. In this case, the probability of an excess of providers (i.e., the summation of the steady state probabilities of the positive numbered states) is 0.754, while the probability of an excess of customers (i.e., the summation of the steady state probabilities of the negative numbered states) is 0.026. Therefore, this corresponds to a case where the *provider's side of the platform* is relatively more congested.

Example 2. $\Lambda^c = 2$, $\Lambda^p = 1$, $K^p = 1$, $\delta^c = 0.437519$, $\delta^p = 0.1$, $t^c = 0.0936184$, and $t^p = 0.4$. In this case, $M_{\mathbb{P}} = 0.875039 > M_{\mathbb{O}} = 0.840748 > M_{\mathbb{C}} = 0.1$. Fig. 15a shows the resulting CTMC under Regime $\mathbb{P}$. In this case, the probability of an excess of providers is 0.014, while the probability of an excess of customers is 0.863. Therefore, this corresponds to a case where the *customer's side of the platform* is relatively more congested.

The key takeaway is that when one of the asymmetric regimes (i.e., Regime $\mathbb{P}$ or Regime $\mathbb{C}$) is optimal, either side could be relatively more congested under the optimal regime.

F. Proof of Proposition 2

We prove Proposition 2 for the optimality of Regime $\mathbb{C}$; the proof for the optimality of Regime $\mathbb{C}$ is analogous, with customers and providers switching roles. We will first show that when $t^p \in \left(\frac{\Lambda^c}{K^p + \Lambda^c}, 1\right)$, the match rates under Regimes $\mathbb{P}$ and $\mathbb{O}$ are bounded above by $\delta^c \Lambda^c$. Subsequently, we will show that when $t^c \to 0$ and $\delta^c \Lambda^c < \delta^p \Lambda^p$, the match rate under Regime $\mathbb{C}$ is greater than $\delta^c \Lambda^c$.

We begin with Regimes $\mathbb{P}$ and $\mathbb{O}$. Note that under both regimes, providers receive information and patient providers' delay sensitivity is high, i.e., $t^p \in \left(\frac{\Lambda^c}{K^p + \Lambda^c}, 1\right)$. Under this setting, based on (9), the bounding state for providers in both regimes is 0. Accordingly, in both regimes, impatient customers do not join the system as there will always be a delay for customers in the system. Hence, only patient customers can potentially join the system, and accordingly, the patient customers' maximum arrival rate $\delta^c \Lambda^c$ becomes the best possible match rate under Regimes $\mathbb{P}$ and $\mathbb{O}$.

Next, we show that the match rate under Regime $\mathbb{C}$ is greater than $\delta^c \Lambda^c$. We show this by examining the utility of providers when they join with rate $\delta^c \Lambda^c$ (i.e., with probability $J_{\mathbb{P}}(t^p|\emptyset) = \frac{\delta^c \Lambda^c}{\delta^p \Lambda^p} < 1$): if their utility is greater than 0 when they join with this probability, the equilibrium joining probability will be higher than $\frac{\delta^c \Lambda^c}{\delta^p \Lambda^p}$. This suffices for our proof, since the providers' arrival rate is the match rate of the system under Regime $\mathbb{C}$.

Using the Markov Chain for Regime $\mathbb{C}$ and replacing the providers' arrival rate by $\delta^c \Lambda^c$, we obtain expressions for the steady state probabilities, the resulting expected delay and providers' expected utility as follows:

$$\pi_{\mathbb{C}}(0) = \frac{1 - \delta^c}{1 + b^c - \delta^c b^c}, \tag{72}$$

$$\mathbb{E}[W_{\mathbb{C}}^p(\emptyset)] = \frac{1}{\Lambda^c(\delta^c - 1)\,(\delta^c b^c - b^c - 1)}, \tag{73}$$

$$\mathbb{E}[U_{\mathbb{C}}^p(\emptyset)] = 1 - t^p \left(\frac{K^p}{\Lambda^c\,(\delta^c - 1)\,(\delta^c b^c - b^c - 1)} + 1 \right), \tag{74}$$

where $b^c = \lfloor \delta^c \Lambda^c \omega^c \rfloor$.

Note that the expected delay in Eq. (73) is decreasing in b^c (and the utility in Eq. (74) is increasing in b^c), and b^c contains a floor function. Accordingly, replacing b^c by $\delta^c\Lambda^c\omega^c - 1$ yields an upper bound on the expected delay in Eq. (73), and hence a lower bound on the utility in Eq. (74):

$$\mathbb{E}[U^p_{\mathbb{C}}(\emptyset)] \geq 1 - t^p\left(\frac{K^p}{\Lambda^c(1-\delta^c)\left(1-\delta^c\left(\Lambda^c\delta^c\omega^c - 1\right) + \left(\Lambda^c\delta^c\omega^c - 1\right)\right)} + 1\right). \tag{75}$$

We now show that this lower bound on the utility is larger than 0 by taking its limit when t^c approaches 0 from above:

$$\lim_{t^c\to 0^+} 1 - t^p\left(\frac{K^p}{\Lambda^c(\delta^c - 1)\left(\delta^c\left(\frac{\Lambda^c\delta^c(1-t^c)}{t^c} - 1\right) - \left(\frac{\Lambda^c\delta^c(1-t^c)}{t^c} - 1\right) - 1\right)} + 1\right) = 1 - t^p > 0. \tag{76}$$

This completes the proof.

G. Proof of Proposition 3

We show how to obtain the limiting match rate under each information regime.

***Regime* $\mathbb{O}$.** Under Regime $\mathbb{O}$, customers receive occupancy information; hence, their behavior is independent of other customers' behavior. Accordingly, an unbounded increase in their arrival rate does not impact whether they join at state 0. Furthermore, due to negligible delays, patient providers are incentivized to join at state 0. Accordingly, the providers' bounding state is guaranteed to be positive, leaving us with two settings where the customers' bounding state is either *positive* or *zero*, depending on the customers' delay sensitivity t^c. If t^c is small enough, the bounding state for customers is positive. If t^c is sufficiently large, the bounding state for customers is 0.

- When customers' delay sensitivity is small enough to yield a non-zero bounding state (i.e., $t^c \leq \frac{\Lambda^p}{1+\Lambda^p}$; this condition can be derived from Eq. (9) by setting the bounding state to be exactly one), their unbounded arrival rate causes the probability that the system is at the customers' bounding state to approach one:

$$\pi_{\mathbb{O}}(b^c_{\mathbb{O}}) = \frac{(\rho^c)^{b^c_{\mathbb{O}}}}{\frac{1-(\rho^p)^{b^p_{\mathbb{O}}}}{1-\rho^p} + \rho^c\frac{1-(\rho^c)^{b^c_{\mathbb{O}}}}{1-\rho^c}}, \tag{77}$$

$$\lim_{\Lambda^c\to\infty} \pi_{\mathbb{O}}(b^c_{\mathbb{O}}) = 1, \tag{78}$$

  which follows from the observation that $\rho^c \to \infty$ and $\rho^p \to 0$ when $\Lambda^c \to \infty$. Finally, note that at the customers' bounding state, providers join with rate Λ^p as there is no delay for them in this state. Accordingly, $M_{\mathbb{O}} = \Lambda^p$ when $\Lambda^c \to \infty$ in this setting.

- When the customers' delay sensitivity is sufficiently large to yield a bounding state of 0 (i.e., $t^c > \frac{\Lambda^p}{1+\Lambda^p}$), we examine the probability of the system being at state 0 (since the customers' bounding state is 0), which approaches one as $\Lambda^c \to \infty$:

$$\pi_{\mathbb{O}}(0) = \frac{1-\rho^p}{1-(\rho^p)^{b^p_{\mathbb{O}}}}, \tag{79}$$

$$\lim_{\Lambda^c \to \infty} \pi_{\mathbb{O}}(0) = 1. \tag{80}$$

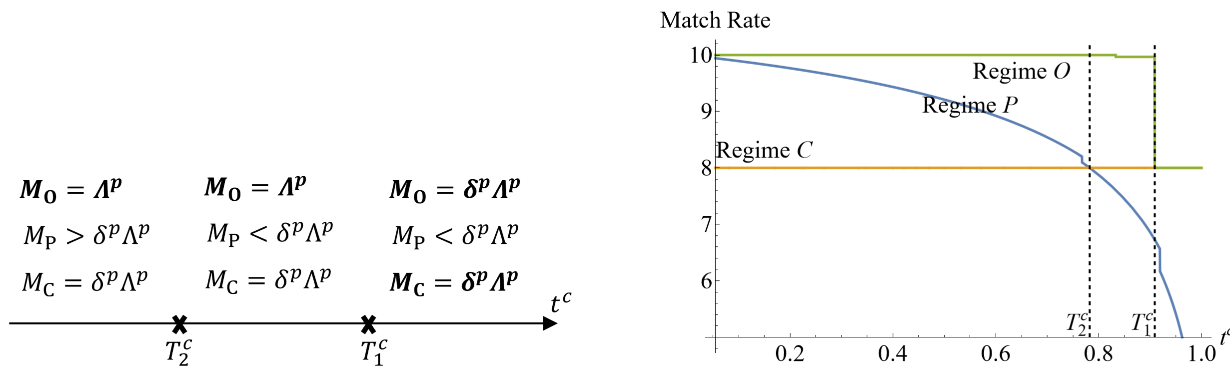
In this state, providers join at rate $\delta^p \Lambda^p$ due to the positive delays. Accordingly, $M_{\mathbb{O}} = \delta^p \Lambda^p$ when $\Lambda^c \to \infty$ in this setting.

Regime $\mathbb{C}$**.** Under Regime $\mathbb{C}$, customers receive occupancy information; hence, their behavior is independent of other customers' behavior. Accordingly, an unbounded increase in their arrival rate does not impact whether they join. However, the unbounded increase in the customers' arrival rate leads to negligible delays for providers (as both $\delta^c \Lambda^c$ and Λ^c are unbounded), which leads to patient providers joining at their maximum possible rate $\delta^p \Lambda^p$. In more technical terms, due to negligible delays, the system is in Case (a) of Lemma 2. Hence, $M_{\mathbb{C}} = \delta^p \Lambda^p$ when $\Lambda^c \to \infty$.

Regime $\mathbb{P}$**.** Under Regime $\mathbb{P}$, customers do not receive state information; hence, their behavior depends on other customers' behavior. Accordingly, an unbounded increase in their arrival rate leads to a mixed strategy joining behavior where only some customers join, putting the system in either Case (b) or (c), as defined by Lemma 2. However, since providers see negligible delays at state 0, the provider's bounding state $b^p_{\mathbb{P}}$ is strictly positive, placing the system in Case (c). Unfortunately, a closed-form solution for their arrival rate (which is also the match rate) does not exist, as shown in Lemma 2. Hence, we use other methods to compare Regime $\mathbb{P}$'s match rate with the other regimes' match rates when $\Lambda^c \to \infty$.

First, we note that the arrival rate that results from the mixed strategy is bounded above by Λ^p (from stability considerations). Second, we check whether customers' mixing results in a match rate lower or higher than $\delta^p \Lambda^p$ by deriving expressions for customers' expected utility if they join with arrival rate $\delta^p \Lambda^p$. If this utility is positive, more customers are willing to join, resulting in a match rate higher than $\delta^p \Lambda^p$; if not, the resulting match rate is lower than $\delta^p \Lambda^p$. Accordingly, there exists a threshold customer delay sensitivity $T_{\mathbb{P}}$ above which customers' match rate is less than $\delta^p \Lambda^p$ and below which their match rate is more than $\delta^p \Lambda^p$. We find $T_{\mathbb{P}}$ by setting customers' utility to 0 when their arrival rate is $\delta^p \Lambda^p$:

$$U^c_{\mathbb{P}}(-) = R^c(1 - T_{\mathbb{P}}(1 + \mathbb{E}[W^c_{\mathbb{P}}(-) | \lambda^c = \delta^p \Lambda^p])) = 0 \tag{81}$$



(a) $\Lambda^c \to \infty$; the optimal regime is highlighted in each region.

(b) Numerical example; $\Lambda^c = 1000 \gg \Lambda^p = 10, \delta^c = 0.6, \delta^p = 0.8, t^p = 0.8, K^p = 0.41$.

Figure 16 **Match rates under market size imbalance when $T_1^c > T_2^c$**

$$\Longrightarrow T_{\mathbb{P}} = \frac{1}{1 + \mathbb{E}[W^c_{\mathbb{P}}(-)|\lambda^c = \delta^p\Lambda^p]} = \left(1 + \frac{1}{\Lambda^p(1-\delta^p)}\left((1-\delta^p)\lfloor\delta^p\Lambda^p\omega^p\rfloor + 1\right)^{-1}\right)^{-1}. \tag{82}$$

As a result, when $t^c > T_{\mathbb{P}}$, $M_{\mathbb{P}} < \delta^p\Lambda^p$ and when $t^c < T_{\mathbb{P}}$, $M_{\mathbb{P}} > \delta^p\Lambda^p$. This completes the proof.

H. Match rates when the condition $T_1^u < T_2^u$ in Proposition 3(b) does not hold

Fig. 16a visualizes the comparison between the three information regimes for different regions of the customers' delay sensitivity t^c when $\Lambda^c \to 1$ and $T_1^u > T_2^u$. Fig. 16b illustrates the comparison in a specific numerical example where Λ^c is much larger than Λ^p (i.e.,$\Lambda^c = 1000$ and $\Lambda^p = 10$).

I. Plots for Regime $\mathbb{O}$'s sub-optimality as provider parameters change

In Fig. 17, we plot the 99% confidence interval of Regime $\mathbb{O}$'s sub-optimality magnitude in response to changes in the providers' parameters. Notably, the sub-optimality magnitude decreases with δ_p (Fig. 17c). However, it tends to be higher for more extreme values of t^p (Fig. 17a) but more intermediate values of Λ_p (Fig. 17b). These changes are similar to those we observed for provider parameters, as discussed in §6 (Fig. 8).

J. Explaining a user class's preference for no information provided to them

In §6.1, our analysis of users' preferences revealed an interesting observation that a user class might be better off when information is hidden from *their own class*. For example, this occurs for customers in our illustrative example in Fig. 9a and when Λ^c is approximately between 7.5 and 9.1 (the region where Regime $\mathbb{C}$'s utility for customers is the lowest of all three regimes).

We can explain the reason by examining the CTMCs in Figs. 1 and 2. Changing from Regime $\mathbb{P}$ to $\mathbb{O}$ has three different effects: (1) The bounding state for providers is higher under Regime $\mathbb{O}$ than under Regime $\mathbb{P}$

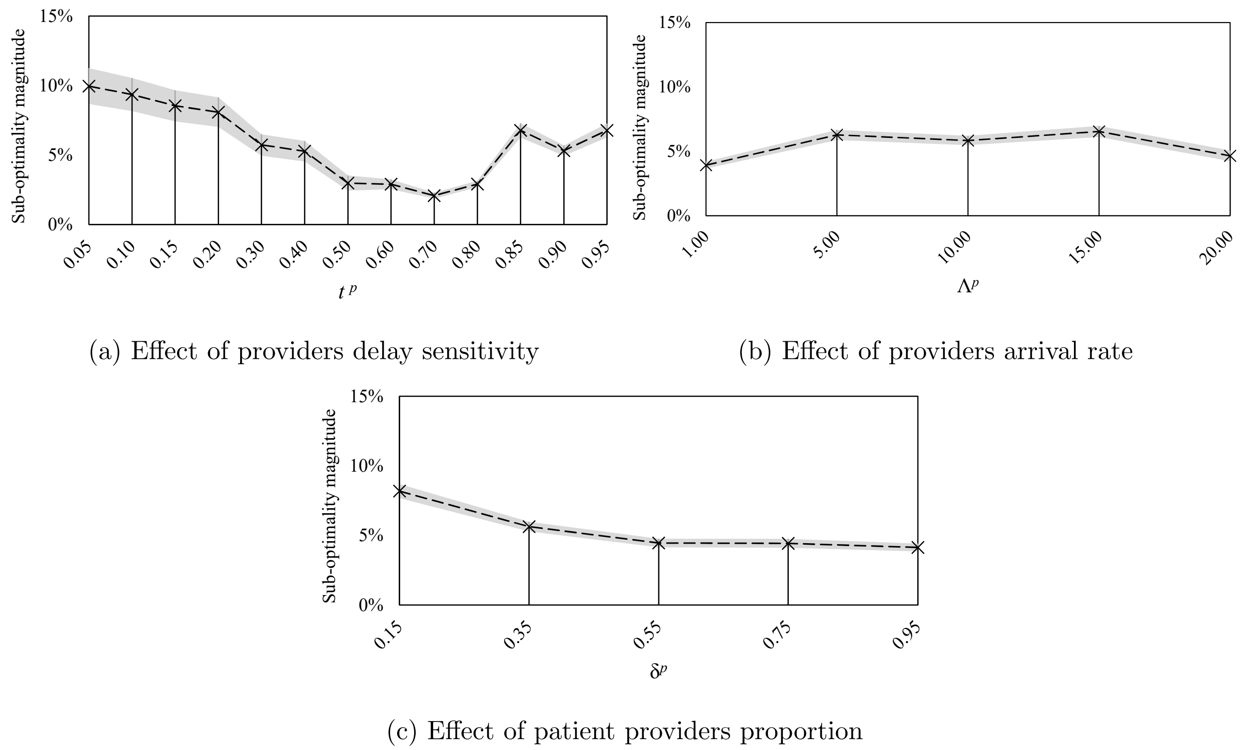


Figure 17 The marginal effect of parameters on Regime $\mathbb{O}$'s sub-optimality magnitude (99% confidence interval)

($b^p_{\mathbb{O}} > b^p_{\mathbb{P}}$). (2) The bounding state $b^c_{\mathbb{O}}$ for customers under Regime $\mathbb{O}$ is finite (whereas the bounding state $b^c_{\mathbb{P}}$ for customers under Regime $\mathbb{P}$ is infinite). Both (1) and (2) reduce the expected delay (and hence increase the utility) of customers. (3) The rate at which providers are cleared from the system when providers are in excess increases under Regime $\mathbb{O}$ (as compared to Regime $\mathbb{P}$) because customers under Regime $\mathbb{O}$ join at a rate $\Lambda^c > \lambda^c_{\mathbb{P}}(\emptyset)$: this causes customers to be less likely to arrive at a system with providers, and therefore, decreases their utility. When the third effect is stronger than the first two, customers obtain higher expected utility under Regime $\mathbb{P}$ than under Regime $\mathbb{O}$.

K. Proof of Proposition 4

Eq. (18) comprises three components: (1) the anticipated time until the first event occurs, (2) the extra expected time if the subsequent event is an arrival of a user of class u, and (3) the additional expected time if the next event is the arrival of class u' (the respective probabilities that the arriving user u' pairs with the focal user u or not are $\frac{1}{|n|+1}$ and $\frac{|n|}{|n|+1}$). The expected waiting time $\mathbb{E}[W^u_I(n)]$ increases in n, necessitating the existence of finite bounding states $b^p_{\mathbb{O}}$, $b^c_{\mathbb{O}}$, $b^p_{\mathbb{P}}$, and $b^c_{\mathbb{C}}$ (like under FCFS). Eqs. (19) and (20) are the boundary conditions.

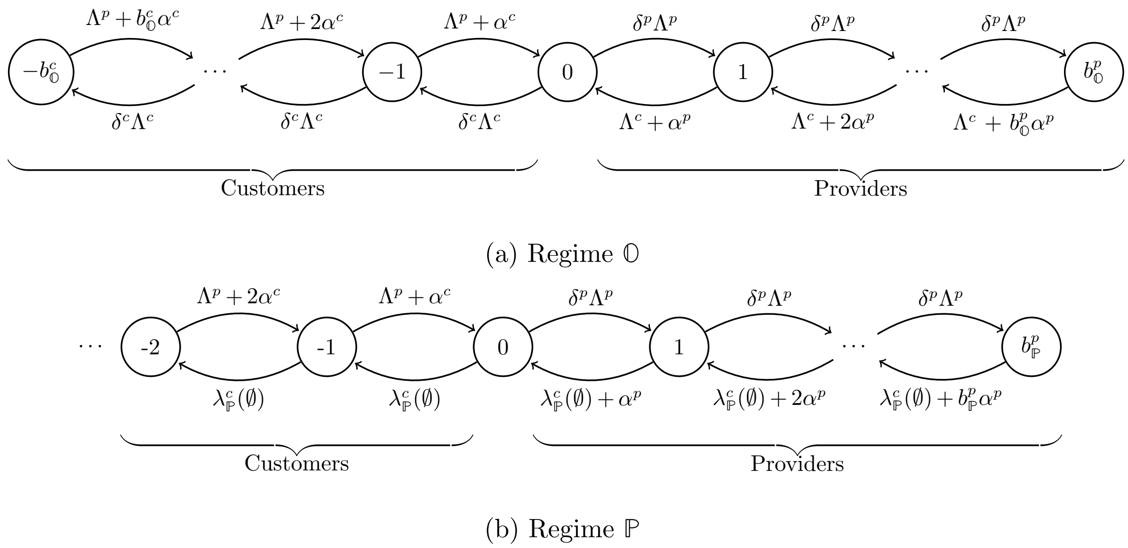


Figure 18 CTMCs for the abandonment models

L. The CTMCs under the abandonment models

Fig. 18 depicts the CTMCs of the abandonment models under Regimes $\mathbb{O}$ and $\mathbb{P}$ (Regime $\mathbb{C}$'s CTMC is analogous to the one for Regime $\mathbb{P}$).

M. Procedures for finding the bounding states and match rates under RSO

Regime $\mathbb{O}$***.*** We explain the procedure from the viewpoint of the providers. The procedure for customers follows similarly. In RSO, like FCFS, providers arrive at rate Λ^p when customers are waiting. However, they only join if the number of providers is below a threshold (the bounding state $b^p_{\mathbb{O}}$). If the expected wait $\mathbb{E}[W^p_{\mathbb{O}}(0)]$ for a provider arriving at state 0 is longer than their willingness to wait (i.e., $\frac{1}{\Lambda^c} > \omega^p$), we set the bounding state $b^p_{\mathbb{O}} = 0$. Otherwise, we iteratively increase $b^p_{\mathbb{O}}$ by one (starting from $b^p_{\mathbb{O}} = 1$), and, in each iteration, we calculate $\mathbb{E}[W^p_{\mathbb{O}}(b^p_{\mathbb{O}} - 1)]$ (the *longest* expected wait time of a provider joining the system belongs to the provider joining at state $b^p_{\mathbb{O}} - 1$) using Eqs. (18)-(20). If $\mathbb{E}[W^p_{\mathbb{O}}(b^p_{\mathbb{O}} - 1)] > \omega^p$, we stop the procedure. Otherwise, we increase $b^p_{\mathbb{O}}$ and repeat the process till we reach the termination condition.

After finding $b^p_{\mathbb{O}}$ and $b^c_{\mathbb{O}}$ based on the above procedure, Eq. (10) yields the match rate $M_{\mathbb{O}}$ under Regime $\mathbb{O}$.

Regime $\mathbb{P}$***.*** Given a customer arrival rate $\lambda^c_{\mathbb{P}}(\emptyset)$, the same procedure used for Regime $\mathbb{O}$ can compute $b^p_{\mathbb{P}}$. We first set $\lambda^c_{\mathbb{P}}(\emptyset) = \delta^c\Lambda^c$ in Eqs. (18)-(20), inducing a finite bounding state $b^p_{\mathbb{P}}$. The resulting CTMC is infinite on the customer's side; therefore, we truncate the customer's side at a suitably large state $(20\Lambda^p\omega_c)$ and numerically approximate the resulting expected wait time for an arriving customer, who experiences an expected wait time of 0 when there is an excess of providers and an expected wait time of $\mathbb{E}[W^c_{\mathbb{P}}(n)]$ (given by Eqs. (18)-(20)) when there are n customers in the system. If the resulting expected wait time is shorter than the customers' willingness to wait $\omega^c = \dfrac{1-t^c}{t^c}$, then the appropriate match rate is $\delta^c \Lambda^c$. Otherwise, we numerically search for the largest value of $\lambda^c_{\mathbb{P}}(\emptyset) \in (0, \delta^c \Lambda^c)$ using Matlab's interior-point method with multiple starts, such that the expected wait time for an arriving customer, calculated as described above, is exactly ω^c. The resulting value of $\lambda^c_{\mathbb{P}}(\emptyset)$ is the required match rate $M_{\mathbb{P}}$.

Regime $\mathbb{C}$**.** The calculations for Regime $\mathbb{C}$ are analogous to Regime $\mathbb{P}$, as explained above, with the roles of customers and providers appropriately switched.