An Asymptotically Optimal Scheme for P2P File Sharing*

Panayotis Antoniadis†, Costas Courcoubetis† and Richard Weber‡

Abstract The asymptotic analysis of certain public good models for p2p systems suggests that when the aim is to maximize social welfare 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. Such a simple scheme eliminates free riding, is incentive compatible and obtains a value of social welfare that is within $o(n)$ of that obtained by the second-best policy of the corresponding mechanism design formulation of the problem. We extend our model to account for file popularity, and discuss properties of the resulting equilibria. The fact that a simple optimization problem can be used to closely approximate the solution of the exact model (which is in most cases practically intractable both analytically and computationally), is of great importance for studying several interesting aspects of the system. We consider the evolution of the system to equilibrium in its early life, when both peers and the system planner are still learning about system parameters. We also analyse the case of group formation when peers belong to different classes (such as DSL and dial-up users), and it may be to their advantage to form distinct groups instead of a larger single group, or form such a larger group but avoid disclosing their class. We finally discuss the game that occurs when peers know that a fixed fee will be used, but the distribution of their valuations is unknown to the system designer.

1 Asymptotically optimal mechanism design

Suppose that peers $1,\dots,n$ are to share the use of a public good. The good can be provided at quantity Q for a cost of $c(Q)$. Peer i has a utility for the good of $\theta_i u(Q)$, 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 $h(\cdot)$. Knowing n and $H(\cdot)$, a social planner wishes to design a mechanism which, as a function of the declared $\boldsymbol{\theta}=(\theta_1,\dots,\theta_n)$, sets Q, determines which peers may use the good and what fees they should pay. These fees are to cover the cost $c(Q)$. Note the assumption that a peer may be excluded from using the good. Given knowledge of this mechanism, each peer declares his θ_i. The mechanism then sets $Q(\boldsymbol{\theta})$ and also decides which peers may use the good. If peer i is excluded from using the good then $\pi_i(\boldsymbol{\theta})=0$. If he is allowed to use it, then $\pi_i(\boldsymbol{\theta})=1$ and he must pay a fee $p_i(\boldsymbol{\theta})$. In (1)–(2) that follow the expectation is taken over $\boldsymbol{\theta}$ and in (3)–(4) it is taken over $\boldsymbol{\theta}_{-i}$, where this denotes all the preferences parameters apart from θ_i. The mechanism design problem is to maximize expected social welfare:

$$\underset{\pi_1(\cdot),\dots,\pi_n(\cdot),\,Q(\cdot)}{\text{maximize}}\; E\left[\textstyle\sum_i \pi_i(\boldsymbol{\theta})\theta_i u(Q(\boldsymbol{\theta}))-c(Q(\boldsymbol{\theta}))\right] \quad (1)$$

subject to a 'feasibility constraint', which says that the expected payments must at least cover the expected cost:

$$E\left[\textstyle\sum_i \pi_i(\boldsymbol{\theta})p_i(\boldsymbol{\theta})-c(Q(\boldsymbol{\theta}))\right]\geq 0\,, \quad (2)$$

'individual rationality' constraints, which say each peer can expect positive net benefit:

$$E\left[\pi_i(\theta_i,\boldsymbol{\theta}_{-i})\left\{\theta_i u(Q(\theta_i,\boldsymbol{\theta}_{-i})-p_i(\theta_i,\boldsymbol{\theta}_{-i})\right\}\right]\geq 0\,, \quad (3)$$

for all i, and 'incentive compatibility' constraints, such that each peer i does best by declaring his true θ_i rather than 'free-riding' by declaring some other θ_i':

$$\begin{aligned}&E\left[\pi_i(\theta_i,\boldsymbol{\theta}_{-i})\left\{\theta_i u(Q(\theta_i,\boldsymbol{\theta}_{-i})-p_i(\theta_i,\boldsymbol{\theta}_{-i})\right\}\right]\\&\quad\geq E\left[\pi_i(\theta_i',\boldsymbol{\theta}_{-i})\left\{\theta_i u(Q(\theta_i',\boldsymbol{\theta}_{-i})-p_i(\theta_i',\boldsymbol{\theta}_{-i})\right\}\right]\,, \quad (4)\end{aligned}$$

for all i and θ_i'. It can be shown that the above reduces to a problem of maximizing (1) subject to a constraint

$$E\left[\textstyle\sum_i \pi_i(\boldsymbol{\theta})g_i(\theta_i)u(Q(\boldsymbol{\theta}))-c(Q(\boldsymbol{\theta}))\right]\geq 0\,, \quad (5)$$

where, $h(\cdot)$ being the probability density function of H,

$$g(\theta_i)=\theta_i-(1-H(\theta_i))/h(\theta_i)\,. \quad (6)$$

This problem can be solved using Lagrangian methods (see the Appendix). That is, there is a nonnegative λ such that it is equivalent to solve the problem

$$\underset{\substack{\pi_1(\cdot),\dots,\pi_n(\cdot)\\ Q(\cdot)}}{\text{maximize}}\; E\Big[\textstyle\sum_i \pi_i(\boldsymbol{\theta})(\theta_i + \lambda g(\theta_i))u(Q(\boldsymbol{\theta}))-c(Q(\boldsymbol{\theta}))\Big]\,. \quad (7)$$

Note that the optimal $\pi_i(\boldsymbol{\theta})$ depends only on θ_i. We can write $\pi_i(\boldsymbol{\theta})=\pi(\theta_i)$, where this is 1 or 0 as $\theta_i+\lambda g(\theta_i)$ is or is not positive. Let us assume henceforth that $g(\theta)$ is nondecreasing in θ. Then $\theta+\lambda g(\theta)$ is nonincreasing so there is a θ^* such that $\pi_i(\theta)=1$ if $\theta>\theta^*$, and $\pi_i(\theta)=0$ otherwise.

The full solution of our problem is, in general, very complex. However, as in [6] we can use (7) to prove Theorem 1 below, showing that when n is large, the problem can be approximated, and a simple mechanism designed,

*This work is partly supported by EU's Fifth Framework project "Market Management of Peer-to-Peer Services" - MMAPPS (RTD No IST-2001-34201)

†Department of Computer Science, Athens University of Economics and Business, Patision 76, Athens 10434, GR

‡Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, UK

which obtains a value of the objective function that is within $o(n)$ of the maximum achievable. The intuition underlying this result is that when n is large, the law of large numbers assures the social planner that with high probability the number of peers who will have preference parameters in the interval $[\theta, \theta+\delta]$ will be within $o(n)$ of $nh(\theta)\delta$. Suppose he finds the mechanism that would be optimal if this were true exactly (i.e., finds optimizing θ and Q in (8) and (9) below). Then a mechanism setting $\pi_i(\eta) = 1\{\eta \geq \theta\}$ and $Q(\boldsymbol{\theta}) = Q$ will be nearly optimal for the original problem.

Theorem 1 *Suppose $u(\cdot)$ is bounded, and Q and θ are the optimizing values of Z and y in the problem*

$$\underset{y\in[0,1],Z\geq 0}{\text{maximize}} \left\{ nu(Z) \int_y^1 \eta h(\eta) d\eta - c(Z) \right\} \qquad (8)$$

subject to

$$n[1-H(y)]yu(Z) - c(Z) \geq 0\,. \qquad (9)$$

Let P be the problem defined by (1) *and* (5)*. Suppose we take as a feasible solution to P the variables $\pi_i(\eta) = 1\{\eta \geq \theta\}$ and $Q(\boldsymbol{\theta}) = Q$. Then the expected social welfare is equal to* (8) *and this in only $o(n)$ less than the maximum possible social welfare achieved by mechanism design.*

Let again θ and Q be the maximizing values of the decision variables in (8)–(9). Notice that all peers who are allowed to use the good pay the same fee of $f = \theta u(Q)$. Peer i may use the good if and only if $\theta_i \geq \theta$. This is the same as the condition that his net benefit should be nonnegative, i.e., $\theta_i u(Q) - f \geq 0$. The expected number of peers for which this holds is $n(1-H(\theta))$ and $Q = n(1-H(\theta))f$. This means that there is actually no need for the planner to intervene in an active manner. The planner should be viewed as the software designer. Once f has been set, the optimum $\pi(\cdot)$ and Q arise simply by peers making their own self-interested decisions.

As a simple illustrative example, suppose that $u(Q) = 0.6Q^{1/2}$, $c(Q) = Q$, and θ_i is uniformly distributed on $[0,1]$, so $H(x) = x$. The solution of (8)–(9) is $\theta = 1/4$, $Q = 0.0126563n^2$ and the social welfare is $0.006328125n^2$. The fee is $0.016875n$. We can compare this to the maximized social welfare that could be achieved if we were unrestricted by constraints of individual rationality and incentive compatibility. This social welfare would be $0.01125n^2$, which is achieved by $Q = 0.01125n^2$. The need to satisfy the constraints leads to a reduction in social welfare of 43.75%. Arguments in [9] show that if it were not possible to exclude participants then the social welfare would tend to infinity at a rate slower than $O(n^2)$ so that a vanishingly small amount of social welfare is obtained relative to that which can be obtained with exclusions.

2 A p2p file sharing system

We apply the above ideas to a problem of peer to peer file sharing by defining the appropriate functions u and c. Suppose that n peers make available files to share with one another. It is the number of distinctly different files which are shared that matters, so we must account for the possibility that more than one peer will make available the same file. Suppose that the utility obtained by peer i when the expected number of distinct supplied files is Q, is $\theta_i u(Q)$, where u is concave in Q. This is a key assumption in our modelling approach. Content availability is a public good: all peers benefit from the number of available files, and content is not consumed by downloading. Next we model the cost $c(Q)$ for provisioning Q. We might imagine that each peer provides the same number of files, say f, choosing these randomly from amongst a set of N (we relax this equal contribution assumption later). Then the expected number of distinct files that will be provided is

$$Q = N(1-(1-f/N)^n)\,,$$

so

$$f(Q) = N\left(1-(1-Q/N)^{1/n}\right)\,.$$

Suppose that each peer incurs a cost in providing files that is proportional to the number he contributes. For simplicity we let the constant of proportionality be one (noting that we could always re-scale the utility function). Thus the total cost is $c(F) = F$, where $F = nf(Q)$, and this is a convex increasing function of Q. Also, for any fixed Q, the cost nf rapidly increases with n to the asymptote of $-N\log(1-Q/N)$. This is greater than Q, the total cost if there were no duplication in the files peers supply [1] Note (see Figure 1) that for a large range of values of Q the cost is almost linear in Q, but then increases rapidly as Q approaches N. For example, for $n = 100$, we find $nf(Q) = Q + 0.00005Q^2 + 3.32 \times 10^{-9}Q^3 + \cdots$. This justifies an approximation $c(F(Q)) = Q$ when Q/N is of moderate size.

A slightly more sophisticated model might imagine that the peers share different numbers of files. Suppose $n\rho_i$ of peers each share i files, each of them choosing his i files randomly from amongst a set of $N = na$ files, $a > 0$. Let m be an upper bound on the number of files that any one peer can share, and $\sum_k \rho_k = 1$. The

[1] An alternative would be that a peer's cost is proportional to the rate at which he serves upload requests. Assuming files are equally popular this means that the total cost incurred by all the peers will be proportional to the product of the number of participating peers (that generate the requests) and the number of unique files, i.e., $c(Q) = (\sum_i \pi_i)Q$. If peers can only access files held within a certain neighbourhood of their location, this might be better modelled as $c(Q) = (\sum_i \pi_i)^\beta Q$, where $0 < \beta < 1$. There is a problem reproving Theorem 1 because the proof that Lagrangian methods work (proved here in the Appendix) no longer holds. This is for future research. We would expect to be able to address a limiting problem in which $u(Q)$ is concave in Q and $c(Q) = [n(1-F(\theta)]^\beta Q$.

expected number of distinct files supplied will be

$$Q = na\left[1 - \prod_{k=1}^{m}\left(1 - \frac{k}{na}\right)^{n\rho_k}\right]$$
$$= na\left[1 - e^{-\sum_k k\rho_k/a}\right] + o(1)\,. \qquad (10)$$

Now $F = n\sum_k k\rho_k$ is the total number of files provided by the peers. So we can again use the same approximation as above:

$$Q(F) = N\left(1 - e^{-F/N}\right)\,. \qquad (11)$$

Note that as $Q(F)$ is concave in F, we have that $\bar{u}(F) = u(Q(F))$ is also concave in F. For a given Q we will require

$$F(Q) = -N\log\left(1 - Q/N\right)\,. \qquad (12)$$

Of course when Q/N is not close to 1, $F(Q) = Q(1 + \frac{1}{2}(Q/N) + o(Q/N))$, so again $c(F) \approx Q$.

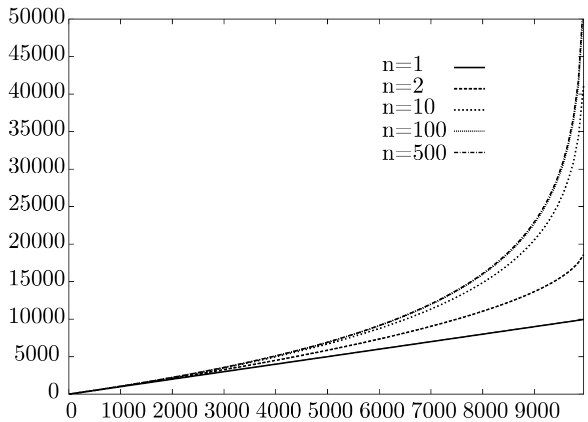

Figure 1: $nf(Q)$ for $N = 10000$; $n = 1, 2, 10, 100, 500$

Both of the above lead to models that are covered by the results of the previous section. The social planner wishes to design a mechanism which maximizes social welfare, subject to its being feasible, individually rational and incentive compatible.

Following the second model above, continuing the assumption that g is nondecreasing, dropping the bar from $\bar{u}(\cdot)$, and noting that $u(Q)$ is bounded by $u(N)$, we apply Theorem 1 and attain to within $o(n)$ of the optimum by choosing θ and F so as to

$$\underset{F,\,\theta}{\text{maximize}}\;\; nu(F)\int_{\theta}^{1}\eta h(\eta)\,d\eta - F$$

such that

$$n[1 - H(\theta)]\theta u(F) - F \geq 0\,.$$

Let F and θ be the maximizing values of the decision variables. We see that each peer who has a preference parameter of at least θ is included and pays the same fixed fee of $\theta u(F)$. Since the cost is linear in F this fee can be paid 'in kind', i.e., without monetary payments: each included peer pays his fee by contributing the same number of files: $F/n(1 - H(\theta))$.

Repeated rounds. In the limiting problem there is no reason that a peer should be other than truthful in representing himself to the system. If he knows that Q unique files will be shared and that the fee is f, then peer i should join if $\theta_i u(Q) \geq f$. In the non-limiting version of the problem, addressed by the mechanism design of Section 1, the individual rationality constraint (3) is in terms of expected value, so for some $\boldsymbol{\theta}_{-i}$ it can be that $\theta_i u(Q(\theta_i, \boldsymbol{\theta}_{-i}) - p_i(\theta_i, \boldsymbol{\theta}_{-i}) < 0$. When this happens, peer i might be tempted to defect and to not pay $p_i(\theta_i, \boldsymbol{\theta}_{-i})$. However, as file sharing system is intended to last for more than one time step, we could operate a 'tit-for-tat'-like protocol, that would penalize such defection, for example, by threatening to exclude peer i at a later time when $\boldsymbol{\theta}_{-i}$ is such that his net benefit would be positive. We are imagining that $\boldsymbol{\theta}$ is not fixed, but varies over time, as from time to time the peers' preference parameters are freshly sampled from H. The effect of the threat would be to make peer i willing to participate on such occasions that he has to accept a short-term negative net benefit, knowing that on average he will benefit, as is guaranteed by (1) and (3). If every peer's preferences parameter varies over time with the distribution H, each will obtain on average $1/n$th of the maximized social welfare.

3 Stability

Suppose that the social planner designs a mechanism on the basis that there are n peers. He expects that $(1 - H(\theta))n$ of them will pay a fee of $f = \theta u(F)$. Since the fee is paid 'in kind' and equates to providing f files, the total number of files that are provided will be $F = (1 - H(\theta))nf$.

Suppose that there are indeed n peers, but initially some of them are dubious that F will be as large as the planner claims. Consequently, some do not participate and the number of files that is initially provided is $F_1 < F$. Once the peers have observed F_1, those peers with $\theta_i > f/u(F_1)$ will realise that it is to their advantage to participate. Their fees will provide F_2 files where

$$F_2 = \left(1 - H\left(\frac{f}{u(F_1)}\right)\right)nf\,. \qquad (13)$$

Write this as $F_2 = \phi(F_1)$ and imagine iterating $F_{k+1} = \phi(F_k)$, $k = 1, 2, \ldots$. In general, there can be more than one root to $F = \phi(F)$. For example, suppose $u(F) = 0.6F^{1/2}$, $f = 5$, $n = 120$, and θ_i is uniformly distributed on $[0, 1]$. Then $\phi(F) = \left(1 - 5/0.6F^{1/2}\right)(120)(5)$. In this example there are two roots, $F = 100.00$ and $F = 320.87$. One can easily prove that if F_1 exceeds the smaller root then F_k tends to the larger root as k tends to infinity. Otherwise $F_k \to 0$. For $F = 100$ the social welfare is 10, whereas for $F = 320.87$ it is 184.4. Thus the greater F, to which the system converges, is also the root for which a greater number of peers participate and the greater social welfare is achieved.

4 Heterogeneous file popularity

We can extend the above ideas to circumstances in which files have different popularities. For example, suppose the following scenario. There are popular files and less popular files. A popular one is requested at twice the rate of an unpopular one (and so generates twice the cost to a peer who provides it), but is also twice as valuable. The total cost corresponding to the total upload rate is[2] $c(F_1, F_2) = 2F_1 + F_2$ and the utility is now, say, $u(2F_1 + F_2)$. The analysis is much as above. Each peer is asked to contribute f_1 type 1 files and f_2 type 2 files. Notice that because the ratio of value to cost is the same for both file types, it is only the value of $f = 2f_1 + f_2$ that actually matters. So the planner has no preference for the precise combination of f_1 and f_2 by which a peer makes his contribution. Moreover, the planner can check that a peer is making his required contribution simply by verifying that the total rate at which files are uploaded from the peer is consistent with f. This agrees with a commonly held belief amongst technologists that the only thing that need be measured to police a peer's contribution to a file sharing network is the rate at which uploads are made from the peer. If the value/cost ratios had been different for the two file types, then it would have been optimal to share only one type of file (the one with greater value/cost ratio). This fact, and notions of equilibrium economics, suggest that all file types that are actually worth sharing will effectively have the same value/cost ratio and hence the upload rate is to be a good measure of a peer's contribution.

5 Group formation

It may sometimes be possible for the global planner to distinguish between types of peers and use this information to model the distributions of their preference parameters more accurately. Suppose, for example, that the population of peers consists of both ISDN dial-up users and DSL users with a uniform distribution of their preference parameters on $[0, 0.5]$ and $[0.5, 1]$ respectively. This reflects the fact that DSL users value more and benefit more from the shared content than do the dial-up users.

As we have seen, the asymptotic problem that must be solved to determine the optimal fixed fees is quite simple. By computational experiments we can gain some important insights into the formation of groups. There are three possible scenarios that the planner could pursue. In the first scenario, dial-up users and DSL users form distinct groups A and B and do not share content. In the second, they form a single group, in which they share content, but their types (DSL or dial-up) are indistinguishable to the global planner. He only knows the initial proportions of users of each type. The third scenario is the same as the second, except that peers now form a group in which they are distinguishable, i.e., they disclose their types to the planner. Which of these is best for each user type? Does the answer depend on the relative number of the users of each type? By solving our fixed fee model we obtain some interesting insight (see Figure 2).

(a) As the proportion of DSL users decreases, the DSL users prefer the second scenario: a large indistinguishable group; their next preferred option is the first scenario, when they form their own distinct group. Preference for the first scenario becomes even more marked as the distribution of the preferences becomes less spread (e.g., uniform on $[0.8, 1]$, rather than on $[0.5, 1]$); conversely, as it becomes more spread they tend to prefer the second scenario.
(b) As their proportion decreases, dial-up users favour the third scenario since at the SW optimum the DSL users offer the majority of the content. The second scenario, of a single indistinguishable group, is not so attractive, as they are forced to pay a substantial fixed fee. When their number is small they gain by the large amount of content made available by the DSL users, compared to the content they would obtain in their own group. This difference becomes negligible when they are the dominant type in the mixture.

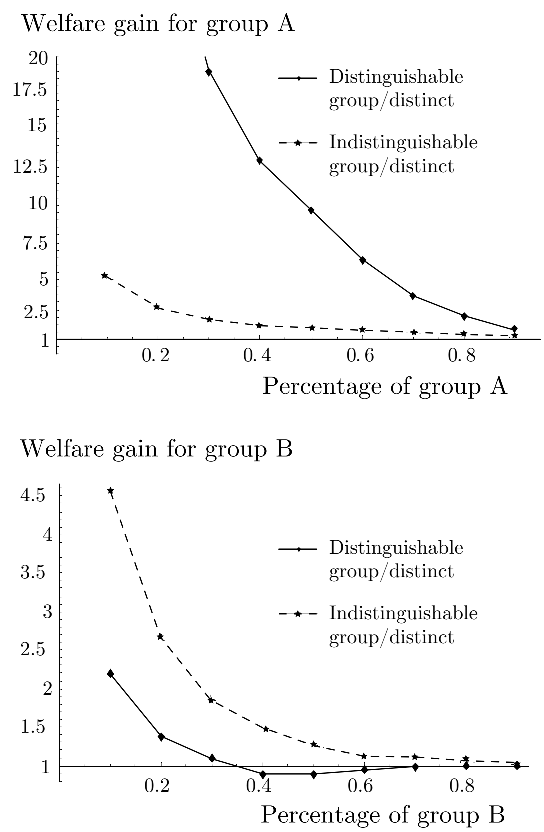

Figure 2: Welfare gain of dial-up (group A) and DSL (group B) users when forming groups in which they are distinguishable and indistinguishable to the planner compared to the welfare they obtain when they form distinct groups.

[2] more precisely, by the model of footnote 1, proportional to

The underlying reason is rather simple: when forming a larger indistinguishable group, DSL users benefit from the larger content available and the fact that the fixed fee, which is the same for all, is less than the 0.5 they would pay if they could be distinguished. Furthermore, dial-up users who can afford it contribute a substantial amount to the common cost, whereas if they can be distinguished, they free-ride on the DSL users by paying zero fees.

An interesting issue for further research is to design a sub-optimal policy that provides incentives for both types to combine into one group. For instance, we could reduce the cost imposed on DSL user by limiting the download rates of dial-up users, or change the fee of DSL users to make it attractive for them to join and declare their type. It would also be interesting to pursue the possibility of permitting more than one level of participation; we could arrange for those peers who desire a greater average rate of uploading to pay a greater fee. Ideally, the allowed options would target the different peer types and provide the incentives for each peer to self-select the option that is targeted at his type. For instance, imagine that each peer has the choice between an effective u of $u(F_1+F_2)$, for a fee of f_2, or $u(\rho(F_1+F_2))$, for a fee of f_1, where $\rho < 1$ and $f_1 < f_2$. Note that adding congestion cost as in the footnote 1 reduces the incentives for building larger groups.

In the previous analysis we assumed that it is equally costly for DSL and dial-up users to share a file. This may be a reasonable approximation when the up-link speed of DSL users is small. But other times this may not be the case. Suppose that our two user types both have θs distributed uniformly in $[0,1]$, but their costs for sharing f files are af and f respectively, $a > 1$. The results of our experiments for this model are similar to those above. The type 1 peers, for whom file sharing is more costly, benefit when all peers disclose their types, whereas the type 2 peers may not. Type 1 peers are better off both because of the larger content selection and because they contribute less.

Best splitting. Suppose the global planner could obtain information that would allow him to partition peers into smaller subgroups, such that the preference parameters in each subgroup are distributed on non-overlapping subintervals of $[0,1]$. Can he always gain by doing this? The answer is no. Suppose that the initial distribution is uniform on $[0,1]$. Then, depending on the growth of the function $u(F)$, a finite splitting of this initial interval may be optimal, and there will be nothing extra to gain from further splitting. For instance, if $u(F) = F^a$, and $a < 1/2$, then splitting $[0,1]$ in half is enough. If $a < 2/3$, then splitting $[0,1]$ into three equal subintervals is enough, and so on.

6 Parameter discovery

We have assumed that n, $u(\cdot)$ and $H(\cdot)$ are known to the social planner when he computes the optimal fee f. It would be an interesting issue for further research to see how, in absence of this knowledge the planner might take advantage of what we have already learned about the form of the optimal policy to design an adaptive policy that learns the parameters. Here, we make only some preliminary remarks. If only n is not known, the planner could set an entrance fee and then observe the size F at which the system stabilizes (after iterations of (13)). Then n can be estimated by $F/(1-H(f/u(F))$.

If only H is not known, and peers actually declare their preference parameters, then the planner might estimate H from the empirical distribution of the declared preference parameters, say $\hat{H}$, and then implement the solution that is optimal for $\hat{H}$. For n large he should have $\hat{H} \approx H$, provided the peers are truth-telling. In a repeated game formulation, in which the preference parameter of a typical peer is repeatedly sampled from $H(\cdot)$, then his average net benefit will $1/n$th of the total social welfare and it will indeed be optimal for him to be truth-telling.

However, if a peer's preference parameters are chosen once for all or remain relatively static for a long time then there may be an incentive for a group of peers to lie about their preference parameters, hoping to fool the social planner into mis-estimating H and doing better for themselves thereby. Consider the following example (for which calculations were done with Mathematica). Suppose that $N = 100000$, $n = 100$, $u(Q(F)) = 2\sqrt{N(1-\exp(-F/N))}$. Suppose that θ_i is one of {0.25, 0.5, 0.75, 1.0} with frequencies 0.4, 0.4, 0.1 and 0.1 respectively. The mechanism which maximizes social welfare takes $\theta = 0$ and $F = 2183.5$. Each of the peers who has $\theta = 0.5$ makes net benefit of 24.6390. However, suppose that the peers who have $\theta = 0.5$ act in concert and arrange for one-quarter of them to untruthfully declare it as 0.75. The central planner will conclude that the frequencies of the four possible parameter values are 0.4, 0.3, 0.2 and 0.1, and for this he maximizes social welfare by taking $\theta = 0$, $F = 2411.14$. Under this mechanism, the peers who have $\theta = 0.5$ will now make a greater net benefit of 24.6975. They profit from being untruthful. (However, if all the peers who have $\theta = 0.5$ untruthfully declare it as 0.75 then they do not do better than if they are all truthful.) Another way to view this example is that it shows there can be more than one Nash equilibrium in the n-person game being played by the peers.

Notes

Another application of the ideas in Section 1 is in our paper on wireless LANs [6]. Our key Theorem 1 is motivated by reading [9] [11], but our proof (in [6]) is perhaps simpler, being a fairly straightforward application of the law of large numbers (though at the price of missing some finer asymptotic detail). We are uniquely able to deal with multiple constraints because of our new method of establishing the applicability of Lagrangian methods for the mechanism design problem, as we explain for one constraint in the appendix below.

Segal [14] has considered a related setup in which a monopolist is trying to maximize his profit by selling units of a single good to n buyers (agents). He designs a mechanism such that the price each buyer pays is a function of all n buyers' declared valuations (our $\boldsymbol{\theta}$). He supposes that the underlying distribution F of these valuations is unknown and shows how to define the mechanism so as n tends to infinity, it achieves the optimal monopoly profit that could be obtained had F been known. Essentially, the formulae are the same, except that θ_i is taken to be a sample from $\hat{F}_i(\cdot|\boldsymbol{\theta}_{-i})$, which is F conditional on knowing $\boldsymbol{\theta}_{-i}$. Like us, he finds that the limiting policy is to offer each buyer the same fixed price, but he does not derive the $o(n)$ error, deal with the additional optimization over Q that our public good model involves, nor explain any application to a problem of p2p resource allocation.

A comparison of complete information and incomplete information schemes in the context of our public good formulation is presented in [2]. The inefficiency of p2p systems has been pointed out in [1] [13] and designing incentives for contribution using reciprocity concepts is discussed in [8] [10] [3] [7] among others. There exist several real p2p applications which use reciprocity-based (e.g., [4]) or minimum contribution (see Direct Connect —http://www.neo-modus.com) mechanisms to provide the necessary incentives to peers to share their resources. Implementation issues such as how the accounting of the information can be performed, how incentive rules and exclusions can be enforced, security issues, etc. are discussed in [5] [12] [15].

We are grateful to Robin Mason for pointing out the important connection of p2p with public good models and fixed fee contributions and Peter Norman for some helpful discussions about mechanism design. A longer version of this paper will be placed at
www.statslab.cam.ac.uk/Reports/2004/2004-01.pdf.

References

[1] E. Adar and B. A. Huberman. Free Riding on Gnutella. *First Monday*, 5(10), October 2000.

[2] P. Antoniadis, C. Courcoubetis, and R. Mason. Comparing Economic Incentives in Peer-to-Peer Networks. *Special Issue on Network Economics, Computer Networks, Elsevier*, forthcoming.

[3] C. Buragohain, D. Agrawal, and S. Suri. A Game Theoretic Framework for Incentives in P2P Systems. In *Proceedings of the Third IEEE International Conference on Peer-to-Peer Computing (P2P 2003), Linkoping, Sweden*, 2003.

[4] B. Cohen Incentives Build Robustness in BitTorrent. *Workshop on Economics of Peer-to-Peer Systems, Berkeley, CA*, 2003.

[5] MMAPPS Consortium. Market Management of Peer-to-Peer Services - Introduction and Overview. Available at `http://www.mmapps.org/`, 2004.

[6] C. Courcoubetis and R. R. Weber. Asymptotics for Provisioning Problems of Peering Wireless LANs with a Large Number of Participants. In *Proceedings of WiOpt'04: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, University of Cambridge, UK*, 2004.

[7] M. Feldman, C. Papadimitriou, J. Chuang, and I. Stoica. Free-Riding and Whitewashing in Peer-to-Peer Systems. *3rd Annual Workshop on Economics and Information Security (WEIS04)*, May 2004.

[8] P. Golle, K. Leyton-Brown, I. Mironov, and M. Lillibridge. Incentives for Sharing in Peer-to-Peer Networks. In *Proceedings of the 3rd ACM conference on Electronic Commerce*, October 2001.

[9] M. F. Hellwig. Public-good Provision with Many Participants. *Review of Economic Studies*, forthcoming.

[10] K. Lai, M. Feldman, I. Stoica, and J. Chuang. Incentives for Cooperation in Peer-to-Peer Networks. *Workshop on Economics of Peer-to-Peer Systems, Berkeley, CA*, 2003.

[11] P. Norman. Efficient Mechanisms for Public Goods with Use Exclusions. *Review of Economic Studies*, forthcoming.

[12] Andy Oram. *"Peer-to-Peer : Harnessing the Power of Disruptive Technologies"*. O'Reilly & Associates, March 2001.

[13] S. Saroiu, P. K. Gummadi, and S. D. Gribble. A Measurement Study of Peer-to-Peer File Sharing Systems. In *Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02)*, 2002.

[14] I. Segal. Optimal Pricing Mechanisms with Unknown Demand. *The American Economic Review*, 93(3), 509–529, 2003.

[15] D. Wallach. A Survey of Peer-to-Peer Security Issues. In *International Symposium on Software Security, Tokyo, Japan*, 2002.

A Appendix

To see that the problem of (1) and (5) can be addressed by Lagrangian methods we reformulate it as

$$\underset{x_1(\cdot),\ldots,x_n(\cdot),Q(\cdot)}{\text{maximize}} \; E\left[\textstyle\sum_i x_i(\boldsymbol{\theta}) - c(Q(\boldsymbol{\theta}))\right]$$

subject to constraints on the nonnegative decision variables, $x_1(\boldsymbol{\theta}),\ldots,x_n(\boldsymbol{\theta}),Q(\boldsymbol{\theta})$ of $x_i(\boldsymbol{\theta}) - \theta_i u(Q(\boldsymbol{\theta})) \leq 0$, for all $i,\boldsymbol{\theta}$ and $-E\left[\sum_i x_i(\boldsymbol{\theta})\frac{g(\theta_i)}{\theta_i} - c(Q(\boldsymbol{\theta}))\right] \leq 0$. The objective function is concave in the decision variables and the left hand sides of the constraints are convex in the decision variables. These are sufficient to guarantee existence of a nonnegative λ such that it is equivalent to solve

$$\underset{x_1(\cdot),\ldots,x_n(\cdot),Q(\cdot)}{\text{maximize}} \; E\left[\textstyle\sum_i x_i(\boldsymbol{\theta})\left(1+\lambda\frac{g(\theta_i)}{\theta_i}\right) - c(Q(\boldsymbol{\theta}))\right]$$

such that $x_i(\boldsymbol{\theta}) - \theta_i u(Q(\boldsymbol{\theta})) \leq 0$ for all $i,\boldsymbol{\theta}$. This may be seen to be equivalent to (7).