Economic Issues in Shared Infrastructures

Richard Weber

Statistical Laboratory, University of Cambridge

Costas Courcoubetis

Department of Computer Science, Athens University of Economics and Business

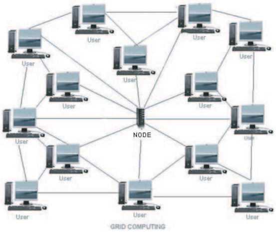
VISA 2009, August 17, Barcelona

Motivation: Shared Infrastructures

Virtual facilities

. . . composed of shared resources, such as computers, routers, and communication links, which are used together to so that agents can perform tasks.

Example: grid computing



Issue for this talk

How do we derive sensible resource allocation policies in shared infrastructures (Grids, communication links) when participants have private information?

Issue for this talk

How do we derive sensible resource allocation policies in shared infrastructures (Grids, communication links) when participants have private information?

Key observation: system operating policies influence strategies of agents.

Issue for this talk

How do we derive sensible resource allocation policies in shared infrastructures (Grids, communication links) when participants have private information?

Key observation: system operating policies influence strategies of agents.

Naive policies (like 'internal market', or 'equal sharing') may not be suitable.

Issue for this talk

How do we derive sensible resource allocation policies in shared infrastructures (Grids, communication links) when participants have private information?

Key observation: system operating policies influence strategies of agents.

Naive policies (like 'internal market', or 'equal sharing') may not be suitable.

Two main problems:

- Since the facility cost must be shared, agents like to free-ride.
- We must resolve contentions for resource in a way incentivizes agents to truthfully reveal private information.

A model of a managed shared infrastructure

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

A model of a managed shared infrastructure

- An infrastructure is composed of resources (links, servers, buffers, etc.)
- It an be operated in different ways, say $\omega \in \Omega$ (by scheduling, routing, bandwidth allocation, etc.)

A model of a managed shared infrastructure

- An infrastructure is composed of resources (links, servers, buffers, etc.)
- It an be operated in different ways, say $\omega \in \Omega$ (by scheduling, routing, bandwidth allocation, etc.)
- On a given day the subset of agents who wish to use the infrastructure is S (a 'state of nature'). This occurs with frequency $\alpha(S)$.

A model of a managed shared infrastructure

- An infrastructure is composed of resources (links, servers, buffers, etc.)
- It an be operated in different ways, say $\omega \in \Omega$ (by scheduling, routing, bandwidth allocation, etc.)
- On a given day the subset of agents who wish to use the infrastructure is S (a 'state of nature'). This occurs with frequency $\alpha(S)$.
- If the infrastructure is operated in manner ω then agent i has benefit

$$\theta_i u_i(\omega)$$

where $u_i(\cdot)$ is pubic knowledge, but only agent i knows θ_i.

A model of a managed shared infrastructure

- An infrastructure is composed of resources (links, servers, buffers, etc.)
- It an be operated in different ways, say $\omega \in \Omega$ (by scheduling, routing, bandwidth allocation, etc.)
- On a given day the subset of agents who wish to use the infrastructure is S (a 'state of nature'). This occurs with frequency $\alpha(S)$.
- If the infrastructure is operated in manner ω then agent i has benefit

  $$\theta_i u_i(\omega)$$

  where $u_i(\cdot)$ is pubic knowledge, but only agent i knows θ_i.
- ω is to be chosen as a function of S and the declared $\theta = (\theta_1, \ldots, \theta_n)$.

Agents pay for operating cost

Agent i is charged a fee $p_i(S, \theta)$.

Fees are to cover a daily operating cost, c, so we require

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

Agents pay for operating cost

Agent i is charged a fee $p_i(S, \theta)$.

Fees are to cover a daily operating cost, c, so we require

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

Agent i wishes to maximize his expected net benefit

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

Agents pay for operating cost

Agent i is charged a fee $p_i(S, \theta)$.

Fees are to cover a daily operating cost, c, so we require

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

Agent i wishes to maximize his expected net benefit

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

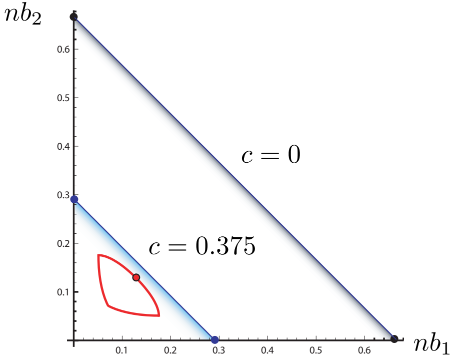
In some situations we may take the fee as money.

In others we may wish to take the fee as a contribution to the pool of resources that is available in the infrastructure.

The efficient frontier

We wish to find Pareto optimal points of the vector

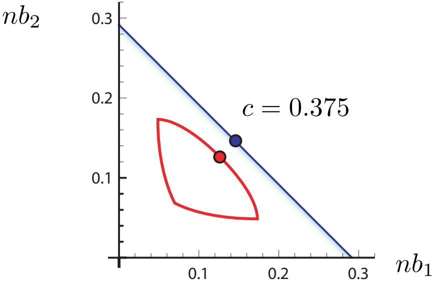
$$(nb_1, \ldots, nb_n)$$



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

The problem of economics for infrastructure optimization is as follows.

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

Our infrastructure optimization problem

The problem of economics for infrastructure optimization is as follows.

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

Our infrastructure optimization problem

The problem of economics for infrastructure optimization is as follows.

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

Do the two things 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: scheduling a server

- Suppose N agents share a single server. Agent i generates a jobs as a Poisson process of rate λ_i, whose service times are exponentially distributed with parameter 1.

Example: scheduling a server

- Suppose N agents share a single server. Agent i generates a jobs as a Poisson process of rate λ_i, whose service times are exponentially distributed with parameter 1.
- Initially, agents contribute resource amounts $y_1, \ldots, y_N$. This results in 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

- Suppose N agents share a single server. Agent i generates a jobs as a Poisson process of rate λ_i, whose service times are exponentially distributed with parameter 1.
- Initially, agents contribute resource amounts $y_1, \dots, y_N$. This results in 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 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, but it has an *a priori* distribution that is public information.

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.

To illustrate ideas we consider a simple infrastructure shared by just 2 participants, both present on all days. On day t, agent i has utility for resource of $\theta_{i,t}u(x)$, where $\theta_{i,t}$ is assumed known to be distributed $U[0,1]$. The infrastructure is described by a single resource parameterized by a number(such as computing cycles), so operating methods choice of allocations:

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

To illustrate ideas we consider a simple infrastructure shared by just 2 participants, both present on all days. On day t, agent i has utility for resource of $\theta_{i,t}u(x)$, where $\theta_{i,t}$ is assumed known to be distributed $U[0,1]$. The infrastructure is described by a single resource parameterized by a number(such as computing cycles), so operating methods choice of allocations:

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

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

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

We call this the ‘first best’.

A ‘second-best’ mechanism can be constructed as follows. If agent i declares θ_i then he is charged a fee

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

He obtains $x_i = 1$ if $\theta_i > \theta_2$ and $\theta_1 \geq \bar{\theta}$.
Note that the resource is given wholly to one agent.

It turns out that

- Agents are incentivized to truthful.

It turns out that

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

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

It turns out that

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

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

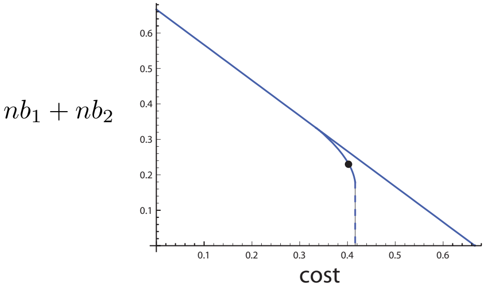
- The expected social welfare is decreasing in θ_0.

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

$$E\left[\sum_{i=1}^{2} \theta_i u(x_i) - p(\theta_i)\right]$$

  subject to covering cost c.

Second-best versus first-best



Expected social welfare as a function of c, compared to the first-best value.

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.

(a)

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

This makes $p_1(\theta_1, \theta_2) + p_2(\theta_1, \theta_2) = c$. We call this 'ex-post' cost-covering.

Other mechanisms can also work.

(a)

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

This makes $p_1(\theta_1, \theta_2) + p_2(\theta_1, \theta_2) = c$. We call this 'ex-post' cost-covering.

(b)

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

This makes incentive compatibility and rationality ex-post.

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

The resource is shared differently.

The optimal policy is found by solving a Lagrangian dual problem

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

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

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

Note that λ is a tuning parameter. As λ increases the fee structure changes, so that greater cost can be covered. The social welfare decreases, but is maximal 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 $x_i(\theta_1, \theta_2) = \theta_i^2/(\theta_1^2 + \theta_2^2)$.

The role of operating policy

- The resource is not allocated in the 'most efficient' way. That would be $x_i(\theta_1, \theta_2) = \theta_i^2/(\theta_1^2 + \theta_2^2)$.

This is one of our most important lessons:

To optimally incentivize participation in shared infrastructures, and make the most of the resources available, one should appreciate that both (i) fee structure, **and** (ii) operating methods, must both 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 allocation policies need to take account of need to give right 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 allocation policies need to take account of need to give right 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 allocation policies need to take account of need to give right 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!!!