Controlling Congestion when Consumers Choose Their Service Time

Extended Abstract MSOM 2018

In many markets the value that consumers get from a service depends on the amount of time they spend there and they may therefore choose how long they want to stay. At the same time, consumers dislike waiting to be served and prefer to wait as little as possible. For example, consumers going to a gym decide how long they want to spend running on treadmills but dislike waiting for the machines to become available; drivers searching for parking choose how long to park, but incur a disutility from searching for a parking place; customers choose how long to stay at a restaurant, but do not like waiting to be seated; and tourists determine how many hours they want to spend at an art museum, but are unhappy waiting in a long line to enter the exhibition. All of these examples are inherently different from service systems in which the service times are exogenous or from systems in which service providers control service speeds, which implies that firms or social planners may benefit from using different mechanisms to control congestion in these settings.

One way to control congestion, which is commonly used in practice and widely studied in the literature, is to charge a per-use fee for service. Consumers pay a fixed fee every time they enter service and the price serves as a lever on the basis of which consumers choose whether to request service. Naor (1969) and many subsequent papers with various extensions show that per-use fees are effective in controlling the congestion in the system.

Even though the literature has primarily focused on per-use pricing as a means to control congestion, it was almost always under the assumptions that service times are not chosen by consumers and that customers prefer as short of a service as possible. Exceptions include papers on discretionary services, where customers prefer longer service times, but even in these settings consumers do not choose how long to stay and the service provider is able to influence service
speed (see, for example, Anand et al. (2011) and Tong and Rajagopalan (2014)). Settings in which
consumers have service time needs and may prefer longer service times are fundamentally different
than settings in which customers value faster service and speed is controlled by servers who may
be able to exert effort that results in faster service. These service settings may lend themselves to
better mechanisms for controlling congestion. It is therefore suggestive that traditional strategies
that are commonly used to control congestion may not work as well in these settings and that it
may be more appropriate to use other, less studied, strategies instead.

Two such strategies involve instituting time limits and charging a fee per unit of time in service.
These strategies that cannot be implemented in traditional service systems, in which service times
are controlled by the firm, may work well in systems where consumers have individual service
time needs. In fact, these strategies are frequently observed in practice in such settings both
individually and combined. For example, in congested metropolitan areas parking meters are often
set to limit the parking time allowed while at the same time charge consumers a fee per hour parked.
Many gyms set time limits for the use of exercise equipment and the users must stop exercising, if
other costumers are waiting to use them. Recently, museums have placed limits on the time that
visitors could spend when visiting popular exhibitions—The Guggenheim museum limited Doug
Wheeler’s Synthetic Desert III show to 10 minutes and The Hirshhorn Museum in Washington DC
along with five other venues have limited Kusama’s Infinity Mirrors exhibition to only 30 seconds
per visitor per room (Finkel, 2017).

We investigate how well these strategies work for controlling congestion as well as the implications
for firm profits, consumer and social welfare. We consider a single service provider that offers
service to consumers who receive service needs according to a Poisson process. Consumers have
heterogeneous service needs and a consumer’s type denotes the optimal time that she would like
to use the service. While consumers enjoy spending time in service and receive value according to their type, they dislike waiting and incur a cost per unit of time spent waiting.

We characterize customers’ joining behavior and show that three types of behavior are possible in equilibrium: joining and using the service for the desired time, joining and using the service up to the time limit set by the service provider, and balking. How is the firm able to control congestion by setting a time limit? First, consumers who have moderate service needs may choose to enter service, but use it only for the allowed time, thereby limiting time spent in service. Second, in contrast to the usual balking behavior discussed in the literature, according to which consumers with low service values decide not to request service, time limits may result in balking of high value consumers. Specifically, consumers who wish to use the service for a significantly longer time than the limit may choose to balk, if the value they receive from using the service for a short period of time is too low for their needs or does not justify the price. Imagine, for example, drivers searching for parking. A couple who is on their way to watch a movie will likely give up a space that limits the parking duration to 30 minutes. However, a person searching for a place to have dinner may opt for a quick bite at a fast food restaurant instead of a longer sit-down dining experience.

We examine the optimal policy (price and time limit combination) that maximizes firm’s revenue and consumer surplus. We compare the per-unit of time rate with the more commonly analyzed per-use fee and show that in settings where customer have service time needs the price per-unit of time dominates, because it is better at exploiting the consumer heterogeneity in service time needs. We show that to maximize consumer surplus, setting a time limit is sufficient--it is never optimal to charge a positive price. Finally, we identify conditions under which the firm would want to set both a price per unit of time and a strict time limit in order to maximize its revenue.
References:


