Self-regulation of a queue: A tutorial

Moshe Haviv
Department of Statistics
and Federmann Center for the Study of Rationality
The Hebrew University of Jerusalem
91905 Jerusalem
Israel

February 24, 2016

1 Extended abstract

It is well-known that in a queueing system, customers, who mind only their selfish interests, join a queue at a rate which is higher than it is socially desired. The reason behind that is that when customers assess the costs and rewards associated with joining, they mind only their own, while ignoring the additional costs, known as externalities, they inflict on others due to their joining. A central planner, who acts on behalf of society at large, minds these extra costs. In particular, he usually wishes that some (but of course not all) of those who plan to join, will not do so. As reaching this goal by force or by brutally pushing some of the customers out of the queue is usually undesired, we look for other means to get the same effect. Moreover, in the case where this goal is achievable, we prefer those means which come with minimal measures as possible.

The tutorial will commence with describing the simplest queueing model of M/M/1 where all arriving customers are homogeneous with respect to their linear waiting cost parameter and their reward associated with service completion. We look into two cases which vary with the information possessed by the customers: (1) the unobservable case where the queue-length upon arrival is not known to the decision maker, and (2) the observable case,

where it does. Customers need to decide whether to join or not. In the case where customers decide by themselves, they are in fact engaged in an non-cooperative game among themselves. The social planer is facing an optimization problem but he is aware of the fact that any decision he takes will results in some customers' behavior, which in turn will determine the social utility.

1.1 The unobservable case

Assuming a large enough arrival rate, the resulting Nash equilibrium is to join with some probability. In fact, with a probability that makes customers indifferent between joining and not joining. In particular, the resulting consumer surplus is zero. A social planner can do better by prescribing a smaller joining probability, one that maximizes the consumer surplus among all joining probabilities.

The next question now is how to change the rules of the game, i.e., to impose some regulating means, such that under the new rules, the joining probability in the Nash equilibrium in the regulated game, coincides with the socially optimal probability. Moreover, if there exist a few such regulation schemes, one would prefer the minimalist one in terms of the means it imposes.

The most obvious scheme is to impose the right entry fee, one in which the resulting reward from the individual customer point of view is reduced to the level such the new equilibrium joining probability coincides with the social optimal one (under the original rewards). The disadvantages of this scheme is that it involves money transfers and that the entry fee is sensitive to the model parameters and in particular, they need to be known to the social planner prior to the imposition of this fee. Another option is to a sign a contract with the entering customers asking them to pay some function of a to-be-realized value of some random variable, e.g., their service times. This will lead to the desired regulation as long as the mean payment coincides with the abovementioned entry fee. Such schemes, on top of being perceived as being more fair, are sometime (as we will show) more robust to the model's parameters. Of course, they still are based on money transfer. Another scheme which leads to regulation is by auctioning priority. Specifically, customers, on top of having the option not to join, can pay for a priority level: The more they pay, the higher is their priority (regardless of time of arrival). As it turns out, this extra competition among the customers, leads to a Nash equilibrium joining rate which coincides with the socially optimal one.

The tutorial will suggest a new regulating scheme which will be based on designing a new queue regime which serves the purpose of regulating the system. Moreover, it comes with no money transfers (in particular, no need for contracts) and is robust to the model parameters and in fact, there will be no need to know them by the regulator. This regime is called *preemptive random priority*. Under this scheme each arrival draws a random priority parameter from some continuous distribution (it does not matter what is this distribution). In case a customer joins, he is served in a preemptive priority way where his priority is as drawn. Customers inspect their (random) priority parameters and then decide whether to join or not. We show that this scheme regulates the system (on top of being robust to the model parameters and not being based on money transfers). We will also describe some variations of this scheme. They may require the knowledge of some of the model parameters but come with some other, to be described, appealing properties.

1.2 The observable case

In this case customers inspect the queue-length upon their arrival and only then decide whether to join or not. When customers mind only their own utility they join if and only if the queue-length is below (an easy to find) threshold. This threshold in fact implies a dominant strategy from the customers' side. Again, customers ignore the externalities they inflict on those who are behind them. Thus, a central planner, who does not, prescribes a smaller threshold.

As in the unobservable case, one wishes to align customers' decisions with the society's. One option is to impose entry fees (which might beflat or queue-length dependent) such that the value of service from the customers' point of view will be reduced so as they will decide to join if and only if it is socially optimal to do so.

The tutorial will concentrate on regulating schemes which are based on changing the entrance regime which is otherwise assumed to be first-come first-served (FCFS). One scheme is to change this regime to not-FCFS. Here an arrival is placed anywhere but not in the rear of the queue (if only one customer is in the system, his service is preempted and the arrival commences service immediately). Now customers face the dilemma when to renege once the queue ahead of them is too long. As it turns out, they do so at the socially optimal threshold. Another option is to set a number of prices such that the more one pays (if one joins), the higher is one's priority. Again, under the equilibrium path, one never joins when the number in the system is above the socially optimal threshold. We will also describe some new regulating schemes which are based on asking joining customers to select a waiting slot from which they cannot move from later on. The server always serves the lowest occupied slot. Once again, regulation is achieved here as well. Some variations of this scheme will be dealt with as well.