An introduction to the economics of resource allocation and provision: the case of grid computing and planetlab

Panayotis (Panos) Antoniadis

(PhD, Athens University of Economics and Business)

Network and Performances Analysis Group, LIP6

September 2006

Why economics?

- Economics are relevant in cases where
  - Resources are scarce and/or
  - Resource provision is costly
  - Need to provide incentives for "correct" behavior
- Many such resources in networking
  - bandwidth, CPU, memory, content,
  - … and combinations (grid, file sharing, etc)
- Resource allocation questions
  - how to share bandwidth under congestion? who should have priority for using the PlanetLab infrastructure one week before the Infocom deadline?
- Resource provision (p2p systems)
  - why share files in Kazaa? why allow people to use my WLAN?

Yes but …

- The problem is very difficult and final solutions depend on many parameters (often not included in the models)
- A large body of work on incentive mechanisms for the Internet (e.g. bandwidth, BGP routing, p2p file sharing, grid computing) but very few applications of the theory produced
- Even simple models difficult to solve and corresponding mechanisms difficult to implement
- Obstacles
  - User acceptance
  - Complexity
  - Information
  - Enforcement (e.g. accounting)

The case of the Grid

- Originally a term used for distributed computations (SETI@home)
  - but then extended to include .. almost everything (DataGrid, SensorGrid, …)
- Key is sharing and aggregation of unused resources
  - Take advantage of the bursty nature of demand (e.g. scientific experiments), multiplexing
- High interest on using market mechanisms for commercial exploitation or efficiency
  - Depending on the point of view the majority of economic approaches are applicable in this context
  - Markets, bartering economies, public goods
- Important special characteristics
  - High value, unpredictability of demand, technology restrictions (e.g. middleware, parallelization), independent rational agents (need incentives!), complementarities between resources

The goal of this presentation

- Introduction to some basic concepts from economics and game theory for designing incentive mechanism in network applications
  - Economic modelling, markets, auctions, social welfare maximization

- Literature review of the incentive mechanisms proposed for grid computing and planetlab

- Discuss the relation with the OneLab Project

Outline

- The economics of resource allocation
- *The case of grid computing and planetlab*

- The economics of resource provision
  - focus on public goods
- *The case of grid computing and planetlab*

- Discussion: the case of federation (OneLab)

Fundamental concepts of (micro-) economics

- Consumers and producers .. or peers
- Utility and cost .. modelling, abstraction
- Social welfare
  - total utility minus total cost
  - other definitions as well (e.g. maximize min utility)
- Incentive mechanisms
  - Reward desirable behavior (e.g. contribution of resource)
  - Punish undesirable behavior (e.g. over consumption)
  - Prices, rules, reciprocity (e.g. parking, taxi tariff, …)
- Rationality
  - Agents maximize their net benefit (utility/profit minus cost)
- Central planner
  - Defines the goals and designs incentive mechanisms
  - Tries to enforce them (explicitly or implicitly)

Game theory?

- Games: models of interactive decision making
  - Players are given a set of possible strategies
  - Final outcome for each one depend on the choices made by others
  - strategic form: a player chooses his plan of action once and for all covering all possible contingencies
  - extensive form: explicit description of sequential structure of the decision problems
  - one-shot, repeated games

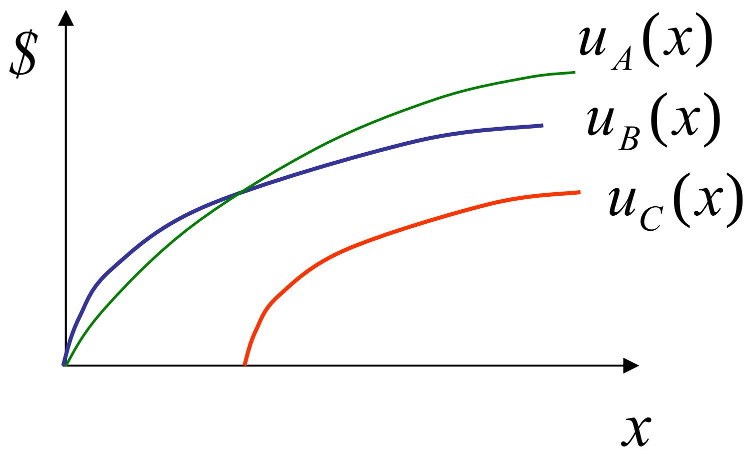
- Game theory: study the "result" (equilibrium) of different games under different strategies employed
  - more general than micro-economics
  - different solution concepts
    - E.g. nash equilibrium (no player wants to change his strategy given the others' strategies)

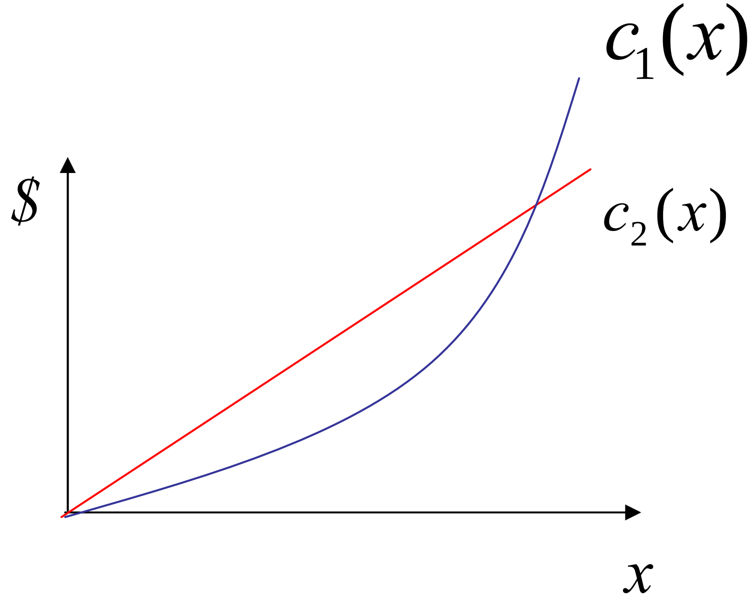
The utility and cost function

- Consumers are characterized by the utility function $u(x)$
  - translate into "monetary units" the benefit (or satisfaction)
  - Often concave

- Producers are characterized by their cost function $c(x)$
  - expresses the cost for producing resource x
  - Often convex

$$u_A(10) = 5, u_B(10) = 2$$





Incentive mechanisms

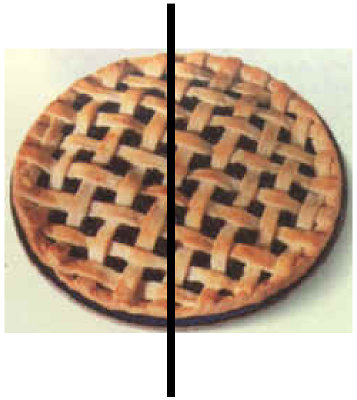
- Prices
  - Markets (free, regulated, open, close)
  - Auctions (first-price, second-price, sealed bid, …)
- Rules
  - Bartering
  - Fixed contributions
- Also less “strict” ones
  - Priority to well-behaved users
  - Reputation … or even social incentives

What is the goal?

- Maximize social welfare
  - Requires information about the utility and cost functions
- Fairness
  - No distinctions among different users
  - Different definitions possible in some cases
    - E.g. Max-min, proportional in the case of a network
- *Profit maximization*
  - When there is no central planner ☺

Resource allocation: share a pie

User A



User B

$u_A(x) = ?$

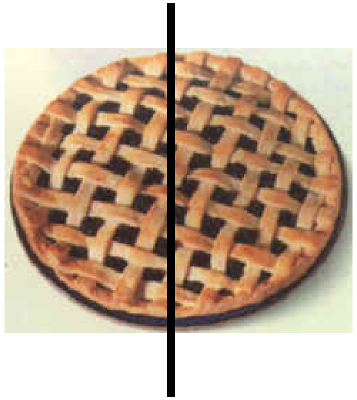
$u_B(x) = ?$

Fairness: give 0.5 to each

x={0, 0.25, 0.5, 0.75, 1}

Resource allocation: share a pie

User A



User B

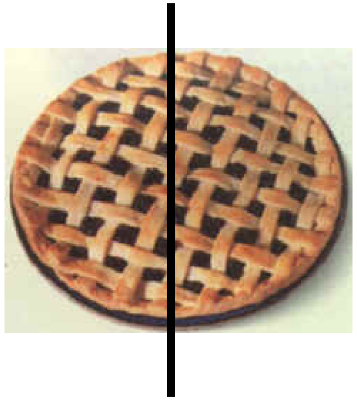
$u_A(x) = u_B(x)$

Fairness: give 0.5 to each

If users are identical welfare is also maximized

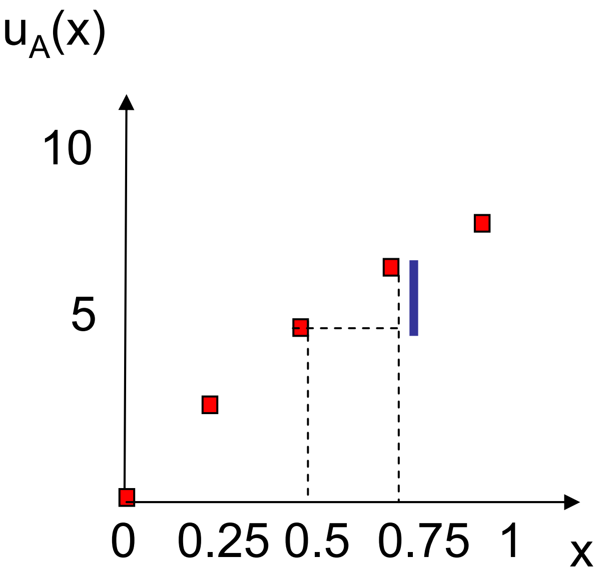
Resource allocation: share a pie

User A



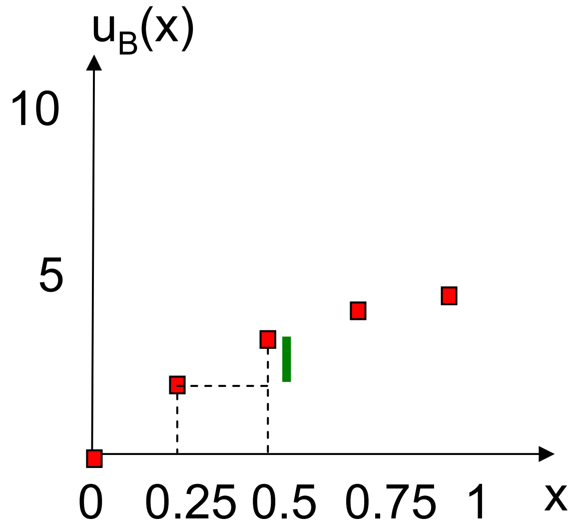
User B

Fairness: give 0.5 to each



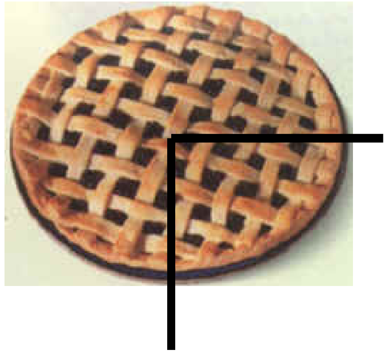
But otherwise …

inefficiency

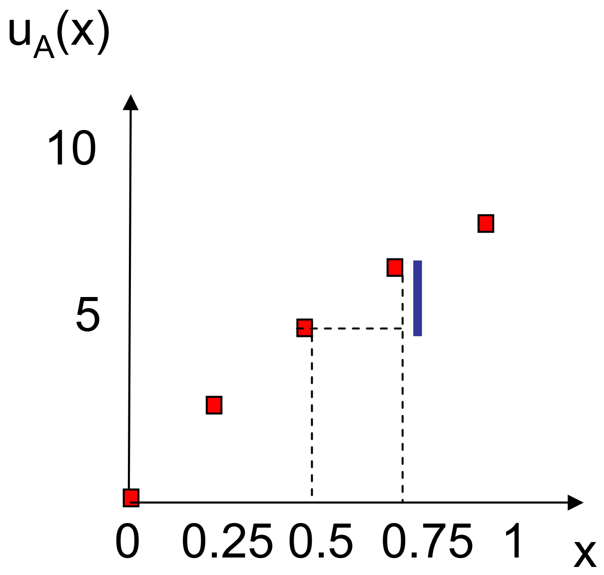


Resource allocation: share a pie

User A

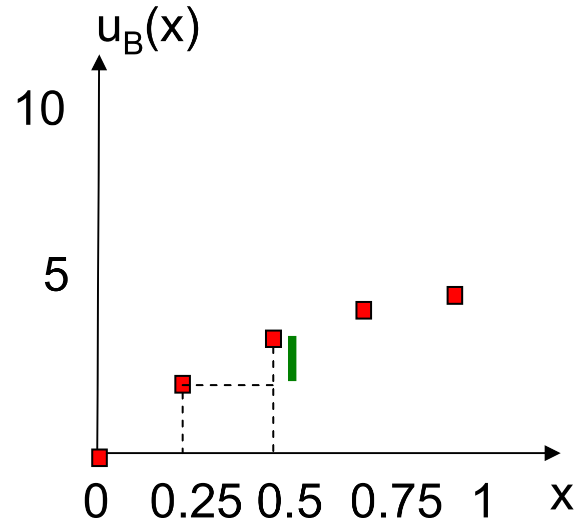


User B



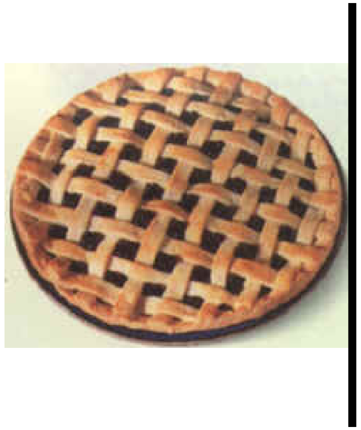
SW maximization:

$x_A = 0.75$, $x_B = 0.25$

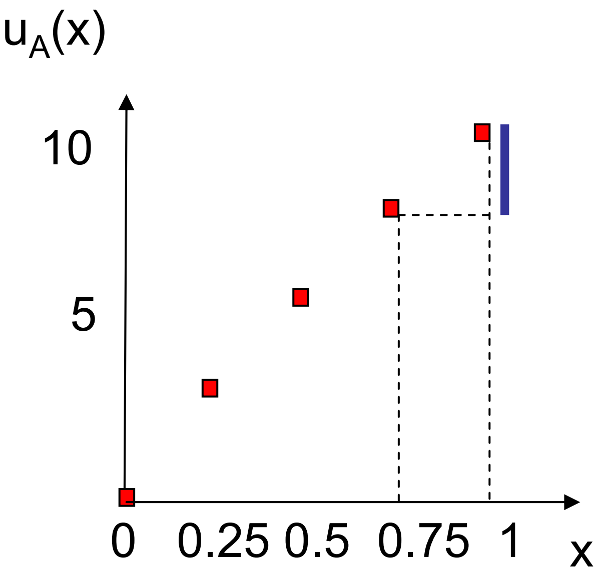


Resource allocation: share a pie

User A

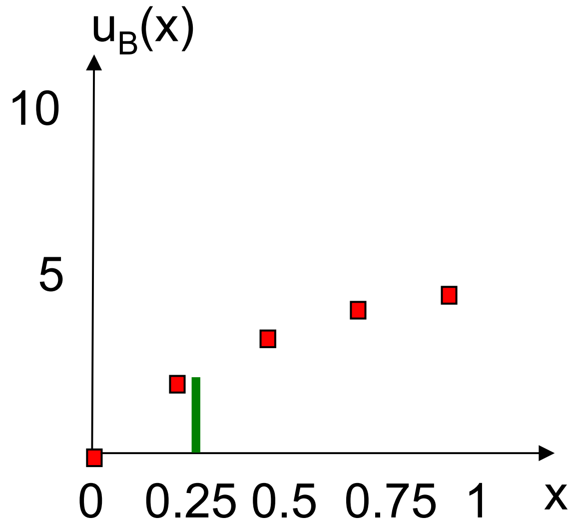


User B



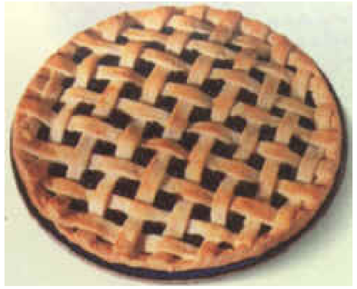
SW maximization:
$x_A = 1$, $x_B = 0$

Could be unfair!



Resource allocation: share a pie

User A



User B

$u_A(x) = ?$

$u_B(x) = ?$

.. and needs information

A simple economic model

- n consumers, 1 resource of total capacity C
- Allocation vector $x = \{x_i\} = \{x_1, \ldots, x_n\}$

$$\sum x_i \leq C$$

- Utility function of user i, $u_i(x_i)$
- Social welfare = $\sum_{i=1}^{n} u_i(x_i)$
- Economic efficiency: maximize social welfare subject to the capacity constraint

Mathematical solution

- Maximization problem:

$$\max_{\{x_i\}} \sum_i u_i(x_i) \quad s.t. \quad \sum_i x_i \le C \qquad (1)$$

- We must maximize the Lagrangian

$$\max_{\{x_i\}} L(\lambda, \{x_i\}) = \sum_i u_i(x_i) - \lambda(\sum_i x_i - C)$$

The optimal point of (1) is characterized by $\lambda, \{x_i\}$ for which:

$$\sum_i x_i = C, \; \frac{\partial u_i}{\partial x_i} = \lambda$$

($n+1$ equations with $n+1$ unknown variables)

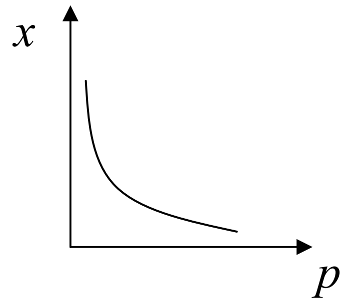
Complete information

- Implement the solution of (1)
  Set $x_i = x_i^*$, for all i (difficult in practice)
- Set a suitable price p (users should pay p for each unit of the resource)
  - Given this price user i will solve:

$$\max_{x_i}\left[u_i(x_i) - p\ \ x_i\right] \longrightarrow \frac{\partial u_i}{\partial x_i} = p$$

  - $x_i^*(p)$ = demand function

$$u(x) = \log x,\ \ x(p) = 1/p$$



  - $p^* = \lambda$

Incomplete information

- User i has clearly the incentive to misreport his utility function
  - In our example to declare *higher* one

- Two main approaches to elicit private information
  - Market mechanisms
  - Mechanism design

- Otherwise … fairness

Market mechanisms (tatonnement)

1. Planner sets price p^t, users post their demands $x_i^t(p^t)$
2. Planner computes excess demand $z^t = \sum_i x_i^t - C$
3. Planner updates price: $p^{t+1} = p^t + \alpha z^t, \alpha < 1$

Under general conditions, $p^t \to \lambda$

where λ is the Lagrange multiplier in (1)

Observe:

- The system planner does not need to know the utilities of the users

Mechanism Design

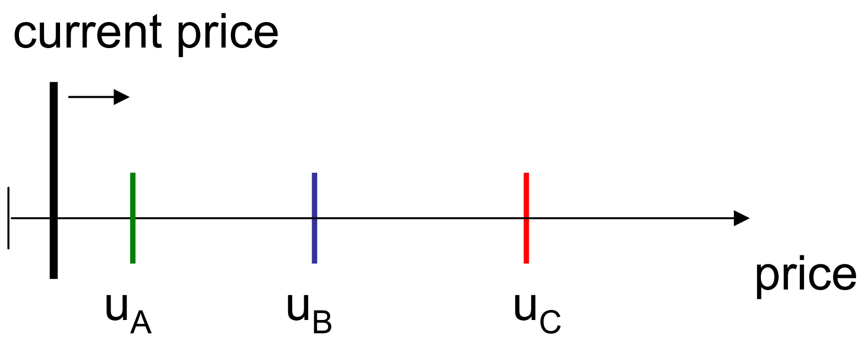
- Mechanism:
  1. The system planner **decides on an allocation mechanism**
  2. The users, knowing the mechanism, **report their valuations** wishing to maximize their net benefit
  3. The system planner executes the mechanism and **announces the allocation** and the prices paid by each user
- Inverse game theory
  - Design the game that would lead to the desired equilibrium
  - Auctions, variants of the VCG mechanism
- Many applications in networking
  - spectrum allocation, multicast cost sharing, BGP routing
- "Strategyproof" when users have the incentive to declare their true valuations

Auctions

- Consumers compete (bid) for resources
  - One item, multiple items, combinatorial, divisible vs. indivisible,..
- Highest bids win
  - (the item or a larger share of the resources)
- Single step vs. repeated, ascending vs. descending, open vs. sealed-bid
- Several approaches for price determination
  - First-price, second-price, …
- English, Dutch, Vickrey, double, reverse auctions
- Very popular for selling paintings (English), allocating spectrum (many different approaches), stock market (double auctions) …
- Undesirable strategies: bid shading, collusion

The English auction

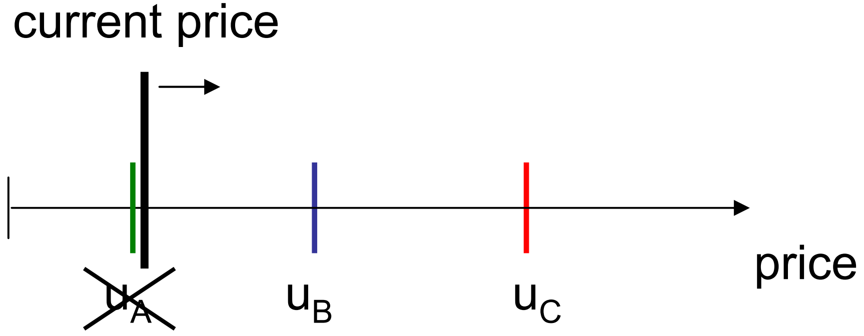
- One item, repeated, open bid, first-price
- The most popular auction for selling paintings, etc.
- The winning bid is increased until only one user wishes to pay for it
- Example: three users A,B: u_A = 3, u_B=7, u_B=10



The English auction

- One item, repeated, open bid, first-price
- The most popular auction for selling paintings, etc.
- The winning bid is increased until only one user wishes to pay for it
- Example: three users A,B,C: u_A = 3, u_B=7, u_C=10

User A leaves

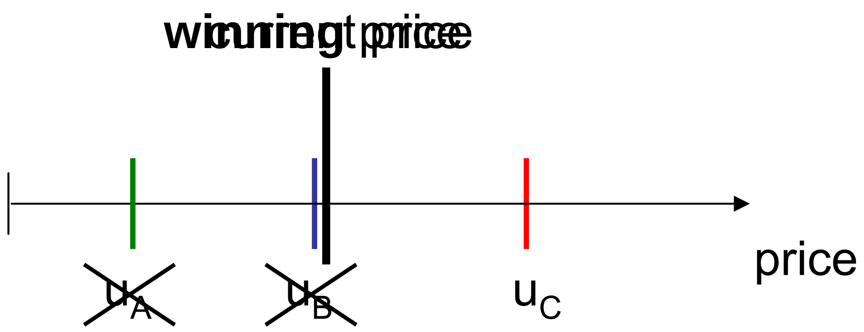


The English auction

- One item, repeated, open bid, first-price
- The most popular auction for selling paintings, etc.
- The winning bid is increased until only one user wishes to pay for it
- Example: three users A,B,C: u_A = 3, u_B=7, u_C=10

User B leaves

User C wins the item for price = 7 (+ ε)

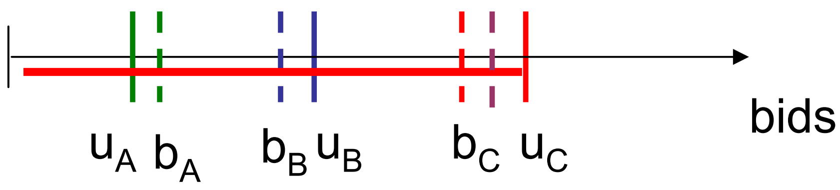


The Vickrey auction

- Single step, sealed bid, second-price

price = second highest bid

Could loose the item

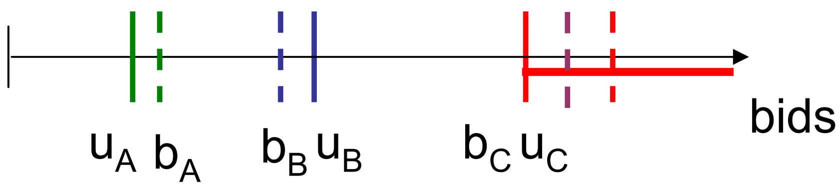


The Vickrey auction

- Single step, sealed bid, second-price
- Same properties as the English Auction
  - incentive compatible, maximizes social welfare
  - .. but in a single step
  - However, profit could be very small (see spectrum auctions in New Zealand) … “The lovely but lonely Vickrey Auction” [Crampton et. al]

price = second highest bid

Could pay more than u_C



Divisible goods: auction-based approach

- Users bid and the resource is allocated proportionally

$$x_i = \frac{b_i}{\sum_{j=1}^{n} b_j} C$$

- Then users can change their bids and so on .. (similar to tatonnement)
  - Under certain assumptions leads to an equilibrium (and maximizes welfare)
  - See work of Frank Kelly et. al

The VCG mechanism

- A generalization of the **V**ickrey Auction with the work of **C**larke and **G**rooves for public good problems
- Each user i pays the corresponding opportunity cost:
  - The loss of value of the rest users due to his existence

$$p_i = \sum_{j \neq i} u_j(x_{-i}^*) - \sum_{j \neq i} u_j(x^*)$$

- Price does NOT depend on the bid of the corresponding user → incentive compatibility
- .. but no budget balance (costs might not be covered)
- And susceptible to coalitions

How can we choose the suitable approach?

- Demand
  - Elasticity, predictability, burstiness, relation to supply
- Strategy space
  - E.g. ability to collude
  - behavior over time
- Type and number of resources involved
  - Single item, multiple items, divisible or not, …
  - The case of non-rivalrous resources
- Implementation
  - Required overhead, complexity, enforcement
- Policy
  - What is the goal?

Grid computing as a resource allocation problem

- Abstraction of a scheduler (or more) sharing a CPU to incoming jobs

Some older approaches:

- [Spawn] [Huberman et. al]
  - reservations for time slices of specific hosts
  - sealed-bid, second-price auctions on each host
- [Popcorn] [Regev and Nisan]
  - Intermediary (the market) to match buyers and sellers of CPU time (java operations)
  - Double auctions (sealed-bid and repeated)

(Double auctions)

- Single step or repeated
- Used in stock markets
- Main idea:
  - Buyers bid with highest buying bid
  - Sellers bid with lowest selling bid
  - These bids could be updated (in rounds or automatically) until an equilibrium is reached or not
  - When two prices match we have a trade
    - At the corresponding price
    - Or at a global price set by the planner (e.g. the smallest matched price or the highest unmatched price, etc.)

… and more recent work

- [Tycoon] [Lai, Huberman, and Fine, HP Labs]
  - Resources are allocated via continuous bids (first-price)
  - Not obvious the efficiency achieved (depends on user strategies)
- [Lai, Feldman, and Zhang]
  - Proportional share based on bids (iterative scheme)
  - Study of equilibrium when budget is fixed and different assumptions about parallelization
- [Bellagio/SHARE/ICE/EGG] [Chun, Parkes et. al]
  - Focus on combinations (important for planetlab!)
  - Different bidding languages and combinatorial auctions proposed
  - Approximations for winner determination
  - Currency management!

(Combinatorial auctions)

- Multiple items selled
- Users bid for different combinations of these items
- Example
  - items (a,b,c,d)
  - b = {5, (a,b)} (pay 5 for any bundle that contains a,b)
  - b = {5, (a,b)} XOR {10, (c,d)}
- Bidding language: expressiveness vs. complexity (e.g. winner determination)
- NP-complete to find optimal allocation even for single bundle bids

E.g. Bellagio/SHARE

- Already deployed in planetlab (?)
- Second-price price determination rule (SHARE)
- Virtual currency (budget distribution based on contribution)
- auctions cleared every hour
- XOR language
- Reservation period less than 64 hours

Resource provision

Reminder (resource allocation) …

- Maximization problem:

$$\max_{\{x_i\}} \sum_i u_i(x_i) \quad s.t. \quad \sum_i x_i \le C \qquad (1)$$

- We must maximize the Lagrangian

$$\max_{\{x_i\}} L(\lambda, \{x_i\}) = \sum_i u_i(x_i) - \lambda(\sum_i x_i - C) \qquad (1)$$

The optimal point of (1) is characterized by $\lambda, \{x_i\}$ for which:

$$\sum_i x_i = C, \quad \frac{\partial u_i}{\partial x_i} = \lambda$$

($n+1$ equations, n unknown variables)

Resource provision – how big pie

- Prices at the equilibrium can play the role of "signals" for increase or decrease of the provisioned capacity of the good

If $V(C) = \max_{\{x_i\}} \sum_i u_i(x_i) \quad s.t. \quad \sum_i x_i \leq C$

with Lagrange multiplier λ,

then $\dfrac{\partial V(C)}{\partial C} = \lambda$

So if the marginal cost of C increase is mc ,

$\lambda > mc \Rightarrow$ increase of C

$\lambda < mc \Rightarrow$ decrease of C

Recall: the value of λ equals the equilibrium price in the market

Distributed provision – A “free” market

A standard result from economics:

- Many producers with different (convex) cost functions
- Every participant in the market is small, can not affect prices
- Equilibrium: stable point where production = demand, price p

- Under competition amount of resources produced is the optimal and welfare is maximized
  - without any information required (but requires tatonnement)
  - Price = marginal cost = marginal utility (marginal cost prices)

- But important limitations
  - tatonnement is difficult to implement
  - strategic behavior is possible when market small
  - dependence on the form of the utility and cost functions

Exchange economy

- Exchange economy of k commodities (k time slots) .. the case of grid computing.
- Each agent i has an initial endowment $w^i = (w_1,\ldots,w_k)$ of computational power for each of the k slots (it could be that they are equal for all slots)
- And utility $u_i(x_1^i,\ldots,x_k^i)$ where x_j^i is the amount of resources consumed in slot j.
- According to demand, (relative) prices for the resources at different slots will be formed which would lead to social welfare maximization (under competition)
- Limitations
  - Not appropriate in cases where demand of specific slots is not predictable in advance (the most common case in this context)
  - Does not address the issue of combinations of resources
  - Nor the issue of dimensioning (investment) …

Less formal approaches

- Ideally, we would like to formulate a model that captures all three decisions for **resource investment, provision, and consumption of a single agent** and provides insights for the suitable mechanism that would maximize the overall efficiency … not available!
- Many existing approaches propose the used of "closed" markets (using virtual currencies) of different types
  - In order to consume players should contribute!
  - But no obvious way to choose the market that would lead them to optimal choices
- Important issues
  - Currency management (Inflation, deflation, distribution, re-distribution)
  - Wide space of undesirable strategies (e.g. "saving")
  - The role of demand

Contribution rules

- Simplest approach for incentivizing resource provision: all participating users should contribute a fixed amount of resources
  - Very simple to conceive
  - Already used in real systems (e.g. planetlab)
  - Captures externalities (public good aspect)
  - Optimal efficiency under some assumptions (see next)
- But
  - does not provide a solution for efficient resource allocation under contention
  - Enforcement is not always easy (requires monitoring)

Public goods

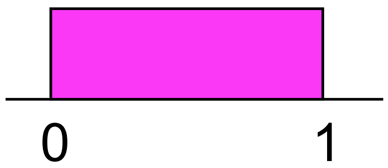
- Non-rivalrous resources
  - E.g. street light, content
- Extreme form: non-excludable
  - E.g. national defense
- Incentive problem in provisioning: the free-rider problem
  - User i prefers the other player to contribute
  - Free market fails to provision optimum amount of public goods

A non-excludable public good

- n agents bargain to provision a public good
- Q = quantity of public good, all agents enjoy it
- $c(Q)$ = cost of public good, agent i pays p_i

$$\theta_i u(Q) - p_i$$ = agent's i net benefit

- θ_i iid, has distribution F



- Examples:

$$u(Q) = Q^{1/2}, \quad c(Q) = Q^2$$

$$Q \in \{0,1\}, \quad u(Q) = Q, \quad c(Q) = cQ$$

Example

Build a bridge

$$Q \in \{0,1\}, \quad u(Q) = Q, \quad c(Q) = \underset{\leq 1}{a}\, Q, \quad \theta_i \text{ iid uniform on } [0,1]$$

- First-best policy:

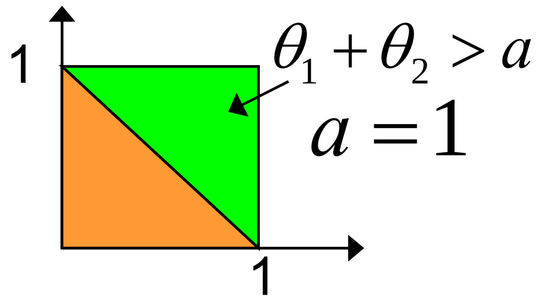
$$\max_{Q(.)} \sum \theta_i u(Q(\theta)) - c(Q(\theta))$$

- Solution (n=2):

$$Q(\theta) = 1 \text{ if } \theta_1 + \theta_2 > a, \text{ use any } p_1 \leq \theta_1, p_2 \leq \theta_2, \text{s.t. } p_1 + p_2 = a$$

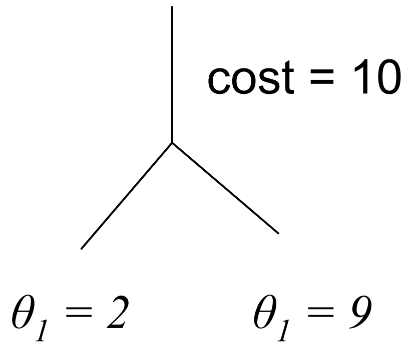
$$Q(\theta) = 0 \text{ if } \theta_1 + \theta_2 \leq a$$

$$p_i(\theta) = \frac{\theta_i}{\theta_i + \theta_j} a$$



Example (2)

- Why should agents declare their actual θs ?
- If $p_i(\theta) = \dfrac{\theta_i}{\theta_i + \theta_j} a$ then agent with highest θ_i gains by declaring less -> SW loss
- Fairness could also lead to inefficiency (SW loss)
  - E.g. $\alpha = 10,\ \theta_1 = 2,\ \theta_2 = 9$
  - User 1 refuses to pay 5 units and user 2 cannot pay 10 .. *so* $Q=0$ even if $\theta_1 + \theta_2 > \alpha$
  - The multicast cost sharing problem



The VCG mechanism

$$p_i = \sum_{j \neq i} \left[\theta_j u_j(Q^*_{-i}) - c(Q^*_{-i}) \right] - \sum_{j \neq i} \left[\theta_j u_j(Q^*) - c(Q^*) \right]$$

If

$\sum_{j \neq i} \theta_j < a \quad (Q^*_{-i} = 0)$ ⟶ = 0

$\sum \theta_j > a \quad (Q^* = 1)$

(only when agent i is "necessary" for provisioning the good …

$$= \sum_{j \neq i} \theta_j - a \Rightarrow p_i = a - \sum_{j \neq i} \theta_j$$

… he pays the "necessary" amount for covering the cost)

It is easy to see that $\sum p_i \leq a$

no cost recovery

Second best

- The best allocation policy given that
  - User information is not visible (we require incentive compatibility)
  - Provision is private (we require cost recovery)

- Impossibility Theorem (Myerson-Satterhwaite (1983))
  - Second Best (SB) < First Best (FB)

The role of exclusions

- Optimal incentive policies are impractical to evaluate in most situations
  - Need for good approximations
- Existing results for specific models suggest that as $n \to \infty$

$$\frac{SB}{FB} \to 0$$

- If **exclusions** are possible, then

$$\frac{SB}{FB} \to \alpha > 0$$

  - Incentive payments converge to fixed contributions
  - we can obtain a general theorem when the system is large!

A limit theorem
[Courcoubetis and Weber, 2005]

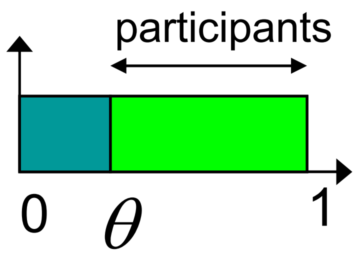
Suppose

- $u(Q) = AQ^{\alpha}$, and $c(n,Q) = Bh(n)Q^{\beta}, \beta \geq 1 > \alpha$
- and Q^*, θ^* maximize

$$P = \max_{\theta \in [0,1], Q \geq 0} nu(Q) \int_{\theta}^{1} yf(y)dy - c(n,Q)$$

subject to

$$\overbrace{n[1 - F(\theta)]}^{\text{\# of participants}} \overbrace{\theta u(Q)}^{\text{fixed fee}} - c(n,Q) \geq 0$$



Then the simple mechanism $\pi_i(\theta) = 1\{\theta_i \geq \theta^*\}$, $Q(\theta) = Q^*$,

$p_i(\theta) = \theta^* u(Q^*)$, achieves $P \leq SB \leq (1 + O(n^{-1}))P$

The role of size

- Why size helps?
  - in a large network it is hard to get people pay more than a minimum
- As the number of peers gets larger
  - a peer feels that his own declaration will have a negligible effect on the final system size
  - hence his strongest incentive is to only reduce his payment
  - therefore he declares the minimum possible theta which corresponds to the minimum fixed fee by agreeing to participate.

Applications

- File Sharing
  - public good = content availability
  - that is, number of total distinct files shared
- P2P WLANs
  - peers share wireless access to the internet
  - public good = coverage
- Grid computing
  - Public good = total grid capacity
- Planetlab
  - Public good = the size of the network (?)

My PhD thesis (by the way ☺)

- "Economic modelling and incentive mechanisms for efficient resource provision in p2p systems"
  - Focus on file sharing (many differences with grid computing)
- Main contributions
  - Modeled resource provision in p2p file sharing as a public good problem
  - Demonstrated the nice properties of fixed contribution rules
  - Studied the case of group formation
  - Proposed a memory-less enforcement mechanism focusing on peer availability
  - Formulated a novel economic model for the proposed enforcement mechanism which
    - Provided the means to evaluate it in terms of economic efficiency achieved and compare it with simple reciprocity rules (1 upload for 1 download) studied extensively in the literature

A Scientific Grid

- Many small organizations/universities that have sporadic but very demanding jobs to execute
  - Assume the main resource of interest is CPU time
- The value comes from completing them in some reasonable time .. no strict deadlines
- High degree of parallelization
- Constant connectivity and strong identities
- Dedicated resources
- Scheduling
- Problem: economically efficient resource provision of resources

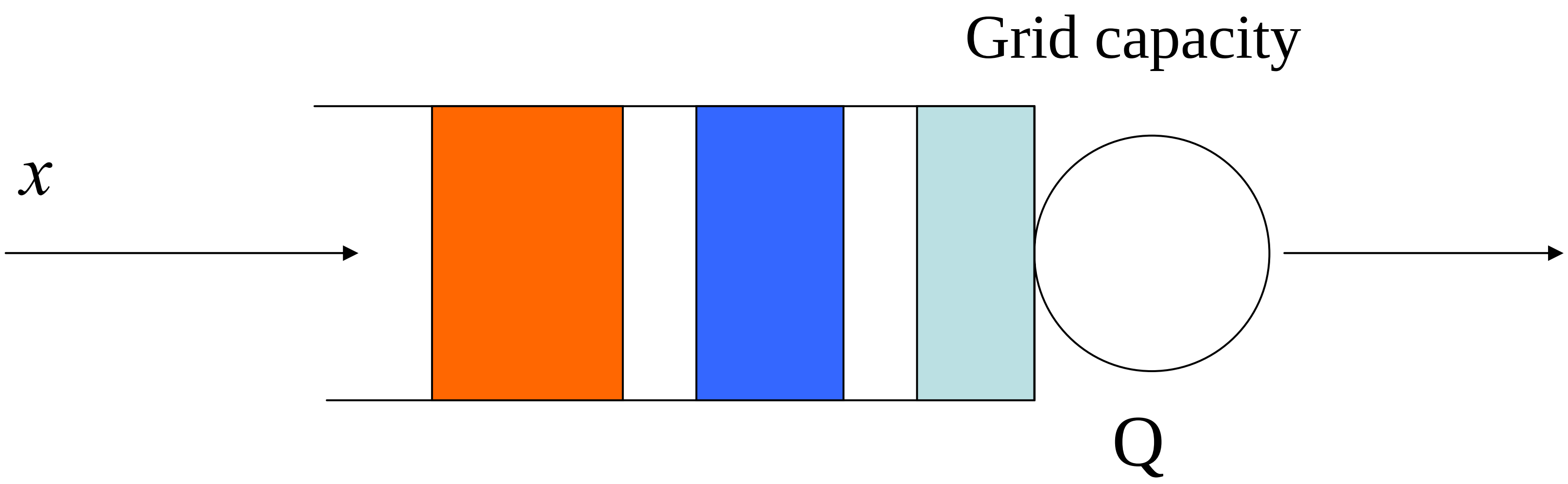
The public good perspective

- CPU and storage space are rivalrous .. but they are also renewable
- When there is no congestion, they could be treated as non-rivalrous in time
- Model a scientific grid as a server with capacity Q built by the clients themselves
  - The more is Q the less will be the average delay for everyone
- Focus on resource investment (instead of service provision)

A mathematical abstraction of a Scientific Grid

In scientific grids the main utility generator is delay

And delay depends somehow on total grid capacity built by the individual contributions/investments



Planetlab

- Different ways to see it
  - Large testbed
  - Monitoring the Internet (telescope)
  - A deployment platform
- More general than grid computing
  - grid computing could be an application
- Combinations are very important
  - No focus on aggregated computational power or bandwidth … but on heterogeneity of resources (realistic environment)
- Strong public good aspect
  - Small players are important .. but incentives also
- But contention cannot be disregarded
  - Infocom deadline problem ☺

Federation

- Positive externalities
  - Resources
  - Services
  - Multiplexing
  - Network externalities
- More diversity, realistic environment
- Incentives?
  - Intra- vs. inter-

First thoughts

- Key issue assess the “value” generated by a planetlab community (in order to compare and define peering agreements)
  - Size?
  - Amount of resources?
  - Heterogeneity?
  - Static or according to market dynamics (e.g. demand)?
- What would be the form of agreements?
  - Currency exchange rates?
  - Inter-system consumption rules? SLAs?
- Hidden information?
- Accounting?
- Policy issues?
- What we have learned from ISP peering/transit agreements?
  - Very complicated problem ☺

Summary

- We reviewed some main economic principles used (or proposed) for the allocation of resources in the Internet focusing on the case of grid computing
  - Efficiency vs. fairness
  - Markets and mechanism design (auctions)
  - Incentive compatibility
  - Public goods and fixed contribution
- Issues not covered
  - Game theoretical aspects
  - Enforcement
  - QoS, SLAs
  - Policy issues