Managing users in a peer-to-peer storage system

LÁSZLÓ TOKA, ATTILA VIDÁCS

High Speed Networks Laboratory, Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics

{toka, vidacs}@tmit.bme.hu

Keywords: peer-to-peer networks, storage, backup, game theory, incentives, pricing

When managing a peer-to-peer storage system, where users store their data on the disks of their peers for security, availability and accessibility reasons, the lack of incentives for peers to contribute to the service calls for suitable solutions. We suggest two different approaches: either each peer's use of the service be limited to her contribution level (symmetric schemes), or that storage space be bought from and sold to peers by a system operator that seeks to maximize profit. Using a non-cooperative game model to take into account user selfishness, we study these mechanisms with respect to the social welfare performance measure, and give necessary and sufficient conditions for one scheme to socially outperform the other.

1. Introduction

Along with the convergence in every field of the information technology, various kinds of digital content are now likely to be created and accessed from several types of devices connected to different networks. Therefore, an appropriate system for storing users' data should offer a wide range of services, such as ease of access, protection against device failures, versioning and short transfer time from the system to a given device. In this context, the possibility of storing data online appears as a promising solution.

To exploit the great opportunity, many companies propose online data storage service, most of them offering a given storage capacity (up to 25 GB) for free, with the possibility of extending that quota to a higher value for a fixed price (1$/GB/year). However, while running such a storage service implies owning huge storage capacities and affording the associated energy and warehouse costs, one can imagine using smaller but numerous storage spaces of the users themselves.

In a peer-to-peer (p2p) storage system, the participants are at the same time the providers and the users of the service: each participant offers a part of her disk space to provide the service to the others, and benefits from storing her data onto the system. The added value of the service then comes from the protection against failures provided by the system, from the ease of data access, from the versioning management that may be included, and from the difference in the amount of data stored into the system versus offered for service.

An online storage service is valuable only if data availability is assured, therefore to cope with disk failures and with participants disconnecting their disks from the system, data replicates must be spread over several (sufficiently reliable) peers to guarantee that data are almost always available. A functional p2p storage system needs the participants to offer sufficient fractions of their disk space to the system, and to remain online for long periods. However, both of these requirements imply costs (or at least constraints) for participants, who may be reluctant to devote their resources to the system instead of using them for their own needs. The work presented in this paper focuses on the incentives to make participants contribute to the system. We consider that users behave selfishly, i.e., are only sensitive to the quality of service they experience, regardless of the effects of their actions on the other users, therefore the framework of *Non-Cooperative Game Theory* [7] is particularly well-suited to study the interactions among peers.

While the economic aspects of p2p file sharing networks have already been extensively studied (see [2, 4,8,9] and references therein), there are, to our knowledge, no works tackling the economics of p2p storage networks, although a basic difference exists: in file sharing systems when a peer provides some files to the community, she gives value for *all* users since they all can access the data she proposes; in a p2p storage system, on the other hand, the storage space offered by a user can be shared among different peers but each part is then devoted to only *one* user.

The existing literature on p2p storage systems mainly focuses on security, reliability and technical feasibility issues [3,6,10], whereas the incentive aspect has received little attention. Only solutions that do not imply financial transactions are considered in current works, therefore to create some incentives to participate, the counter payment for providing service is usually the service in question as well. This approach leads to a scheme where every peer should contribute to the system in terms of service at least as much as she benefits from others [5,11]; we call such a mechanism a *symmetric* scheme.

We also investigate solutions based on monetary exchanges: users can "buy" storage space for a fixed unit price, and "sell" their own memory space to the system at another unit price. It is known from economics that when those unit prices are fixed by the supply and demand curves (as in a perfect market), the selfish user choices lead to a socially efficient situation. However, it is more likely here that the system is managed by a profit-maximizing entity that fixes prices so as to maximize revenue, therefore we study the *profit oriented price-based* scheme.

Our target is to decide which scheme is more convenient for the society. We consider the social welfare, i.e., the total value that the system has for all participants, as the measure of comparison. Under some assumptions on the user utility functions, we derive a necessary and sufficient condition for symmetry-based systems to outperform revenue-oriented management. We obtain that user heterogeneity tends to favor pricing-based schemes that are more flexible, and, above a given user heterogeneity threshold, even a monopoly-managed system will be socially better than a system imposing symmetry.

The paper is organized as follows. The next section presents the user preference model and the two incentive mechanisms mentioned earlier. After that, we define the social welfare performance measure and compute its value for those two schemes. We compare them in the subsequent section to determine the management scheme that is best suited to a given society; and the last section contains our conclusions.

2. Our model

2.1 Data availability, redundancy and transfers

In a p2p storage system the availability of the stored data is considered as the most important factor in user's appreciation. There are no direct means to guarantee that a given user disk storing a specific file will be online 100% of the time, but to ensure data availability, the system can introduce several tools, such as data replication and redundant coding. We suppose here that the system, when detecting that a peer has gone offline, triggers a recovery of the data stored in that peer from the replicas in the system, and a new storage of those data onto other peers. Likewise, when a peer comes back online, then a new data load will be transferred onto her offered storage space, independently what and whose data she was storing before.

Such a data protection mechanism implies data transfers, and therefore non-monetary costs due to resource consumption (CPU, bandwidth utilization, etc.). A peer i is concerned by those data transfers in two situations: when she comes back online after an offline period (receives new data load), and when other peers enter and leave the system (upload traffic if user i stores replicates of the leaving user's data, download traffic when user i has to store more data). The mean data transfer associated with the first situation is thus proportional to the amount of capacity C_i she offers to the system, and to the mean number of online-offline cycles per unit of time: denoting by t_i^{on} (resp. t_i^{off}) the mean duration of online (resp. offline) periods of user i. The corresponding mean amount of transferred data is then proportional to $C_i/(t_i^{on}+t_i^{off})$. The mean amount of data transferred to and from user i per unit of time in the second situation is proportional to the weighted (by the offered capacity) mean $\bar{\mu}$ of peer status changes per unit of time. This term appears only at those peers who offer storage space, and only during the time they are online.

Consequently, the transfer cost perceived by user i for offering capacity C_i with the mean availability π_i is expressed by, $C_i\pi_i\left(\delta_i/t_i^{\rm on}+\gamma_i\bar{\mu}\right)$, where δ_i and γ_i are parameters that reflect the user characteristics such as sensitivity, access bandwidth, or hardware profile.

2.2 User preferences

We describe user preferences by a utility function which reflects the benefit of using the service by storing C_i^s data amount in the system, the cost of offering storage space $C_i^o := \pi_i C_i$ for other users, and the monetary transactions, if any. The utility U_i of user i is of the form

$$U_i\left(C_i^s, C_i, t_i^{\rm on}, t_i^{\rm off}, \epsilon_i\right) = V_i(C_i^s) - \underbrace{O_i(C_i\pi_i) - C_i\pi_i\left(\delta_i/t_i^{\rm on}+\gamma_i\bar{\mu}\right)}_{P_i(C_i,t_i^{\rm on},t_i^{\rm off})} - \epsilon_i,$$

where

• $V_i(C_i^s)$ is user i's valuation of the storage service, i.e., the price she is willing to pay to store the amount C_i^s of data in the system.

• $P_i(C_i,t_i^{on},t_i^{off})$ is the overall non-monetary cost of user i for offering capacity C_i to the system with availability π_i. It consists of two distinct costs:

• the opportunity cost $O_i(C_i\pi_i)$ of offering storage capacity for other users (during online periods) instead of using it for her own needs;

• $C_i\pi_i\left(\delta_i/t_i^{\rm on}+\gamma_i\bar{\mu}\right)$ data transfer costs due to the data protection mechanism implemented by the system as described in the previous subsection.

• ϵ_i is the monetary price paid by user i for the service taken. This term is 0 in case of a symmetric scheme, and otherwise equals to the price difference between the charge for storing her data into the system and the remuneration for offering her disk space.

2.3 Incentive schemes for cooperation

We assume here that users selfishly choose strategies to maximize their utilities and apart from C_i^s and C_i, each user i can also decide about her behavior related to availability π_i. We describe the two types of incentive mechanisms that we intend to compare later. Both schemes may imply the existence of a central authority or clearance service to supervise peer behavior and/or manage payments: as the model aims to give hints for commercial applications, we do not try to avoid such a centralized system control.

Symmetric schemes

As evoked in the introduction about management solutions without pricing, the principle of those schemes is that users are invited to contribute to, at least as much as they take from, the other users, i.e., it is imposed to each user i that $C_i^o \geq C_i^s$. The availability of the peer is checked (e.g., at randomly chosen times) to ensure that $C_i^o = \pi_i C_i$ exceeds the peer's service use C_i^s.

Payment-based schemes

We consider a simple payment-based mechanism where users can "buy" storage space in the system for a unit price p^s (per byte and per unit of time) and "sell" some of their disk capacity for a (uptime-average) unit price p^o. The (possibly negative) monetary amount that user i is charged is then $\epsilon_i = p^s C_i^s - p^o C_i^o$. We assume that the prices are set by a system operator so as to maximize her revenue, knowing *a priori* the reactions of the users. The operator can thus drive the outcome of the game to the most profitable situation for herself, and in this sense, she acts as the leader of a Stackelberg (or leader-follower) game [7]. In a real implementation of the mechanism, the operator may not perfectly know the user reactions, but an iterative groping of prices can converge to the profit-maximizing ones.

2.4 User behavior related to availability

As given before, a user i has four strategic variables, namely her offered C_i and stored C_i^s capacities, and her mean online t_i^{on} and offline t_i^{off} period durations. Based on the utility equation, when C_i^s and C_i are fixed, the utility of each user is increasing in t_i^{on}, so t_i^{on} will be set by the selfish user to the reachable maximum value $\bar{t}_i^{on}$, which is constrained by uncontrolled events (power black-out, accidents, hardware failures, etc.) that may force the user off the network. Notice that this selfish decision is profitable to the whole network: longer online periods mean fewer data protection transfers and therefore smaller costs for the system (the parameter $\bar{\mu}$ in the utility equation being small).

Note also that by introducing $p_i^{\min} := \delta_i / \bar{t}_i^{\text{on}} + \gamma_i \bar{\mu}$ the transfer costs simply write as $C_i^o p_i^{min}$.

2.5 User supply and demand functions

Supply and demand functions are widely used in economics, and are derived from the utility of consumers and cost functions of providers, respectively. For a user i we call supply function (resp. demand function) the function $s_i(p)$ (resp. $d_i(p)$) such that for a given $p \geq 0$, $s_i(p)$ (resp. $d_i(p)$) is the amount of storage capacity that user i would choose to sell (resp. buy) if she were paid (resp. charged) a unit price p for it.

Our illustrative results consider quasi-linear form of affine supply and demand functions and quadratic valuation and opportunity cost functions. Under this consideration, a user i is entirely described by four parameters *(see Figure 1)*:

- two price thresholds, namely p_i^{min} and p_i^{max}, that respectively represent the minimum value of the selling unit price and the maximum value of the buying unit price;
- two price sensitivities a_i and b_i, that respectively correspond to the increase of sold capacity when selling price grows and the decrease of bought storage space along with the growth of the buying unit price.

The total supply function $S := \sum_{i \in \mathcal{I}} s_i$ is then a (piecewise affine) increasing convex function on the interval [$\min_i p_i^{min}$, $\max_i p_i^{min}$]. Likewise, the total demand function $D := \sum_{i \in \mathcal{I}} d_i$ is decreasing and convex on [$\min_i p_i^{max}$, $\max_i p_i^{max}$], as illustrated in *Figure 2*.

3. Performance of incentive mechanisms

We call the performance measure, used in the following to compare incentive schemes, (social) welfare and denote the sum of the utilities of all agents in the system by W:

$$W := \sum_{i \in \mathcal{I}} V_i(C_i^s) - P_i(C_i^o).$$

Notice that no prices appear in the social welfare expression, since all system agents are included, the operator as well that receives or gives payments, if any, and whose utility is her revenue.

Figure 1. User reactions to prices and user's valuation

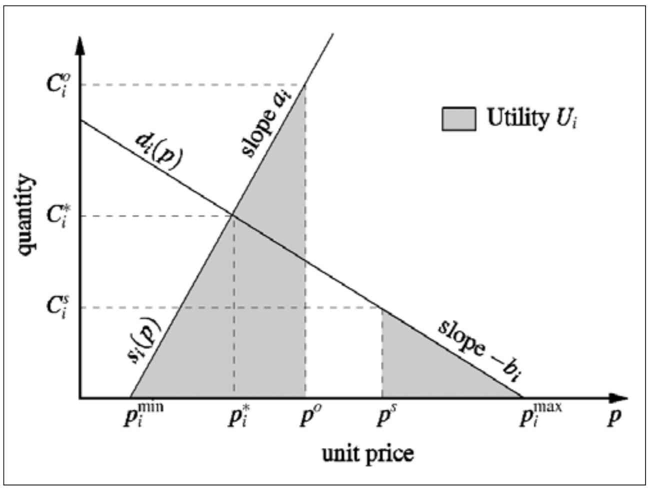
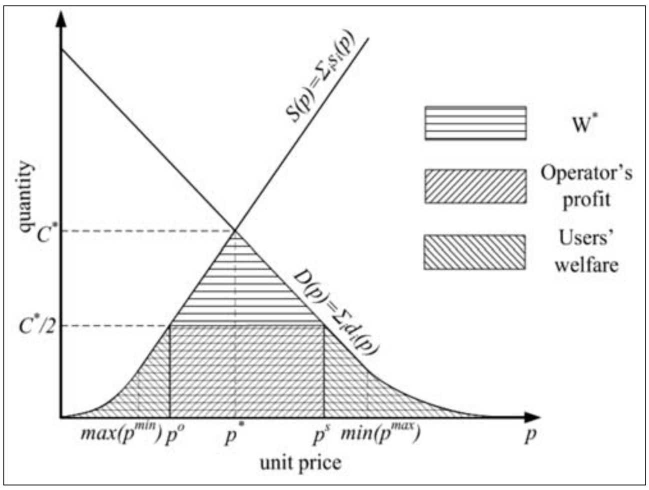


Figure 2. Total supply S and demand D functions, maximum social welfare (lined zone) and surplus repartition (hatched zones) in the case of revenue-driven monopoly under Assumption 1



3.1 Optimal value of social welfare

The optimal situation that the system can attain corresponds to a maximization problem of W, subject to the feasibility constraint $\sum_i C_i^o \geq \sum_i C_i^s$; it can be solved by the Lagrangian method. The maximal value W^* as well as the so-called "shadow price" p^* are illustrated in *Figure 2*. Remark that this optimal situation is reached with a special payment-based scheme where $p^o = p^s = p^*$.

3.2 Performance of symmetric schemes

Under a symmetry-based management scheme, each user i chooses C_i^o and C_i^s so as to maximize her own valuation, respecting the $C_i^o \geq C_i^s$ constraint.

In [13] it is shown that user i maximizes her utility at the point $C_i^s = C_i^o = C_i^*$, as illustrated in Figure 1: this corresponds to the case where every user "exchanges" capacity at the *virtual unit price* p_i^*. However, compared to the socially optimal situation, each and every user loses a part of her utility, thus the system is suboptimal (in terms of social welfare). The loss of welfare depends on the heterogeneity of the users' p_i^*s; if all users have the same p_i^*, then symmetric management schemes maximize social welfare.

3.3 Performance of profit-oriented pricing schemes

When a profit-driven monopoly is employed, the system operator strives to extract the maximum profit out of the business by tuning the prices p^s and p^o.

Figure 2 plots the total supply (S) and the total demand (D) as functions of the unit selling price p^o and the unit buying price p^s, respectively. Our first remark is that p^o and p^s will be chosen such that the demand would be *exactly* satisfied by the supply: otherwise it is always possible for the operator to decrease p^o (if oversupply) or increase p^s (if overdemand) to strictly improve its revenue. The operator revenue with such prices is then the area of the rectangle displayed in Figure 2, embedded within the triangle-shape zone whose area is the maximum value of social welfare.

To be able to predict the maximal profit generating strategy of the monopoly, we make the following assumption regarding user price thresholds.

Assumption 1

The repartitions of price thresholds $p_{\min}$ and $p_{\max}$ are such that their variances on the user set are below a certain level (for details, see [13]).

Moreover, user profile values a_i (resp. b_i) are independent and identically distributed, and each user's a_i and b_i are independent.

Proposition 1

Under Assumption 1, a profit-oriented pricing yields the social welfare W_{mon} such that

$$W^* - W_{mon} = \frac{1}{8} C^{*2} \left(\frac{1}{\sum_i a_i} + \frac{1}{\sum_i b_i} \right)$$

4. Which management to prefer?

When we compare the outcomes of the two schemes, we immediately have the following result.

Proposition 2

Under Assumption 1, symmetric schemes socially outperform profit-oriented pricing mechanisms if and only if

$$\frac{1}{4} C^{*2} \left(\frac{1}{\sum_i a_i} + \frac{1}{\sum_i b_i} \right) \geq \sum_i (a_i + b_i)(p^* - p_i^*)^2.$$

In other words, if the global shadow price and the users' virtual prices are alike, the symmetric scheme reaches higher social welfare.

Proposition 2 combines the four user heterogeneity factors, namely the price thresholds $p_{\min}$, $p_{\max}$ and the price sensitivities a, b, to determine the best mechanism in terms of social welfare. Whereas the right-hand term is the (weighted) variance of the p_i^*, the left-hand term is hard to interpret.

We thus suggest to take a look at the particular cases where user heterogeneity lies entirely on price sensitivities (resp. on price thresholds).

4.1 Homogeneous price thresholds

We consider here that users only differ by their price sensitivities a_i and b_i, and they have the same price thresholds $p_i^{\min}$ and $p_i^{\max}$. This case has been studied in a previous work, we therefore recall the main results and refer the interested reader to [12] for details: it is proven that

$$W_{sym} = \left(\frac{1}{\sum_i a_i} + \frac{1}{\sum_i b_i} \right) \sum_i \left[\frac{1}{\frac{1}{a_i} + \frac{1}{b_i}} \right] W^*,$$

$$W_{mon} = \frac{3}{4} W^*,$$

which yields the following comparison.

Proposition 3

Under our assumptions, symmetric schemes socially outperform profit-oriented pricing mechanisms if and only if $\left(\frac{1}{\sum_i a_i} + \frac{1}{\sum_i b_i} \right) \sum_i \frac{1}{\frac{1}{a_i} + \frac{1}{b_i}} \geq \frac{3}{4}.$

Moreover, if the couples (a_i, b_i) are independently chosen for all users and identically distributed, then the inequality holds if and only if

$$\frac{\mathbb{E}[f(a,b)]}{f(\mathbb{E}[a], \mathbb{E}[b])} \geq \frac{3}{4}$$

when the number of users tends to infinity (law of large numbers), with $f : (x,y) \mapsto \frac{1}{1/x + 1/y}$.

Since the function f is strictly concave, from Jensen's inequality the left-hand term is always smaller than 1, and decreases as the dispersion of (a, b) increases. Remark that when (a, b) are deterministic then the left-hand term equals to 1 and symmetric schemes are better than profit-oriented ones, as we remarked above.

Let us have a look at Proposition 3 for two simple examples of (a, b)'s distribution, assuming that a and b are independent variables.

Uniform distribution:

If a (resp. b) is uniformly distributed over $[0, a_{\max}]$ (resp. $[0, b_{\max}]$), the inequality always holds.

Exponential distribution:

If a (resp. b) follows exponential distribution with parameter μ_a (resp. μ_b), either a symmetric or a profit-oriented mechanism is socially preferable depending on the relative values of μ_a and μ_b.

4.2 Homogeneous price sensitivities

We now consider the case where the price thresholds p_i^{min} and p_i^{max} are user-specific, but the price sensitivities a_i and b_i are identical for every user. The couples (p_i^{min}, p_i^{max}) are independent and identically distributed among users; moreover p_i^{min} is independent of p_i^{max} for all users.

Proposition 4

Under these assumptions, managing mechanisms based on symmetry are always socially better (in terms of social welfare) than profit-oriented pricing mechanisms.

5. Conclusion

In this work we have addressed the problem of user incentives in a peer-to-peer storage system. Using a game theoretical model to describe selfish reactions of all systems actors (users and the operator), we have studied and compared the outcomes of two possible managing schemes, namely symmetry-based and profit-oriented payment-based policies.

Not only the size of the offered storage space was targeted with incentives, but as the availability and reliability are particularly important issues in storage systems, the churn as well. By comparing the social welfare in the two cases, under some assumptions on user preferences, we exhibited a necessary and sufficient condition for a type of management to be preferable to the other: it appears that profit oriented payment-based schemes may be socially better than symmetric ones under some specific circumstances, namely if the heterogeneity among user profiles is high.

Acknowledgement

László Toka hereby thanks Patrick Maillé for his continuous and extremely helpful support during his work, and to the HTE Education Committee for considering his MSc thesis valuable to reward and publish.

Authors

ATTILA VIDÁCS was born in Budapest, in 1973. He received the MSc and PhD Degrees from the Budapest University of Technology and Economics at the Department of Telecommunications and Media Informatics, Faculty of Electrical Engineering and Informatics, in 1996 and 2001, respectively. During 1997, he worked as a visiting researcher at the Research and Development Center of the Nippon Telegraph and Telephone Corp., Tokyo, Japan. Currently he is Associate Professor at the Dept. of Telecommunications and Media Informatics, Budapest University of Technology and Economics. His research interests are in the field of teletraffic modeling and traffic engineering, sensor networking, and dynamic spectrum access networks.

LÁSZLÓ TOKA graduated in 2007 at Budapest University of Technology and Economics and received his MSc degree in Telecommunications on the Faculty of Electrical Engineering and Informatics. He spent the last two years of his study in France and obtained the engineer diploma of Telecom Bretagne and Eurecom. He also participated in the pre-doctoral education courses of the Networks and Distributed Systems Department at the University of Nice Sophia-Antipolis and received a research master degree. During his undergraduate years László participated in research projects focusing on economic modeling of distributed IT systems and networks. After graduating he followed up on his research domain and enrolled as a PhD candidate at Telecom Paris and at the Budapest University of Technology and Economics.

References

[1] E. Adar and B. Huberman, Free riding on gnutella. Tech. Rep., Xerox parc, 2000.

[2] P. Antoniadis, C. Courcoubetis, R. Mason, Comparing economic incentives in peer-to-peer networks. Computer Networks, 46(1):133–146, 2004.

[3] C. Batten, K. Barr, A. Saraf, S. Treptin, pStore: A secure peer-to-peer backup system. Tech. Rep., MIT Lab. for Computer Science, 2001.

[4] C. Courcoubetis and R. Weber, Incentives for large peer-to-peer systems. IEEE JSAC, 24(5):1034–1050, 2006.

[5] L. Cox and B. Noble, Samsara: Honor among thieves in peer-to-peer storage. In Proc. of 19th ACM Symposium on Operating Systems Principles (SOSP'03), 2003.

[6] P. Druschel and A. Rowstron, PAST: A large-scale, persistent peer-to-peer storage utility. In HotOS VIII, pp.75–80., 2001.

[7] D. Fudenberg and J. Tirole, Game Theory. MIT Press, 1991.

[8] P. Golle, K. Leyton-Brown, I. Mironov, M. Lillibridge, Incentives for sharing in peer-to-peer networks. In Proc. of 3rd ACM Conf. on Electronic Commerce (EC'01), pp.264–267., 2001.

[9] K. Lai, M. Feldman, I. Stoica, J. Chuang, Incentives for cooperation in peer-to-peer networks. In Proc. of 1st Workshop on Economics of Peer-to-Peer Systems, 2003.

[10] M. Lillibridge, S. Elnikety, A. Birrell, M. Burrows, M. Isard A cooperative internet backup scheme. In Proc. of 1st Workshop on Economics of Peer-to-Peer Systems, 2003.

[11] B. Stefansson, A. Thodis, A. Ghodsi, S. Haridi, MyriadStore. Tech. Rep., Swedish Inst. of Comp. Science, 2006.

[12] L. Toka, P. Maillé, Managing a Peer-to-Peer Backup System: Does Imposed Fairness Socially Outperform a Revenue-Driven Monopoly? In LNCS Proc. of 4th International Workshop on Grid Economics and Business Models (GECON), 2007.

[13] L. Toka, Peer-to-peer sharing of hard disks: study of incentive mechanisms, MSc thesis, 2007.