Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Computer Networks 46 (2004) 133–146
COMPUTER NETWORKS
www.elsevier.com/locate/comnet

Comparing economic incentives in peer-to-peer networks ☆

Panayotis Antoniadis [a,*], Costas Courcoubetis [a], Robin Mason [b]

[a] Department of Computer Science, Athens University of Economics and Business, Patison 76, 10434 Athens, Greece
[b] Department of Economics, University of Southampton, Highfield, Southampton SO17 1BJ, UK

Available online 12 April 2004

Abstract

Users who join a peer-to-peer network have, in general, suboptimal incentives to contribute to the network, because of the externalities that exist between them. The result is an inefficient network where the overall levels of contribution are less than would be the case if each peer acted in the interests of the entire network of peers. Incentives provided in the form of prices or contribution rules that require no money transfers can play an important role in reducing these inefficiency effects. The problem in designing such incentive schemes is information: Designing an optimal incentive scheme requires complete knowledge of the types and preferences of the individual peers and their identities. In this paper we discuss the above issues in terms of a simple but representative example by introducing the basic economic concepts and models. We then investigate the practical issue of designing several simpler incentive schemes requiring less information and compare their efficiency loss to the optimal. We show using numerical analysis that these schemes converge to a fixed proportion of the full information optimal as the number of peers in the network becomes large. This result means that it is not necessary to collect large amounts of information, or to undertake complicated calculations, in order to implement the correct incentives in a large peer-to-peer network.

Keywords: Economics; p2p networks; Public goods; Incomplete information; Incentives; Mechanism design

1. Introduction

Users who join a peer-to-peer network have, in general, incorrect incentives to contribute to the network, because of the externalities that exist between them. The result is an inefficient network: The overall levels of contribution are less than would be the case if each peer acted in the interests of the entire network. To attain full efficiency using an incentive scheme, a network manager must (i) have complete information about the payoffs of all peers who join the network, (ii) be able to offer each peer a personalised scheme, and (iii) where necessary, enforce (or make payments to ensure) participation. None of these conditions are likely to be achievable in practice: network managers are typically poorly informed about peers' payoffs; even if managers have good information, they may find it difficult to implement incentive schemes that are based on the identities

☆ This work is partly supported by EU's Fifth Framework project 'Market Management of Peer-to-Peer Services'—MMAPPS (RTD No. IST-2001-34201).
* Corresponding author.
E-mail addresses: antoniad@aueb.gr (P. Antoniadis), courcou@aueb.gr (C. Courcoubetis), robin.mason@soton.ac.uk (R. Mason).

1389-1286/$ - see front matter
doi:10.1016/j.comnet.2004.03.021

of peers; and coercion or extensive payments may be infeasible.

In this paper, we show in the context of a simple model that the inefficiency that arises because of these practicalities becomes small as the number of peers in a network becomes large. To do this, we determine the 'first-best' incentive scheme—one that uses complete information and is personalised. This scheme sets prices, one for each peer per unit of contribution that the peer makes to the network. [1] Equivalently, the scheme can use a rule, i.e., set a contribution level for each peer. The prices and contribution levels are calculated using knowledge of all peers' payoffs. The first-best incentive schemes raise the three issues that are highlighted above. The first-best rule may involve some peers receiving a negative payoff from being part of the network. We therefore consider rules when peers must be given incentives to participate. First-best prices are personalised. So next we consider prices that are uniform across peers, but that still use complete information about peers' payoffs. First-best rules and prices use full information about peers' payoffs. Hence we consider incentive schemes that can be used when the network manager does not know the actual payoffs of peers and cannot personalise. We derive 'average' rules and prices, as well as a 'fixed fee' scheme that requires peers to contribute (the same) minimum level to join the network.

As is to be expected, network efficiency is highest under the first-best scheme. We show, using simulation analysis, the loss in efficiency that results from departing from the first-best. We show, however, that for certain schemes, the degree of inefficiency decreases as the number of peers in the network becomes large. Our simulation results suggest a ranking of the incentive schemes in large networks: the full information rule with participation incentives yields the highest efficiency; the fixed fee scheme yields lower efficiency, but slightly more than the average rule; uniform and average prices yield the lowest level of efficiency. Also the overall efficiency of non-personalised rules increases as the variance of the peers' payoffs decreases. This suggests that if peers, in terms of their payoffs, can be better classified (by using some objective characteristic like the access speed of their modem) in groups having a higher degree of homogeneity, then simple incentive schemes will perform even better.

There are two reasons why these results are relevant: First, many peer-to-peer networks are large, for example, the number of peers on Gnutella and Kazaa runs into the millions. Secondly, the results mean that it is not necessary to collect large amounts of information, or to undertake complicated calculations, in order to implement the correct incentives in a peer-to-peer network. Instead, a relatively simple scheme achieves a high level of efficiency: All that is required is for peers to be charged a uniform contribution to join the network. The level of the contribution can be worked out using a simple calculation. Also rules have the practical advantage that they require no actual money transfers between the peers themselves or between peers and a third party. The latter is attractive in a large, decentralised system in which implementing a currency can be difficult.

Many existing p2p file sharing applications have recently started to apply simple system-specific rules to provide for the suitable incentives to peers to contribute. In later versions of Kazaa, for example, the contribution of each peer is computed and, according to its level, peers have a corresponding priority in case of congestion. However, such simple compensating rules still allow for a significant degree of free-riding. In many real cases, more strict, reciprocity-based, e.g. [3], or minimum contribution mechanisms (see Direct Connect—http://www.neo-modus.com) are being employed. Actually, our fixed-fee scheme resembles the entrance rules required by Direct Connect groups. [2]

[1] If a peer makes several different contributions—for example, files shared and bandwidth—then a price is set for each aspect of the contribution. Note that prices can be positive or negative i.e., the peer might pay or be paid by the network manager, depending on the nature of externalities between peers. See Section 2.

[2] For example, in many groups, a peer should share music files of a minimum total specified size in order to join. The quality of the files shared is checked through some means of social control and peers that do not meet their obligations are expelled.

The inefficiency of p2p systems has already been pointed out by several researchers (see [1,6, 8,15]). But these papers did not provide an appropriate economic modeling framework where these important incentive issues could be formally defined and compared. Most of the relevant work in the economics literature like [7,13] concerns the provisioning of public goods. These models cannot be directly used for p2p networks. For instance, they assume the possibility of money transfers between peers. In our paper we make this important connection and explain how to interpret important results from mechanism design for public good provisioning in the context of p2p networks.

We must stress to the reader that this paper focuses on economic aspects of p2p. It leaves out significant implementation details such as how the accounting of the information will be performed, how incentive rules and exclusions will be enforced, security issues, etc. Some references discussing such issues are [11,14,16]. Thus, some of the schemes compared for their economic efficiency may be harder to implement than others.

The rest of the paper is structured as follows: The next section introduces the basic model that identifies the incentive problem in a peer-to-peer network. Section 3 describes various schemes that can be used to correct incentives. We distinguish between the various constraints that schemes might recognise participation incentives (Section 3.2), anonymity (Section 3.3), and information limitations (Section 3.4). Section 4 examines numerically the levels of efficiency achieved by these schemes for different network sizes. Section 5 summarises the paper's findings and suggests avenues for future work.

2. An economic model of peering

In this section, we develop a simple model of file sharing as a canonical example of a peer-to-peer activity. We discuss in Section 5 how to extend the analysis beyond this basic example. Each peer i decides on the number of files f_i it shares. The payoff of the agent from the p2p network with N peers and an N-vector $\boldsymbol{f} = (f_1, \ldots, f_N)$ of shared files is

$$u_i(\boldsymbol{f}) = \theta_i v\left(\sum_{j=1}^{N} \sqrt{f_j}\right) - f_i. \quad (1)$$

Here, the 'benefit' function $v(\cdot) \geqslant 0$ is assumed to be continuously differentiable, increasing and strictly concave in its argument. The benefit function is the same for all peers; and each peer faces the same cost for each file shared, normalised to 1. Peers differ in the payoff parameter θ_i: These are drawn from a distribution F with support normalised to the unit interval $[0, 1]$. For any N-vector $\boldsymbol{\theta}$ of payoff parameters, without loss of generality order the peers so that $\theta_1 \leqslant \theta_2 \leqslant \cdots \leqslant \theta_N$.

In this model, file sharing is a public good: The copying of a file by one peer does not prevent another peer also from copying it; and, for the moment, peers cannot be prevented from copying files offered by others. (We consider the possibility of peer exclusion later.) [3] The standard problem with public goods is that the (Nash) equilibrium in which agents determine their contribution levels to maximise their own utility is, typically, inefficient relative to the social optimum, in which contributions are set to maximise the sum of all utilities, due to free-riding. In our case of a p2p network, we show that in equilibrium, the number of shared files is too low, relative to the efficient level.

There are several features of the payoff function in Eq. (1) to comment on

- any difference between the peers is captured in the payoff parameter θ_i. The separability that this assumption enforces simplifies the expressions that are derived later, except where we indicate, in Section 3.4, the assumption can be dispensed with, at the cost of added complexity;
- the argument of the benefit function is the sum of the square roots of the files shared by each peer. This models file duplication: a contribution of f_i files by peer i results in a 'useful'

[3] In the formulation, it is clear that f_i can be any action that imposes positive externalities.

amount of content which is less than f_i, since some of these files are already made available by other peers. Again, this assumption simplifies calculations, but is otherwise not crucial. We could assume that the argument of the benefit function is $\psi(f_1, f_2, \ldots, f_N)$ for some function ψ. In order for solutions to the various maximisation problems to exist, we require that $\partial\psi/\partial f_i \neq \partial\psi/\partial f_j$ for $i \neq j$ (in the differentiable case), otherwise, ψ can be arbitrary;
- each peer has the same cost function for sharing files: linear with no fixed cost. Except for the analysis in Section 3.4, this assumption is not crucial for the qualitative results we obtain. It allows payments to be done in 'kind' (i.e., files) instead of actual money which would be then converted to files according to a more complicated cost function;
- peers exhibit some degree of altruism since they obtain benefit from their own contributions. This modelling assumption is consistent with observed peer behaviour in existing content sharing systems, where most peers contribute some (low) amount of files with no explicit external incentives.

Peer i seeks to maximise its payoff: $\max_{f_i} u_i(\boldsymbol{f})$, taking as given the files shared by all other peers. A Nash equilibrium is comprised of a vector of files shared such that all peers are simultaneously maximising their payoffs. The first-order condition for peer i's maximisation problem is

$$\frac{\theta_i}{2\sqrt{f_i}} v'\left(\sum_{j=1}^{N} \sqrt{f_j}\right) - 1 \leqslant 0, \quad f_i \geqslant 0, \tag{2}$$

where $v'(\cdot)$ denotes the derivative of the benefit function. Let the equilibrium choice of peer i be denoted $\hat{f}_i$. In equilibrium, only peers with high enough values of θ_i will make strictly positive contributions. Let the payoff parameter of the marginal peer, i.e., the peer with the lowest value of θ_i out of those that make positive contributions, be denoted θ_e, where $e \in \{1, \ldots, N\}$; and let $E \equiv \{e, \ldots, N\}$. For those peers who make positive contributions, i.e., with $i \in E$, the equilibrium contribution is

$$\hat{f}_i = \left(\frac{\theta_i}{2} v'\left(\sum_{j=e}^{N} \sqrt{\hat{f}_j}\right)\right)^2. \tag{3}$$

Hence

$$\sum_{j=e}^{N} \sqrt{\hat{f}_j} = \frac{\Theta_e}{2} v'\left(\sum_{j=e}^{N} \sqrt{\hat{f}_j}\right), \tag{4}$$

where $\Theta_e \equiv \sum_{j=e}^{N} \theta_j$. Eq. (4) is an implicit equation for the variable $\hat{F} \equiv \sum_{j=e}^{N} \sqrt{\hat{f}_j}$; with the concavity assumption on $v(\cdot)$, there is a unique solution for $\hat{F}$ (which, it should be noted, depends on the value of Θ_e—we return to this point below), and hence a unique value $\hat{\boldsymbol{f}}$ for the N-vector of equilibrium contributions.

Finally, the identity of the marginal peer is determined by the indifference condition

$$\theta_e v(\hat{F}) - \hat{f}_e = 0. \tag{5}$$

Due to integer constraints, this indifference condition may not hold exactly; if it does not, the marginal consumer is identified as the 'last' peer with a non-negative payoff.

In summary, in the Nash equilibrium, $N - e$ out of the N peers—those with the highest payoff parameters—make strictly positive contributions, given by Eq. (3); the others share no files. The total number of files shared in equilibrium is given by the solution to Eq. (4).

Contrast this characterisation of the Nash equilibrium with the solution that arises when the network is run by a benevolent and fully-informed manager who chooses the contributions of peers to maximise the total payoff of all peers:

$$\max_{\{f_1,\ldots,f_N\}} \sum_{i=1}^{N} \left(\theta_i v\left(\sum_{j=1}^{N} \sqrt{f_j}\right) - f_i\right).$$

The first-order condition for peer i is

$$\frac{\theta_i}{2\sqrt{f_i}} v'\left(\sum_{j=1}^{N} \sqrt{f_j}\right) - 1 + \frac{\sum_{j\neq i} \theta_j}{2\sqrt{f_i}} v'\left(\sum_{j=1}^{N} \sqrt{f_j}\right) \leqslant 0, \quad f_i \geqslant 0. \tag{6}$$

Denote the resulting number of shared files f_i^* for $i \in \{1, \ldots, N\}$. Note that in this solution, if it is optimal for any peer to contribute files, then it is optimal for all peers to share files (since the marginal costs and benefits of file sharing are the same for all peers).

An important difference between Eqs. (2) and (6) is the presence in the latter of externalities, which can be measured by

$$\frac{\sum_{j\neq i}\theta_j}{2\sqrt{f_i}}v'\left(\sum_{j=1}^{N}\sqrt{f_j}\right)$$

for peer i. In the next section, we analyse approaches to ensure that peers consider these externalities when deciding how many files to share.

The contribution of each peer is

$$f^* = \left(\frac{\Theta}{2}v'(F^*)\right)^2, \tag{7}$$

where $\Theta \equiv \sum_{j=1}^{N}\theta_j$ and $F^* \equiv N\sqrt{f^*}$. This gives the implicit equation for the total number of shared files in this case:

$$F^* = \frac{N\Theta}{2}v'(F^*). \tag{8}$$

(There is a unique solution to this equation, due to the strict concavity of $v(\cdot)$.) Comparison of Eq. (8) with Eq. (4) shows that $F^* \geqslant \hat{F}$, since $\Theta \geqslant \Theta_e$ and $N \geqslant 1$.

In summary, in the 'social optimum', in which every peer is concerned about the payoffs of all peers in the network, each peer shares f^* files, where $f^* \geqslant \max_i \hat{f}_i$; consequently the total number of files shared in the social optimum is greater than in equilibrium. This means that the total payoff attained in the social optimum, denoted S_{SO}:

$$S_{SO} \equiv \sum_{i=1}^{N}\left(\theta_i v\left(\sum_{j=1}^{N}\sqrt{f_j^*}\right) - f_i^*\right)$$

is greater than the total payoff attained in equilibrium, denoted S_{NE}:

$$S_{NE} \equiv \sum_{i=1}^{N}\left(\theta_i v\left(\sum_{j=1}^{N}\sqrt{\hat{f}_j}\right) - \hat{f}_i\right),$$

i.e., $S_{SO} \geqslant S_{NE}$. In short, equilibrium is inefficient.

3. Incentive schemes

In this section, we derive analytical expressions for a number of incentive schemes that might be used to correct the externalities and resulting inefficiencies identified in the previous section. We start by considering the 'first-best', when the scheme designer [4] has complete and perfect information about the payoff parameters of all peers and is able to set personalised incentives for each peer. We then move on to less ideal situations, in which participation incentives must be given; personalisation is not possible and information is incomplete.

3.1. The first-best

The simplest method of ensuring full efficiency of the peer-to-peer network is to require each peer to share f^* files. In summary

Scheme 1 (*First-best rule*). Each peer shares f^* files, where f^* is given by Eq. (7).

An equivalent approach uses prices to give peers the correct incentives. In the simple case of file sharing that we are considering, the prices are *subsidies* paid [5] to peers to encourage them to share files. If negative externalities (e.g., congestion) were present, then payments might occur in the opposite direction: The peers would be charged to discourage the activity generating negative externalities. By comparing Eqs. (2) and (6), it is apparent that the appropriate price for peer i is

$$p_i \equiv \frac{\sum_{j\neq i}\theta_j}{2\sqrt{f^*}}v'(F^*);$$

[4] In our view, the scheme designer coincides with the designer of the software of a p2p application and does not need to be implemented by a physical entity participating in the actual operation of the system.

[5] Such subsidies may be paid by a third party who indirectly benefits from the efficient operation of the system. Although this is a standard procedure in the provision of public goods, it may be impractical in many realistic p2p contexts where such a payment system is hard to implement. In any case we use it for comparison purposes.

substituting in the expressions for f^* and F^*, this reduces to the very simple form

$$p_i = \frac{\Theta_{-i}}{\Theta}, \tag{9}$$

where $\Theta_{-i} \equiv \sum_{j\neq i} \theta_j$. This leads to

Scheme 2 (*First-best prices*). Peer i is paid a price p_i per file shared, where p_i is given by Eq. (9).

There are three facts to note about the first-best schemes: First, both use complete information—the rule and prices depend on the vector $\boldsymbol{\theta}$ of payoff parameters. [6] Secondly, the rule may involve some peers (those with low values of θ) receiving a negative payoff from being part of the network. This raises the issue of whether those peers can be forced to be part of the network, or whether they must be given incentives to join. Thirdly, the first-best prices are personalised—each peer faces a different subsidy for each file that it shares. (Such prices are often referred to as *Lindahl prices* in economics—see for example [10].) The first-best rule, on the other hand, is not personalised—each peer shares the same number of files.

3.2. Complete information rules with participation incentives

In this section, we consider rules that use information about peers' payoffs and provide incentives for peers to join the network. The problem facing the manager is to choose contribution levels to maximise total payoffs,

$$\max_{\{f_1,\dots,f_N\}} \sum_i \left(\theta_i v\left(\sum_{j=1}^N \sqrt{f_j}\right) - f_i\right);$$

subject to the constraints that each peer must receive a non-negative payoff:

$$\theta_i v\left(\sum_{j=1}^N \sqrt{f_j}\right) \geqslant f_i \quad \forall i. \tag{10}$$

The Lagrangian for this constrained maximisation problem is

$$L \equiv \sum_i \theta_i(1+\lambda_i)v\left(\sum_{j=1}^N \sqrt{f_j}\right) - \sum_i (1+\lambda_i)f_i. \tag{11}$$

The first-order conditions are

$$\sqrt{f}_i = \frac{\sum_j \theta_j \rho_j v'\left(\sum_{l=1}^N \sqrt{f_l}\right)}{2\rho_i}, \tag{12}$$

where $\rho_i \equiv 1 + \lambda_i$. In these equations the λ_i are non-negative. They are also decreasing for the following reason. For the i's where the participation constraint (10) binds, we have that $\lambda_i > 0$ and also that $f_i = \theta_i v(\cdot)$. Combining this with Eq. (12), we obtain that $\theta_i(1+\lambda_i)^2$ must be constant. Since we assumed that $\theta_1 \leqslant \theta_2 \leqslant \cdots \leqslant \theta_N$, we must also have that the λ_i's are decreasing in i.

The above suggests that rules take the following form: for some threshold k,

$$\tilde{f}_i = \begin{cases} \theta_i v(\tilde{F}), & i < k, \\ \theta_k v(\tilde{F}), & i \geqslant k, \end{cases} \tag{13}$$

where $\tilde{F} \equiv \sum_{j=1}^N \sqrt{\tilde{f}_j}$. Given these levels of contributions, the threshold k can then be chosen to maximise total payoffs.

Hence we have the following incentive scheme:

Scheme 3 (*Rules with participation incentives*). Peer i shares $\tilde{f}_i$ files, where $\tilde{f}_i$ is given by Eq. (13).

3.3. Non-personalised, complete information prices

Now suppose that we have complete information about the payoff parameters of the peers, but we are unable to establish their identities—so that personalisation of an incentive scheme is not possible. Another reason is that implementing a personalised incentive scheme may be too costly or infeasible. This raises no problems for first-best rules—the first-best rule is not personalised in the

[6] It is an interesting feature of the solution that information about the benefit function is not always required. The first-best prices do not depend on the form of the benefit function $v(\cdot)$. This feature arises because of the way in which peer heterogeneity appears in the problem: The payoff parameter multiplies a function which is the same for all peers. The first-best rule does, however, depend on the form of $v(\cdot)$.

case of our model. The matter is more complicated for prices. Clearly, there are several ways to calculate a uniform price. We shall assess the following:

$$p = 1 - \frac{\max\{\theta_1, \ldots, \theta_N\}}{\Theta}, \tag{14}$$

i.e., the personalised, complete information (first-best) price received by the peer with the highest payoff parameter. The numerical analysis in Section 4 confirms that all peers wish to participate given this price. [7]

Scheme 4 (*Non-personalised price*). Each peer is paid a price p per file shared, where p is given by Eq. (14).

3.4. Incomplete information schemes

We now turn to the case in which there is incomplete information about the peers' payoff parameters. To be specific, the number of peers N is known, as is the distribution from which the N peers' payoff parameters are drawn (they are i.i.d. random variables with distribution F) and the form of the benefit function $v(\cdot)$. But only peer i knows the realisation of its payoff parameter θ_i; peer i does not know the payoff parameter of peer j, and we assume that the agent who designs the incentive scheme observes none of the payoff parameters.

This information structure rules out immediately all of the schemes considered above. To start the analysis, we consider adaptations of these schemes that, in Section 4, will be compared to a scheme that is derived from full consideration of the incomplete information problem.

The first-best rule 'share f^* files' depends on the realised $\boldsymbol{\theta}$, since f^* depends on Θ (see Eq. (7)). Making this dependence explicit by writing $f^*(\Theta)$, an average rule can be computed as

$$\bar{f} \equiv \int \cdots \int f^*(\Theta)\, \mathrm{d}F^N(\theta), \tag{15}$$

where F^N is the probability distribution of the random vector $\boldsymbol{\theta}$. An example helps to make this clear. Suppose that the benefit function $v(\cdot)$ is iso-elastic, $v(x) = x^\alpha$, where $\alpha \in (0, 1)$. Then

$$f^*(\theta) = \left(\frac{\alpha\Theta}{2} N^{\alpha-1}\right)^{2/(2-\alpha)},$$

and

$$\bar{f} \equiv \int \left(\frac{\alpha\Theta}{2} N^{\alpha-1}\right)^{2/(2-\alpha)} \mathrm{d}F^N(\theta),$$

and can be calculated explicitly by making an assumption about the distribution of payoff parameters (e.g., independently drawn from the uniform distribution on $[0, 1]$), or by using simulation.

Scheme 5 (*Average rule*). Each peer shares $\bar{f}$ files, where $\bar{f}$ is given by Eq. (15).

Of course, it may be that some peers would receive a negative payoff if they contributed $\bar{f}$ files; hence under this rule, they will not take part in the network. This will affect any calculation of efficiency (see Section 4).

A similar procedure can be applied to prices, to give an average price

$$\bar{p} \equiv \int \cdots \int \frac{\Theta_{-i}}{\Theta}\, \mathrm{d}F^N(\theta). \tag{16}$$

Scheme 6 (*Average price*). Each peer is paid a price $\bar{p}$ per file shared, where $\bar{p}$ is given by Eq. (16).

Notice that the average rule depends on both the functional form of $v(\cdot)$ and the distribution from which peers' payoff parameters are drawn. The average price depends only on the latter.

These adaptations of the previous schemes do not take into account explicitly additional factors that arise in the presence of incomplete information. In this case, it is likely that the scheme can be improved (in terms of the network efficiency achieved) by eliciting information from peers. To do this, each peer must be given the appropriate incentive to reveal its private information truthfully. The task, then, is to design a *mechanism*. In a mechanism, peers report their payoff parameter: $\hat{\theta}_i$

[7] The choice of another definition for a uniform price, e.g., using instead of the maximum, the average of the θ_i's, or their minimum, does not affect the qualitative results presented.

denotes the report of peer i, $\hat{\boldsymbol{\theta}}$ the vector of peers' reports. The mechanism specifies a rule for providing shared files that has three components:

1. $\phi_i(\hat{\boldsymbol{\theta}})$ is the contribution rule that specifies the number of files that peer i must share, for $i \in \{1,\ldots,N\}$;
2. $\mu_i(\hat{\boldsymbol{\theta}}) \in \{0,1\}$ is the inclusion rule that specifies whether peer i is included in the network, for $i \in \{1,\ldots,N\}$.

Note that we continue to assume that the files shared by each peer can be verified by the mechanism designer. While verifying the number of files shared may be relatively easy, it is likely to be more difficult to verify relevant characteristics (such as quality or popularity) of files. We do not consider the issue of files of variable quality (and so the problem of peers sharing low quality or unpopular files to minimise their contribution to the network). We return to this issue in the conclusion.

Let $\mathbb{E}$ denote the expectations operator with respect to $\boldsymbol{\theta}$; and $\mathbb{E}_{-i}$ denote the expectations operator with respect to $\theta_{-i} = (\theta_1,\ldots,\theta_{i-1},\theta_{i+1},\ldots,\theta_N)$. In order for the mechanism to induce truthful reporting by peer i, it must be that

$$\mathbb{E}_{-i}\left[\theta_i\mu_i(\boldsymbol{\theta})v\left(\sum_{j=1}^{N}\sqrt{\phi_j(\boldsymbol{\theta})}\right) - \phi_i(\boldsymbol{\theta})\right] \geqslant \mathbb{E}_{-i}\left[\theta_i\mu_i(\hat{\theta}_i,\boldsymbol{\theta}_{-i})v\left(\sum_{j=1}^{N}\sqrt{\phi_j(\hat{\theta}_i,\boldsymbol{\theta}_{-i})}\right) - \phi_i(\hat{\theta}_i,\boldsymbol{\theta}_{-i})\right] \quad (17)$$

for all i and θ_i, $\hat{\theta}_i \in [0,1]$. In words, all peers must gain an expected payoff from reporting their type honestly that is at least as large as the expected payoff received when they are dishonest. In order for peer i to be willing to participate in the network and mechanism, it must be that

$$\mathbb{E}_{-i}\left[\theta_i\mu_i(\boldsymbol{\theta})v\left(\sum_{j=1}^{N}\sqrt{\phi_j(\boldsymbol{\theta})}\right) - \phi_i(\boldsymbol{\theta})\right] \geqslant 0 \quad (18)$$

for all i and $\theta_i \in [0,1]$. In words, all peers must gain an expected payoff from reporting their type honestly that is at least as large as their outside option, which is assumed to yield zero payoff.

The objective is to maximise the expected total payoff subject to these constraints. That is, the problem facing the mechanism designer is to choose the functions $\{\phi_i(\boldsymbol{\theta}),\mu_i(\boldsymbol{\theta})\}_{i=1,\ldots,N}$ to

$$\mathbb{E}\left[\sum_{i=1}^{N}\left(\theta_i\mu_i(\boldsymbol{\theta})v\left(\sum_{j=1}^{N}\sqrt{\phi_j(\boldsymbol{\theta})}\right) - \phi_i(\boldsymbol{\theta})\right)\right] \quad (19)$$

subject to Eqs. (17) and (18).

Denote the expected total payoff from this program S_{II}.

We shall not solve the full problem (19) here [10]. There are two features of the mechanism that solves the problem that we wish to comment on, before developing an approximate solution: First, note that any mechanism for this problem must be 'budget-balanced', in the sense that it is not possible for more files to be shared than have been contributed by the peers. [8] Myerson and Satterthwaite [12] demonstrate that no mechanism can be efficient, individual rational and budget balanced. Hence we know that the solution to problem (19) yields less total surplus than the social optimum (analysed in Section 2): $S_{\mathrm{II}} \leqslant S_{\mathrm{SO}}$. Secondly, the mechanism that solves problem (19) will be, in general, very complex. In most cases, a large amount of information has to be passed from the peers to the mechanism designer; and the subsequent calculation is complicated.

Nevertheless, recent analysis by Hellwig [7] and Norman [13] (who build on the earlier work of Mailath and Postlewaite [9]) shows that in public good provisioning models which include our peer-to-peer file sharing model that allows peers to be excluded, both factors are reduced in importance when the number of participants is large. To be precise, as N increases, the second-best policy obtained by solving S_{II} converges to a fixed proportion of the social benefit S_{SO} obtained by the

[8] In more general public good provision problems, budget balance requires that the total financial payments made by agents cover the cost of providing the required level of the public good.

first-best policy. [9] Hellwig and Norman show that a necessary feature of the mechanism that achieves this degree of efficiency is that some peers are excluded from the network. In fact, Norman's results suggest that the following mechanism suffices in our case when N is large: Each peer must contribute a uniform minimum number of files to join the network, where this number depends on the declarations $\hat{\boldsymbol{\theta}}$ of all the peers. This scheme has two advantages: Its simplicity in terms of the form of the policy, being a simple contribution rule; and the fact that prices are not involved. The latter is attractive in a large, decentralised system in which implementing a currency can be difficult.

There is still a considerable computational burden to calculating the minimum contribution required by Norman's mechanism. Rather than undertaking that task, we use the result of Courcoubetis and Weber [4], where it is shown that for a large class of models including the one in this paper, a simple fixed fee policy which is independent of the declarations of the peers, is enough to get us within $o(N)$ of S_{II}. In many interesting cases, this fixed fee can be easily computed by solving a simple optimisation problem. In our case, assuming for simplicity that the distribution F of the θ's is uniform on [0, 1], we need to solve the following mathematical program:

$$\begin{aligned} \max_{\Phi,\bar{\theta}} \quad & Nv(\Phi)\int_{\bar{\theta}}^{1} x\,\mathrm{d}F - c(\Phi) \\ \text{s.t.} \quad & N(1-F(\bar{\theta}))\bar{\theta}v(\Phi) = c(\Phi), \end{aligned} \tag{20}$$

where $c(\Phi)$ is the total number of files that must be provided for achieving an effective number of shared files Φ. We come back to how to express this cost as a function of Φ after we provide some intuition for (20).

This program has a very simple interpretation. It sets two variables: the total effective number of shared files Φ (i.e., the argument of the benefit function $v(\cdot)$); 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 $\bar{\theta}v(\Phi)$ files (i.e., the benefit of the marginal peer). Peers with a payoff parameter greater than $\bar{\theta}$ are willing to join; peers with $\theta < \bar{\theta}$ are not. The total contribution is covered by the contributions of the peers which will participate, which are $N(1-F(\bar{\theta}))$ on the average. The average value of file sharing per peer is $v(\Phi)\int_{\bar{\theta},1} x\,\mathrm{d}F$. Hence (20) maximises the expected social welfare over the choice of fixed fee policies. The optimal policy will correspond to the optimal values of the two variables Φ and $\bar{\theta}$.

We can reduce 20 into a simpler form. First observe that since all participating peers contribute equally some amount f, $\Phi = m\sqrt{f}$, where m is the number of final participants. But also $c(\Phi) = mf$, and hence $c(\Phi) = \Phi^2/m$. By doing integration by parts and using the cost constraint we obtain the equivalent program

$$\begin{aligned} \max_{\Phi,\bar{\theta}} \quad & v(\Phi)\int_{\bar{\theta}}^{1} (1-F(x))\,\mathrm{d}x \\ \text{s.t.} \quad & N(1-F(\bar{\theta}))\bar{\theta}v(\Phi) = \frac{\Phi^2}{N(1-F(\bar{\theta}))}, \end{aligned} \tag{21}$$

where we substituted m by its average $N(1-F(\bar{\theta}))$.

Denote the expected total payoff from this program a S_{CW}. According to Courcoubetis and Weber [4], the difference between S_{CW} and S_{II} is $o(N)$, and so becomes negligible as N becomes very large. This result motivates us to consider the following, 'fixed fee' scheme.

[9] This is not the case when peers cannot be excluded from the network. In this case, 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 $1/\sqrt{N}$, hence a total of $\sqrt{N}$ for the N peers) in any incentive compatible scheme. Since a relatively small part of the cost can be covered by such contributions (the cost of the optimal size system increases in most cases faster than $\sqrt{N}$), the system evolves into a much smaller size than the optimal one. Actually, $S_{\mathrm{II}}/S_{\mathrm{SO}} \to 0$ as N increases. In contrast to that, if exclusions are allowed and mechanisms with a minimum participation fee are possible to implement, then an $O(1)$ revenue can be obtained from each peer. This is because a positive percentage of peers will participate and each participant will contribute at least the minimum fixed amount (in addition to some incentive payment which will depend on his declaration $\hat{\theta}_i$, and which becomes negligible when N gets large). In this case a significantly larger total contribution of files can be achieved which allows for $S_{\mathrm{II}}/S_{\mathrm{SO}} \to b$, for some $0 < b < 1$.

Scheme 7 (*Fixed fee*). All peers who join the network are required to share $\bar{\theta}v(\Phi)$ files, where $\bar{\theta}$ and Φ are given by the solution to the program (21).

4. Simulation assessment of incentive schemes

In previous sections, we have defined seven incentive schemes:

1. first-best rule,
2. first-best prices,
3. rules with participation incentives,
4. non-personalised price,
5. average rule,
6. average price,
7. fixed fee.

In this section, we first calculate these schemes for a specific functional form for the benefit function: $v(x) = x^{\alpha}$, where $\alpha = 0.5$; and a specific distribution function for the peers' payoff parameters (uniform on the unit interval, with i.i.d draws). The objective is to assess the performance of the different schemes, in terms of the total payoff that they yield, as the number of peers increases. Then, in order to show the robustness of our results, we use a lognormal distribution, and we vary its variance to investigate how peer heterogeneity may affect the results.

The procedure we use is as follows:

1. Fix the number of peers at N.
2. Calculate the average rule and price in Schemes 5 and 6. [10]
3. Draw N values $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_N)$ randomly from the specified distribution.
4. For this realisation of peers' valuations, calculate rules/prices in Schemes 1–4.
5. For this realisation of peers' valuations, calculate total payoffs for all schemes.
6. Return to step and repeat 100 times.
7. Average the total payoffs achieved over the draws of $\boldsymbol{\theta}$.
8. Increase the number of peers by 1 and return to step 2.

The outcome of the simulations is summarised in Figs. 1 and 2. In these figures, the total payoff achieved by each scheme, averaged over realisations of peers' payoff parameters, is expressed as a proportion of the first-best total payoff, as the number of peers varies. So, the first-best rule or price would be represented by a flat line at 1 for all network sizes. As expected, all other schemes return a total payoff strictly less than the first-best

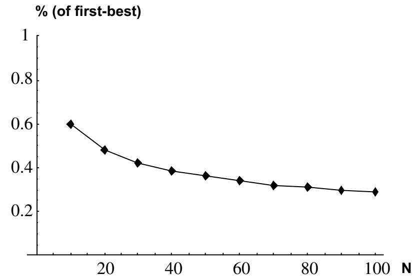


Fig. 1. Total payoffs in the Nash equilibrium.

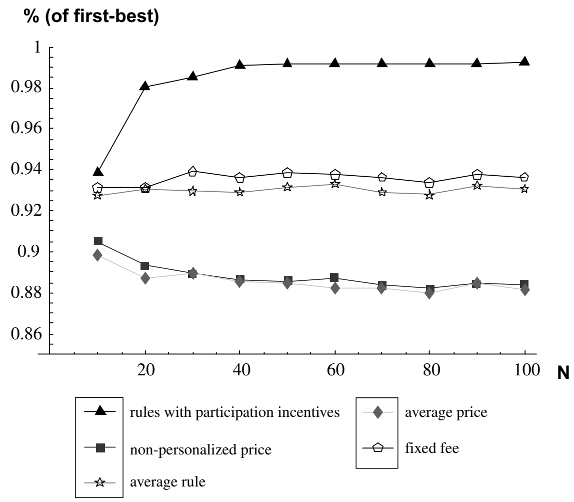


Fig. 2. Comparison of incentive schemes.

[10] Since these do not depend on the realisation of the particular θ's.

level; hence the lines for these schemes lie strictly below 1. Fig. 1 confirms the observation of Section 2: without an incentive scheme, the externalities that exist between peers leads to inefficiency. The figure shows the total payoff in the Nash equilibrium, as a proportion of the first-best level. In a small network (10 peers), the degree of inefficiency is marked: the Nash equilibrium achieves only 60% of the first-best total payoff. As the network grows, the inefficiency becomes worse and worse; by the time the network includes 100 peers, the Nash equilibrium total payoff is just 30% of the first-best. This illustrates the intuitive property that free-riding is worse when there are many agents. Each peer anticipates that it has little effect on the total number of files shared when the network is large. Consequently the incentive to contribute nothing and rely only on the contributions of others grows with network size.

There are several features that should be noted from Fig. 2: First, the rules with participation incentives achieve close to full efficiency even for moderate network sizes (20 peers or more). This indicates that (within this set of calculations, at least) there is little efficiency loss arising from the need to give participation incentives to peers. The intuition for this result is that as the system gets larger, the participation constraint becomes easier to satisfy. This is because the value of the shared content to peers increases and peers with the same θ are willing to contribute more in order to participate. For very large N, a very small fraction of peers will be reluctant to pay the fixed fee contributed by the rest of the peers as defined by the optimal policy. Hence the participation constraint of the optimisation problem becomes irrelevant, and the solution converges to the first-best.

The fixed fee scheme yields a strictly lower level of utility. This demonstrates the Myerson–Satterthwaite result: incomplete information leads to efficiency loss. (A major difference between Schemes 3 and 7 is, of course, that the former uses full information about peers' payoff parameters, which is not available for the latter.) The efficiency loss is relatively small, however, around 6–7%. An intuition for this can be gained by considering the sources of efficiency loss. One is that participating peers do not contribute the efficient amount of files (f^*); another is that some peers are excluded. The size of the second inefficiency is limited, however, by the fact that it is peers with low payoff parameters that are excluded; and they make little difference to the overall level of efficiency. Further, from the results of Hellwig [7] and Norman [13], as the number of peers becomes large, the fixed fee schemes approaches the full second-best mechanism. A major reason for this is that calculations based on averages or expectations—like the calculation of the fixed fee—become more and more accurate as the number of peers becomes large, for standard statistical reasons. There is a second, more subtle argument as we discussed in the previous section, which explains why, although in the full second-best problem, incentive compatibility constraints must be included—see Eq. (17), these incentive constraints are effectively ignored in the calculation of the fixed fee. Hence the numerical results indicate that these constraints become less important (in efficiency terms) as the number of peers grows. Since our fixed fee policy becomes asymptotically second-best optimal, it is also optimal among the set of all fixed fee policies.

Thirdly, the average rule scheme (number 5) does well for efficiency, but appears systematically to yield a strictly lower total payoff than Schemes 3 and 7. This indicates that the approximation used in this scheme (taking a straight expectation of the first-best rule) is inferior to the fixed fee approximation. This is to be expected: The fixed fee method derives an instrument to maximise average payoff; the approximation in Scheme 5 averages over an instrument that maximises total payoff under complete information. The former is better suited to the situation in which there is incomplete information—a fact reflected in the figure.

Finally, prices of any flavour (other than first-best) yield lower levels of efficiency than the other schemes. The degree to which they under-perform appears to be driven by the specification of our model: in particular, the linear 'cost' term in Eq. (1). With this specification, small errors between the first-best price and an approximate price are translated into very large differences in the number

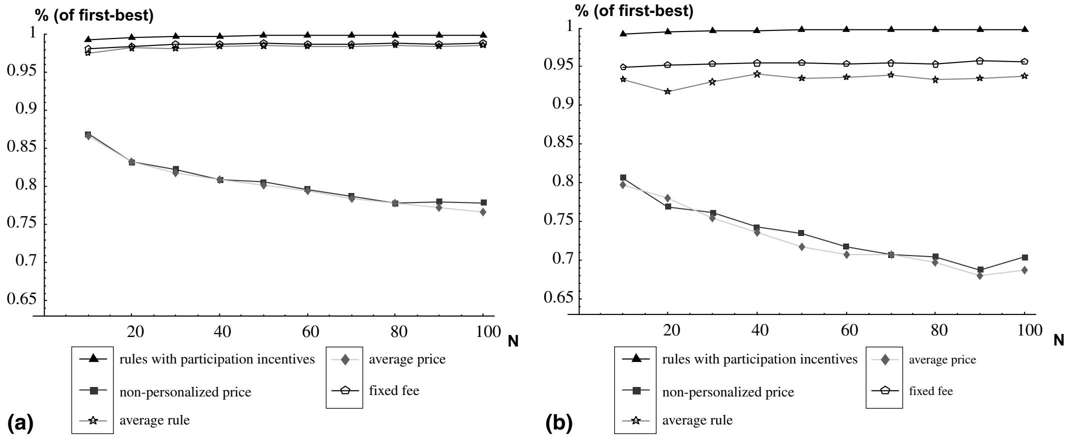


Fig. 3. Comparison of incentive schemes using the lognormal distribution: (a) $(\mu, \sigma) = (0.5, 0.6)$ and (b) $(\mu, \sigma) = (0.5, 0.8)$.

of files contributed by peers. Hence the resulting efficiency level of approximate (non-personalised or average) prices are low. With alternative cost specifications (especially cost functions that are strictly convex), the performance of approximate price schemes improves. [11] In fact, in a variety of such models with convex cost functions, as N increases, uniform prices perform close to the first-best.

In the next results we use a lognormal distribution of the peer payoffs. The reason is to investigate the effects of the variance of the θ's to the performance of the various schemes. The results are in Fig. 3, for two different values of the variance of the distribution. These results can be easily explained. Smaller variance makes uniform rules more effective since information loss is less important. This suggests that if peers in terms of their payoffs can be better classified (by using some objective characteristic like the access speed of their modem) in groups having a higher degree of homogeneity, then simple fixed fee schemes will perform better than having all peers in a single group.

[11] This observations raises the question: why choose a quasi-linear payoff function? The reason is that this allows us to match the model directly to the public good provision models used by Hellwig [7] and Norman [13], and hence to apply their results in our context.

5. Conclusions

In this paper, we have identified that participants in a peer-to-peer network have too little incentive to contribute to the network. The resulting inefficiency motivates our analysis of a variety of incentive schemes. The different schemes are distinguished by the constraints that are considered, relative to the first-best situation in which the network manager has full information, can coerce participation in the network, and is able to personalise incentives. Numerical evaluation of the various schemes shows that there can be significant differences in the efficiency levels that they achieve. A major finding is that a simple fixed fee scheme can achieve a high level of efficiency in large networks. This suggests that a complicated mechanism that relies on information reporting and complicated payments to peers is not necessary to run an efficient peer-to-peer network. Instead, peers need only contribute a standard number of files to the network on joining. A similar fixed contribution approach has been studied in the context of building a p2p system from wireless LANs in [5]. Another interesting finding is that reducing the variance of the peers' preferences is beneficial in the performance of such simple schemes and may be worth the implementation cost.

In order to investigate the gains of differentiating peers into subgroups we can easily extend (21) and calculate the different participation fees that should be used in the various subgroups. Content should be shared among all participants independently of the subgroup to which they belong.

There are many more avenues for further work. This paper should be seen as the early stages of a wider project to assess and evaluate incentive schemes in peer-to-peer systems. With the current model, we intend to continue the numerical evaluation of schemes, using alternative functional forms for the benefit and cost function and different distributions for peers' payoff parameters. These extensions will allow us, among other things, to assess how important peer heterogeneity is for the design of incentive schemes. We intend also to extend the model, in three directions especially: First, to applications other than file sharing in order to examine how details of incentive schemes might vary across different situations—an initial model for a novel p2p WLAN application is in [2]. Secondly, we intend to model activities that incur congestion (e.g., requests for files, within the file sharing application). Finally, the most difficult challenge is to implement the ideas in this paper in a fully decentralised p2p network, in which it can be difficult to verify the number, and especially the characteristics (e.g., quality or popularity), of the files shared by peers. Our goal in this paper has been to demonstrate the general difficulty of setting incentives in p2p networks, especially when there is limited information about peers' payoffs; and to show how some of the complications are reduced in large networks. In future work, we plan to analyse mechanisms when there is incomplete information about the characteristics of files shared.

References

[1] E. Adar, B. Huberman, Free riding on Gnutella, First Monday 5 (10) (2000).

[2] P. Antoniadis, C. Courcoubetis, E.C. Efstathiou, G.C. Polyzos, B. Strulo, Peer-to-peer wireless consortia: modelling and architecture, in: Third IEEE International Conference on Peer-to-Peer Computing (P2P 2003), 2003.

[3] B. Cohen, Incentives build robustness in BitTorrent, in: Proceedings of the Workshop on Economics of Peer-to-Peer Systems, 2003, Berkeley, CA.

[4] C. Courcoubetis, R. Weber, Asymptotics for provisioning problems with a large number of participants, AUEB Technical Memorandum, 2003. Available from <http://nes.aueb.gr/research/>.

[5] C. Courcoubetis, R. Weber, Asymptotics for provisioning problems of peering wireless LANs with a large number of participants, in: Proceedings of WiOpt'04: Modelling and Optimization in Mobile, Ad Hoc and Wireless Networks, University of Cambridge, UK, 2004.

[6] P. Golle, K. Leyton-Brown, I. Mironov, M. Lillibridge, Incentives for sharing in peer-to-peer networks, in: Proceedings of ACM Conference on Electronic Commerce, 2001.

[7] M.F. Hellwig, Public-good provision with many participants, Review of Economic Studies 70 (3) (2003) 589–614.

[8] K. Lai, M. Feldman, I. Stoica, J. Chuang, Incentives for cooperation in peer-to-peer networks, in: Proceedings of the Workshop on Economics of Peer-to-Peer Systems, 2003, Berkeley, CA.

[9] G. Mailath, A. Postlewaite, Asymmetric information bargaining problems with many agents, Review of Economic Studies 57 (1990) 351–368.

[10] A. Mas-Colell, M.D. Whinston, J.R. Green, Microeconomic Theory, Oxford University Press, New York, 1995.

[11] MMAPPS Consortium, Market management of peer-to-peer services—White Paper, 2003. Available from <http://www.mmapps.org/>.

[12] R.B. Myerson, M.A. Satterthwaite, Efficient mechanisms for bilateral trading, Journal of Economic Theory 29 (1983) 265–281.

[13] P. Norman, Efficient mechanisms for public goods with use exclusions, Review of Economic Studies, in press.

[14] A. Oram (Ed.), Peer-to-Peer: Harnessing the Power of Disruptive Technologies, O'Reilly, Sebastopol, CA, 2001.

[15] S. Sariou, P.K. Gummadi, S.D. Gribble, A measurement study of peer-to-peer file sharing systems, in: Proceedings of Multimedia Computing and Networking 2002 (MMCN '02).

[16] D. Wallach, A survey of peer-to-peer security issues, in: International Symposium on Software Security, Tokyo, Japan, 2002.

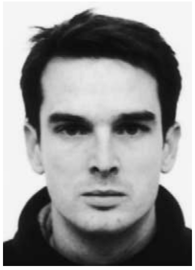
Panayotis Antoniadis received his B.S. (1997) and M.S. (1999) degree in Computer Science from the University of Crete, Heraklion, Greece. For his diploma thesis, he received the first 'Ericsson Award of Excellence in Telecommunications'. Since 2000, he is a Ph.D. candidate at the Department of Informatics of the Athens University of Economics and Business and a member of Network Economics and Services Laboratory.

His current research interests are in network economics with emphasis in the economic modelling and incentive mechanisms for p2p services and multicasting, network management, distributed systems, and high-speed scheduling. Homepage: http://nes.aueb.gr/~antoniad/.

Costas Courcoubetis was born in Athens, Greece. He received his Diploma (1977) from the National Technical University of Athens, Greece, in Electrical and Mechanical Engineering, his M.S. (1980) and Ph.D (1982) from the University of California, Berkeley, in Electrical Engineering and Computer Science. From 1982 until 1990 he was Member of the Technical Staff (MTS) in the Mathematical Sciences Research Center, Bell Laboratories, Murray Hill, NJ, and from 1990 until 1999 he was Professor in the Computer Science Department at the University of Crete in Heraklion, Greece, and headed the Telecommunications and Networks Group at the Institute of Computer Science, FORTH. Since Fall 1999 he is a Professor in the Computer Science Department at the Athens University of Economics and Business.

His current research interests are economics of networks with emphasis in the development of pricing schemes that reduce congestion and enhance stability and robustness, quality of service and management of integrated services, performance and traffic analysis of large systems, applied probability models. Other interests include the combination of e-commerce technologies with telecommunications, and formal methods for software verification.

He is a co-author with Richard Weber of 'Pricing Communication Networks and Modeling' (Wiley, 2003).



Robin Mason is a Reader in the Economics Department at the University of Southampton, UK. He graduated from Cambridge in Natural Sciences, from where he also obtained a Ph.D. in Economics. He has held research fellowships at the University of Oxford and Cambridge. His research interests lie in the area of industrial organisation, and in particular the economics of networks (especially telecoms and the Internet). He is currently an associate editor of the Journal of Industrial Economics, and a member of the panel of economic advisers to Ofcom, the regulator for the UK communications industries.