An Asymptotically Optimal Scheme for P2P File Sharing

Panayotis Antoniadis — Athens University of Economics and Business
Costas Courcoubetis — Athens University of Economics and Business
Richard Weber — Centre for Mathematical Sciences, Cambridge, UK

Contents

A model for p2p file sharing

- Resource allocation in p2p file sharing is a *public good* problem
  - all peers benefit from the contribution of any single peer
  - downloading a file by one peer does not prevent another peer from downloading the same file (no congestion effects)
  - but contribution is costly
  - positive externality creates an incentive to *free-ride* on efforts of others
  - a peer's incentive is to offer a few files in the common pool and requests lots of downloads from others

peer i:
$$\begin{cases} \text{benefit} = \theta_i u(Q), & \theta_i \text{ is iid with distribution } H \\ \text{cost} = f_i = \text{payment in "kind"} \end{cases}$$

n peers

Implications

- Implication: "free market" solution is inefficient
  - each peer maximizes own net benefit
  - actions affect others
  - hence private optimum differs from social optimum
- Need regulation: use prices or rules to influence behaviour
  - incentives for each peer reflect the effect it has on others
  - example of a rule: downloads = uploads
- Problem: optimal design requires information on user types
  - under full info: personalized price/rule for each peer
  - "first-best" policy
- Existing approaches based on heuristics
  - reciprocity based punishments/rewards

What to do?

- How can the system/planner/network manager get the required information to design optimal contribution rules?
  - if lucky, can gather personalized data about users
  - otherwise, users **must be given incentives** to reveal relevant information to planner
- Mechanism Design: set prices/rules to encourage users to act truthfully, maximize social welfare
  - Well-developed economic theory; but solutions typically
    - **very complex, dependent on fine details**
    - **require large amounts of info to be passed to centre**
    - **"second-best" policy**
- **Approximations?**

Large systems are simpler!

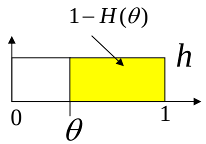
- Size helps!
  - simplifies mechanism, limits per capita efficiency loss
- Theorem: A very simple mechanism
  "contribute f if join, 0 otherwise"
  is nearly optimal when the network is large
- Why?
  - in a large network it is hard to get people pay more than a minimum
- Other major benefits:
  - Low informational requirements, easy to apply in a large class of examples

Some formulas for SW…

No contributions, system of size *Q*

$$SW = \left(n \underbrace{\int_0^1 yh(y)dy}_{E[\sum \theta_i]} \right) u(Q) - c(Q)$$

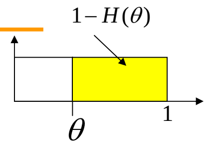
Fixed contributions covering cost, system of size *Q*



$$SW = \left(n \int_\theta^1 yh(y)dy \right) u(Q) - c(Q)$$

$$\underbrace{n[1-H(\theta)]}_{\text{expected number of participants}} \; \underbrace{\theta u(Q)}_{\text{fee}} - c(Q) \geq 0$$

Theorem



Let Q^*, θ^* maximize

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

$$\textit{subject to} \quad n[1-H(\theta)]\theta u(Q) - c(Q) \geq 0$$

Then, the policy:

each participating peer contributes $f = \theta^* u(Q^*)$

achieves $P \leq P_{SB} \leq P + o(n)$

P_{SB} = efficiency of second-best policy

Example

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

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

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

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

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

File sharing

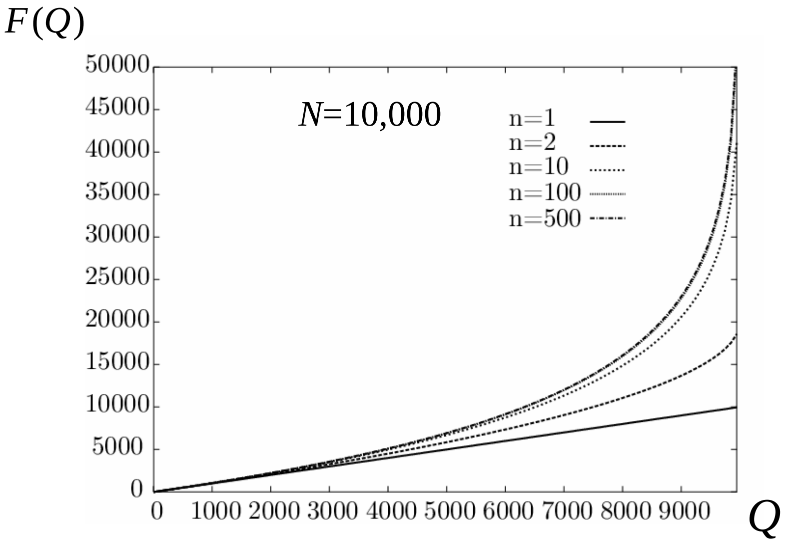
- Q : expected number of distinct files
- peer i : utility = $\theta_i u(Q)$, cost = f_i = number of files provided to the system
- f_i randomly chosen from N files

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

- Solve $\max_{\theta\in[0,1],F\geq 0}\left(n\int_\theta^1 yh(y))dy\right)u(Q(F))-F$

$$subject\ to\quad n[1-H(\theta)]\theta u(Q(F))-F\geq 0$$

The function *F(Q)*



Heterogeneous file popularity

- General case: $u(F_1,F_2), c(F_1,F_2)$

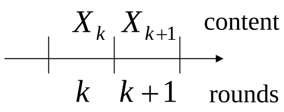
  F_1 : popular content

  F_2 : less popular content

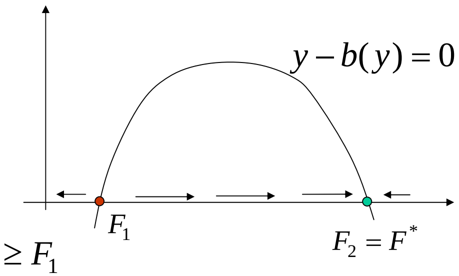
  - specify contributions f_1^*, f_2^*
- Interesting case: $u(aF_1+F_2),\ c(F_1,F_2)=bF_1+F_2$
- Then, provide both types only if $a=b$
- Optimum contribution is a scalar f^*
  - a peer can provide any combination $f_1, f_2,$ s.t. $af_1+f_2=f^*$
  - measuring rate of uploads is a good proxy

Stability

- Assume contribution f^* fixed
- Participation decision: based on prior expectation regarding total content availability *F*
- Will F^* be reached?



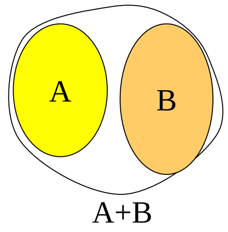
$$X_{k+1} = b(X_k, f^*)$$



stability if $X_0 \geq F_1$

Group formation (1/5)

- 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 (2/5)

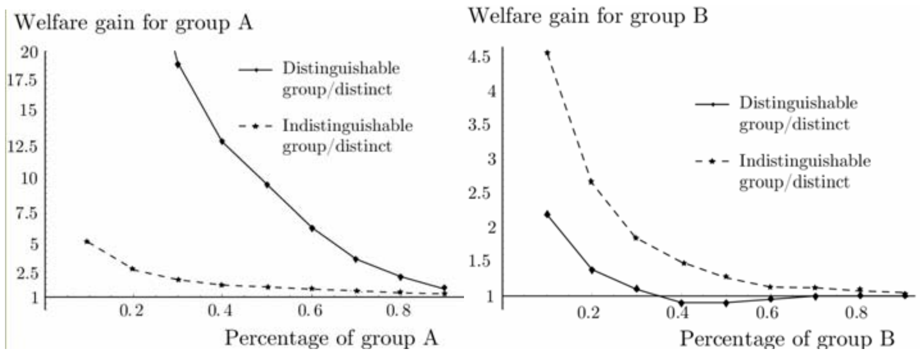
- Group A: $\theta_i^A \sim [0,0.5]$ (e.g. ISDN users)
- Group B: $\theta_i^B \sim [0.5,1]$ (e.g. DSL users)

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

Group formation (3/5)

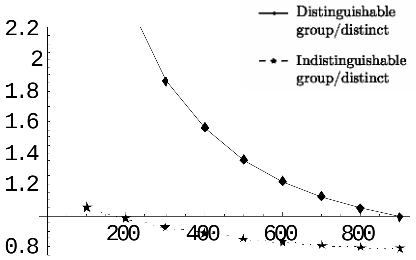
*As a function of the **percentage of each group** in the mix*



Group formation (4/5)

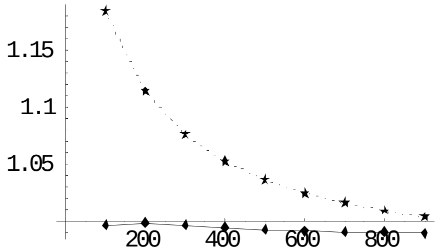
Adding a "congestion" cost: $c(F) = m^a F$, m = # of participants

Group A



N

Group B



N

Group formation (5/5)

- How to provide better incentives for both types to combine and reveal their types?
  - reduce cost of heavy users by limiting upload rates
  - reduce fees of heavy users
- Offer sets of tariffs (versioning)
  - allow self-selection
- Model difference in cost for uploading
  - higher-cost peers benefit in a larger group when types can be distinguished

Adaptation

- What if $H(\cdot)$ not known?
- In general incentive to shade declarations
- Repeated game formulation: in each round, peer i samples θ_i from H, declares θ_i
  - truth-telling equilibrium

Conclusions

- Fixed contribution schemes can be efficient
- Result to tractable optimization problems
- Provide useful insight to many interesting questions
- Information regarding user types may be strategic
- Open issues:
  - more complex cost structures
  - adaptation
  - multiple round games
  - practical application
- Check also ...
  - Market Management of P2P Systems (MMAPPS)
    - http://www.mmapps.org
  - AUEB Network Economics and Services Group
    - http://nes.aueb.gr/p2p.html