‘To Share or not to Share’
An Analysis of Incentives to Contribute in Collaborative File Sharing Environments

Kavitha Ranganathan[1] , Matei Ripeanu[1] , Ankur Sarin[2] and Ian Foster[1,3]

[1] *Department of Computer Science, The University of Chicago, Chicago, IL 60637, USA*
[2] *Irving.B.Harris School of Public Policy, The University of Chicago, IL 60637, USA*
[3] *Math and Computer Science Division, Argonne National Laboratory, Argonne, IL, 60439, USA*

1. Introduction

Projects developing infrastructure for the pooling of distributed resources (data, storage, or computation) [1, 2] often assume that resource owners have committed their resources and that the chief task is to integrate and use them efficiently. Such projects frequently ignore the question of whether individual resource owners are willing to share their personal resources for the overall good of the community. However, experiences [3-6] with peer-to-peer (P2P) file sharing systems like Gnutella, Napster, and Kazaa suggest that users are not altruistic. In Gnutella, for example, 70% of all users do not share files, and 50% of all requests are satisfied by the top 1% sharing hosts. Thus, *incentive* mechanisms that motivate users to contribute resources may be critical to eventual success of such systems. Various approaches to incentives have been proposed, including pricing or micro-currency schemes [7] and so-called “soft incentive” or non-pricing schemes [8]. However, the effectiveness of these different schemes is not well understood.

In this paper we take a step towards understanding the performance of incentive schemes by defining and applying an analytic model based on Schelling’s Multi-Person Prisoner’s Dilemma (MPD) [9]. We use both this framework and simulations to study the effectiveness of different schemes for encouraging sharing in distributed file sharing systems. We consider three such schemes: the soft-incentive, reputation-based *Peer-Approved* and *Service-Quality*, and the *Token-Exchange* pricing scheme.

After introducing the MPD model (Section 2), we use it to explain the rational behavior of users in a P2P file sharing community without incentives (Section 3). We then analyze user behavior when *Peer-Approved* is used as the incentive mechanism and find that it is effective in incentivizing rational users to share more files (Section 4).

We then shift to the use of simulations so as to look beyond the assumptions of the MPD model. We measure the effectiveness of the reputation-based soft-incentive *Peer-Approved* scheme and compare it with that of the *Token-Exchange* pricing scheme (in Section 5). We find that even simple soft-incentive schemes can motivate users of P2P file-sharing systems to increase contributions in a way that benefits all users, including themselves.

2. Let Us All Starve Together

The P2P sharing problem can be modeled as a Prisoner’s Dilemma [10]. In the classical Prisoner’s Dilemma (CPD), two players choose simultaneously whether or not to co-operate. Each is rewarded if both cooperate, but at a lower rate than the penalty received if one cooperates and the other does not. Hence the dilemma: the rational choice of not cooperating leaves both worse off than if both had co-operated.

Not all assumptions underlying the CPD apply to a P2P environment. First, we must extend the framework to more than two participants. Second, players in a typical P2P setting can observe actions of others and make choices influenced by others. Hence, the assumption that players act simultaneously needs to be relaxed. Third, the incentives to contribute/co-operate may depend on how many other users are contributing; this number needs to be incorporated into the theoretical model. The resulting Multi-Person Prisoner’s Dilemma (MPD) framework provides a more realistic model (albeit with other limitations) of a P2P environment.

The following four conditions define a MPD [9]:

1. There are ***n*** players in the system, each with the same binary choice and payoffs.
2. Each player has the same preferred choice, which does not change, no matter what other players do.
3. A player is always better off if more among the others choose the un-preferred alternative.
4. For a certain $k > 1$, if k or more players choose the un-preferred alternative, they are better off than if all players had chosen the preferred alternative.

Figure 1 illustrates the situation graphically, showing the payoff curves for a player that chooses the preferred (P, upper line) or un-preferred (U, lower line) alternative, as a function of the number of other players, from 0 to *n*, that choose the un-preferred alternative (to share files in our case). We assume that there are *n+1* players in total, and hence *n* "others." E.g., at $x = {}^{n}/_{3}$, a third of the other players choose the un-preferred alternative and two thirds choose the preferred alternative; P_x and U_x are the pay-off to the player that chooses the preferred or un-preferred alternatives, respectively.

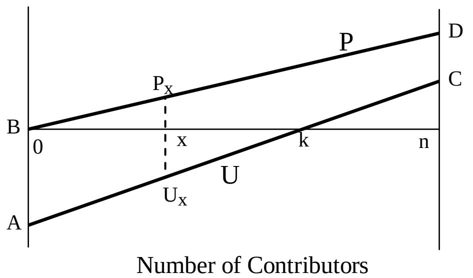


Figure 1: MPD model. Payoff curves for a player that chooses the preferred (P, upper line) and un-preferred (U, lower line) alternative, as a function of the number of other players that choose the un-preferred alternative.

While the payoff functions can be linear or curved, depending on the specific problem, the MPD definition implies that the vertical order of the four end-points of the payoff curves remain the same. That is, the payoffs are the least if everyone else chooses the preferred alternative (point A), and the highest if everyone else chooses the un-preferred alternative (point D). Point C should be higher than point B, so that after some critical value ***k***, it is more profitable to join the un-preferred group than when everyone was in the preferred group (point B).

3. Free Riding as Multi-Person Prisoner's Dilemma

P2P file-sharing systems can be modeled as MPDs. We make the following simplifying assumptions: there are always ***n*** users in the system, and each user has one (unique) file that each of them can either decide to share with other peers (the un-preferred option) or keep only to self (the preferred option). All files are the same size, are equally popular and require unit bandwidth to download/upload. The benefit to a user participating in the system is the access gained to files made available by other users in the system. Note that, in the absence of an incentive mechanism, contributors and non-contributors derive the same benefit. (We exclude altruism as a possible benefit for contributors.) On the other hand, a contributor incurs a cost, namely the bandwidth consumed when responding to requests.

Given the above assumptions, suppose there are ***c*** contributors and ***r*** free-riders/non-contributors (*c+r=n*). Also suppose ***f*** files are requested in the system during a time unit. Thus, assuming files have similar popularity, the expected number of file requests (Π) reaching a contributor **x** is:

$$\Pi = \frac{Num.\ of\ \ files\ at\ x \times Num.\ of\ requests}{Total\ num.\ of\ \ files\ in\ the\ system} = \frac{f}{c}$$

Hence, the expected *cost* for a contributor is a function of ${}^{f}/_{c}$, while for a free rider it is always 0, as a free rider does not share files.

In the absence of incentives, all users have access to all files in the system and thus receive the same *benefit*. This benefit is a function of the total number of files available in the system, which in this case is a function of *c*, the number of contributors. We make this function logarithmic to model the intuition that the incremental benefit for gaining access to a new file decreases as the number of files available in the system increase. (However, we could use any increasing function without changing our model.) Thus, the net payoff for a contributor $payoff_c = log(c)-(f/c)$ and the net payoff for a free-rider $payoff_r = log(c)$.

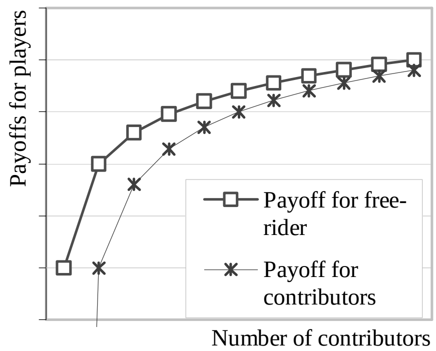


Figure 2: Payoffs for contributers and free-riders.

Figure 2 plots the net payoff curves for free riders and contributors for *n* = 100 and *f* = 10. The Y axis represents the net payoff. Note that the vertical order of the four end-points of the payoff curves is consistent with the MPD definition. $Payoff_r$ dominates $payoff_c$,: i.e., at any given state of the system, a user receives higher payoff if they do not contribute. System equilibrium [11] (the state where no one has an incentive to deviate from their action given the choices made by others) is thus on the extreme left, where nobody contributes. Note that this equilibrium is *inefficient* since users could have obtained higher payoffs had they all made the opposite choice and chosen to share their files. Thus, in the absence of incentives, the rational choice of not contributing leaves users worse off than if they had contributed.

4. Incentive Mechanisms

File sharing can be 'incentivized' using either *pricing policies* that involve an explicit payment for every file transferred or *non-pricing policies* (also called soft-incentive schemes) that encourage sharing in other ways.

We describe three schemes here: *Token-Exchange*, which is similar to a pricing scheme, and two non-pricing schemes, *Peer-Approved* and *Service-Quality*.

- ***Token-Exchange***. In this scheme, a consumer must transfer a token to the supplier prior to a file download. To enable newcomers to use the system, each first-time user might be allotted a fixed number of tokens, but once these run out, the user has to serve files to earn tokens. The Mojo Nation system [12] was implemented along these lines.

This scheme is similar to a pricing scheme with fixed prices, as a user must decide, for each potential download, whether the file in question is worth a token. The token transfer and validation cost incurred on each file exchange can be high, depending on how the token 'currency' is implemented. In the following, we assume the existence of the required currency mechanism, and focus on the incentive policies that may be layered on it.

- ***Peer-Approved***. In this scheme, a reputation system is used to maintain ratings for users, who are allowed to download files only from others with a lower or equal rating. This strategy motivates users to increase their rating in order to gain access to more files. User ratings can be based on different metrics: e.g., the number of files advertised by a user or the number of file-requests served by a user.

First-time users without files to share should be allowed to download a small number of files so that they can enter the system and build their rating.

This scheme is more flexible than *Token-Exchange* in that a user need not take a decision every time they want a file. Moreover, it has been suggested that non-pricing schemes may be more practical to implement in certain kinds of P2P networks [8] than direct payments between users. Past work also suggest that users may prefer (and thus accept more quickly) schemes that do not require payments or decisions for each transaction [13]. However, *Peer-Approved* needs a secure and reliable mechanism for maintaining user reputations [14, 15], and such a mechanism can be expensive to operate. In the following, we assume the existence of such a mechanism and focus on the incentive policies that may be layered on it.

- ***Service-Quality***. This scheme also uses a reputation mechanism. In contrast to *Peer-Approved*, users advertise all their files and may send download requests to any other user. However, each user assigns incoming requests to service classes according to the sender's reputation.

Combinations of these schemes are also possible. For example, in the Paris Metro pricing scheme [16], suggested initially for providing differentiated services in packet networks, a number of service quality classes are defined and users are assigned to a class based on how much they are willing to pay for the service.

5. Performance of Peer-Approved Policy

We want to compare, from the perspectives of effectiveness and fairness, soft-incentive policies such as *Peer-Approved* and *Service-Quality* with pricing policies such as *Token-Exchange*.

As a first step, we analyze the performance of *Peer-Approved* using the MPD model. (We plan to study *Service-Quality* and other soft-incentive schemes in the future.) We use our model to determine if such an incentive scheme would indeed motivate rational users to share more.

As our MPD model has limitations (binary choices for users, lack of heterogeneity in user characteristics, and inability to capture dynamics), we also perform simulations to evaluate *Peer-Approved* under a wider range of possible resource distributions and user behaviors. We also use these simulations to compare *Peer-Approved* and *Token-Exchange*.

5.1. Theoretical Analysis

Let all users have $D=h*d$ files. In order to conform to the binary choice model, we assume two choices for users. *Full contributors* advertise all their files, while *partial contributors* advertise only a fixed fraction ($^1/_h$), i.e., d files. A user's rating is the number of files advertised. Note that we cannot reasonably model true free riders in this scenario, as the *Peer-Approved* policy of only allowing downloads from other users with a lower or equal rating would immediately exclude true free riders from the system. Thus, we assume that all users advertise at least part of the files available locally.

The analysis follows as in Section 3. Table 1 describes the variables used.

Table 1: Variables used in analysis

*D=h*d*	Num. of files advertised by a full contributor
d	Num. of files advertised by a partial contributor
f	Total num. of **f**iles requested at unit time
u	Num. full contributors (**u**n-preferred alternative)
p	Num. partial contributors (**p**referred alternative)
n	Total num. of users (*n=p+u*)

Recall that ***u*** users choose the un-preferred alternative (to share all *d*h* files) and ***p*** choose the preferred alternative (to share only *d* files). If ***f*** file requests are made per unit time, the expected number of file requests (Π_x) reaching a user ***x*** is:

$$\Pi_x = \frac{Num.\ of\ \ files\ advertised \times Num.\ of\ \ requests}{Total\ num.\ of\ \ files\ advertized\ in\ the\ system}$$

Thus, the expected cost for a full contributor is:

$$c_c = \frac{hd * f}{hdu + dp} = \frac{hf}{hu + p}.$$

Similarly, the expected cost for a partial contributor is $c_p = \frac{f}{hu + p}$ and the expected benefit for both is *log(hdu+dp)*. Figure 3-left plots the payoff curves for both kind of users, and by the same logic used earlier, users tend to conglomerate at an inefficient equilibrium on the extreme left.

We now introduce the *Peer-Approved* incentive policy, according to which (in this scenario) full contributors only serve other full contributors, whereas partial contributors serve all users. The expected number of requests originating from a full contributor is $\frac{uf}{n}$ and the expected number of these requests reaching a certain full contributor is $\frac{uf}{n} \times \frac{h}{hu + p}$, which is in fact the expected cost for a full contributor. The cost for a partial contributor is unchanged: $\frac{f}{hu + p}$, since requests originating from anywhere can access a partial contributor. The number of files accessible to a full contributor is *hdu+dp,* which is the total number of files advertised in the system, while the number of files available to a partial contributor is *dp* as the only files accessible to a partial contributor are those accessible by other partial contributors.

One extreme case (Figure 3, center) emerges when the benefit users perceive from the files available in the system increases slowly with the number of files. In this case, introducing the incentive scheme moves the equilibrium to the right: a number of users find it is in their advantage to contribute. (Remember that without an incentive scheme, at equilibrium, users did not contribute to the system regardless of perceived benefits. In this case, the equilibrium was at the leftmost point in our graphs: Figure 3, left.)

The relative 'strength' of the benefit function will determine how much more to the right the equilibrium shifts: i.e., how many more users are motivated to contribute. In the case where the benefit function is the logarithm of the number of files accessible (Figure 3, right), a user faced with a choice will, at any point on the X-axis, choose to contribute, thus shifting the system more and more to the right and to its *efficient* equilibrium. (An efficient equilibrium is defined in the Pareto sense [17]: no one can be made better off without making someone else worse off.) At this point, everyone shares all their files, resulting in higher payoffs for all.

5.2. Simulations

We have shown that an incentive scheme can significantly improve the efficiency of equilibrium state of a P2P file sharing system. However, to

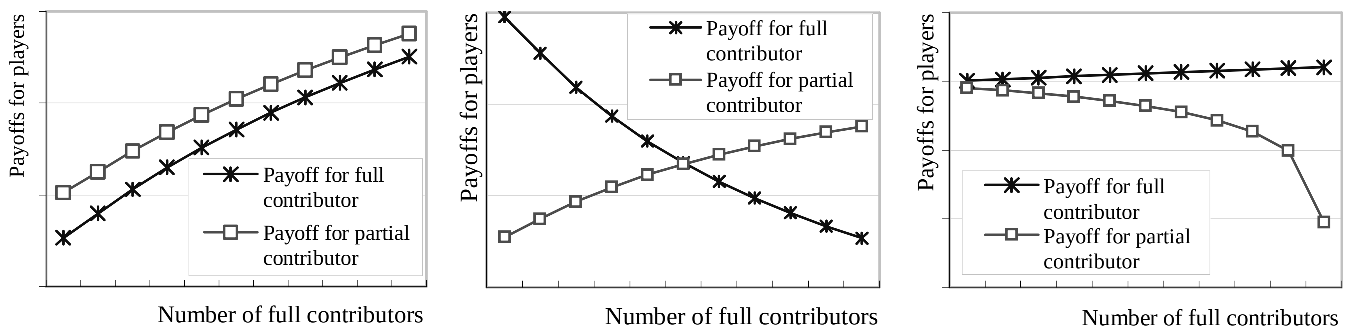


Figure 3. Net payoff curves for full and partial contributors when: (left) no incentive mechanism in place, situation is similar to Figure 2; (middle) Peer-Approved incentive and the benefit function increases very slowly with the number of available files;(right) same scenario, incentives mechanism in place, but the benefit function increases faster – a log function in this case. (These plots will have the same shape regargless of the values used. Here we use: n=100, d=1, f=10, h=2).

overcome the limitations in our analytical model (binary choices for users, lack of heterogeneity in user characteristics, and inability to capture dynamic situations) we use simulations to study a more general case in which a heterogeneous set of users can (incrementally and dynamically) change the number of files they share, depending on perceived benefits. Simulations also allow us to compare *Peer-Approved* to (the pricing-like) *Token-Exchange* in this dynamic and more realistic scenario.

We assume a fixed number of users with limited storage and bandwidth and an initial state in which files are placed at users according to a distribution function. Files are assumed to be equally popular. Each user initially advertises only a percentage of their files, again according to a distribution. At each iteration, ***f*** users request one file each. No individual files are modeled, and requests are assigned to a peer selected at random. Hence, a peer advertising more files will receive proportionally more requests than a peer advertising fewer files.

A request is satisfied if and only if the requesting user meets the criteria for the incentive scheme in use: i.e., if the user has a token to spend in the case of *Token-Exchange* or a rating that is not less than the server rating in case of *Peer-Approved*. Note that in these two schemes, unlike in *Service-Quality*, the peer advertising a file cannot block an eligible user from downloading that file. Hence, even though there is no immediate cost to advertising a file, an advertised file will attract requests that cannot be denied. Hence, advertising a file has a potential cost associated with it.

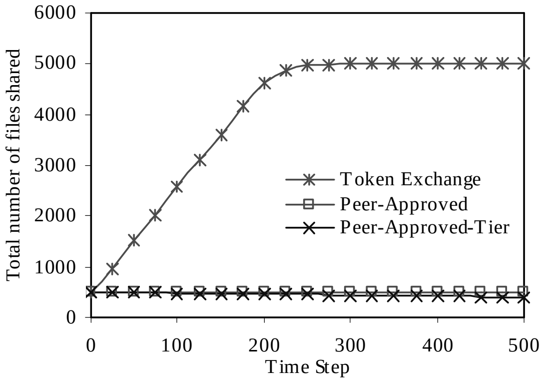
In the case of *Peer-Approved*, the rating of a user is the number of files currently advertised by that user. For *Token-Exchange*, each user is initially assigned a small number of tokens.

Given the above scenario, we model 'rational' user behavior in two ways. Firstly, a user that is denied access is motivated to advertise one more file. For example, when *Peer-Approved* is the incentive scheme in place, the user is denied a file if they have lower ratings than the server peer. In order to gain access to a wide variety of files, the user is then motivated to increase their rating, which they do by advertising more files. In the case of *Token Exchange*, a user is denied a file if their tokens are exhausted. The only way to gain additional files is to gain more tokens, which can only be achieved if someone downloads a file. So again the user is motivated to advertise more files.

Secondly, a user reduces the number of shared files if too much of their own bandwidth is consumed by others. At each iteration, users keep track of how much local bandwidth was used (i.e., how many downloads were served). If this quantity exceeds a threshold, the user reduces the number of advertised files, which then leads to fewer download requests in the future.

Thus, depending on the perceived benefits and costs of file-sharing, at each iteration the user is motivated either to stay in the current status or to increase or decrease (by one) the number of files advertised.

Since the overall goal of an incentive scheme is to motivate users to share, or in our case to advertise more files, we measure the success of a scheme by the total number of files advertised.



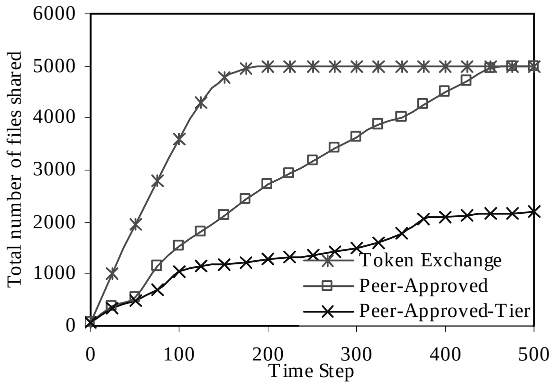


Figure 4: Simulation results for initial (top plot) Uniform file distribution and (bottom plot) Zipf file distribution.

We present our simulation results in Figure 4, which shows the performance (expressed in terms of the total number files shared) of *Peer-Approved, Peer-Approved-Tier* (a variation of Peer-Approved in which there are only a limited number of user rating categories), and *Token-Exchange*, under two different initial file-sharing distributions.

In the uniform file distribution scenario (top), every user initially has 50 files and shares/advertises 5 files each. In the non-uniform scenario (bottom) each user has 50 files and advertises according to a

Zipf distribution [18] (N= 50, α=2). In both cases, users have the same bandwidth and storage space.

We see that in the case of a uniform initial distribution of shared files, all users start with the same rating and hence can access files from all other users. Thus the rating schemes do not motivate users to advertise more files.

However, in the non-uniform case, user ratings initially vary. Thus, in the case of *Peer-Approved* lower-ranked users are motivated to advertise more files, albeit more slowly than in the case of *Token-Exchange. Peer-Approved-Tier* is even slower to converge to an equilibrium because users are distributed into a smaller number of rating slots and thus have access to more files than in *Peer-Approved*, and so are slower to advertise more of their files.

We believe that the non-uniform case is a more realistic scenario in present day P2P file-sharing systems such as Gnutella and Kazaa, in which a handful of users account for most files served. Thus *Peer-Approved* could be a useful incentive scheme in such scenarios, since without involving direct payments, its performance is comparable to a pricing scheme like *Token-Exchange*. However, *Token-Exchange* has the advantage of converging faster in most settings considered.

6. Conclusions

We have presented a model based on the Multi-Person Prisoner's Dilemma (MPD) for studying the free-riding problem in P2P file sharing systems. We used both this model and simulations to analyze the effectiveness of different incentive schemes designed to motivate increased user contributions. We compared one such scheme, the reputation-based *Peer-Approved*, with a *Token-Exchange* based scheme. Our results support the intuition that these simple incentive schemes can be used effectively to counter selfish user behavior.

We leave a number of important issues open. First, the mechanisms required to support the incentive schemes that we study can impose significant communication costs on a system. Although we perceive the costs associated with the different schemes to be relatively similar, we have not analyzed those costs in detail, nor have we investigated alternative incentive schemes that might involve lower costs. We would like to compare the communication costs of different schemes in order to quantify, ultimately, tradeoffs between cost and effectiveness. Second, our simulation study considers only two simple incentive schemes and in relatively standard settings. We plan to study additional incentive schemes, focusing on schemes that have low overhead, are easily deployable, and are acceptable by users.

Acknowledgements

We thank Adriana Iamnitchi and other members of our group for helpful discussions and feedback. This work was supported in part by the National Science Foundation under contract ITR-0086044 (GriPhyN).

References

1. Foster, I., C. Kesselman, and S. Tuecke, *The Anatomy of the Grid: Enabling Scalable Virtual Organizations.* International Journal of Supercomputing Applications, 2001. **15**(3).
2. Oram, A., *Peer-to-Peer: Harnessing the Power of Disruptive Technologies.* 2001: O'Reilly & Associates.
3. Ripeanu, M., I. Foster, and A. Iamnitchi, *Mapping the Gnutella Network: Properties of Large-Scale Peer-to-peer Systems amd Implications for System Design.* IEEE Internet Computing Journal, Special issue on Peer-to-peer Networking, 2002. **6**(1).
4. Saroiu, S., P.K. Gummadi, and S.D. Gribble. *A Measurement Study of Peer-to-Peer File Sharing Systems*. in *Multimedia Computing and Networking Conference (MMCN), San Jose, CA*. 2002.
5. Adar, E. and B.A. Huberman, *Free Riding on Gnutella*. 2000, First Monday.
6. Leibowitz, N., M. Ripeanu, and A. Wierzbicki. *Deconstructing the Kazaa Network*. in *3rd IEEE Workshop on Internet Applications (WIAPP'03)*. 2003. Santa Clara, CA.
7. Golle, P., K.Leyton-Brown, and I.Mironov. *Incentives for Sharing in Peer-to-Peer Networks*. in *ACM Conference on Electronic Commerce*. 2002.
8. Krishnan, R., M.D. Smith, and R. Telang, *The Economics of Peer-To-Peer Networks*. 2002, Working Paper, Carnegie Mellon University.
9. Schelling, T.C., *Micromotives and Macrobehavior*. 1978: W.W.Norton & Company.
10. Hardin, R., *Collective Action*. 1982, Baltimore: John Hopkins University Press.
11. Osborne, M.J. and A. Rubinstein, *A Course in Game Theory*. 1994, Cambridge, Massachusetts: MIT Press.
12. Wilcox-O'Hearn, B. *Experiences Deploying a Large Scale Emergent Network*. in *International Workshop on Peer-to-Peer Systems (IPTPS'02)*. 2002. Cambridge, MA.
13. Szabo, N. *Micropayments and Mental Transaction Costs*. in *2nd Berlin Internet Economics Workshop*.
14. Ngan, T., D. Wallach, and P. Druschel. *Enforcing Fair Sharing of Peer-to-Peer Resources*. in *2nd International Workshop of Peer-To-Peer Systems (IPTPS'03)*. 2003. Berekely, CA.
15. Kamvar, S.D., M.T. Schlosser, and H. Garcia-Molina. *EigenRep: Reputation Management in P2P Networks*. in *The 12th International World Wide Web Conference (WWW-12)*. 2003. Budapest, Hungary.
16. Odlyzko, A.M. *Paris Metro Pricing: The minimalist differentiated services solution*. in *7th International Workshop on Quality of Service (IWQoS '99)*. 1999.
17. Varian, H.R., *Microeconomic Analysis*. 1992: W.W.Norton & Company.
18. Adamic, L., *Zipf, Power-Laws, and Pareto - a Ranking Tutorial*, http://ginger.hpl.hp.com/shl/papers/ranking/.