Incentives for Large P2P Systems

Costas Courcoubetis
Athens University of Economics and Business

joint work with
Richard Weber
and Panos Antoniadis

Outline

- Motivation and modeling
- Public goods and Mechanism Design
- Equal contribution schemes
- Applications in file sharing and WLANs
- Contribute while consuming
- Conclusions

The free riding issue in p2p systems

- Peers exploit their resources to provide services for the benefit of entire community
  - Positive and (sometimes) negative externalities
- Free riding is the rational strategy
- Suitable incentive mechanisms could increase efficiency
  - Different applications have different requirements
- Two major challenges:
  - Economic modeling (optimization)
  - Enforcement of incentive mechanisms

Important characteristics of p2p systems

- Public good aspect
- Complicated cost modeling
- Heterogeneity
- Size
- Highly dynamic environment
- Cheap pseudonyms
- Unpredictable quality of service and hidden action
- Centralized vs. distributed implementation
- Rationality vs. altruism

Enforcement issues

- Incentive mechanisms require some sort of accounting of peers' past transactions.
- A very challenging problem when the system designer cannot rely on
  - trusted software
  - ability to monitor transactions
    - false trading
  - persistent identities
    - whitewashing
    - the "sybil attack"
  - central authority to store and certify accounting information
- The majority of research on p2p economics focuses on ways to enforce simple reciprocity rules in terms of actual downloads/uploads under the above restrictions

Economic modelling

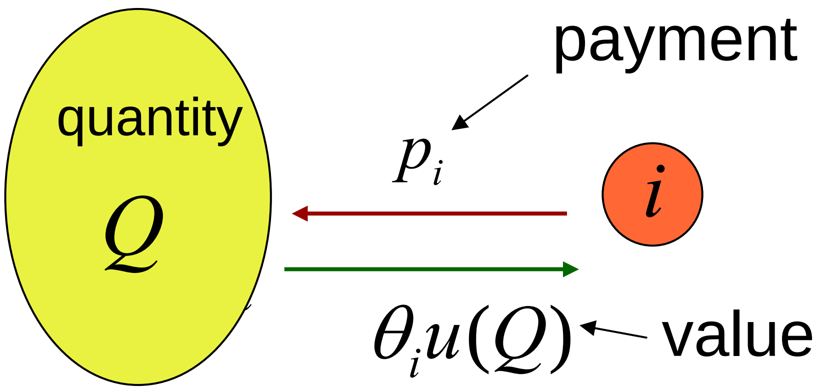
- Utility
  - Value of service provisioning
  - Size of the system (externalities)
- Cost
  - Initial contribution (e.g. disk space, content)
  - Peer availability (e.g. amount of time on-line)
  - Service provision (e.g. number of uploads)
- Is service provisioning costly?
  - Operating system and/or bandwidth manager may minimize certain aspects of cost of service provisioning
  - But when resources are congested (e.g. bandwidth) or consumable (e.g. battery) service provision becomes dominant (as in the case of ad-hoc networks)

A non-excludable public good model

n agents participate in the provision of a public good

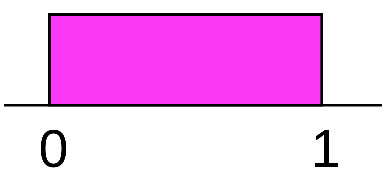
$c(Q)$

= cost



θ_i : private information of agent *i*, distribution F

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



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

Maximizing Social Welfare

$$\max_{Q} \; E \left(\sum_i \theta_i\right) u(Q) - c(Q)$$

- Difficulty: need to raise payments to cover cost!
- Free market: inefficient due to externalities
- Need regulation: design participation rules to get payments (like in an auction)
- Problem: optimal design requires information on user types
  - Full info: personalized rule for each peer
    - "first-best" policy
  - Incomplete info: Mechanism Design

Mechanism Design

- Design participation rules to encourage users to act in an incentive-compatible way, maximizing social welfare

->"Second-Best" policy

A. Post system policy rules:

  - System size: $Q(\theta_1, \ldots, \theta_n)$
  - Participation: $\pi_i(\theta_1, \ldots, \theta_n)$
  - Payment: $p_i(\theta_1, \ldots, \theta_n)$

B. Agents respond by declaring their θ_is,….

- Well-developed economic theory; but solutions are typically very complex, dependent on fine details, require large amounts of info to be passed to centre

Example 1

$$u(Q) = \frac{2}{3} Q^{1/2},\ c(Q) = Q,\ \theta_i \text{ iid uniform on } [0,1]$$

"First Best policy": $\max_Q \quad u(Q) \sum_1^n \theta_i - c(Q)$

$$E\, SW_{FB}^* = n^2 / 36 + n / 108$$

But we do not know the θ_is !!

"Second Best policy":

With exclusions: $SW_e^{SB} = \frac{3}{128} n^2 + \frac{7}{148} n + O(1)$

Without exclusions: $SW_{ne}^{SB} = a n^{3/2} + O(n^{1/2})$

Example 1 (cont.)

Other simpler policies? **equal contributions?**

Policy A: post fixed fee ϕ and system size Q

Participation condition: $\theta_i u(Q) \geq \phi \quad SW_A = 3n^2/128$

Policy B: post system size Q, share cost equally $c(Q)/X$

Participation: $\theta_i u(Q) - c(Q)E[1/X] \geq 0$

$SW_B = 3n^2/128 - n/384 + O(1)$

Policy C: post fee ϕ, build maximum facility from participant fees

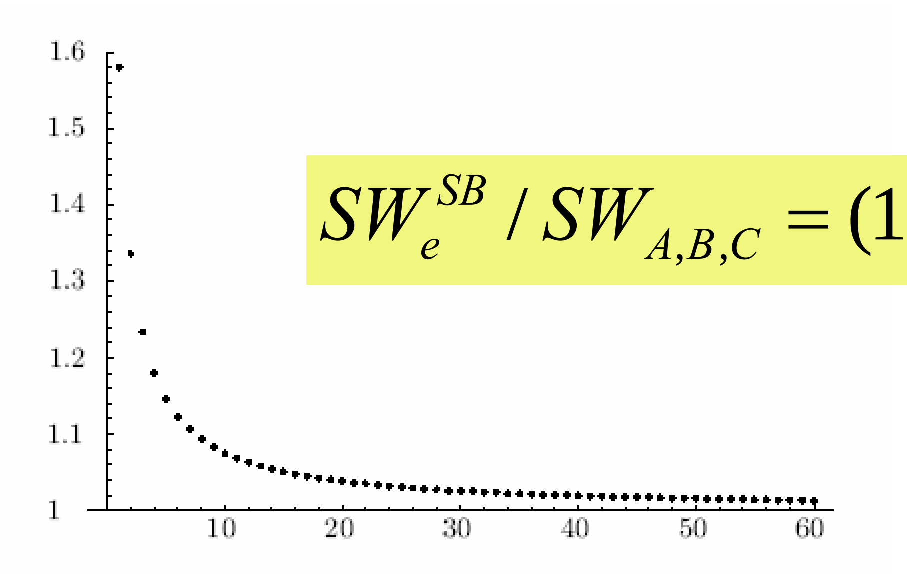
Participation: $\theta_i Eu(X\phi) - \phi \geq 0$

$SW_C = 3n^2/128 + 7n/1536 + O(1)$

Policy D: post $Q(X)$, charge $c(Q(X))/X$

Example 1 (cont.)

Conclusions:



$$SW_e^{SB} / SW_{A,B,C} = (1 + O(1/n))$$

Nice if it holds in general!!

Theorem?: A very simple mechanism
"contribute f if join, 0 otherwise"
is nearly optimal when the network is large

Mechanism Design and Public Goods

- An example
- Some basic theory
- Our approximation Theorem

Example 2: Building a bridge

0,1 0,c

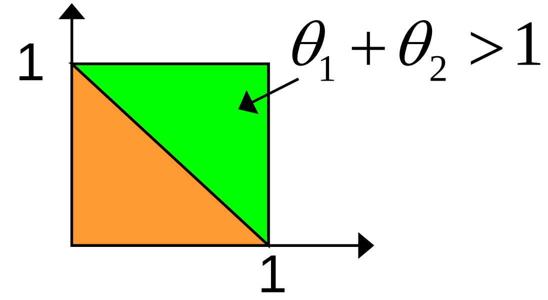
$$Q \in \{0,1\}, \quad u(Q) = Q, \quad c(Q) = cQ, \quad \theta_i \text{ iid uniform on } [0,1]$$

- First-best policy: $\max_Q \left(\sum \theta_i\right) u(Q) - c(Q)$
- Solution (n=2, c=1):

$$Q(\theta) = 0 \text{ if } \theta_1 + \theta_2 \le 1$$

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

$$p_1(\theta) = \frac{\theta_1}{\theta_1 + \theta_2}$$



Example 2 (cont.)

- Why should agents declare their actual θs ?
- If $p_i(\theta) = \frac{\theta_i}{\theta_i + \theta_j} c$ then agent with highest θ_i gains by declaring less -> SW loss
- Which is the best allocation policy ("**second-best**") ?
- Impossibility Theorem (Myerson-Satterhwaite (1983))
  - Second Best (SB) < First Best (FB)

Allocations in Second-Best policies

For each $\theta = (\theta_1, \dots, \theta_n)$

What quantity $Q(.)$

Who participates $\pi_1(.), \dots, \pi_n(.) = 0{,}1$

What payment $p_1(.), \dots, p_n(.)$

- Feasible: $c(Q(\theta)) \leq \sum p_i(\theta)$
- incentive compatible: $E_{\theta_{-i}}[\theta_i u(Q(\theta_i, \theta_{-i}) - p_i(\theta_i, \theta_{-i})]$
$$\geq E_{\theta_{-i}}[\theta_i u(Q(\hat{\theta}_i, \theta_{-i}) - p_i(\hat{\theta}_i, \theta_{-i})]$$
- Individually rational: $E_{\theta_{-i}}[\theta_i u(Q(\theta_i, \theta_{-i}) - p_i(\theta_i, \theta_{-i})] \geq 0, \quad \forall \theta_i$

The optimization problem

- Problem:

$$\max_{\substack{Q(.),\\ \pi_1(.),\dots,\pi_n(.),\\ p_1(.),\dots,p_n(.)}} E[\sum \theta_i \pi_i(\theta) u(Q(\theta)) - c(Q(\theta))]$$

subject to

- feasibility $c(Q(\theta)) \leq E[\sum p_i(\theta)]$
- individual rationality $E_{\theta_{-i}}[\theta_i u(Q(\theta_i, \theta_{-i}) - p_i(\theta_i, \theta_{-i})] \geq 0, \quad \forall \theta_i$
- incentive compatibility ….

Using standard MD theory

Solve:

Maximize total value – cost

such that

Maximum Incentive payment $\geq$ cost

MD theory allows us to write this as function of $Q(.)$ and $\{\pi_i(.)\}$

Using standard MD theory

- Solve

$$\max_{Q(.),\{\pi_i(.)\}} E_\theta[\sum_1^n \theta_i \pi_i(\theta) u(Q(\theta)) - c(Q(\theta))]$$

- subject to

$$E_\theta[\sum_1^n \pi_i(\theta)(\theta_i - \frac{1-F(\theta_i)}{f(\theta_i)})u(Q(\theta)) - c(Q(\theta))] \geq 0$$

- Use Lagrangian methods -> λ

Solution

To find $Q(.), \pi_i(.)$: pointwise maximization of

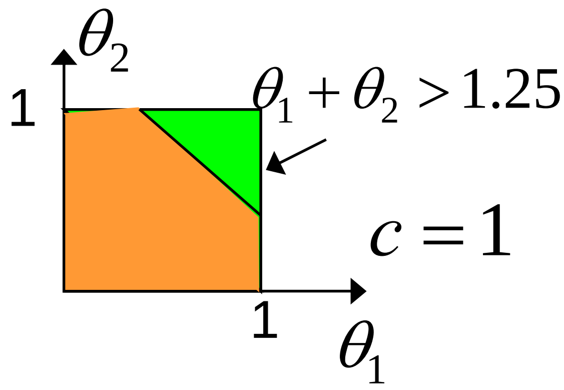
$$\max_{Q,\{\pi_i\}} \frac{\sum_1^n \pi_i(\theta)(\theta_i + \lambda g(\theta_i))}{1+\lambda} u(Q) - c(Q)$$

$$\pi_i(\theta_i) = 0 \text{ if } \theta_i + \lambda g(\theta_i) < 0 \text{ i.e., } \theta_i < \overline{\theta}(\lambda)$$

$$Q(\theta) = \arg\max_Q \frac{\sum_{\theta_i > \overline{\theta}} (\theta_i + \lambda g(\theta_i))}{1+\lambda} u(Q) - c(Q)$$

Back to the bridge construction

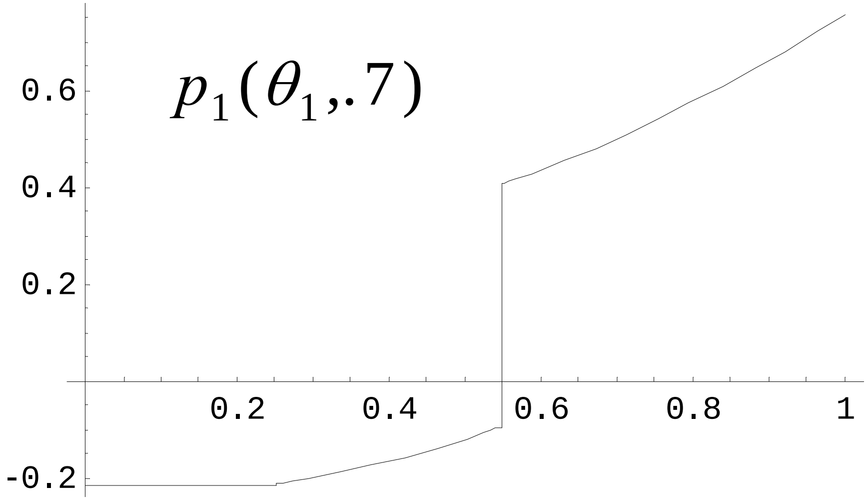
$$Q(\theta) = 1 \text{ iff } \sum_{1}^{n} \theta_i - \frac{\lambda}{1+\lambda} \sum_{1}^{n} \frac{1 - F(\theta_i)}{f(\theta_i)} \geq c$$

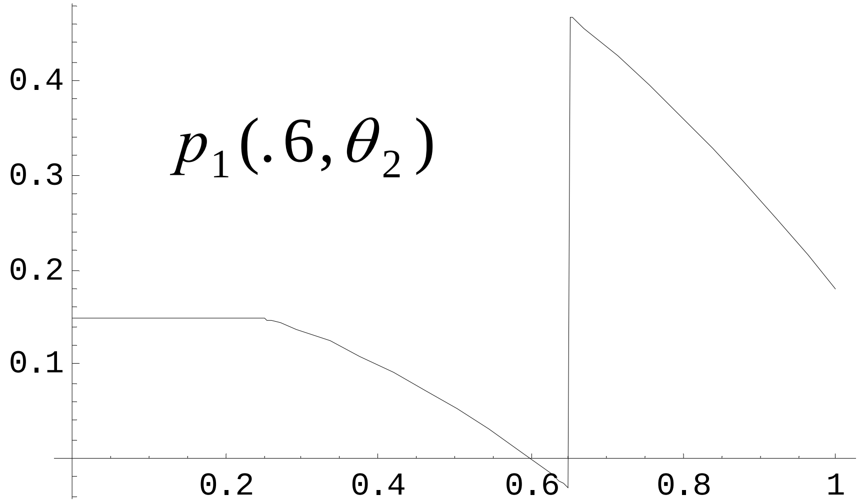


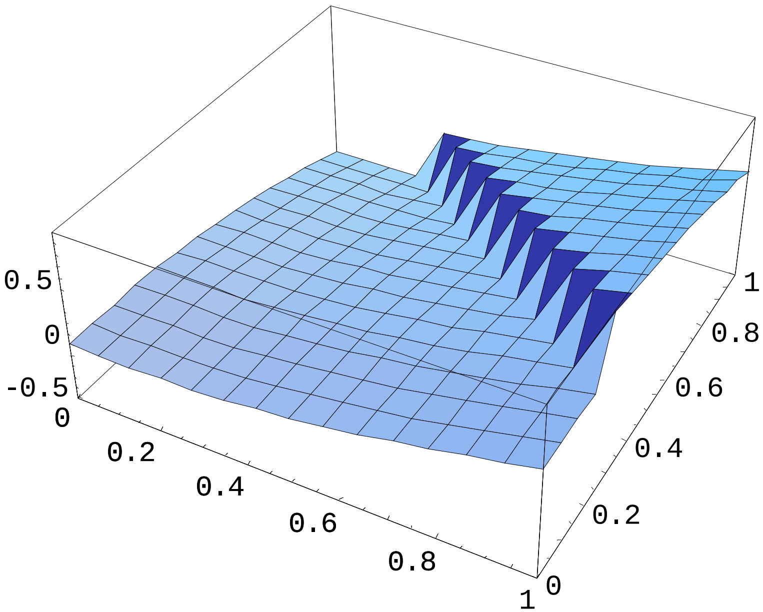
$c = 1$

- Incentive payments: extremely complex!
  - functions of complete vector
  - involve money transfers between agents
  - no known simple approximation

The payments for n=2







$$Q = 1 \text{ if } \theta_1 + \theta_2 \geq 1.25$$

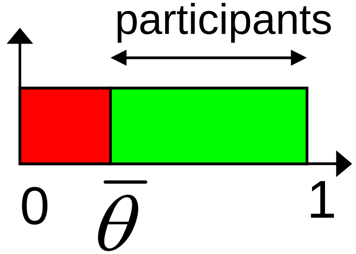
A limit theorem

Suppose

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

and $Q, \bar{\theta}$ maximize

$$P^* = \max_{\bar{\theta} \in [0,1], Q \geq 0} E \sum_{\theta_i \geq \bar{\theta}} \theta_i u(Q) - c(n,Q)$$



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

Then the equal contribution mechanism $\phi = \bar{\theta} u(Q)$

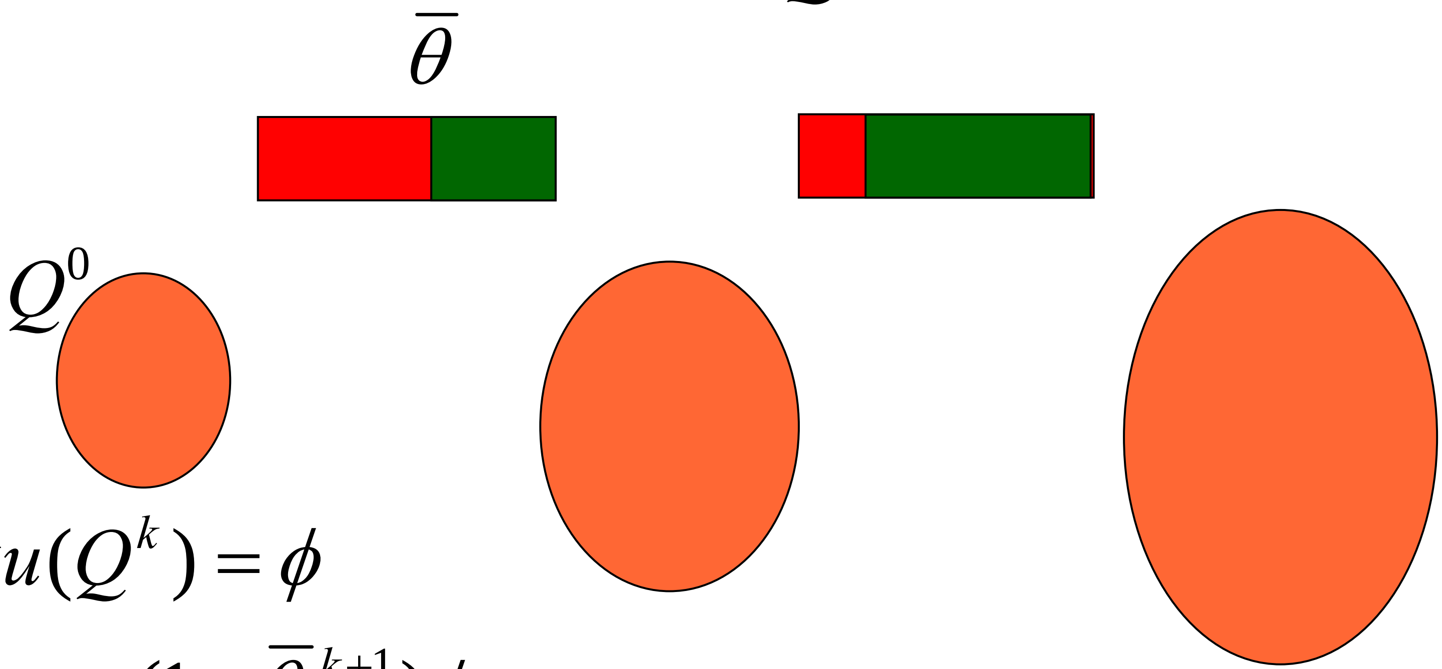
achieves $P^* \leq SB \leq (1 + O(n^{-1}))P^*$

Why large systems are simpler

- 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.

Stability

- Assume contribution ϕ is fixed, initial condition Q^0
- Where will the system converge?
- Under what conditions will Q^* be reached?



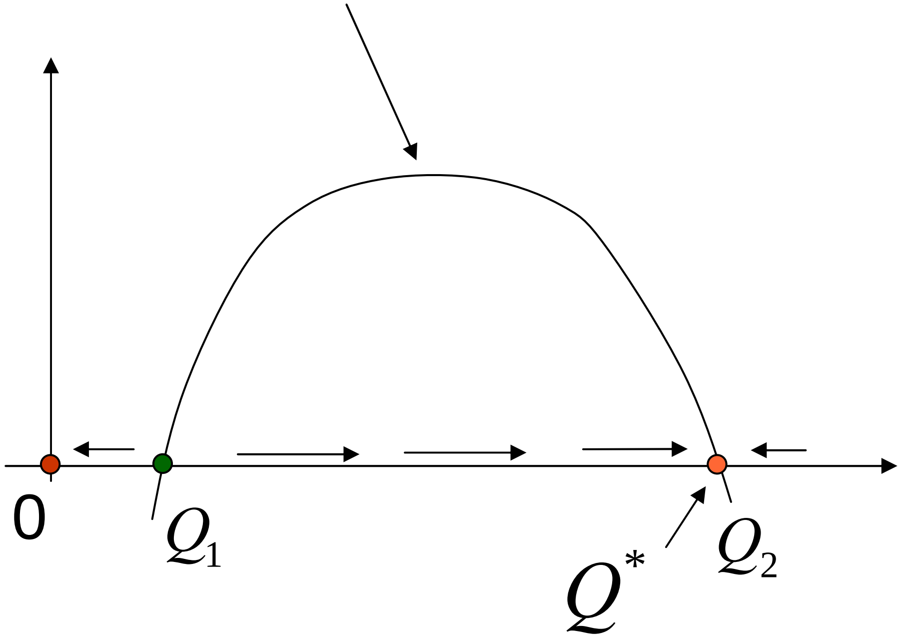
$$\begin{cases} \bar{\theta}^{k+1} u(Q^k) = \phi \\ Q^{k+1} = n(1-\bar{\theta}^{k+1})\phi \end{cases}$$

Stability

- Assume contribution ϕ is fixed, initial condition Q^0
- Where will the system converge?
- Under what conditions will Q^* be reached?

Let Q_1, Q_2 = roots of $Q - n(1-\phi/u(Q))\phi = 0$

Reach $\mathrm{Q}^* = Q_2$ if $Q^0 \geq Q_1$



Applications

- **File sharing**
  - Q = content availability
  - p_i = number of files shared per unity of time
  - Not necessarily copyright infringement (but requires a global indexing –e.g. earth coordinates)

- **WLAN sharing**
  - Q = coverage: probability to obtain roaming service at a random point
  - p_i = rate of roaming requests served by peer i

File sharing

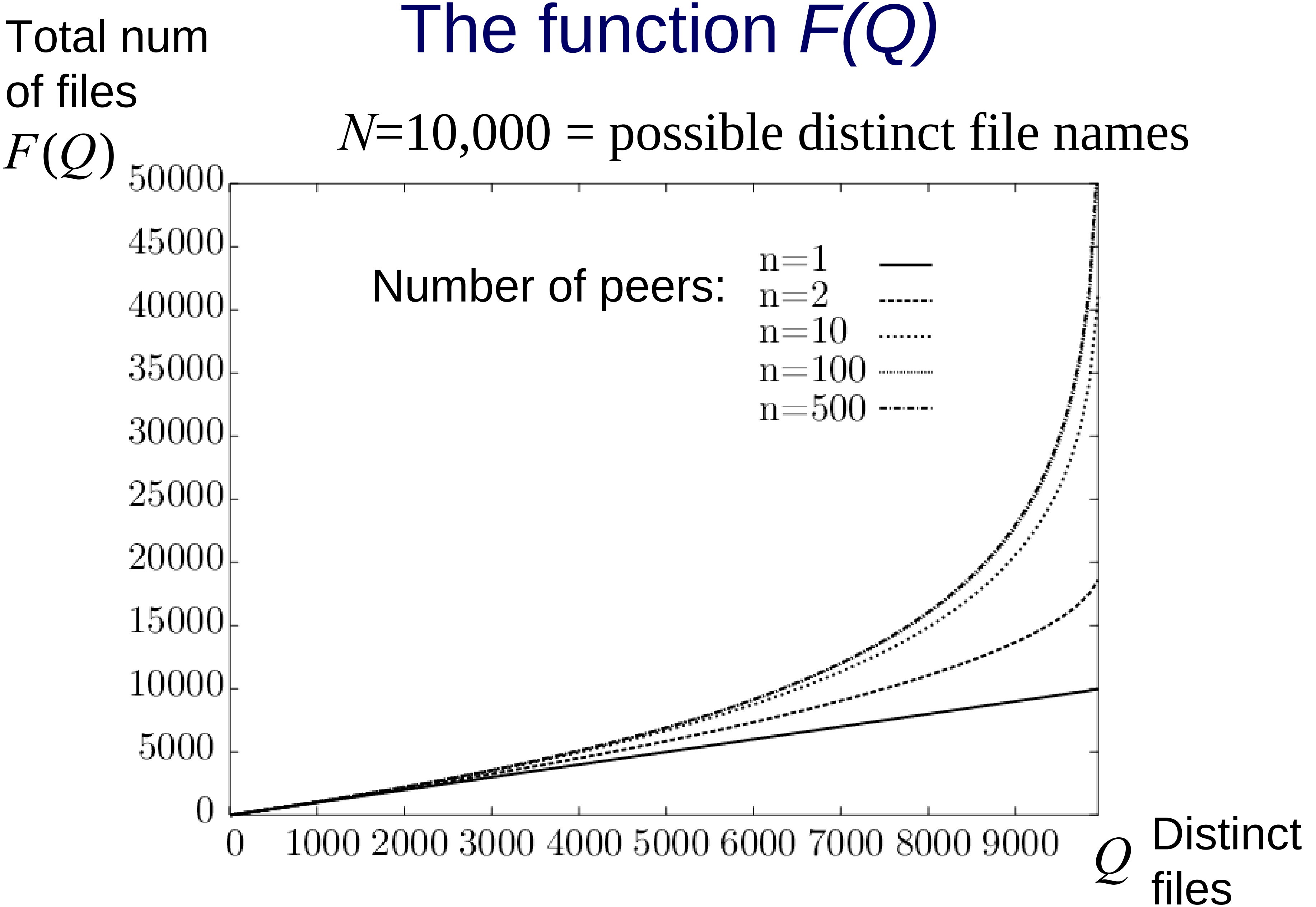
- Q: expected number of **distinct** files
- peer *i* :
  - utility = $\theta_i u(Q)$,
  - cost = f_i = number of files provided to the system
  - f_i randomly chosen from *N* possible file names
- Rewrite equations in terms of F = total # of files

$$Q(F) \approx N(1 - e^{-F/N}), \text{ where } F = \sum f_i$$

- Compute optimum fixed contributions as before

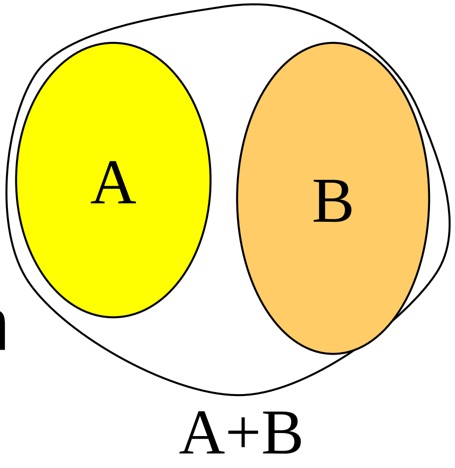
The function *F(Q)*

N=10,000 = possible distinct file names



Group formation (1/3)

- Assume peers of different sub-types
- Type A: $\theta_i^A \sim [0,0.5]$ (e.g. ISDN users)
- Type B: $\theta_i^B \sim [0.5,1]$ (e.g. DSL users)
- Do they gain more by
  - forming 2 distinct groups vs forming a larger group?
  - being distinguished by the system in the larger group?



Group formation (2/3)

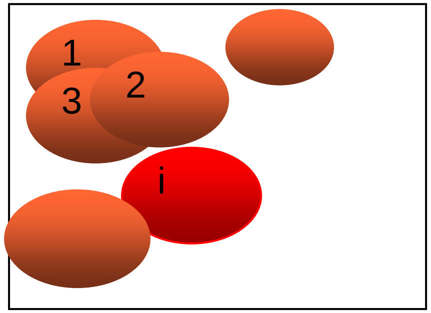
- Group A: $\theta_i^A \sim [0,0.5]$ (e.g. ISDN users)
- Group B: $\theta_i^B \sim [0.5,1]$ (e.g. DSL users)

Assume that the percentage of each group in the mix is 50% (n=1000)

Welfare	Group A	Group B	Total
Distinct groups	3296	35156	38452
Indistinguishable	6976 (+ 111%)	44792 (+ 27%)	51768
Distinguishable	31249 (+ 848%)	31250 (-11%)	62500

Peering WLANs

- Total utility $\sum_{1}^{n} \theta_i \pi_i(\theta) u(Q)$
- Total cost $aQ \sum_{1}^{n} \pi_i(\theta)$



- Total cost depends on final number of participants!
- Asymptotic problem:

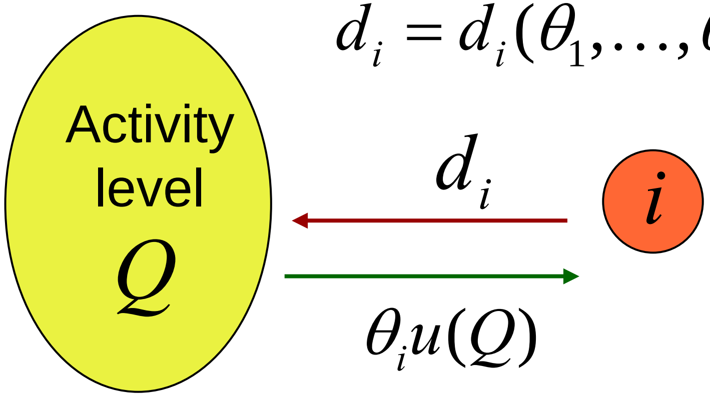
$$P^* = \max_{\bar{\theta} \in [0,1], Q \geq 0} nu(Q) \int_{\bar{\theta}}^{1} \theta dF(\theta) - n(1 - F(\bar{\theta}))c(Q)$$

$$P^* \leq SB \leq (1 + O(n^{-1/2}))P^*$$

Contribute while consuming

- Cost = time a peer stays ON
- A peer contributes to system activity while he is ON
  - Only time a peer allows uploads is when he is ON
- But peers would like to do their downloads and disconnect!
- System control: reduce download rate to force peers to stay ON for longer time
- Tradeoff: content availability vs. slow performance
- Can we use the same machinery to analyze this??

The model with download time



$d_i = d_i(\theta_1, \dots, \theta_n) =$ download time for peer *i*
= prob the peer is ON

A peer downloads a file with rate $x_i = 1$ or $x_i = 0$

- The value of the file depends on system activity $Q = \sum d_i x_i$
- Maximize social welfare: second-best policy (MD)
- Equal contribution mechanism: $d_i(\theta) = d$

$$\max_d E \sum_{i \in I} (\theta_i u(Q) - d)$$

$$\text{where } Q = d \sum_{i \in I} x_i \text{ and } I = \{i : \theta_i u(Q) - d > 0\}$$

Results

- Convergence towards $Q_d > 0$ for all d s.t. $d / u(0) < 1$
- For $u(Q) = aQ^{1/2}$, SW maximization results in

$$SW = 27/512n^2, \bar{\theta}_d = 1/4, d^* = 3a^2n/64$$

- Extensions:
  - for continuous choice of the x_i s
  - Increasing x_i may have negative effects

Conclusions

- Simple incentive model: externalities <-> public goods
- Equal contribution schemes perform well for large n
  - Easy to enforce
  - Simple computation of optimal fee
  - Contribution in kind
- But our approach = “first order term” in actual problem
- Reasonable extensions:
  - Adaptation
  - Dynamics
  - Robustness

Extra topics

Enforcement issues (reminder)

- Incentive mechanisms require some sort of accounting of peers' past transactions.
- A very challenging problem when the system designer cannot rely on
  - trusted software
  - ability to monitor transactions
    - false trading
  - persistent identities
    - whitewashing
    - the "sybil attack"
  - central authority to store and certify accounting information
- The majority of research on p2p economics focus on ways to enforce simple reciprocity rules in terms of actual downloads/uploads under the above restrictions

Real life applications

- Kazaa
  - Simple reputation mechanisms with priority under contention as an incentive
  - Enforced by the software -> failed!
- Direct Connect
  - Fixed contribution rules (focus on content availability)
  - Centralized monitoring and punishment (exclusion based on IP addresses)
- BitTorrent
  - Direct exchange of resources (i.e. upload bandwidth)
  - A very nice example of a memory-less mechanism
  - But doesn’t address the issue of content availability
  - Has attracted a lot of attention lately!

Proposed solutions in the literature

- Token based currencies
  - Require central or distributed bank to check for double spending
  - Self-minted currencies need reputation
- Public accounts
  - Require account holders, cryptography, substantial communication overhead
  - Additional incentive issues
- Reputation mechanisms
  - Suffer from whitewashing and false trading
  - Extensive research in this area. Basic concepts:
    - Treat newcomers badly (the social cost of cheap pseudonyms)
    - Don’t trust ratings of unknown peers

Our approach: contribute while consuming

- Recall that
  - Peer contribution: **number of files** per **unity of time**
  - We assume uploading is costless while downloading
  - Asymptotically optimal rule: fixed contribution (but difficult to enforce over time)
- Enforcing entity = service provider (i.e. uploader)
  - Ensures that the downloader **shares a predefined number of valid files** while offering service
  - Upload using a **fixed throughput** so as not to be completed too fast and thus increase peer availability
- Contribution of each peer will depend on her request rate and content availability
- Additional incentive issues arise (we discuss them later)

System Requirements

- Super peers
  - Realistic assumption (see Kazaa)
  - Run a distributed index for search
  - Act as seeds of content
  - Tune system parameters (e.g. fixed upload throughput)
- Protocols for broadcasting the value of the upload throughput and for checking validity of files
- Two types of attacks on validity
  - illegal file names (flush invalid files from the index using a service like CDDB)
  - corrupted files (check while downloading or before upload)

Example

$$u(Q) = 0.6Q^{1/2}, \quad c(Q) = Q, \quad \theta_i \text{ uniform in } [0,1]$$

$$\max_{\theta \in [0,1], Q \geq 0} \left(n \int_{\theta}^{1} y dy \right) 0.6Q^{1/2} - Q$$

$$s.t. \quad n[1-\theta]\theta \; 0.6Q^{1/2} - Q \geq 0$$

Solution: $\theta^* = 1/4, \quad Q^* = 0.0126n^2, \quad SW = 0.006328n^2$

- satisfaction of cost coverage constraint: reduction of SW by 43%