Incentive Schemes in Memory-less P2P Systems

Costas Courcoubetis

Athens University of Economics and Business

joint work with

Panos Antoniadis and Richard Weber

Workshop on Autonomic Communication

October 2005

Outline

- Important issues for addressing free riding in p2p systems
- Recent results from public good theory
- A memory-less incentive mechanism
- Economic modeling and evaluation

The free riding issue in p2p systems

- Peers exploit their unused resources to build a generic service for the benefit of everybody
  - 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
  - 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
- Centralized vs. distributed implementation
- Hidden action
- Rationality vs. altruism

Economic Modelling

- Contribution (and cost)
  - Initial contribution (e.g. hardware, bandwidth, content)
  - Peer availability (e.g. amount of time on-line)
  - Service provision (e.g. number of uploads)
- Utility
  - Amount of resources acquired through service provision
  - Size of the system (public good aspect)
- Different weights according to application/technology
  - Operating system and/or bandwidth manager could minimize the cost of service provision in certain cases and then initial contribution and availability become the dominant factor of contribution (e.g. grids, content sharing)
  - On the contrary, when resources are congested (e.g. bandwidth) or consumable (e.g. battery) service provision becomes dominant (as in the case of ad-hoc networks)

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 focus on ways to enforce simple reciprocity rules in terms of actual downloads/uploads under the above restrictions

We focus on content availability

- Definition: amount of files shared per unity of time
- Content availability is a public good
  - A file is not consumed by downloading but contributing files to the common pool is costly
- We assume
  - Service provision (i.e. uploading) has negligible cost
    - particularly true while consuming resources (see later)
  - There is no congestion
    - We focus on the ‘long tail’ of the content
    - Existing systems focus on bandwidth sharing for downloading popular items (e.g. Bittorrent)

Maximizing Social Welfare

- "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

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?**

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:

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

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

Allocations

- For each $\theta = (\theta_1, \ldots, \theta_n)$
  - what quantity $Q(.)$?
  - what contributions $p_1(.), \ldots, p_n(.)$?
- 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: $E_{\theta_{-i}}[\theta_i u(Q(\theta_i, \theta_{-i}) - p_i(\theta_i, \theta_{-i})] \geq 0, \quad \forall \theta_i$

Allocations (2)

- First-best: maximizes Social Welfare (SW) under complete information (is trivially feasible)

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

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

- Second-best: maximizes SW under incomplete information, i.e.,
  - subject to
    - feasibility
    - incentive compatibility
    - individual rationality

Large systems are simpler!

- Size helps!
  - simplifies mechanism, limits per capita efficiency loss
- Theorem: A very simple mechanism

  "contribute f if join, 0 otherwise"

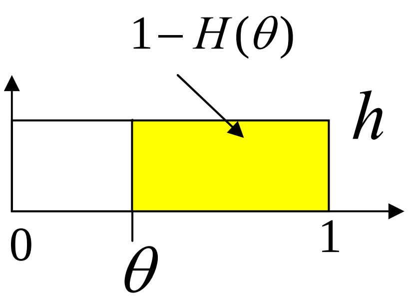
  is nearly optimal when the network is large
- Why?
  - in a large network it is hard to get people pay more than a minimum
- Other major benefits:
  - Low informational requirements, easy to apply in a large class of examples

Some formulas for SW…

No contributions, system of size *Q*

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

Fixed contributions covering cost, system of size *Q*

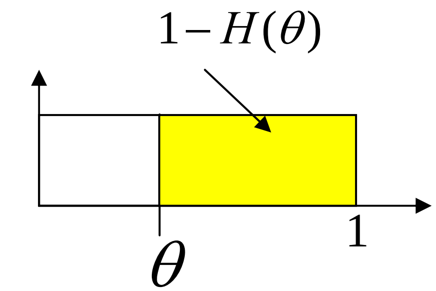


$$\text{fee} = \theta u(Q)$$

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

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

Theorem



Let Q^*, θ^* maximize

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

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

Then, the policy:

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

achieves $P_{SB} \leq P(1 + O(1/n))$

P_{SB} = efficiency of second-best policy

Example

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

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

Solution:

$$f = 0.0168n, Q^* = 0.0126n^2, \theta^* = 1/4, SW = 0.006328n^2$$

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

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
  - p_i = area covered by access points of peer i

# Other equal contribution schemes

- Fix Q, share cost among participants

$$f = c(Q)/m$$

- Fix fee f, build a facility according to how many decide to participate

$$c(Q) = mf$$

- Define $Q(m)$, charge $c(Q(m))/m$

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)

Economic modelling

- Average download time d (depends on fixed upload throughput)
- Peer chooses her type x_i between 0 and θ_i according to d and Q
- Request rate $g(x_i)$ depends on type
- Assume initial amount of content Q_0 (made available by super peers)
- Peer's cost depends on the percentage of time she stays on-line downloading (the cost for providing a fixed amount of files f is considered sunk)

A simple model

- Each peer chooses his type x_i given Q, d

$$\max_{x_i, x_i \le \theta_i} x_i u(Q_0 + Q) - ag(x_i)\frac{Q_0 + Q}{N} d$$

where $Q(x) : Q = \sum g(x_i)\frac{Q_0 + Q}{N} fd$

For $u(Q) = Q^{1/2}, g(x) = x^2, Q_0 = 0$

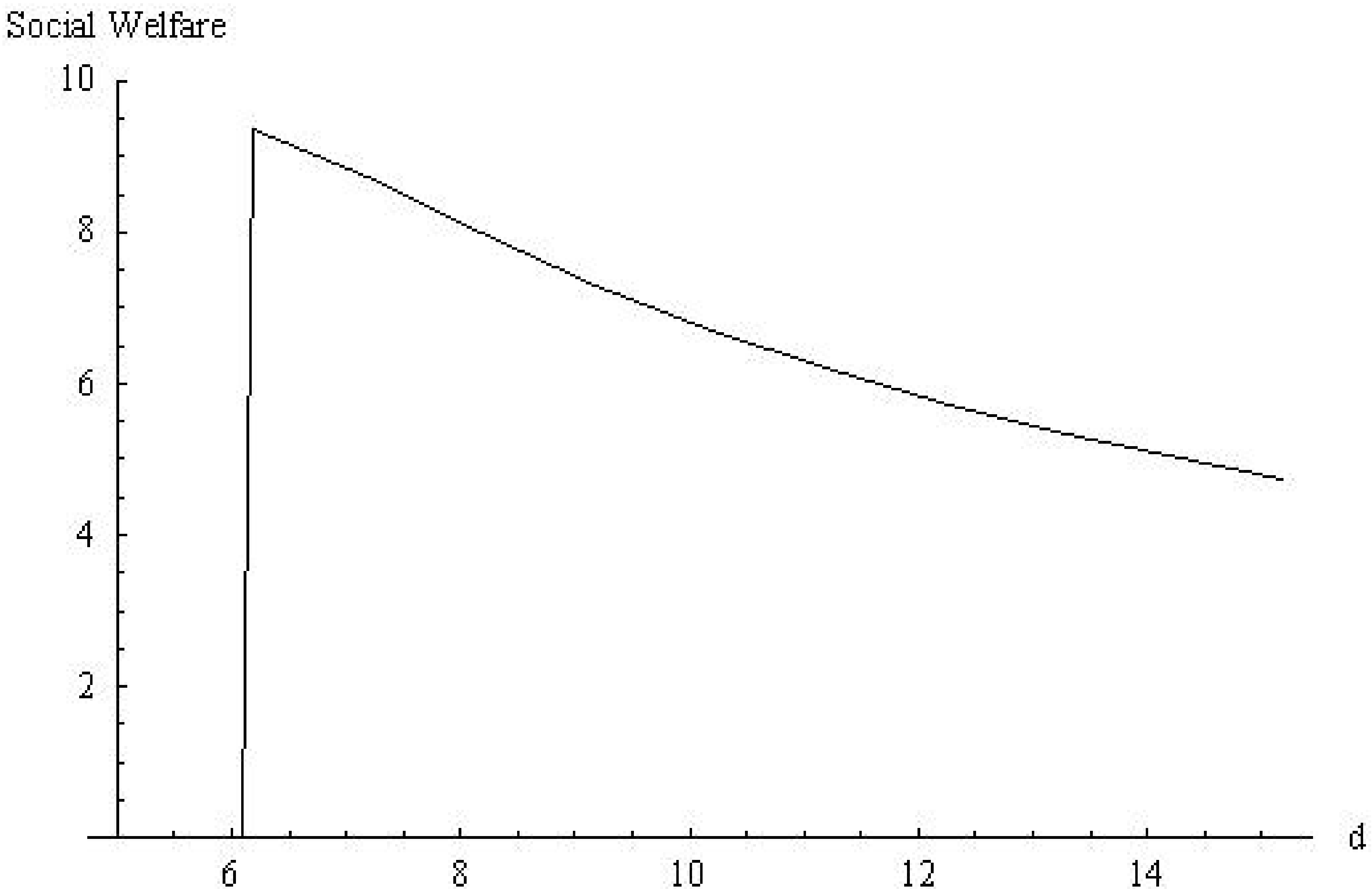
- we can compute the equilibrium
- the SW is maximized for $d = 3.06N / nf$

The basic issues

- Properties of equilibria
- Convergence of myopic strategies
  - super modularity?
- How the equilibrium is affected by d
  - how to adapt d to maximize efficiency
- More refined models of interaction
  - extend the above results for a general class of models

$$\theta_i u(y_i) - a(1 - e^{-y_i d}),\ y_i(x)$$

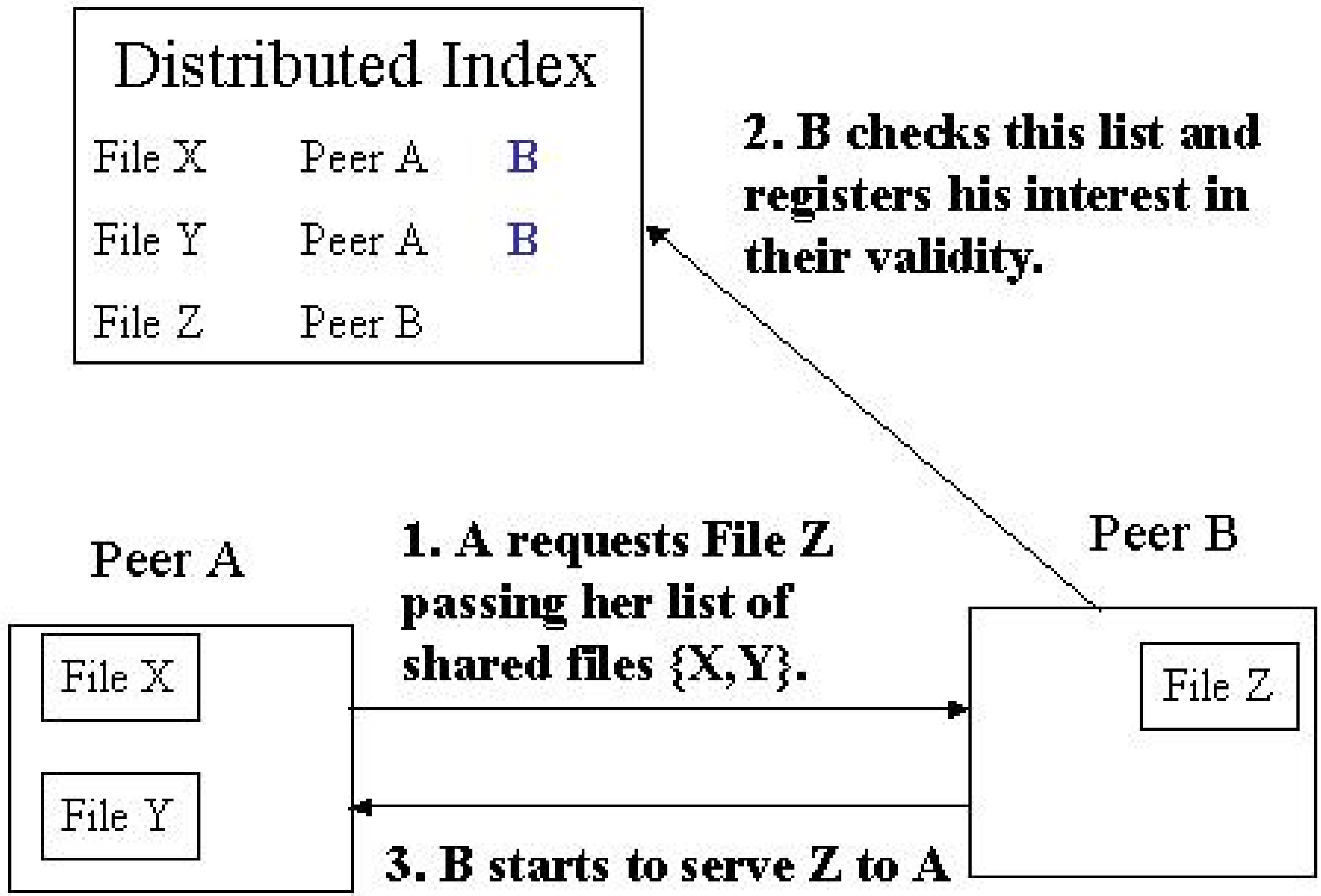
Simulation results (1)



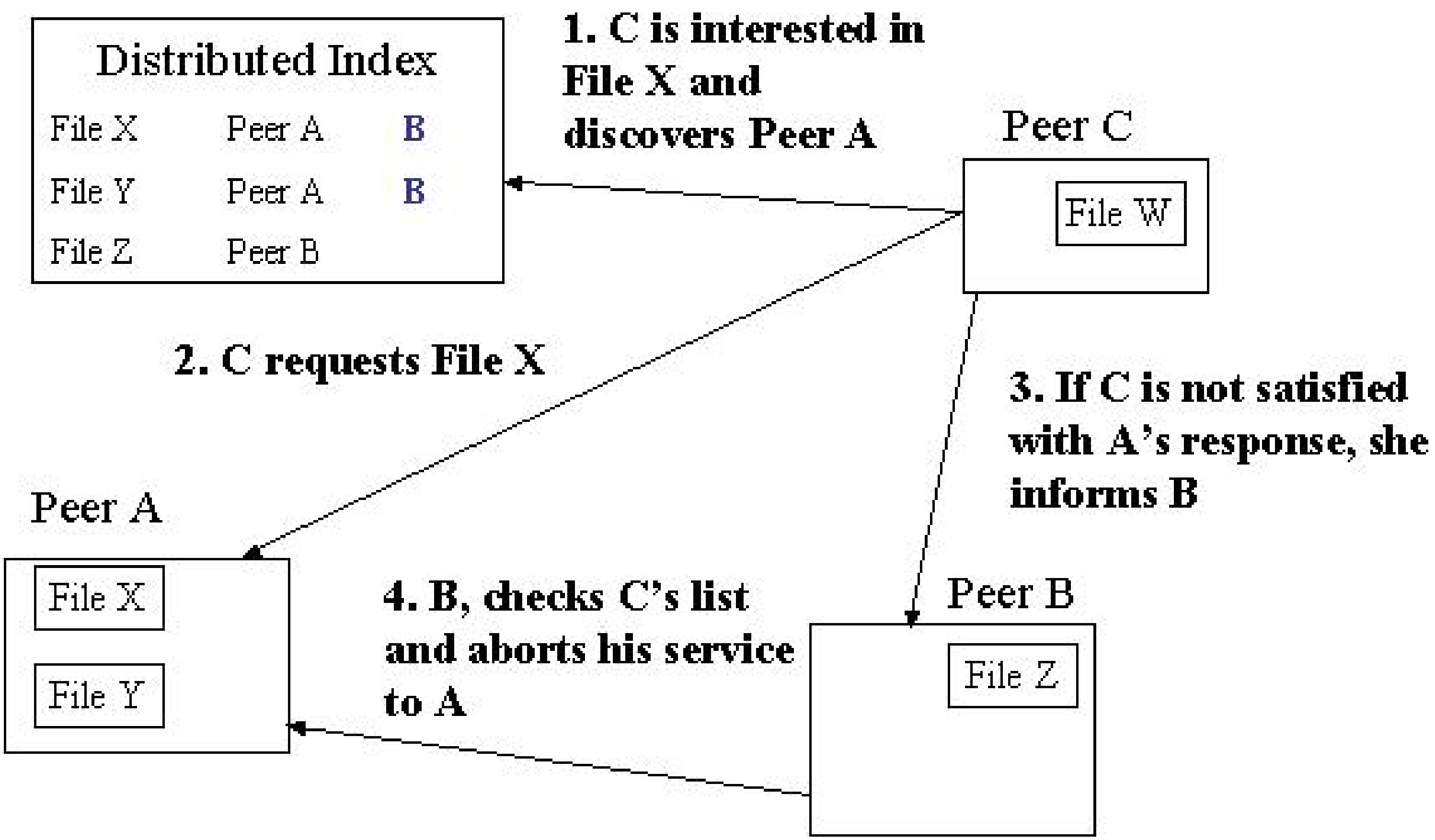
System Requirements

- 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)

Checking for file validity (1)



Checking for file validity (2)



Incentive Issues

- Failing to serve
  - Possible strategy to avoid uploading
  - BUT not rational under our assumptions
- Incentives for enforcement
  - Checking for file validity
    - There is a cost and system designer should try to minimize this
  - Fixing upload throughput
    - Not rational to increase or decrease it under our assumptions
    - If its decrease is considered a valid strategy there are ways to address it
  - There should be a majority of peers who wish to play by the rules
    - a requirement for all distributed enforcement mechanisms

Conclusions

- As access bandwidth increases and large amounts of content is readily available in people’s PCs peer availability will become a very important factor for the success of p2p file sharing systems

- We propose a memory-less mechanism to provide the suitable incentives
  - With appropriate tuning of its basic parameters (number of files and upload throughput) can achieve comparable to the optimal efficiency
  - Introduces a new class of incentive mechanisms that require contribution to the public good while consuming resources and has some very interesting related modeling issues

Future work

- Study the application of both our modelling (public good) approach and enforcement (contribute while consuming) in other contexts
  - A public good model for scientific grids
  - A “contribute while consuming” mechanism for ad-hoc networks

- **http://nes.aueb.gr/p2p.html**