Regulating Arrivals to a Queue

When Customers Know their Demand

Moshe Haviv

Department of Statistics and Center for the Study of Rationality
The Hebrew University of Jerusalem

Brisbane, July 2013

The basic queueing model (M/M/1)

- single server
- first come first served (FCFS)
- Poisson arrival rate λ
- exponential service rate $\mu > \lambda$ (mean of $\frac{1}{\mu}$)
- value of service R
- cost per unit of wait C

Some facts

- mean service time $1/\mu$
- utilization level $\rho = \lambda/\mu < 1$
- mean time in the system

$$W = \frac{1}{\mu(1-\rho)}$$

Some facts

- mean service time $1/\mu$
- utilization level $\rho = \lambda/\mu < 1$
- mean time in the system

$$W = \frac{1}{\mu(1-\rho)}$$

- mean time in the system for a stand-by customer

$$\frac{1}{\mu(1-\rho)^2}$$

equals the total added time to the society due to the marginal arrival

Some facts

- mean service time $1/\mu$
- utilization level $\rho = \lambda/\mu < 1$
- mean time in the system

$$W = \frac{1}{\mu(1-\rho)}$$

- mean time in the system for a stand-by customer

$$\frac{1}{\mu(1-\rho)^2}$$

  equals the total added time to the society due to the marginal arrival

Example: assume $\lambda = 0.9$ and $1/\mu = 1$
$\Rightarrow \rho = 0.9$
$\Rightarrow$ mean time in the system 10
$\Rightarrow$ mean socially added time 100 (for 1 unit of service!)

Some facts

- mean service time $1/\mu$
- utilization level $\rho = \lambda/\mu < 1$
- mean time in the system

$$W = \frac{1}{\mu(1-\rho)}$$

- mean time in the system for a stand-by customer

$$\frac{1}{\mu(1-\rho)^2}$$

  equals the total added time to the society due to the marginal arrival

Example: assume $\lambda = 0.9$ and $1/\mu = 1$
$\Rightarrow \rho = 0.9$
$\Rightarrow$ mean time in the system 10
$\Rightarrow$ mean socially added time 100 (for 1 unit of service!)

You care for the 10, not for the 100. This is why queues are too long.

To queue or not to queue Edleson and Hildebrand, '75

assume

$$R - \frac{C}{\mu} > 0 \quad \text{and} \quad R - \frac{C}{\mu(1-\rho)} < 0$$

if nobody joins, one better joins. If all join, one better do not join.

To queue or not to queue Edleson and Hildebrand, '75

assume

$$R - \frac{C}{\mu} > 0 \text{ and } R - \frac{C}{\mu(1-\rho)} < 0$$

if nobody joins, one better joins. If all join, one better do not join.

(Nash) equilibrium: join with probability p_e where

$$R - \frac{C}{\mu(1 - p_e \rho)} = 0$$

In equilibrium, all are indifferent between joining or not.

To queue or not to queue Edleson and Hildebrand, '75

assume

$$R - \frac{C}{\mu} > 0 \;\; \text{and} \;\; R - \frac{C}{\mu(1-\rho)} < 0$$

if nobody joins, one better joins. If all join, one better do not join.

(Nash) equilibrium: join with probability p_e where

$$R - \frac{C}{\mu(1-p_e\rho)} = 0$$

In equilibrium, all are indifferent between joining or not.

social optimization: join with probability p_s where

$$p_s = \arg\max_{0<p<p_e} p\lambda(R - \frac{C}{\mu(1-p\rho)})$$

To queue or not to queue Edleson and Hildebrand, '75

assume

$$R - \frac{C}{\mu} > 0 \text{ and } R - \frac{C}{\mu(1-\rho)} < 0$$

if nobody joins, one better joins. If all join, one better do not join.

(Nash) equilibrium: join with probability p_e where

$$R - \frac{C}{\mu(1 - p_e \rho)} = 0$$

In equilibrium, all are indifferent between joining or not.

social optimization: join with probability p_s where

$$p_s = \arg \max_{0<p<p_e} p\lambda(R - \frac{C}{\mu(1-p\rho)})$$

$$R - \frac{C}{\mu(1 - p_s\rho)^2} = 0$$

In social optimization, the society is indifferent whether the marginal customer joins or not.

Some facts

- The equilibrium arrival rate: $\lambda_e = \mu - \frac{C}{R}$.
- The socially optimal arrival rate: $\lambda_s = \mu - \sqrt{\frac{C\mu}{R}}$.
- Either rate is not a function of the potential rate.
- $$\lambda_s < \lambda_e \Rightarrow \text{long queues}$$
- The consumer surplus is zero in equilibrium.
  It is $(\sqrt{R\mu} - \sqrt{C})^2$ in social optimization.

Regulating by an entry fee (Pigouvian tax)

socially optimal entry fee T:

$$R - T - \frac{C}{\mu(1 - p_s \rho)} = 0$$

$$\Downarrow$$

$$T = R - CW = R - \sqrt{\frac{CR}{\mu}}$$

Regulating by an entry fee (Pigouvian tax)

socially optimal entry fee T:

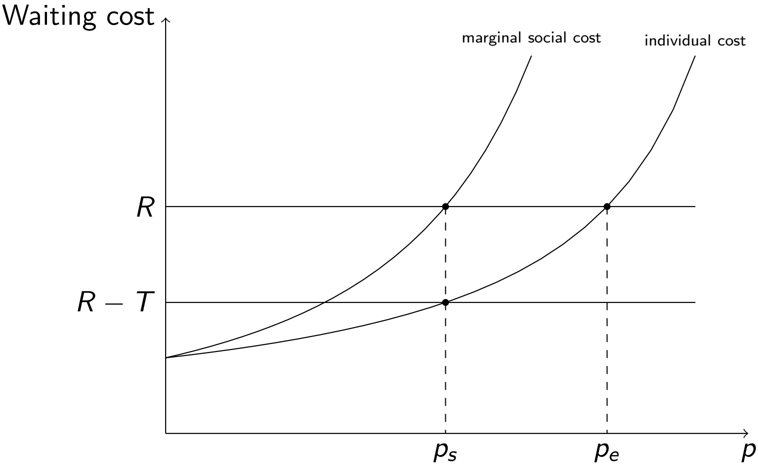
$$R - T - \frac{C}{\mu(1 - p_s \rho)} = 0$$

$$\Downarrow$$

$$T = R - CW = R - \sqrt{\frac{CR}{\mu}}$$

$$T = \frac{C}{\mu(1 - p_s \rho)^2} - \frac{C}{\mu(1 - p_s \rho)}$$

$T =$ externalities the marginal joiner inflicts under the socially optimal scenario


Waiting cost
marginal social cost
individual cost
R
R − T
p_s
p_e
p

Regulating by increasing waiting costs

the same effect is achieved with an added holding fee h:

$$R - \frac{C + h}{\mu(1 - p_s \rho)} = 0$$

$$\Downarrow$$

$$h = \sqrt{RC\mu} - C$$

Regulating contracts

A contract: if you join, pay $f(X)$ for some unknown random variable X.

If $\mathrm{E}(f(X))$ coincides with the externalities under social optimal joining rate, this scheme leads to regulation.

$f(X) =$ the expected externalities given X.

Regulating contracts

A contract: if you join, pay $f(X)$ for some unknown random variable X.

If $\mathrm{E}(f(X))$ coincides with the externalities under social optimal joining rate, this scheme leads to regulation.

$f(X) =$ the expected externalities given X.

Possible random variables:

- time in the system
- queue length upon arrival
- queue length upon departure
- service time

Expected Externalities

W = time in the system (service inclusive)

$$C \frac{\lambda_s W}{\mu(1 - p_s \rho)} = C \left(\sqrt{\frac{R\mu}{C}} - 1 \right) W$$

Expected Externalities

W = time in the system (service inclusive)

$$C\frac{\lambda_s W}{\mu(1-p_s\rho)} = C\left(\sqrt{\frac{R\mu}{C}}-1\right)W$$

L_a = number in the system upon arrival (inclusive)

$$C\frac{L_a}{\mu(1-p_s\rho)} - C\frac{L_a}{\mu} = C\left(\sqrt{\frac{R}{C\mu}}-\frac{1}{\mu}\right)L_a$$

Expected Externalities

W = time in the system (service inclusive)

$$C\frac{\lambda_s W}{\mu(1-p_s\rho)} = C\left(\sqrt{\frac{R\mu}{C}}-1\right)W$$

L_a = number in the system upon arrival (inclusive)

$$C\frac{L_a}{\mu(1-p_s\rho)} - C\frac{L_a}{\mu} = C\left(\sqrt{\frac{R}{C\mu}}-\frac{1}{\mu}\right)L_a$$

L_d = number in the system upon departure (exclusive)

$$\frac{C}{\mu(1-p_s\rho)}L_d = \sqrt{\frac{CR}{\mu}}L_d$$

Expected Externalities

W = time in the system (service inclusive)

$$C\frac{\lambda_s W}{\mu(1-p_s\rho)} = C\left(\sqrt{\frac{R\mu}{C}}-1\right)W$$

L_a = number in the system upon arrival (inclusive)

$$C\frac{L_a}{\mu(1-p_s\rho)} - C\frac{L_a}{\mu} = C\left(\sqrt{\frac{R}{C\mu}}-\frac{1}{\mu}\right)L_a$$

L_d = number in the system upon departure (exclusive)

$$\frac{C}{\mu(1-p_s\rho)}L_d = \sqrt{\frac{CR}{\mu}}L_d$$

S = service time

$$C\frac{\lambda_s}{2(1-p_s\rho)}S^2 + C\frac{(p_s\rho)^2}{(1-p_s\rho)^2}S$$

Quadratic fees Kelly, '91

W = waiting time
Charge $aW^2 + bW$.

Any a, b with $a\mathrm{E}(W^2) + b\mathrm{E}(W) = T$ will do

For example, $a = C\mu/2$ and $b = -1$

Quadratic fees Kelly, '91

W = waiting time
Charge $aW^2 + bW$.

Any a, b with $a\text{E}(W^2) + b\text{E}(W) = T$ will do

For example, $a = C\mu/2$ and $b = -1$

These a and b are free of R!

This is the unique function $f(W)$ with $\text{E}(f(W)) = T$ which is free of R

A similar scheme with L_a

Some facts

- customers internalize the externalities they inflict on others

Some facts

- customers internalize the externalities they inflict on others
- all the consumer surplus goes to the central planner

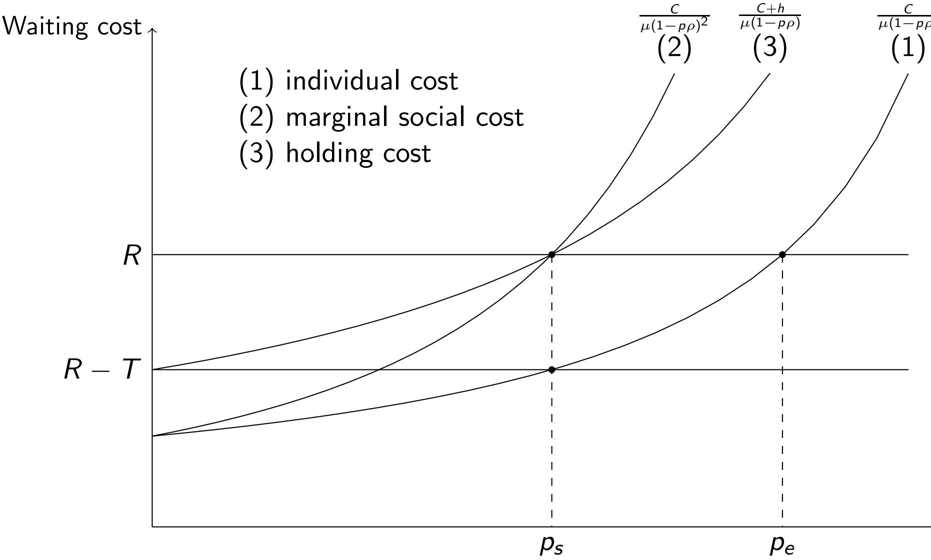
$$(\sqrt{R\mu} - \sqrt{C})^2$$

Some facts

- customers internalize the externalities they inflict on others
- all the consumer surplus goes to the central planner

$$(\sqrt{R\mu} - \sqrt{C})^2$$

- customers are ending up with nothing as they possess no private information


Waiting cost
(1) individual cost
(2) marginal social cost
(3) holding cost
$\frac{C}{\mu(1-p\rho)^2}$
(2)
$\frac{C+h}{\mu(1-p\rho)}$
(3)
$\frac{C}{\mu(1-p\rho}$
(1)
R
$R-T$
p_s
p_e

Regulating by pessimism

- p_e equilibrium joining probability
- p_s socially optimal joining probability

Regulating by pessimism

- p_e equilibrium joining probability
- p_s socially optimal joining probability

Interestingly, $$\frac{1}{\mu(1-p_e\rho)} = \frac{1}{\mu(1-p_s\rho)^2}$$

and hence,

$$R - \frac{C}{\mu(1-p_s\rho)^2} = 0$$

Under a socially optimal joining probability, a stand-by customer is indifferent between joining or not. So is the society: He inflicts no externalities. But society does not mind order of service

Regulating by pessimism

- p_e equilibrium joining probability
- p_s socially optimal joining probability

Interestingly, $$\frac{1}{\mu(1-p_e\rho)} = \frac{1}{\mu(1-p_s\rho)^2}$$

and hence,

$$R - \frac{C}{\mu(1-p_s\rho)^2} = 0$$

Under a socially optimal joining probability, a stand-by customer is indifferent between joining or not. So is the society: He inflicts no externalities. But society does not mind order of service

If all think they are stand-by customers, then p_s is an equilibrium. **Problem:** contradicts standard assumptions in games and economics: all being last cannot be common knowledge....

When customers know their demand

M/G/1, $g(x)$ density of service time

customers know their demand and decide whether to join or not

When customers know their demand

M/G/1, $g(x)$ density of service time

customers know their demand and decide whether to join or not

$W_x(y)=$ mean time for a y job, when x is the threshold
$L_x=$ mean number in the system

assumption: some threshold strategy is a best response

When customers know their demand

M/G/1, $g(x)$ density of service time

customers know their demand and decide whether to join or not

$W_x(y)$= mean time for a y job, when x is the threshold
L_x= mean number in the system

assumption: some threshold strategy is a best response

equilibrium threshold: $R - CW_{x_e}(x_e) = 0$
x_e is a best response against x_e.

socially optimal threshold:

$$x_s = \arg\max_x \{\lambda G(x) R - C L_x\}$$

socially optimal threshold:

$$x_s = \arg\max_x \{\lambda G(x) R - C L_x\}$$

introduce fees making an x_s customer indifferent between joining or not against all using threshold x_s

flat entry fee T:

$$R - T - C W_{x_s}(x_s) = 0$$

linear holding fee h:

$$R - (C + h) W_{x_s}(x_s) = 0$$

linear service fee w:

$$R - C W_{x_s}(x_s) - w x_s = 0$$

socially optimal threshold:

$$x_s = \arg\max_x \{\lambda G(x) R - C L_x\}$$

introduce fees making an x_s customer indifferent between joining or not against all using threshold x_s

flat entry fee T:

$$R - T - C W_{x_s}(x_s) = 0$$

linear holding fee h:

$$R - (C + h) W_{x_s}(x_s) = 0$$

linear service fee w:

$$R - C W_{x_s}(x_s) - w x_s = 0$$

$$h W_{x_s}(x_s) = w x_s = T$$

the externalities that an x_s customer inflicts on a y customer:

$$\frac{C}{g(x_s)}\frac{d}{dx}W_{x_s}(y),\ 0 \le y \le x_s$$

it is socially optimal if externalities are being internalized

the externalities that an x_s customer inflicts on a y customer:

$$\frac{C}{g(x_s)} \frac{d}{dx} W_{x_s}(y), \; 0 \leq y \leq x_s$$

it is socially optimal if externalities are being internalized

total externalities=socially optimal entry fee:

$$T = \frac{C}{g(x_s)} \int_{y=0}^{x_s} \frac{d}{dx} W_{x_s}(y) g(y) \, dy.$$

the externalities that an x_s customer inflicts on a y customer:

$$\frac{C}{g(x_s)} \frac{d}{dx} W_{x_s}(y),\ 0 \leq y \leq x_s$$

it is socially optimal if externalities are being internalized

total externalities=socially optimal entry fee:

$$T = \frac{C}{g(x_s)} \int_{y=0}^{x_s} \frac{d}{dx} W_{x_s}(y) g(y)\, dy.$$

in case of no externalities:

- $x_s = x_e$
- $T = h = w = 0$

Comparing schemes

Level of Regulation (LoR): mean payment divided by mean waiting cost.

axiom: The lower the LoR the better

Comparing schemes

Level of Regulation (LoR): mean payment divided by mean waiting cost.

axiom: The lower the LoR the better

all customers pay most under the holding fee scheme

Comparing schemes

Level of Regulation (LoR): mean payment divided by mean waiting cost.

axiom: The lower the LoR the better

all customers pay most under the holding fee scheme

a y-customer, $0 \leq y \leq x_s$, prefers holding fees to service fees iff

$$W_{x_s}(y)/y \leq W_{x_s}(x_s)/x_s$$

Examples

- First come first served (FCFS)
- Processor sharing (PS)
- Non-preemptive priority to short jobs (SJF)
- static preemptive priority to short jobs (PSJF)

FCFS

all pay more under flat entry fee

holding fee: affine function between $(0, W_{x_s}(0))$ and (x_s, T).

service fee: linear function between $(0, 0)$ and (x_s, T).

$\Rightarrow$ All prefer service fees on holding fees.

Preemptive SJF Hassin and Haviv, '03

It is socially optimal that:

- only short jobs join
- short jobs receive priority

Preemptive SJF Hassin and Haviv, '03

It is socially optimal that:

- only short jobs join
- short jobs receive priority

Static preemptive priority is given to short jobs (PSJF)
Customers, knowing their service times, decide whether to join or not

Equilibrium behavior: Join if and only if service is shorter than or equal to x_s

an x_s customer inflicts no externalities: His/her and the society's interests coincide all pay more under flat entry fee

denote $\rho(x) = \lambda \int_{t=0}^{x} tg(x)\, dt$

$$W_{x_s}(y) = \frac{y}{1-\rho(x_s)}, \quad \text{linear}$$

$$\Downarrow$$

holding fees and service fees coincide (in mean)

$$C\frac{x_s}{(1-\rho(x_s)^2)} = R$$

$$h = \frac{T(1-\rho(x_s))}{x_s} \quad , \quad w = \frac{T}{x_s}$$

SJF (M/M/1)

Non-preemptive shortest job first

1. if $\rho \leq (1 + 9e^{-2.5}) \Rightarrow$ all prefer service fees on holding fees
2. otherwise, for small and large value of y, service fees are preferred. For mid values of y, holding fees.

M/G/1: an odd number of intervals where the preferences alternate.

Regulating by auctioning priorities Hassin, '85

One who pays x overtakes (preemptively) all those who pay y, $y < x$.

Decision problem: To join or not to join. If join, how much to pay?

Regulating by auctioning priorities Hassin, '85

One who pays x overtakes (preemptively) all those who pay y, $y < x$.

Decision problem: To join or not to join. If join, how much to pay?

Equilibrium: join with probability p_s (as socially optimal!)

Q: And how much to pay?
A: Mix with density along $[0, a]$ where

$$a = \frac{C}{\mu(1 - p_s \rho)^2} - \frac{C}{\mu}$$

and distribution function $F(x)$,

$$R - x - \frac{C}{\mu(1 - p_s(1 - F(x))\rho)^2} = 0, \quad 0 \leq x \leq a$$

Regulating by auctioning priorities Hassin, '85

One who pays x overtakes (preemptively) all those who pay y, $y < x$.

Decision problem: To join or not to join. If join, how much to pay?

Equilibrium: join with probability p_s (as socially optimal!)

Q: And how much to pay?
A: Mix with density along $[0, a]$ where

$$a = \frac{C}{\mu(1 - p_s\rho)^2} - \frac{C}{\mu}$$

and distribution function $F(x)$,

$$R - x - \frac{C}{\mu(1 - p_s(1 - F(x))\rho)^2} = 0, \quad 0 \leq x \leq a$$

Proof: The interests of the one who enters and pays nothing (and is always last and inflicts no externalities), and that of society's coincide

Each pays the externalities he/she inflicts (Pigouvian tax)

Observable queues Naor, '69

Equilibrium: Join if and only if the number upon arrival is smaller than n_e.

$$n_e = \max_n \{R - \frac{C(n+1)}{\mu} \geq 0\}$$

Observable queues Naor, '69

Equilibrium: Join if and only if the number upon arrival is smaller than n_e.

$$n_e = \max_n \{R - \frac{C(n+1)}{\mu} \geq 0\}$$

Multiple equilibria but a unique threshold-based equilibrium

Observable queues Naor, '69

Equilibrium: Join if and only if the number upon arrival is smaller than n_e.

$$n_e = \max_n \{R - \frac{C(n+1)}{\mu} \geq 0\}$$

Multiple equilibria but a unique threshold-based equilibrium

Socially optimal strategy: Join if and only if upon arrival the number in system is smaller than n_s

$n_s \leq n_e$ (and equality iff $n_e = 1$) $\Rightarrow$ long queues

A right entry toll coincide the new n_e and the old n_s

Regulating by not-FCFS Hassin, '85

Change the entrance policy to not-FCFS: An arrival is placed anywhere (with preemption) but the last position

Regulating by not-FCFS Hassin, '85

Change the entrance policy to not-FCFS: An arrival is placed anywhere (with preemption) but the last position

The individual decision problem: to renege if queue ahead is too long

Equilibrium: Renege when at position $n_s + 1$

Explanation: The one at the back does not inflict any externalities. His utility coincides with the society's

THANK YOU