Economic Issues in Shared Infrastructures

Costas Courcoubetis

Department of Computer Science,
Athens University of Economics and Business

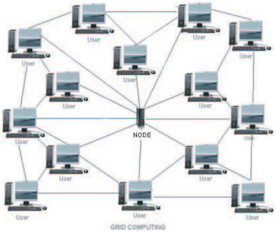
Richard Weber

Statistical Laboratory, University of Cambridge

VISA 2009, August 17, Barcelona

Resource management in virtual facilities

Resource sharing within virtual infrastructures is made complex because of the details of technology specificities.



Mathematics/economics can help to highlight some key issues.

Example: scheduling a server

- A single server is to be shared amongst n agents.
  Agent i generates a jobs at rate λ_i.

Example: scheduling a server

- A single server is to be shared amongst n agents. Agent i generates a jobs at rate λ_i.
- Initially, agents contribute resource amounts $y_1, \ldots, y_n$, building a server of rate $\sum_k y_k$. Under FCFS scheduling all jobs have mean waiting time $1/(\sum_k y_k - \sum_k \lambda_k)$.

Example: scheduling a server

- A single server is to be shared amongst n agents.
  Agent i generates a jobs at rate λ_i.
- Initially, agents contribute resource amounts $y_1, \ldots, y_n$, building a server of rate $\sum_k y_k$.
  Under FCFS scheduling all jobs have mean waiting time $1/(\sum_k y_k - \sum_k \lambda_k)$.
- Agent i suffers a delay cost, so his net benefit is, say,

$$nb_i = \lambda_i r - \theta_i \lambda_i \frac{1}{\sum_k y_k - \sum_k \lambda_k} - y_i$$

  θ_i **is private information** of agent i.

The key issue in this talk

Agents (users) have **private information** (about the value of the tasks they wish to carry out).

This creates a problem for efficiently sharing resources.

- Agents will attempt to **free-ride**.
- Obvious policies (like 'internal market', or 'equal sharing') may not be suitable.

The key issue in this talk

Agents (users) have **private information** (about the value of the tasks they wish to carry out).

This creates a problem for efficiently sharing resources.

- Agents will attempt to **free-ride**.
- Obvious policies (like 'internal market', or 'equal sharing') may not be suitable.

How one chooses to share a facility's resources will influence what agents reveal of their private information.

We would like agents to truthfully reveal their privately held information since then we can operate the facility more efficiently.

A model of a managed shared infrastructure

- An infrastructure as composed of **resources**.
  (links, servers, buffers, etc.)

A model of a managed shared infrastructure

- An infrastructure as composed of **resources**.
  (links, servers, buffers, etc.)
- It can be **operated in various ways**, $\omega_1, \omega_2, \ldots$
  (by scheduling, routing, bandwidth allocation, etc.)

A model of a managed shared infrastructure

- An infrastructure as composed of **resources**.
  (links, servers, buffers, etc.)
- It can be **operated in various ways**, $\omega_1, \omega_2, \ldots$
  (by scheduling, routing, bandwidth allocation, etc.)
- The subset of agents who wish to use resources of the infrastructure, say S, differs from day to day.

A model of a managed shared infrastructure

- An infrastructure as composed of **resources**.
  (links, servers, buffers, etc.)
- It can be **operated in various ways**, $\omega_1, \omega_2, \ldots$
  (by scheduling, routing, bandwidth allocation, etc.)
- The subset of agents who wish to use resources of the infrastructure, say S, differs from day to day.
- If operated in manner ω on day t then agent i has **utility**

$$\theta_{i,t} u_i(\omega)$$

$u_i(\cdot)$ is pubic knowledge, but **only agent i knows $\theta_{i,t}$**.

A model of a managed shared infrastructure

- An infrastructure as composed of **resources**.
  (links, servers, buffers, etc.)
- It can be **operated in various ways**, $\omega_1, \omega_2, \ldots$
  (by scheduling, routing, bandwidth allocation, etc.)
- The subset of agents who wish to use resources of the infrastructure, say S, differs from day to day.
- If operated in manner ω on day t then agent i has **utility**

$$\theta_{i,t} u_i(\omega)$$

  $u_i(\cdot)$ is pubic knowledge, but **only agent i knows $\theta_{i,t}$**.
- ω is to be chosen as a function of S and of the declared $\theta_t = (\theta_{1,t}, \ldots, \theta_{n,t})$.

Agents pay for resources

Agents may contribute resources to a shared infrastructure, like $\boldsymbol{y_1(\theta_1), \ldots, y_n(\theta_n)}$.

Agents pay for resources

Agents may contribute resources to a shared infrastructure, like $y_1(\theta_1), \ldots, y_n(\theta_n)$.

Other times agents pay fees. In this case, we should like them to cover some daily operating cost, c,

$$E_{S,\theta}\Big[p_1(S,\theta) + \cdots + p_n(S,\theta)\Big] \geq c$$

Agents pay for resources

Agents may contribute resources to a shared infrastructure, like $y_1(\theta_1), \ldots, y_n(\theta_n)$.

Other times agents pay fees. In this case, we should like them to cover some daily operating cost, c,

$$E_{S,\theta}\Big[p_1(S,\theta) + \cdots + p_n(S,\theta)\Big] \geq c$$

Agent i wishes to maximize his **expected net benefit** :

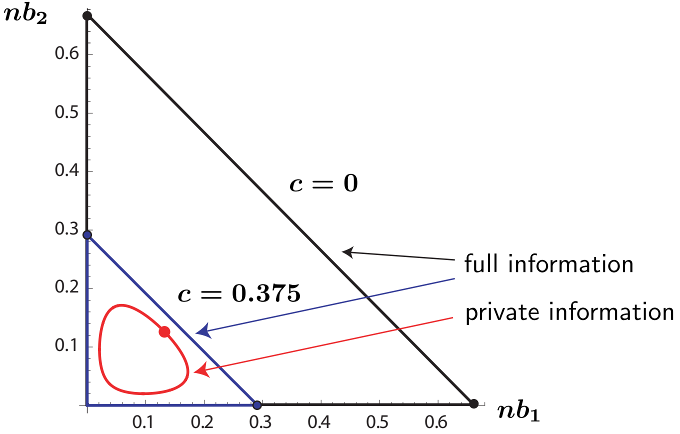
$$nb_i(\theta_i) = E_{S,\theta}\Big[\theta_i u_i(\omega(S,\theta)) - p_i(S,\theta) \,\Big|\, \theta_i\Big]$$

He may be untruthful in declaring his θ_i.

The efficient frontier

We wish to find Pareto optimal points of the vector

$$(nb_1, \ldots, nb_n) = E_\theta\left[nb_1(\theta_1), \ldots, nb_n(\theta_n)\right]$$

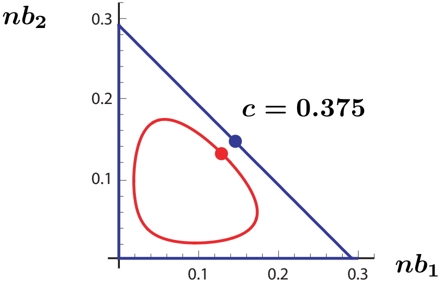


Maximum social welfare

Suppose we wish to find the particular point that maximizes

$$nb_1 + \cdots + nb_n =$$

$$E_{S,\theta}\left[\theta_1 u_1(\omega(S,\theta)) + \cdots + \theta_n u_n(\omega(S,\theta))\right] - c$$



We call this the 'social welfare'.

Our infrastructure optimization problem

Our infrastructure optimization problem:

- Say how the infrastructure will be operated for all possible subsets of users S.

Our infrastructure optimization problem

Our infrastructure optimization problem:

- Say how the infrastructure will be operated for all possible subsets of users S.
- Say what fees will be collected from users.

Our infrastructure optimization problem

Our infrastructure optimization problem:

- Say how the infrastructure will be operated for all possible subsets of users S.
- Say what fees will be collected from users.

Do the above, as function of declared θ_i, so that:

1. Users find it in their best interest to truthfully reveal their θ_i.
2. Users will see positive expected net benefit from participation.
3. Expected total fees cover the daily running cost, say c.
4. Expected social welfare (total net benefit) is maximized

Example: a scalar resource

- 2 participants, both present on all days.

Example: a scalar resource

- 2 participants, both present on all days.
- On day t, agent i has utility for resource of $\theta_{i,t}u(x)$, assumed known to be distributed $U[0,1]$.

Example: a scalar resource

- 2 participants, both present on all days.
- On day t, agent i has utility for resource of $\theta_{i,t}u(x)$, assumed known to be distributed $U[0,1]$.
- The infrastructure provides a single resource, parameterized by a number (such as computing cycles), so operating methods correspond to allocations:

$$\{\omega\} \equiv \{x_1, x_2 \,:\, x_1 + x_2 \leq 1\}$$

Example: a scalar resource

- 2 participants, both present on all days.
- On day t, agent i has utility for resource of $\theta_{i,t} u(x)$, assumed known to be distributed $U[0, 1]$.
- The infrastructure provides a single resource, parameterized by a number (such as computing cycles), so operating methods correspond to allocations:

$$\{\omega\} \equiv \{x_1, x_2 \,:\, x_1 + x_2 \leq 1\}$$

Suppose $u_i(x) = x$. Focus on one day t; with $\theta_i = \theta_{i,t}$.

$$E_{\theta_1,\theta_2}\left[\max_{\substack{x_1,x_2 \\ x_1+x_2\leq 1}} \{\theta_1 u_1(x_1) + \theta_2 u_2(x_2)\}\right] = E\left[\max\{\theta_1, \theta_2\}\right] = \tfrac{2}{3}$$

We call this the '**first best**'.

The second-best solution

A '**second-best**' is with fee structure:

$$p_i(\theta_i) = \begin{cases} 0\,, & \theta_i < \theta_0 \\ \frac{1}{2}(\theta_i^2 + \theta_0^2)\,, & \theta_i \geq \theta_0 \end{cases}$$

Agent i will not wish to participate if $\theta_i < \theta_0$, since his net benefit cannot be positive.

The entire resource is allocated to the agent declaring the greatest θ_i, provided this is $> \theta_0$.

Thus, the resource is given wholly to one agent, but perhaps to neither.

But both agents may pay.

The ‘solution’

This solution has the advantages that

- Agents are incentivized to truthful.

The ‘solution’

This solution has the advantages that

- Agents are incentivized to truthful.
- The sum of the expected payments is

$$E\Big[p_1(\theta_1) + p_2(\theta_2)\Big] = 1/3 + \theta_0^2 - (4/3)\theta_0^3\,.$$

The ‘solution’

This solution has the advantages that

- Agents are incentivized to truthful.
- The sum of the expected payments is

$$E\Big[p_1(\theta_1) + p_2(\theta_2)\Big] = 1/3 + \theta_0^2 - (4/3)\theta_0^3 \,.$$

- The expected social welfare is decreasing in θ_0.

  But by taking $1/3 + \theta_0^2 - (4/3)\theta_0^3 = c$ we maximize the social welfare of

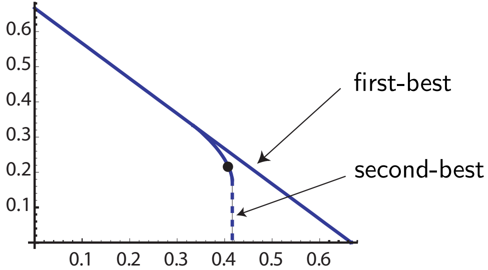
$$nb_1 + nb_2 = E\left[\sum_{i=1}^{2} \theta_i u_i(x_i) - p_i(\theta_i)\right]$$

  subject to covering cost c.

Second-best versus first-best

Expected social welfare as a function of c.

$nb_1 + nb_2$



cost $= c$

For $c \in [0.333, 0.416]$ the second-best falls short of the first-best. There is no way to cover a cost greater than $\frac{5}{12} = 0.416$.

Other mechanisms can also work.

Other mechanisms can also work.

- We can ensure $p_1(\theta_1, \theta_2) + p_2(\theta_1, \theta_2) = c$.

  $$p_1(\theta_1, \theta_2) = \frac{1}{2}c + \frac{1}{2}(\theta_1^2 + \theta_0^2)1_{\{\theta_1 > \theta_0\}} - \frac{1}{2}(\theta_2^2 + \theta_0^2)1_{\{\theta_2 > \theta_0\}}$$

  We call this 'ex-post' cost-covering.

Other mechanisms can also work.

- We can ensure $p_1(\theta_1, \theta_2) + p_2(\theta_1, \theta_2) = c$.

  $$p_1(\theta_1, \theta_2) = \frac{1}{2}c + \frac{1}{2}(\theta_1^2 + \theta_0^2)1_{\{\theta_1 > \theta_0\}} - \frac{1}{2}(\theta_2^2 + \theta_0^2)1_{\{\theta_2 > \theta_0\}}$$

  We call this 'ex-post' cost-covering.

- We can ensure ex-post incentive compatibility and rationality. I.e., that an agent only pays if he gets resource, and is happy after-the-fact with the θ_i he declared.

  $$p_1(\theta_1, \theta_2) = \max(\theta_0, \theta_2)1_{\{\theta_1 > \max(\theta_0, \theta_2)\}}$$

A concave utility

Suppose $u_i(x) = \sqrt{x}$

Now the resource is shared differently.

The optimal policy is found by solving a Lagrangian dual problem

$$\min_{\lambda \geq 0} \left\{ E_{\theta_1,\theta_2} \left[\max_{\substack{x_1,\, x_2 \geq 0 \\ x_1 + x_2 \leq 1}} \sum_{i=1}^{2} h_\lambda(\theta_i) u_i(x_i) \right] - (1+\lambda)c \right\} .$$

where $h_\lambda(\theta_i) = (\theta_i + \lambda(2\theta_i - 1))$ and

$$x_i(\theta_1, \theta_2) = \frac{h_\lambda(\theta_i)^2}{\sum_{j=1}^{2} h_\lambda(\theta_j)^2}$$

As λ increases the fee structure changes, so that greater cost can be covered. The social welfare decreases, but is maximized subject to the constraint of covering the cost.

The role of operating policy

The resource is not allocated in the ‘most efficient’ way. That would be

$$x_i(\theta_1, \theta_2) = \frac{\theta_i^2}{\theta_1^2 + \theta_2^2} \qquad \left(\text{not } \frac{h_\lambda(\theta_i)^2}{h_\lambda(\theta_1)^2 + h_\lambda(\theta_2)^2}\right)$$

The role of operating policy

The resource is not allocated in the 'most efficient' way. That would be

$$x_i(\theta_1, \theta_2) = \frac{\theta_i^2}{\theta_1^2 + \theta_2^2} \qquad \left(\text{not } \frac{h_\lambda(\theta_i)^2}{h_\lambda(\theta_1)^2 + h_\lambda(\theta_2)^2}\right)$$

This is a key lesson:

If one wishes to optimally incentivize participation in shared infrastructures, and to make the most of the resources available, then both the (i) fees, and (ii) policies for ' resource sharing, must play a part in providing the correct incentives to users.

Conclusions

- In most realistic resource allocation problems there is private information to participants.

Conclusions

- In most realistic resource allocation problems there is private information to participants.
- Resource management must give appropriate incentives.
  To encourage agents who value the resource more to say so, and so be willing to contribute more towards the cost, we need to reward them better than an internal market would do.
  But figuring out exactly how to do this is not a simple task!

Conclusions

- In most realistic resource allocation problems there is private information to participants.
- Resource management must give appropriate incentives.
  To encourage agents who value the resource more to say so, and so be willing to contribute more towards the cost, we need to reward them better than an internal market would do. But figuring out exactly how to do this is not a simple task!
- Simple-minded sharing policies (like proportional sharing) may not to produce sufficient incentives for participants to contribute resources.

Conclusions

- In most realistic resource allocation problems there is private information to participants.
- Resource management must give appropriate incentives.
  To encourage agents who value the resource more to say so, and so be willing to contribute more towards the cost, we need to reward them better than an internal market would do. But figuring out exactly how to do this is not a simple task!
- Simple-minded sharing policies (like proportional sharing) may not to produce sufficient incentives for participants to contribute resources.
- Many new interesting problems!!!

Optimal queue scheduling

Instead of declaring contributions they are willing to make, we can imagine that agents (equivalently) declare their θ_i.

Suppose $\theta_1 < \theta_2 < \cdots < \theta_n$.

As a function of these declarations we take contributions of the form $y(\theta_i)$ from some subset of agents $i = 1, \ldots, j$ (a set with smallest θ_i).

Optimal queue scheduling

Instead of declaring contributions they are willing to make, we can imagine that agents (equivalently) declare their θ_i.

Suppose $\theta_1 < \theta_2 < \cdots < \theta_n$.

As a function of these declarations we take contributions of the form $y(\theta_i)$ from some subset of agents $i = 1, \ldots, j$ (a set with smallest θ_i).

We employ a priority scheduling policy in which priority is always given to the current job belonging to the agent with greatest θ_i.

Optimal queue scheduling

Instead of declaring contributions they are willing to make, we can imagine that agents (equivalently) declare their θ_i.

Suppose $\theta_1 < \theta_2 < \cdots < \theta_n$.

As a function of these declarations we take contributions of the form $y(\theta_i)$ from some subset of agents $i = 1, \ldots, j$ (a set with smallest θ_i).

We employ a priority scheduling policy in which priority is always given to the current job belonging to the agent with greatest θ_i.

Under this scheme, an agent with too great a θ_i will find unprofitable to consider participating.

$y_i(\theta_i)$ is increasing in θ_i, and is determined by an incentive compatibility condition.