

Incentives for Content Availability in Memory-less Peer-to-Peer File Sharing Systems

PANAYOTIS ANTONIADIS

Athens University of Economics and Business

and

COSTAS COURCOUBETIS

Athens University of Economics and Business

and

BEN STRULO

BT Research

In this paper we address the issue of content availability in p2p file sharing systems. Content availability is a public good: the copying of a file by one peer does not prevent another peer also from copying it; but contributing files to the common pool is costly. The asymptotic analysis of certain public good models for p2p file sharing suggests that when the aim is to maximize social welfare, a fixed contribution scheme in terms of the number of files shared per unity of time can be asymptotically optimal as the number of participants n grows to infinity. However, the enforcement of such an incentive scheme is not straightforward in a realistic p2p system, where no trusted software or central entity accounting for peers' transactions can be assumed and peers are free to change their identity with no cost. We present a realistic version of the fixed contribution scheme, which does not require the use of system memory but relies only on the time peers are consuming resources to ensure that they contribute adequately. We describe the functionality that should be supported for enforcement and discuss the additional incentive issues that arise in this context, proposing some practical solutions to address them. We also formulate a suitable economic model to estimate the efficiency-loss of the proposed mechanism (compared to the one achieved using the theoretically optimal schemes under complete and incomplete information) and provide some insights for the correct tuning of its basic parameters. Our first results indicate that the proposed mechanism constitutes a good compromise between economic efficiency and implementability and should lead to some interesting and practical solutions for providing incentives for content availability in p2p systems.

Categories and Subject Descriptors: C.2.4 [Computer-Communication Networks]: Distributed Systems; J.4 [Social And Behavioral Sciences]: Economics

General Terms: Design, Economics

Additional Key Words and Phrases: cheap pseudonyms, public goods, content availability

1. INTRODUCTION AND RELATED WORK

Since the first formal identification of the 'free riding' issue in p2p file sharing systems by Adar and Huberman [2000], significant research focus has been given to the incentive mechanisms that could be used to eliminate or constrain opportunistic behaviour in p2p systems.

Detailed modelling of the economic transactions carried out in a file sharing system is, in general, a very complex task. The main reason is that participating peers should contribute different types of resources (bandwidth, storage, CPU cycles, content) with differing characteristics, the provision of which generates complex costs such as legal risks and time spent using the system. Moreover, in

Permission to make digital/hard copy of all or part of this material without fee for personal or classroom use provided that the copies are not made or distributed for profit or commercial advantage, the ACM copyright/server notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.

© 2005 ACM /2005/-0011\$5.00

certain cases, there exist several intangible value generators from participating in such a system, which often involve altruism, community building, “fighting the system”, and more. Actually, some of these might be part of the reason¹ why the theoretical results of economic theory are not always compatible with the performance of real p2p applications, which seems to be acceptable even without explicit incentives for cooperation. Golle et al. [2001] made a first effort to model the utilities and costs associated with the participation in a p2p file sharing system. Using game theoretic analysis, they propose the use of micropayments for achieving the desirable equilibria.

As far as economic efficiency is concerned (i.e. maximization of social welfare), besides the complicated modelling, an additional challenging issue is the lack of information concerning the types and preferences of the individual peers, which is required for the computation of the optimal allocation of resources and cost in a p2p system. The existence of externalities rules out as inefficient the adoption of a free-market approach using market-defined prices for controlling a p2p system. ‘First-best’ policies typically require that each peer face a different price for its use of the p2p network, or face an individual constraint on its request rates and files shared. However, recent asymptotic results [M. F. Hellwig 2003; Courcoubetis and Weber 2004] suggest that when content availability and provision is the main value and cost generator respectively, a fixed contribution scheme in terms of the number of files shared can be asymptotically optimal as the number of participants n grows to infinity. Most importantly, such contributions can be computed using a simple optimization problem, provided that the distribution of peers’ types is known. Buragohain et al. [2003] follow a game theoretic approach to study the equilibria and corresponding efficiency achieved in a p2p file sharing system, based on a similar utility and cost model to ours, assuming that the system can enforce a level of reciprocity.

In addition to this challenging economic resource allocation problem, a p2p system designer has to deal with the inability to rely on trusted software² or on central entities that can monitor and account for peers’ transactions to ensure that they contribute and consume the amount of resources dictated by an underlying economic model. A significant part of the research literature in p2p economics, studies the game theoretic and implementation issues related to the effective accounting of peers’ transactions (by means of reputation, credits, etc.) in a fully distributed and untrusted environment (e.g. [Feldman et al. 2004; Vishnumurthy et al. 2003] among many others). Such approaches adhere to the principle that all peers should benefit from the system as much as (or proportional to what) they contribute in terms of actual downloads/uploads, and thus require some sort of user memory to be able to identify and punish the potential free riders, or to restrict peers in their downloading. Whitewashing and false trading are some of the most important challenging issues that need to be tackled in this context. Friedman and Resnick [2001] demonstrated the need for imposing costs to newcomers in a memory-less system in order to discourage whitewashing. Feldman et al. [2004] propose mechanisms based on reputation to address the issue of false trading and other related attacks.

BitTorrent [Cohen 2003] is an example of a real world application focusing on bandwidth provisioning for content distribution, which actually implements a reciprocative incentive scheme without relying on past transactions of peers but on a direct exchange of resources. Because the incentive scheme does not rely on tracking the long term behaviour of peers it is simple to implement and largely immune to problems of false trading and whitewashing. However, because it is designed to incentivise only the provision of bandwidth, it does not tackle our objective of improving content availability.

On the other hand, Direct Connect³ is another p2p application which mainly relies on central control to enforce minimum contribution rules, expelling misbehaving peers based on their IP addresses. Most of these rules dictate a minimum total size of content shared by each peer and

¹an additional reason could be the low costs for offering mostly idle resources, compared to the value that these systems offer to their participants, while Krishnan et al. [2003] identify the reduction of congestion as a sufficient incentive for peers to contribute without any additional incentives provided by the system.

²Kazaa is a characteristic example of a real world application that tried to implement a reciprocative incentive mechanism (giving priority to peers that contribute more in terms of uploads/downloads ratio), which failed due to a hacked version of its software —see <http://www.k-lite.tk/>

³<http://www.neo-modus.com/>

to that respect, Direct Connect follows the general principles of our public good model, i.e. the requirement for each peer to share a minimum amount of files, without any restriction on the number of requests issued.

We wish to design a system that focuses on content availability (as Direct Connect) without requiring the existence of any sort of user memory or the ability to permanently expel peers from the group, following the implementation principle of BitTorrent. Thus, we hope to improve the economic efficiency of the system without suffering from whitewashing and false trading and without the implementation requirements of longer term tracking of peer behaviour.

In the following, we present in more detail the assumptions and objectives of our system and propose a memory-less enforcement mechanism distinguishing between the number of files shared and the time that these files are shared (both enforced while a peer is downloading files). Next, we describe the requirements on a system implementing this mechanism and discuss the additional incentive issues that arise in this context. We then formulate a suitable public good model and provide a theoretical justification for the mechanism proposed, showing that the efficiency loss due to our restricted implementation environment is acceptable when the system parameters are appropriately tuned. Finally, we briefly discuss the weaknesses of the proposed mechanism and present our conclusions.

2. CONTENT AVAILABILITY

Recent articles in the popular press (initiated by [Anderson 2004]) discuss the importance of the ‘long tail’ of content; that large part of the set of content in which individual files are not popular, but which together constitute the majority of the total requests. The provision of this part of the content in a p2p system requires different types of incentives than the ones usually discussed in the p2p economics literature, which look at uploading cost rather than the cost for contributing to the overall content availability. The result is that mostly popular items are traded and normal peers rarely choose to contribute their own files but prefer to redistribute existing ones.

We therefore choose to consider a system where the probability of a certain file being requested is low but the overall value of satisfying such requests is much greater than for popular items (since it is more difficult to find unpopular or rare items, in many cases even if one wishes to pay for them). Thus we will attempt to give incentives for providing any content item regardless of request rate.

Furthermore, we also need to give incentives to remain in the system and providing content, not simply to have provided at some time in the past. As available bandwidth increases and download times reduce significantly, we believe that peer availability will become a very important aspect for the success of p2p file sharing systems, and especially for those providing items of medium size, like music files, high definition photographs, etc.

Note that, in such a system, we consider uploading cost to be of limited importance. This is especially so since, anticipating one of the main characteristics of our mechanism, we design the system to aim to provide a particular upload throughput for all requests at all times. Since all peers use the same upload throughput, we have a homogeneous system in which the prioritization of TCP ACKs largely eliminates the cost of uploading while downloading (see [Feldman et al. 2003]). We summarize in the following our main assumptions:

- Uploading while downloading imposes negligible costs
- Congestion on the upload link of an individual peer is rare
- All files have a similar (low) rate of requests, either because they are unpopular or because they are distributed in proportion to their popularity
- Peers act rationally in their own self-interest and not maliciously.

3. ENFORCEMENT MECHANISM

For a system to punish peers that do not share a minimum number of files per unit of time, it needs to calculate the total time a peer has stayed in the system (during an enforcement period) and make sure that the required number of valid files were shared —and served upon request—

throughout that time. Based on this information, it can then appropriately penalise misbehaving peers during the next enforcement period. A more realistic approach could treat the actual uploads provided by a peer as a proxy for the number of files shared and use, for example, expiring receipts signed by peers when downloading content to be able to enforce the required level of sharing at each enforcement period. But such a system would be susceptible to false trading (peers signing fake receipts for their friends or even for themselves creating multiple identities) and whitewashing.

A robust memory-less enforcement mechanism needs to rely only on the time a peer is consuming resources in the p2p system (namely, downloading a file) and thus use the uploading peer (the *uploader*) as the ‘enforcing entity’. So, we propose a system where the uploader first checks whether a candidate *downloader* shares the amount of files required before providing a requested file. Moreover, he should ensure that these files are accessible to the rest of the group (e.g. by checking with the search mechanism for their availability). If the downloader refuses to actually upload one of his files to a *requestor*, the latter will inform the uploader to stop the transfer. The same would happen if the file sent by the downloader is not valid as discussed in Section 4.1.

The second attribute of a peer’s contribution is the time that she spends in the system sharing the required amount of files. Since we have assumed that peers can be forced to contribute only during the time they download files, this time is directly related to the upload throughput offered by the uploader. So, in order for the system to enforce a certain contribution by peers, it also needs to constrain the minimum average download time that they face. This could be easily implemented using a simple and lightweight protocol that broadcasts the fixed throughput that all peers will be using for uploading files thus allowing it to be dynamically tuned.

Increased average download time would have a positive impact both on the content availability achieved and on the probability that a file of a downloader is requested (in which case he would be forced to also upload a file to continue downloading). But it also reduces the utility of peers by requiring them to wait longer for their downloads to complete. We study such trade-offs concerning this crucial parameter of our mechanism in Section 5.3.

3.1 System requirements

We describe an abstract protocol which passes the necessary information between the peers to achieve our enforcement. Initially, as depicted in Figure 1(a), Peer A discovers files of interest through the system index and sends a request to an appropriate uploader, Peer B. This request includes a list of files currently being served by A as evidence of her contribution. Peer B needs to check the validity of this list. He does this by querying the index himself. If he fails to find them in the index he will refuse to serve A. If he does find them, he adds a note to these index entries indicating his current interest in the validity of these services. He then begins to serve to A.

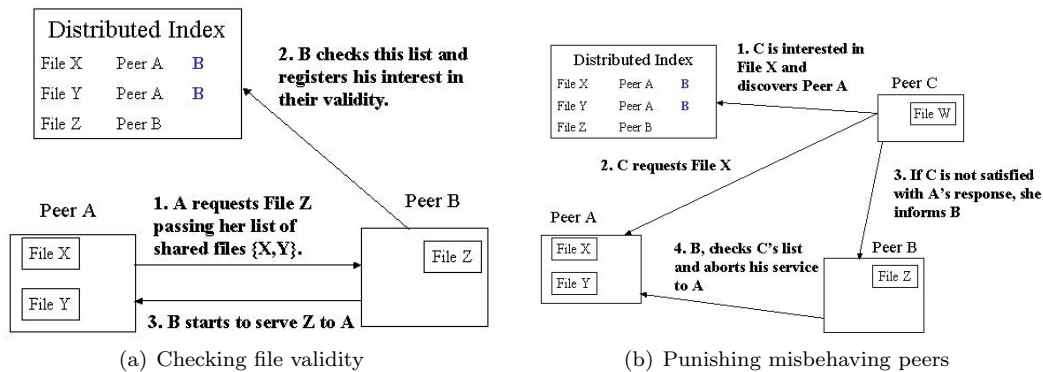


Fig. 1. Basic functionality of the enforcement protocol

During the time in which B is serving A, it is possible for another peer C to wish to request a file from A. C will discover via the index that B has an interest in the validity of these services. Then if the service to C is not performed adequately, (the files are not actually served, or are

corrupt or mis-labeled, etc.) then C can tell B (see Figure 1(b)). At this point, B will punish A by aborting the download of file Z, after checking the validity of files shared by C (to ensure that this was not the reason he wasn't served).

3.2 Super-peers

Many p2p systems rely on the existence of some "super-peers" with a significant amount of resources available and mainly altruistic motives to participate and contribute to the system. It is not at all unrealistic to assume the existence of some peers which will have the resources to play this role (Liang et al. [2004] present measurements in the Kazaa network depicting this fact).

In our case, we rely on such peers to act as seeds for the content and to provide a certain amount of files in order to form the roots of the envisioned trees of downloaders which would further increase the availability of content in the system. Otherwise, our mechanism would require "cycles" of content requests to be formed resembling a direct exchange mechanism. Additionally, our super-peers are responsible for computing useful system information (such as the size of the system, the number of files shared, etc.) and tuning the most important system parameters, like the fixed upload throughput. Thus, the existence of super peers makes the implementation of this functionality more realistic and helps avoid specific incentive issues that arise, as we discuss next.

4. INCENTIVE ISSUES

Does the proposed mechanism actually provide the required incentives to peers under our system assumptions? If peers choose not to advertise shared files then they will be unable to download. However, can peers pretend to be sharing but actually fail to do so?

4.1 Validity of files shared

Since we don't require a peer to actually upload a file while downloading, we need to do our best to check the validity of the files shared as early as possible. The first check is that the files are actually shared in the distributed index. Thus peers cannot claim to be sharing without running the risk of receiving a request.

A cheating peer could advertise files with meaningless or deliberately obscure names thereby reducing the chances of receiving request. They need not actually have the claimed content available. We can imagine processes, perhaps run by super-peers, which flush invalid titles from the index, perhaps by using something similar to the CDDDB service (either centralized or implemented in a distributed way). This will make it difficult for peers to advertise such names. The uploading peer could also perform such checks.

A cheating peer may also advertise files with legal names but corrupted or invalid content. This could also be checked by the uploading peer (or more likely by future requestors of the file) using techniques like the ones proposed by Liang et al. [2005]. However, if the probability of receiving a request is low, this may still be a profitable attack. However, there are a number of mechanisms we can use to reduce this risk. For example

- super-peers could perform random checks on shared files, perhaps focusing on new peers, or newly shared content. If a trusted subset of super-peers is available then passing this check could be rewarded with a signed certificate that could be used in future exchanges.
- peers could be incentivized to contribute their own (possibly rare) tastes into the community and thus become members of social networks of interest. Or there could be social incentives such as improved social reputation for the provision of high quality content.

This paper is mainly focused on how to incentivize peers to provide content for a sufficient time to improve its availability. Thus we do not provide a complete solution to this question of content validity and more research needs to be done in this area. We discuss this further in Section 6.

4.2 Failing to serve

This form of misbehaviour should be detectable as soon as one of the shared files is requested by a third peer. The peer can choose at this point to abort his download. If the downloading peer refuses to serve such requests then his download will be aborted in any case. But the misbehaving

peer can enter into the system at a later time to complete this aborted download. This resembles the free riding issue in Bittorrent, where potential free riders can receive a —limited— service due to ‘optimistic unchoking’ (see [Qiu and Srikant 2004]). In our case, however, this is not a rational decision, since we have assumed that uploading a file does not incur any additional cost and a downloader, having to pay anyway the cost for sharing the required amount of files and for waiting for his download to complete, doesn’t have the incentive not to serve an incoming request.

4.3 Incentives for enforcing

Our enforcement protocol has two parts: a) checking file validity and aborting uploads to misbehaving peers and b) constraining the upload throughput offered.

The first one incurs a certain fixed cost to the peers executing it. However, we expect this cost to be small (and a system designer should ensure that it is minimized by careful design and by moving functionality to super peers) and, at least for some peers, it may be outweighed by the direct benefit of having an audit done; i.e. that other peers are prevented from abusing the system. Moreover, it seems that human nature includes a bias towards punishing cheats even at some cost to the punisher, as described by Fehr and Gächter [2002].

The second one, fixing the upload rate to a prescribed value, carries no cost and thus peers will have no incentive to increase it, especially since its enforcement will be hardwired into the application. Decreasing the upload rate is also not a rational strategy under our assumptions. However, if it is considered, for some reason, a valid attack, the system has the following options (amongst others) to address it:

- A peer receiving lower than expected throughput could treat the corresponding peer as “failing to serve” with some probability (depending perhaps on the probability that a peer’s upload capacity is congested —in which case the lower throughput should be excused)
- using the notation in Figure 1, Peer C could inform B in order to adjust his upload throughput to the value A offers to C.

Finally, colluding peers that agree to upload to each other with higher than the prescribed upload rate will not gain much when they also interact with normal or super peers, since they will still have to wait for these downloads to finish. If they are not interacting with any other peers then they effectively form their own isolated system which is always an option in any case.

5. ECONOMIC MODELLING

Suppose that peers $1, \dots, n$ are to share the use of a public good: the expected number of distinct files made available in the system at some arbitrary time. The good can be provided at quantity Q for a rate of cost $c(Q)$. If N is the maximum number of distinct valid files, we use Q/N , the probability that a random request is satisfied, to express the content availability achieved in a system of size Q . Peer i has a utility for the good of $\theta_i u(Q/N)$, where θ_i is a ‘preference parameter’ which is known only to peer i , but which is a random sample from a distribution on $[0, 1]$, with distribution function $H(\cdot)$ and density function $h(\cdot)$. Here, the ‘benefit’ function $u(\cdot) \geq 0$ is assumed to be continuously differentiable, increasing and strictly concave in its argument (i.e. $u(x) = x^\beta$, where $\beta < 1$ is a positive constant). The benefit function is the same for all peers; peers differ in the payoff parameter θ_i .

In order to build the public good Q , each peer has to share f_i files for a fraction t_i of time ($0 \leq t_i \leq 1$). Then, at some arbitrary time the total average number of not necessarily distinct files shared F , will be $F = \sum_{i=1}^n (f_i t_i)$. Due to duplication, the number of available distinct files Q will be in general a concave function of F , which when Q/N is not very close to 1 and all files are equally popular, could be approximated by F , as shown in [Antoniadis et al. 2004]. That is, in our range of parameters, $Q(F) \approx F$, and we can use Q instead of F . In the following we make this simplifying assumption⁴, which is not crucial for the qualitative results we obtain.

Since we do not account for uploading costs, no limitation is posed on the rate with which peers request and download files. Moreover, the number of files shared is considered a ‘sunk’ cost,

⁴otherwise we can compose the functions $u(Q)$ and $Q(F)$ into another benefit function where our public good is F .

incurred by a peer before entering the system and includes mainly the costs for acquiring (e.g. ripping a CD) and storing the content. In our analysis we have assumed that the system designer has determined beforehand a fixed and common number of files f required to be stored on each peer, based on system parameters like the expected number of participating peers, the maximum number of distinct files N , the average size and value of files shared, etc.

So, during system operation, the rate of cost peers have to contribute for building the public good Q is only due to the fraction of time t_i they have to stay on-line sharing the fixed amount of f files, which we assume that it is linear in t_i and the same for all peers. So, $c(Q) = \alpha \sum_{i=1}^n t_i$, where α converts time units to monetary units, and Q will be equal to $\sum_{i=1}^n t_i f$, and thus $c(Q) = \alpha Q/f$.

In the following we present and compare three alternative incentive mechanisms that could be used by a system designer in order to maximize the social welfare of the system, corresponding to different assumptions for the informational and implementation limitations faced.

5.1 First-Best

Under complete information (when the payoff parameter θ_i of every peer i is known) and unlimited enforcement capabilities, the system designer should decide on the optimal amount of Q built and the fraction of time t_i that peer i should stay on-line, solving the following optimization problem

$$\begin{aligned} & \underset{\{t_1, \dots, t_n\}, Q}{\text{maximize}} \quad \sum_{i=1}^n \theta_i u(Q/N) - c(Q) & (1) \\ & \text{s.t.} \quad 0 \leq t_i \leq 1 \ \&\& \ \theta_i u(Q/N) \geq \alpha t_i, \ \forall i. \end{aligned}$$

When the optimal size of the system Q^* is computed using (1), it is then trivial to compute a set of feasible t_i^* s (i.e. such that $\theta_i u(Q^*/N) \geq \alpha t_i^*$, $\forall i$).

The solution of this problem will form our ‘benchmark’ for assessing the efficiency of our proposed mechanism, since the efficiency achieved solving (1) is the maximum possible, what is called in economics ‘first-best’.

5.2 Fixed contribution mechanism

We study now the case where the system designer has incomplete information on peers’ valuations. Assume the following: Knowing n and $H(\cdot)$, a social planner wishes to design a mechanism which, as a function of the declared $\theta = (\theta_1, \dots, \theta_n)$, sets Q , and determines which peers may use the good and how many files they should contribute⁵.

Recent asymptotic results [M. F. Hellwig 2003; Courcoubetis and Weber 2004] suggest that, when n gets large and exclusions are possible, a fixed contribution mechanism (each peer must contribute a uniform minimum number of files to join the network) suffices to lead the system to the second-best: the maximum efficiency that could be achieved under the ‘feasibility’, ‘incentive compatibility’, and ‘individual rationality’ constraints.

In this case, the system designer should choose $\bar{\theta}$ and Q according to the following

$$\begin{aligned} & \underset{Q, \bar{\theta}}{\text{maximize}} \quad nu(Q/N) \int_{\bar{\theta}}^1 x dH - c(Q) & (2) \\ & \text{s.t.} \quad n(1 - H(\bar{\theta}))\bar{\theta}u(Q/N) = c(Q) \end{aligned}$$

This program has a very simple interpretation. It sets two variables: the total effective number of shared files Q and the identity of the marginal peer who is just indifferent between joining the network and not. All peers who join the network are required to contribute $\bar{\theta}u(Q/N)$ to the total cost $c(Q)$ (i.e., the benefit of the marginal peer). Peers with $\theta < \bar{\theta}$ will not be willing to join. The total contribution is covered by the contributions of the peers which will participate, which are $n(1 - H(\bar{\theta}))$ on the average. The average value of file sharing per peer is $u(Q/N) \int_{\bar{\theta}}^1 x dH$. Hence (2) maximizes the expected social welfare over the choice of fixed fee policies. The optimal policy

⁵Note the assumption that a peer may be excluded from using the good. If no exclusions are possible, as n increases, the power of incentives decreases (each peer feels that his contribution will have an insignificant effect to the overall provisioning), and hence is ready to contribute very little (in the order of $1/\sqrt{n}$) in any incentive compatible scheme.

will correspond to the optimal values of the two variables Q and $\bar{\theta}$. Solving (2) with $c(Q) = \alpha Q/f$ and the additional constraint that $t^* \leq 1$, we can compute Q^* and then the minimum contribution of each peer would be $t^* = \theta u(Q^*/N)/\alpha$.

Although the fixed contribution mechanism is in theory a very simple and attractive incentive scheme for content availability, it cannot be enforced by a realistic p2p system as discussed in Section 3. For instance, in the limiting situation where the access lines of peers have infinite capacity, and hence the average download time is zero, a rational peer participating in such a system would set $t_i = 0$, without affecting his own use of the system. Thus the inability to incentivize peers to stay in the system longer than needed by their downloads to complete, would result to ‘market failure’.

5.3 Fixed upload throughput mechanism

This is the reason we propose the control of the throughput b with which peers upload files to each other, in order to enforce peers to stay in the system sharing their files for at least the amount of time they consume resources for themselves. We denote by d the average download time of one file, which is in general a decreasing function of b . For simplicity, assume that $d(b) = s/b$, where s is the average file size. In a realistic p2p system, where peers have the ability to make multiple requests and thus download many files in parallel, this value could be much less. On the other hand, it would be increased in cases when peers leave the system before their downloads are completed (when for example the download of their uploader finishes and he decides to leave the system). The detailed modeling of peers’ behaviour using simulations in order to evaluate the importance of our assumptions and simplifications is left for future research.

So, the contribution of peers will depend on their request rate multiplied by the probability that their requests are successful, and thus cannot be made the same for everybody as the fixed contribution scheme suggests. We define $g(\theta)$ to be a function that maps a peer’s type to its request rate. For our analysis we have chosen $g(\theta) = \theta^2$, assuming a convex relation between the type and the request rate of peers. Then we denote by x_i how much a peer is willing to reduce her type and hence her request rate respectively facing an average download rate d in a system of (average) size Q . The resulting fraction of time t_i that peers will stay in the system will be equal to the rate of successful requests, $g(\theta_i - x_i)(Q/N)$, multiplied by the average download time d .

We describe now the iterative procedure that will converge to the equilibrium value of Q . Given f , d , and an initial value Q_0 for Q ⁶ each peer i will choose her optimal strategy x_i^* solving the following local optimization problem

$$\begin{aligned} & \underset{x_i}{\text{maximize}} \quad (\theta_i - x_i)u(Q_0/N) - \alpha g(\theta_i - x_i)(Q_0/N) d & (3) \\ & \text{s.t.} \quad 0 \leq x_i \leq \theta_i \ \&\& \ g(\theta_i - x_i)(Q_0/N) d \leq 1. \end{aligned}$$

This will have an effect on the total amount of distinct files Q_1 , which will now become $Q_1 = \sum_{i=1}^n g(\theta_i - x_i^*)(Q_0/N) d f$. Then each peer will solve again (3) with Q_1 instead of Q_0 to adjust its x_i accordingly, and so on. Simple analysis suggests that there is an equilibrium with $Q > 0$ if and only if

$$\frac{\sum_{i=1}^n g(\theta_i) d f}{N} \geq 1. \quad (4)$$

In particular, if (4) does not hold, then for any value of Q , the resulting Q' , based on the x_i^* s computed by peers using (3) will be less than Q , and thus will eventually drop to zero. So, (4) indicates that there is a constraint between the system parameters, below which the system collapses, and thus provides useful insights for the appropriate tuning of f and d .

In this work, we are mostly interested in the properties of the parameter d (assuming f fixed). In Figure 2(a) we show how the social welfare of the proposed system changes, for a specific realization of H , as d grows above the threshold value suggested by (4). As one can see, this threshold value is actually the optimal one, $d^* = \frac{N}{\sum_{i=1}^n g(\theta_i) f}$, which can be also proven analytically.

⁶In practice this could be the average amount of files made available by super-peers but for our equilibrium analysis we can assume that it is the peers’ expectation for the initial amount of content made available.

However, we need complete information (all the values $g(\theta_i)$) to compute d^* . Under incomplete information, setting $d = \bar{d} = \frac{N}{n E[g(\theta_i)] f}$ might not be the optimal choice, since being close to the threshold point, the probability of market failure by not satisfying (4) for some of the realizations of H increases. One could think of a more conservative tuning of d , moving to the right of the curve shown in Figure 2(a), and thus losing a little in efficiency for some cases but gaining in robustness of the equilibrium point.

We can express this trade-off using, for example, the Chebyshev bound⁷ and defining $d(k) = N / (n E[g(\theta_i)] - k \sqrt{\text{Var}[g(\theta_i)] n}) f$. The more risk averse a system designer is, setting higher values for k and thus reducing the probability that $d(k) < d^*$, the higher would be the average distance of $d(k)$ from d^* , and thus a certain amount of efficiency would be sacrificed. Note that, in practice, market failure could be avoided by dynamically adjusting the value of d when there is evidence (based on measurements) that it is inefficiently set.

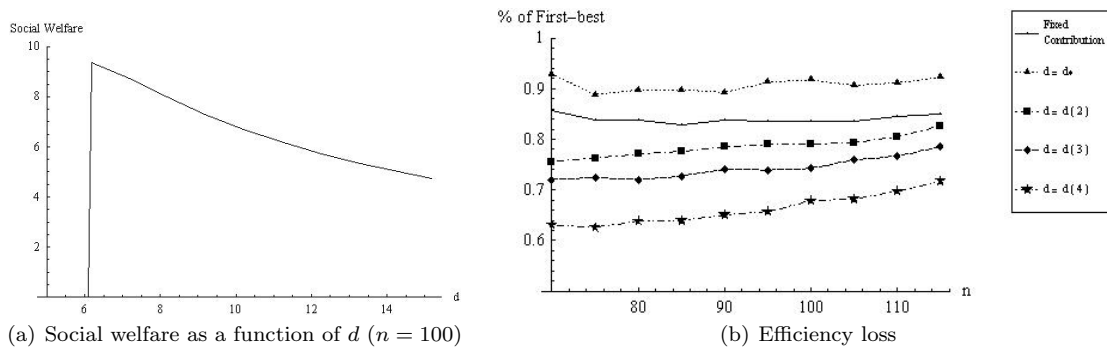


Fig. 2. Evaluation of the ‘fixed upload throughput’ mechanism ($\beta = 0.5$, $\alpha = 0.3$, $N = 10^4$, $f = 50$)

Figure 2(b) depicts the efficiency (% of the first-best) of the ‘fixed upload throughput’ mechanism⁸ as n grows for $d = d^*$, and for $d = d(k)$, $k = 2, 3, 4$ for which, in our computationally feasible range of n , the estimated probability $P[d(k) < d^*]$ is less than 0.03 , $8 \cdot 10^{-3}$, and 10^{-4} respectively. Observe that, for given k , the efficiency loss decreases as n grows, since $d(k)$ comes closer to \bar{d} . So, we expect to be able to control our system closer to the optimum for larger values of n . Finally, notice that the efficiency for $d = d^*$ is higher than this of the ‘fixed contribution’ mechanism because d^* is computed using complete information of peers’ types.

These preliminary results indicate that the efficiency achieved by our proposed mechanism is comparable to the first-best and with appropriate tuning of d can come very close to the one achieved by the ‘fixed contribution’ mechanism. So, our economic model suggests that the ‘price of cheap pseudonyms’ using the proposed mechanism could be significantly constrained, which is encouraging for its applicability in a realistic context. Our on-going work includes further analysis for the optimal setting and tuning of the parameter d and the in-depth understanding of the inter-dependencies between the system parameters using both simulations and analytical tools.

6. WEAKNESSES

The most serious weakness of applications of the proposed mechanism is likely to be attacks based on the provision of invalid content, particularly where the cost of provision of an individual piece of content is high and its probability of request is low. Newcomers may then often be able to get away with advertising invalid content for long enough to successfully complete their downloads.

⁷ $P[|X - \bar{X}| \geq k\sigma] \leq \frac{1}{k^2}$, for a random variable X with standard deviation σ .

⁸ For the computation of the social welfare, for given n , f , N , Q_0 , and d , we did the following: Draw n values $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ randomly from the specified distribution H and for this realisation of peers’ valuations, simulate the iterative procedure described above until a fixed point with 1% precision is reached. Repeat this step 100 times and average the total payoffs achieved over the draws of θ .

However, in other scenarios, where the cost of provision of new content is less high, then this could be a much less attractive attack. We believe that in the future many groups of people will have much content on their PCs, perhaps legally available, and of wide interest within their social network. Then the cost for providing this content will be small (not much larger than the cost of creating a realistic faked file) with any advantage outweighed by the probability of having your downloads stopped.

Other incentive mechanisms could be built on top of our basic incentive scheme, such as the construction of social networks based on taste within which peers will have an incentive to make available the files that best represent their personality. Other psychological incentives, such as encouraging peers to identify with the well-being of the community as a whole, also appear to have an effect in some file-sharing communities and provide interesting avenues for exploration. We believe that the formal discussion of social incentives in the context of p2p applications is an interesting direction for further work.

7. CONCLUSION

In this paper, we have explored how to implement economic incentives in a memory-less p2p system focusing on content availability as the main good provided. Our goal is to study how far one could go with so limited capabilities in terms of system functionality (which in many realistic cases constitute the only ones available).

We identified time spent in the system as an important parameter of the contribution of peers concerning content availability. We then described an incentive mechanism which both controls this time so as not to be too short and provides an incentive to offer a fixed number of files throughout that time. We then presented a theoretical framework for the study of its qualitative characteristics and the estimation of the resulting efficiency. We believe this mechanism is practical and appropriate to improve the economic efficiency of a file-sharing system, at least in respect of content availability, which we consider its most important parameter.

REFERENCES

- ADAR, E. AND HUBERMAN, B. 2000. Free riding on gnutella. *First Monday* 5, 10 (October).
- ANDERSON, C. 2004. The Long Tail. Wired Magazine, <http://www.wired.com/wired/archive/12.10/tail.html>.
- ANTONIADIS, P., COURCOUBETIS, C., AND WEBER, R. 2004. An Asymptotically Optimal Scheme for P2P File Sharing. 2nd Workshop on Economics of Peer-to-Peer Systems, Harvard University.
- BURAGOHAIN, C., AGRAWAL, D., AND SURI, S. 2003. A Game Theoretic Framework for Incentives in P2P Systems. In *Proceedings of the Third IEEE International Conference on Peer-to-Peer Computing (P2P 2003)*, Sweden.
- COHEN, B. 2003. Incentives Build Robustness in BitTorrent. In *Proceedings of the Workshop on Economics of Peer-to-Peer Systems, Berkeley, CA*.
- COURCOUBETIS, C. AND WEBER, R. R. 2004. Asymptotics for Provisioning Problems of Peering Wireless LANs with a Large Number of Participants. In *Proceedings of WiOpt'04 workshop, University of Cambridge, UK*.
- FEHR, E. AND GACHTER, S. 2002. Altruistic Punishment in Humans. *Nature* 415, 137–140.
- FELDMAN, M., LAI, K., CHUANG, J., AND STOICA, I. 2003. Quantifying Disincentives for Collaboration in Peer-to-Peer Networks. 1st Workshop on Economics of Peer-to-Peer Systems, University of Berkeley.
- FELDMAN, M., LAI, K., STOICA, I., AND CHUANG, J. 2004. Robust Incentive Techniques for Peer-to-Peer Networks. ACM E-Commerce Conference (EC'04).
- FRIEDMAN, E. AND RESNICK, P. 2001. The Social Cost of Cheap Pseudonyms. *Economics and Management Strategy* 10, 2, 173–199.
- GOLLE, P., LEYTON-BROWN, K., MIRONOV, I., AND LILLIBRIDGE, M. 2001. Incentives for Sharing in Peer-to-Peer Networks. In *Proceedings of WELCOM'01*.
- KRISHNAN, R., SMITH, M., TANG, Z., AND TELANG, R. 2003. The Virtual Commons: why Free-Riding can be Tolerated in Peer-to-Peer Networks. Workshop on Information Systems and Economics.
- LIANG, J., KUMAR, R., AND ROSS, K. W. 2004. Understanding kazaa. submitted.
- LIANG, J., KUMAR, R., XI, Y., AND ROSS, K. W. 2005. Pollution in P2P File Sharing Systems. In *Proceedings of IEEE Infocom, Miami, FL, USA*.
- M. F. HELLMIG. 2003. Public-Good Provision with Many Participants. *Review of Economic Studies* 70, 589–614.
- QIU, D. AND SRIKANT, R. 2004. Modeling and Performance Analysis of BitTorrent-Like Peer-to-Peer Networks. In *Proceedings of ACM SIGCOMM 2004 Conference, Portland, OR, USA*.
- VISHNUMURTHY, V., CHANDRAKUMAR, S., AND SIRER, E. G. 2003. Karma: A Secure Economic Framework for Peer-to-Peer Resource Sharing. 1st Workshop on Economics of Peer-to-Peer Systems, University of Berkeley.