Managing a Peer-to-Peer Data Storage System in a Selfish Society

Patrick Maillé, László Toka

***Abstract*—We compare two possible mechanisms to manage a peer-to-peer storage system, where participants can store data online on the disks of peers in order to increase data availability and accessibility. Due to the lack of incentives for peers to contribute to the service, we suggest that 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 noncooperative game model to take into account user selfishness, we study those mechanisms with respect to the social welfare performance measure, and give necessary and sufficient conditions for one scheme to socially outperform the other.**

***Index Terms*—Peer-to-peer networks, game theory, incentives, pricing.**

I. Introduction

THE "digital society" that has been soaring since the creation of the Internet implies that all kinds of digital documents are now likely to be created, accessed, and modified from several types of devices. Therefore, an appropriate system for storing the data of a user should offer various services, such as versioning, ease of access, protection against device failures, and short transfer time to a given device.

In that context, the possibility of storing data online appears as a promising solution. Indeed, having access to the Internet becomes easier and easier, with the multiplication of WiFi hotspots, the development of WiMAX and third generation wireless networks, and the appearance of other access modes, such as multi-hop networks that work in an ad-hoc fashion to reach an access point. Let us also highlight the high rise of available transmission rates in access networks, which renders transfer times reasonable, even for large files. Finally, online storage systems are able to cope with document versioning, and to protect data not only against user device failures but also against disk failures, through the use of data replicates stored on different disks.

For those reasons, many companies now propose online data storage services, most of them offering a given storage capacity (between 2 and 25 gigabytes) for free, with the possibility of extending that quota to a higher value for a fixed price per year (the price per year per gigabyte being of the order of 1$). However, while creating such a storage service implies owning huge memory capacities and affording the associated energy and warehouse costs, one can imagine using the smaller but numerous storage spaces of the service users themselves, as is done in peer-to-peer file sharing systems.

In a peer-to-peer storage system, the participants are at the same time the providers and the users of the service: each participant offers some memory capacity (possibly from multiple locations in the network: part of her disk space at home, storing device devoted to the service, ...) to provide the service to the others, and benefits from storing her own 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 to the service.

An online storage service is valuable only if data are available: therefore to cope with disk failures and with participants disconnecting their disk from the system, data replicates must be spread over several (sufficiently reliable) peers to guarantee that data are not lost and are almost always available; the data replication rate then depends on the reliability of the participants. To work properly, a peer-to-peer storage network therefore needs that participants offer a sufficient part of their disk space to the system, and remain online often enough. However, both of those requirements imply costs (or at least constraints) for participants, who may be reluctant to devote some of their storage capacity to the system instead of using it for their own needs.

In this paper, 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. The framework of *noncooperative Game Theory* [1] is therefore particularly well-suited to study the interactions among peers. For a peer-to-peer storage system, it is clear that without any reward for contributing participants, selfishly behaving users will only benefit from the service without providing any part of it[1]. In other words, the only Nash equilibrium of the noncooperative game is the situation where the system actually does not exist due to the lack of offering peers.

The work presented in this paper focuses on the incentives that can be used to make participants contribute to the system, i.e. the changes that can be brought to the game to modify its Nash equilibria. While the economic aspects of peer-to-peer file sharing networks have already been extensively studied (see [3], [4], [5], [6] and references therein), there are to our

P. Maillé is with the Department of Network, Security and Multimedia, TELECOM Bretagne, 2, rue de la Châtaigneraie, 35576 Cesson-Sévigné Cedex, FRANCE. E-mail: patrick.maille@telecom-bretagne.eu

L. Toka is with the Department of Network and Security, EURECOM, 2229 route des Crêtes , 06560 Sophia-Antipolis Cedex, FRANCE and with the Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, 2 Magyar Tudósok krt., 1117 Budapest, HUNGARY. E-mail: toka@eurecom.fr

This work has been partially supported by the TELECOM Institute project DisPairSe.

Manuscript received July 31, 2007; revised March 5, 2008.

[1] Such a behavior, called *free-riding* [2], also appears in peer-to-peer file sharing networks, and is problematic for the survival of those altruism-based networks.

knowledge no references on the economics of peer-to-peer storage networks. Now, the economic models developed for peer-to-peer file sharing systems do not apply to peer-to-peer storage services: in file sharing systems, when a peer provides some files to the community, she adds value to the system for all users since they all can access the data she proposes; in that sense the resource offered to the system is a public good. On the contrary, in a peer-to-peer storage system the memory space offered by a peer is a private good: it can be shared among different users but each part is then devoted to only one user. Therefore the economic implications of those systems are necessarily different.

The existing literature on peer-to-peer storage systems mainly focuses on security, reliability and technical feasibility issues [7], [8], [9], whereas the incentive aspect 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 finally 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 [10], [11]. We call such a mechanism imposing the contribution of each peer to equal her use of the system a *symmetric* scheme.

In this paper, 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 economic theory that when those unit prices are fixed by the supply and demand curves (as in a perfect market [12]), then user selfish choices lead to a socially efficient situation. However, it is more likely here that the system be managed by a profit-maximizing entity that fixes prices so as to maximize revenue. That entity then acts as the leader of a Stackelberg game [1].

The main question addressed in this paper is whether it is socially better to impose a symmetric scheme or to let a profit-maximizing monopoly set prices. The performance measure we consider is social welfare, i.e. the total value that the system has for all participants. Under some assumptions on the peers 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 results presented here are a generalization of our preliminary work [13] that did not consider incentives to stay online and where the only source of heterogeneity came from the price sensitivities.

This paper is organized as follows. Section II introduces the model we consider for user preferences, and for the two incentive mechanisms studied in this paper, namely symmetry-based and profit oriented price-based schemes. In Section III we define the social welfare performance measure and compute its value for those two types of schemes. We compare them in Section IV to determine the management scheme that is best suited to the society, and present our conclusions in Section V.

II. MODEL

A. Content availability management and associated costs

In a peer-to-peer storage system the availability of the stored data is considered as the most important factor in user's appreciation. As the storage disks are users' property, there are no direct means to guarantee that a given user disk storing a specific file will be online 100% of the time. To ensure data availability, the system can introduce several tools, such as data replication and coding [14]. We suppose here that the system 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 into other peers. Likewise, when a peer comes back online, then new data will be transferred into her offered storage space, independently what and whose data she was storing before. Such a scheme is purely reactive (actions are taken when a user departure is detected). One could also imagine using proactive approaches, or a combination of both, to smoothe the incurred traffic [15].

This data protection mechanism implies data transfers, and therefore nonmonetary 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 (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 to 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 data transferred is then proportional to $C_i/(t_i^{\text{on}} + t_i^{\text{off}})$. 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[2]. This term appears only at those peers who offer storage space (proportionally to their offered capacity since the probability that user i be concerned by a peer's departure is proportional to C_i), and only during the time they are online (it is therefore also proportional to the mean availability of user i, $\pi_i := t_i^{\text{on}}/(t_i^{\text{on}} + t_i^{\text{off}})$).

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

B. User preferences

We describe the preferences of a user i in the user set denoted by $\mathcal{I}$ by a utility function, that reflects the benefit of using the service by storing an amount C_i^s of data in the system, the cost of offering storage space $C_i^o := \pi_i C_i$ for

[2]Actually peer i should only be sensitive to the status change rate of all other peers but hers. However we consider here a system with a very large number of users, so that taking the mean of the change rates over all participants but one is equivalent to considering all participants.

other users, and the monetary transactions, if any. We suggest to use a separable additive function.

Definition 1: The utility U_i of a user $i \in \mathcal{I}$ is of the form

$$U_i\left(C_i^s, C_i, t_i^{\text{on}}, t_i^{\text{off}}, \epsilon_i\right) = V_i(C_i^s) - \underbrace{O_i(C_i\pi_i) - C_i\pi_i\left(\delta_i/t_i^{\text{on}} + \gamma_i\bar{\mu}\right)}_{:=P_i(C_i,t_i^{\text{on}},t_i^{\text{off}})} - \epsilon_i, \quad (1)$$

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 an amount C_i^s of data in the system[3]. We assume that $V_i(\cdot)$ is positive, continuously differentiable, increasing and concave in its argument, and that $V_i(0) = 0$ (no service yields no value).

• $P_i(C_i, t_i^{\text{on}}, t_i^{\text{off}})$ is the overall non-monetary cost of user i for offering capacity C_i to the system with mean online and offline durations respectively equal to t_i^{on} and t_i^{off}, i.e. with availability $\pi_i = \frac{t_i^{\text{on}}}{t_i^{\text{on}}+t_i^{\text{off}}}$. It consists of two distinct costs:

- an 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[4], where $O_i(\cdot)$ is assumed positive, continuously differentiable, increasing and strictly convex, and such that $O_i(0) = 0$ (no contribution brings no cost);
- data transfer costs $C_i\pi_i\left(\delta_i/t_i^{\text{on}} + \gamma_i\bar{\mu}\right)$ 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. This term is 0 in case of a symmetric scheme, and otherwise equals the price difference between the charge for storing her data into the system and the remuneration for offering her disk space.

Remark that we implicitly say that the storage space necessary to safely store some data in the system equals the size of those data. This is done without loss of generality, taking into account the redundancy factor r added by the system in users' cost function: a user considered to offer space to store an amount C_i of data actually devotes more of her disk space (rC_i) to the service. Likewise, prices are then per unit of "protected data".

C. Incentive schemes for cooperation

Users selfishly choose strategies that maximize their utility. We assume here that apart from C_i^s and C_i^o, each user i can also decide about her behavior related to availability. In this subsection, we describe the two types of incentive mechanisms that we intend to compare in this paper. Both schemes may imply the existence of a central authority or clearance service to supervise the peers 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.

[3]We assume here that data replication ensures a given availability, so that this availability does not appear in the utility function.

[4]We implicitly assume here that the opportunity cost depends only on the mean capacity offered over time, since during offline periods the user can use the disk space for other purposes than the service.

1) Symmetric schemes: We follow here the ideas suggested in the literature for schemes without pricing. As evoked in the introduction, the principle of those schemes is that users are invited to contribute to, at least as much as they take from, the other users. The availability of the peer is therefore 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.

We assume in this paper that this verification is technically feasible. Determining whether and how it can be done remains an active topic of research and is beyond the scope of this paper, since we only focus here on incentives.

2) 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 (time-average available) disk capacity for a unit price p^o.

The (possibly negative) amount that user i is charged is then

$$\epsilon_i = p^s C_i^s - p^o C_i^o.$$

In this paper, we assume that prices are set by the 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 [1]. In a real implementation of the mechanism, the operator may not perfectly know the user reactions, but an iterative tâtonnement of prices can converge to those profit-maximizing prices.

D. User behavior related to availability

In the game we study, a peer $i \in \mathcal{I}$ 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. Equivalently, we can also consider that the four strategic variables are $C_i^s, C_i^o, t_i^{\text{on}}$, and t_i^{off}. From (1), when C_i^s and C_i^o are fixed, the utility of each user is increasing in t_i^{on}, so t_i^{on} will be set by user i to a maximum value. We denote by $\bar{t}_i^{\text{on}}$ that maximum value, which is only limited 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 (1) being small). Remark also that since t_i^{off} does not appear in (1), there remain only two decision variables, namely C_i^s and C_i^o (that equals $C_i\bar{t}_i^{\text{on}}/(\bar{t}_i^{\text{on}} + t_i^{\text{off}})$). From now we will therefore write $P_i(C_i^o)$ instead of $P_i(C_i, t_i^{\text{on}}, t_i^{\text{off}})$, and will also use the notation

$$p_i^{\min} := \delta_i/\bar{t}_i^{\text{on}} + \gamma_i\bar{\mu}, \quad (2)$$

so that the transfer costs simply write $C_i^o p_i^{\min}$.

E. User supply and demand functions

Supply and demand functions are classically used in economics [12], and are respectively derived from the valuation of consumers and cost functions of providers. Notice however the particularity here that peers can be consumers and providers at the same time.

Definition 2: For a user $i \in \mathcal{I}$, we call *supply function* (resp. *demand function*) the function $s_i(\cdot)$ (resp. $d_i(\cdot)$) such that for all $p \in \mathbb{R}_+$,

$$\begin{aligned} s_i(p) &:= \inf\{q \geq 0 : P_i'(q) \geq p\}, \\ d_i(p) &:= \inf\{q \geq 0 : V_i'(q) \leq p\}, \end{aligned}$$

where g' stands for the derivative function of g, and with the convention $\inf \emptyset = +\infty$.

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.

For the sake of simplicity, our main results in the following consider a particular form of supply and demand functions described in the assumption below.

Assumption A: For all $i \in \mathcal{I}$, the supply and demand functions of user i are affine. More precisely, there exist nonnegative values a_i, b_i, and $p_i^{\max}$ such that

$$s_i(p) = a_i[p - p_i^{\min}]^+, \quad (3)$$
$$d_i(p) = b_i[p_i^{\max} - p]^+, \quad (4)$$

where $p_i^{\min}$ is given in (2), $x^+ := \max(0, x)$, and we assume that $\max_i p_i^{\min} < \min_i p_i^{\max}$.

This actually corresponds to quadratic functions for the valuation and opportunity cost (with $\wedge$ denoting the $\min$):

$$\begin{aligned} O_i(C_i^o) &= \frac{1}{a_i}\frac{C_i^{o2}}{2}, \\ V_i(C_i^s) &= \frac{1}{b_i}\left(-\frac{(C_i^s \wedge b_i p_i^{\max})^2}{2} + b_i p_i^{\max}(C_i^s \wedge b_i p_i^{\max})\right). \end{aligned}$$

Under Assumption A, 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 unit price p^o such that user i sells some of her disk space and the maximum value of the unit price p^s such that she buys some storage space,
- two price sensitivities a_i and b_i, that respectively correspond to the increase of sold capacity with the unit price $p^o \geq p_i^{\min}$ and the decrease of bought storage space with the unit price $p^s \leq p_i^{\max}$.

Consequently, the total supply function $S := \sum_{i\in\mathcal{I}} s_i$ is a (piecewise affine) nondecreasing convex function on the interval $[\min_i p_i^{\min}, \max_i p_i^{\min}]$, and is affine on $[\max_i p_i^{\min}, +\infty)$. Likewise, the total demand function $D := \sum_{i\in\mathcal{I}} d_i$ is nonincreasing, affine on $[0, \min_i p_i^{\max}]$ and convex on $[\min_i p_i^{\max}, \max_i p_i^{\max}]$, as illustrated in Figure 2 displayed in subsection III-C.

III. Social welfare performance of incentive mechanisms

In this section we introduce the performance measure used in this paper to compare incentive schemes, and study its value for the social optimum and the outcomes of the two incentive schemes that are the object of this paper.

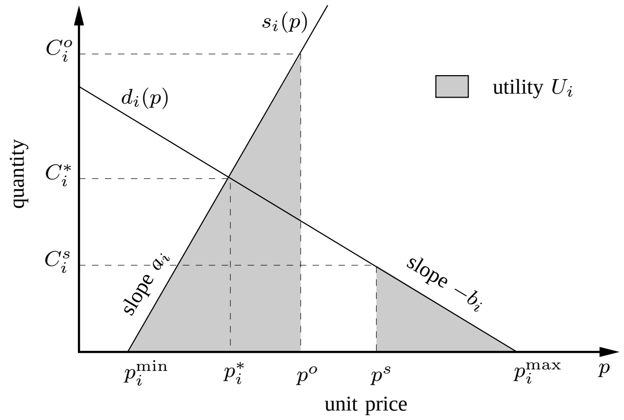


Fig. 1. Reactions to prices and utility of a user $i \in \mathcal{I}$ under Assumption A.

Definition 3: We call *social welfare* (or welfare) and denote by W the sum of the utilities of all agents in the system:

$$W := \sum_{i\in\mathcal{I}} V_i(C_i^s) - P_i(C_i^o). \quad (5)$$

Notice that no prices appear in (5), since all system agents are considered, including the operator that receives or gives payments, if any, and whose utility is her revenue. The operator being a member of the society, all money it exchanges with the users stays within the system and therefore does not influence social welfare.

A. Optimal value of social welfare

The optimal situation (in terms of social welfare) that the system can attain corresponds to the maximization problem $\max_{C_i^s, C_i^o} \sum_{i\in\mathcal{I}} V_i(C_i^s) - P_i(C_i^o)$, subject to the feasibility constraints $C_i^o \geq 0$, $C_i^s \geq 0$ for $\forall i$ and $\sum_i C_i^o \geq \sum_i C_i^s$.

This classical convex optimization problem can be solved by the Lagrangian method: if p^* and C^* are the (unique) solutions of the demand-supply equation

$$C^* := \sum_i s_i(p^*) = \sum_i d_i(p^*), \quad (6)$$

then the maximum social welfare is attained when $C_i^s = d_i(p^*)$, and $C_i^o = s_i(p^*)$ for all $i \in \mathcal{I}$. Under Assumption A, the optimal social welfare W^* is then

$$W^* = \frac{1}{2}\sum_i b_i\left(p_i^{\max 2} - p^{*2}\right) - a_i\left(p^{*2} - p_i^{\min 2}\right). \quad (7)$$

This maximal value W^* as well as the so-called "shadow price" p^* are illustrated in Figure 2 displayed in Subsection III-C. Remark that this optimal situation can be attained with a payment-based scheme where $p^o = p^s = p^*$.

B. 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 $V_i(C_i^s) - P_i(C_i^o)$, subject to $C_i^o \geq C_i^s$. As $P_i(\cdot)$ is increasing in C_i^s, it is in each user's best interest to choose a strategy with $C_i^o = C_i^s$. User i then maximizes her utility at the point $C_i^s = C_i^o = C_i^*$

where $V_i'(C_i^*) = P_i'(C_i^*)$, as illustrated in Figure 1. Under Assumption A, this corresponds to every user "exchanging" capacity $C_i^* = \frac{a_i b_i}{a_i+b_i}(p_i^{\max} - p_i^{\min})$ at the *virtual unit price* $p_i^* := \frac{a_i p_i^{\min} + b_i p_i^{\max}}{a_i+b_i}$. Compared to the socially optimal situation, each and every user loses $\frac{1}{2}(p^* - p_i^*)^2(a_i + b_i)$ of utility if $\max_i p_i^{\min} \le p^* \le \min_i p_i^{\max}$. In that case, the welfare loss of the system is

$$W^* - W_{sym} = \sum_i \frac{a_i + b_i}{2}\,(p^* - p_i^*)^2. \tag{8}$$

Remark that $p^* = \frac{\sum_i (a_i+b_i)p_i^*}{\sum_i a_i+b_i}$ is then the weighted mean of p_i^*, therefore the loss of welfare only depends on the heterogeneity of users' p_i^*. In particular, in the case when all users have the same p_i^*, then symmetric management schemes maximize social welfare.

C. Performance of profit-oriented pricing schemes

We now study a pricing mechanism where the system operator strives to extract the maximum profit out of the business by playing on prices p^s and p^o. Knowing that each user i will sell $s_i(p^o)$ and buy $d_i(p^s)$, the operator faces the following maximization problem.

$$\max_{p^s,p^o} \left(p^s \sum_i d_i\,(p^s) - p^o \sum_i s_i\,(p^o) \right), \tag{9}$$

subject to $p^s \ge 0$, $p^o \ge 0$ and the feasibility constraint $\sum_i s_i(p^o) \ge \sum_i d_i(p^s)$.

Let us examine the best choices for such a profit-driven monopoly. Figure 2 plots two curves: the total supply $S = \sum_i s_i$ and the total demand $D = \sum_i d_i$ as functions of the unit price p. First remark that p^o and p^s must be chosen such that $S(p^o) = D(p^s)$: otherwise it is always possible for the operator to decrease p^o (if $S(p^o) > D(p^s)$) or increase p^s (if $S(p^o) < D(p^s)$) to strictly improve its revenue. The operator revenue with such prices is then the area of the rectangle displayed in the left hand side of Figure 2, embedded within a zone whose area is the maximum value of social welfare. While $p^o > \max_i p_i^{\min}$ and $p^s < \min_i p_i^{\max}$, the largest revenue is attained when $S(p^o) = D(p^s) = C^*/2$. However we are not guaranteed that such p^o and p^s indeed verify $p^o > \max_i p_i^{\min}$ and $p^s < \min_i p_i^{\max}$, nor are we assured that such a choice yields the maximum revenue (that maximum might actually be attained with $p^o < \max_i p_i^{\min}$ or $p^s > \min_i p_i^{\max}$).

To be able to predict the choices of the profit-oriented monopoly, we therefore make the following assumption regarding user price thresholds, that fixes those two points.

Assumption B: The repartition of price thresholds $p^{\min}$ and $p^{\max}$ is such that

$$\max_i p_i^{\min} \quad \le \quad \frac{p^* + \min_i p_i^{\min}}{2}, \tag{10}$$

$$\min_i p_i^{\max} \quad \ge \quad \frac{p^* + \max_i p_i^{\max}}{2}, \tag{11}$$

where $p^* = \frac{\sum_i a_i p_i^{\min} + b_i p_i^{\max}}{\sum_i a_i + b_i}$ from (6). Moreover user profile values a_i (resp. b_i) of all users $i \in \mathcal{I}$ are independent and identically distributed, and a_i and b_i are independent.

Remark that this straightforwardly imply that $\max_i p_i^{\min} \le p^* \le \min_i p_i^{\max}$, so under Assumptions A and B, (8) holds as noticed in the previous subsection.

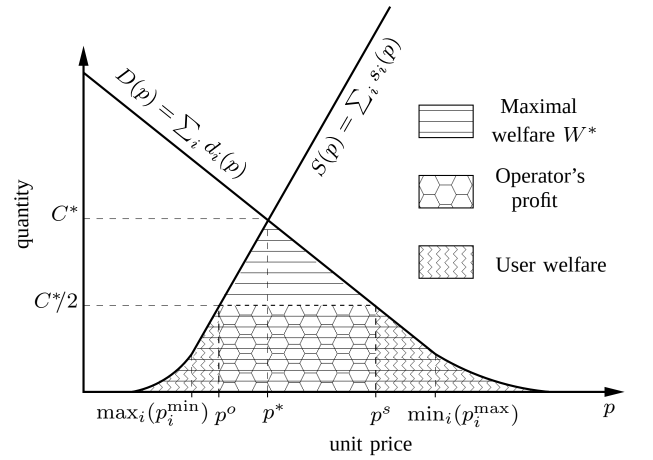


Fig. 2. Total supply S and demand D functions, maximum social welfare and surplus repartition with a revenue-driven monopoly under Assumption A.

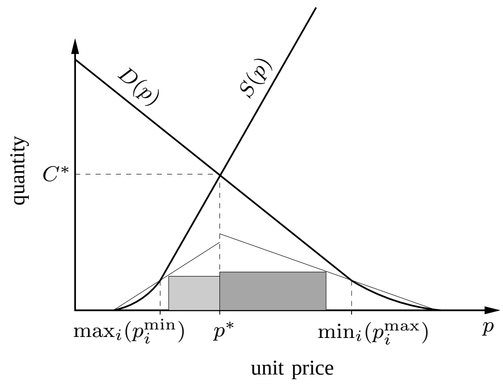


Fig. 3. Illustration of the proof of Proposition 1.

We can now quantify the performance of pricing mechanisms designed to maximize revenue.

Proposition 1: Under Assumptions A and B, a profit-oriented pricing yields a social welfare W_{mon} such that (with C^* given in (6)):

$$W^* - W_{mon} = \frac{1}{8} C^{*2} \left(\frac{1}{\sum_i a_i} + \frac{1}{\sum_i b_i} \right). \tag{12}$$

Proof: We first establish that the monopoly chooses the profit-maximizing unit prices p^o and p^s such that $S(p^o) = D(p^s) = C^*/2$, where p^* is the welfare-maximizing price given in (6). To do so, we compute an upper bound of the revenue that can be attained when choosing the prices in the non-linear part of S or D. Since both functions are convex, we can upper bound them by their cords on $[\min_i p_i^{\min}, \max_i p_i^{\min}]$ for the supply function, and on $[\min_i p_i^{\max}, \max_i p_i^{\max}]$ for the demand function. We extend these segments until the vertical line $p = p^*$ to form two triangles (with the abscissa axis). Under Assumption B, the largest rectangle embedded in each triangle is indeed embedded in the triangle formed by extending the affine parts of S and D, as illustrated in Figure 3. Therefore the sum of their areas is smaller than the revenue yielded by the prices verifying $S(p^o) = D(p^s) = S(p^*)/2$, which are thus the profit-maximizing prices.

As illustrated in Figure 2, the difference $W^* - W_{mon}$ is then simply the area of the hatched triangle above the horizontal line $C^*/2$, which gives (12). ∎

IV. Which management to prefer?

In this section we compare the outcomes of the two practical schemes, i.e symmetric and payment-based schemes. From (8) and (12) we immediately have the following result.

Proposition 2: Under Assumptions A and B, symmetric schemes socially outperform profit-oriented pricing mechanisms if and only if

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

where C^* and p^* are given in (6), and $p_i^* = \frac{a_i p_i^{\min} + b_i p_i^{\max}}{a_i + b_i}$. That condition is equivalent to

$$\frac{1}{4}\left(p^* - \sum_i \alpha_i p_i^{\min}\right)\left(\sum_i \beta_i p_i^{\max} - p^*\right) \geq \sum_i \omega_i (p_i^* - p^*)^2, \quad (13)$$

with the weights for all $i \in \mathcal{I}$: $\alpha_i := \frac{a_i}{\sum_i a_i}$, $\beta_i := \frac{b_i}{\sum_i b_i}$, and $\omega_i := \frac{a_i + b_i}{\sum_i a_i + b_i}$.

Proof: Relation (13) comes after some algebra, using the equalities $p^* = \sum_i \omega_i p_i^*$ and $C^* = \sum_i a_i(p^* - p_i^{\min}) = \sum_i b_i(p_i^{\max} - p^*)$. ∎

Proposition 2 combines the four user heterogeneity factors, namely the price thresholds $p^{\min}, p^{\max}$ and price sensitivities a, b, to determine the best mechanism in terms of social welfare. Whereas the right-hand term of (13) is the variance of the p_i^* with weights w_i, the left-hand term is hard to interpret. We thus suggest to have a look at the particular cases where user heterogeneity lies entirely on prices sensitivities (resp. on price thresholds).

A. Homogeneous price thresholds

We consider here that users only differ by their price sensitivities a_i and b_i. That simplified model has been studied in a previous work [13], we therefore recall the main results and refer the interested reader to [13] for details.

Assumption C: All users $i \in \mathcal{I}$ have the same price thresholds $p_i^{\min}$ and $p_i^{\max}$. Without loss of generality (via a change of abscissa in Figure 2), we can therefore assume that

$$\forall i \in \mathcal{I}, \quad p_i^{\min} = 0 \quad \text{and} \quad p_i^{\max} = p^{\max}.$$

Notice that under Assumptions A and C, Assumption B always holds. It can then be proved (see [13]) that

$$\begin{cases} W_{sym} &= \left(\frac{1}{\sum_i a_i} + \frac{1}{\sum_i b_i}\right)\sum_i \left[\frac{1}{\frac{1}{a_i} + \frac{1}{b_i}}\right] W^*, \\ W_{mon} &= \frac{3}{4}W^*. \end{cases}$$

This yields the following comparison (which can also be directly obtained from Proposition 2 after some algebra).

Proposition 3: Under Assumptions A and C, symmetric schemes socially outperform profit-oriented pricing mechanisms if and only if

$$\left(\frac{1}{\sum_i a_i} + \frac{1}{\sum_i b_i}\right)\sum_i \frac{1}{\frac{1}{a_i} + \frac{1}{b_i}} \geq \frac{3}{4}. \quad (14)$$

Moreover, if the couples (a_i, b_i) are independently chosen for all users and identically distributed, then when the number of users tends to infinity, (14) writes

$$\frac{\mathbb{E}[f(a,b)]}{f(\mathbb{E}[a], \mathbb{E}[b])} \geq \frac{3}{4}, \text{ with } f : (x,y) \mapsto \frac{1}{1/x + 1/y}. \quad (15)$$

Since the function f is strictly concave, from Jensen's inequality the left-hand term of (15) 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 of (15) equals 1 and symmetric schemes are better than profit-oriented ones, as we remarked in subsection III-B.

Let us have a look at (15) for two simple examples of distributions for (a, b), assuming that a and b are independent variables.

• *Uniform distribution*. If a (resp. b) is uniformly distributed over $[0, a_{\max}]$ (resp. $[0, b_{\max}]$),

$$\frac{\mathbb{E}[f(a,b)]}{f(\mathbb{E}[a], \mathbb{E}[b])} = \frac{2}{3}\left(\frac{1}{a_{\max}} + \frac{1}{b_{\max}}\right)\Bigg(a_{\max} + b_{\max} - \frac{a_{\max}^2}{b_{\max}}\ln(1 + \frac{b_{\max}}{a_{\max}}) - \frac{b_{\max}^2}{a_{\max}}\ln(1 + \frac{a_{\max}}{b_{\max}})\Bigg). \quad (16)$$

This expression is minimum when $a_{\max} = b_{\max}$, in which case it equals $8(1 - \ln(2))/3 \simeq 0.82$. Consequently inequality (15) always holds.

• *Exponential distribution*. If a (resp. b) follows an exponential distributions with parameter μ_a (resp. μ_b), i.e. $\mathbb{P}(a > x) = e^{-\mu_a x}$, then we obtain after some calculation

$$\frac{\mathbb{E}[f(a,b)]}{f(\mathbb{E}[a], \mathbb{E}[b])} \geq \frac{3}{4} \quad \Leftrightarrow \quad \alpha \leq \frac{\mu_a}{\mu_b} \leq \frac{1}{\alpha},$$

where $\alpha \simeq 0.179$ is the smallest positive root of $x \mapsto \frac{1+x}{(1-x)^3}(1 - x^2 + 2x\ln(x)) - 3/4$. In that case, either a symmetric or a profit oriented mechanism is socially preferable depending on the relative values of μ_a and μ_b.

B. Homogeneous price sensitivities

We now consider the case where the price thresholds $p_i^{\min}$ and $p_i^{\max}$ can be user specific, but the price sensitivities a_i and b_i are identical for every user.

Assumption D: All users have the same price sensitivity of supply (resp. demand), i.e. $\forall i \in \mathcal{I}, \quad a_i = a \quad \text{and} \quad b_i = b$. Moreover, the couples $(p_i^{\min}, p_i^{\max})$ are independent and identically distributed among users, and $p_i^{\min}$ is independent of $p_i^{\max}$ for all $i \in \mathcal{I}$.

In that case, we establish that one mechanism is always preferable to the other.

Proposition 4: Under Assumptions A, B and D, management mechanisms based on symmetry are always socially better than profit-oriented pricing mechanisms.

Proof: Assumption D implies that for all i, the weights α_i, β_i and ω_i introduced in (13) all equal $\frac{1}{n}$, where n is the number of users. Moreover we have $p^* = \frac{a\tilde{p}^{\min}+b\tilde{p}^{\max}}{a+b}$, where $\tilde{p}^{\min} := \frac{\sum_i p_i^{\min}}{n}$ and $\tilde{p}^{\max} := \frac{\sum_i p_i^{\max}}{n}$. So when n tends to infinity, (13) is equivalent to

$$\frac{ab}{4}\left(\tilde{p}^{\max} - \tilde{p}^{\min}\right)^2 \geq a^2 \text{Var}(p^{\min}) + b^2 \text{Var}(p^{\max}),$$

where Var denotes the variance, and where we used the independence assumption of $p^{\max}$ and $p^{\min}$ to develop the right-hand term.

Since the variance of a real variable with support length y is always smaller than $y^2/4$, and using (10)

$$\text{Var}(p^{\min}) \quad \leq \quad \frac{(p^* - \max p_i^{\min})^2}{4} \leq \frac{(p^* - \tilde{p}^{\min})^2}{4},$$

where the last inequality comes from $\tilde{p}^{\min} \leq \max p^{\min} \leq p^*$. Likewise, $\text{Var}(p^{\max}) \leq (\tilde{p}^{\max} - p^*)^2/4$. Therefore by replacing the optimal shadow price p^* by $(a\tilde{p}^{\min} + b\tilde{p}^{\max})/(a+b)$ and applying the inequality $(a+b)^2 \geq 2ab$, we get

$$a^2 \text{Var}(p^{\min}) + b^2 \text{Var}(p^{\max}) \leq \frac{ab}{4}(\tilde{p}^{\max} - \tilde{p}^{\min})^2,$$

Therefore Relation (13) is always satisfied and symmetric schemes always outperform profit-maximizing pricing schemes. ■

V. Conclusions and future work

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 system 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. 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 model also aimed to reduce churn. By comparing the social welfare level at the outcome 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.

There are different ways to extend the results we have obtained. First of all, in reality the perceived utility of a user should not only depend on the amount of stored data and the associated availability, but also on the rapidity to access those data. Therefore the available bandwidth of a storage space offerer should be taken into account in addition to the amount of space proposed. Another interesting direction would be to consider demand and supply functions that are not affine, but can have any form, or eventually to carry out experiments to estimate the form of those functions. Finally, since a more complete and realistic model may not be solvable analytically, a simulation testbed could be built in order to study the behaviour of a peer-to-peer storage system in a more complex setting and eventually exhibit other phenomena that are not captured by our model.

References

[1] D. Fudenberg and J. Tirole, *Game Theory*. MIT Press, Cambridge, Massachusetts, 1991.

[2] E. Adar and B. Huberman, "Free riding on Gnutella," *First Monday*, vol. 5, no. 10, Oct 2000.

[3] P. Antoniadis, C. Courcoubetis, and R. Mason, "Comparing economic incentives in peer-to-peer networks," *Computer Networks*, vol. 46, no. 1, pp. 133–146, 2004.

[4] C. Courcoubetis and R. Weber, "Incentives for large peer-to-peer systems," *IEEE JSAC*, vol. 24, no. 5, pp. 1034–1050, May 2006.

[5] P. Golle, K. Leyton-Brown, I. Mironov, and M. Lillibridge, "Incentives for sharing in peer-to-peer networks," in *Proc. of 3rd ACM conference on Electronic Commerce (EC'01)*, Tampa, Florida, USA, Oct 2001, pp. 264–267.

[6] K. Lai, M. Feldman, I. Stoica, and J. Chuang, "Incentives for cooperation in peer-to-peer networks," in *Proc. of 1st Workshop on Economics of Peer-to-Peer Systems (P2PECON'03)*, Berkeley, CA, USA, Jun 2003.

[7] C. Batten, K. Barr, A. Saraf, and S. Treptin, "pStore: A secure peer-to-peer backup system," MIT Laboratory for Computer Science, Tech. Rep. MIT-LCS-TM-632, Dec 2001.

[8] P. Druschel and A. Rowstron, "PAST: A large-scale, persistent peer-to-peer storage utility," in *HotOS VIII*, Schloss Elmau, Germany, May 2001, pp. 75–80.

[9] M. Lillibridge, S. Elnikety, A. Birrell, M. Burrows, and M. Isard, "A cooperative internet backup scheme," in *Proc. of 1st Workshop on Economics of Peer-to-Peer Systems (P2PECON'03)*, Berkeley, CA, USA, Jun 2003.

[10] 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)*, Bolton Landing, NY, Oct 2003.

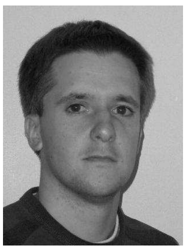
[11] B. Stefansson, A. Thodis, A. Ghodsi, and S. Haridi, "MyriadStore," Swedish Institute of Computer Science, Tech. Rep. T2006:09, May 2006.

[12] D. Besanko and R. R. Braeutigam, *Microeconomics*. Wiley, 2005.

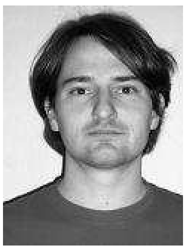
[13] L. Toka and P. Maillé, "Managing a peer-to-peer backup system: Does imposed fairness socially outperform a revenue-driven monopoly?" in *Proc. of 4th International Workshop on Grid Economics and Business Models (Gecon 2007)*, Aug 2007.

[14] R. Rodrigues and B. Liskov, "High availability in DHTs: Erasure coding vs.replication," in *Proc. of 4th International Workshop on Peer-to-Peer Systems (IPTPS)*, Ithaca, USA, Feb 2005.

[15] A. Duminuco, E. Biersack, and T. En-Najjary, "Proactive replication in distributed storage systems using machine availability estimation," in *Proc. of 3rd International Conference on emerging Networking EXperiments and Technologies (CoNEXT)*, New York, USA, Dec 2007.



Patrick Maillé graduated from Ecole polytechnique and TELECOM ParisTech, France, in 2000 and 2002, respectively. He has been an assistant professor at the Network, Security, Multimedia department of TELECOM Bretagne since 2002, where he obtained his PhD in applied mathematics in 2005. His research focuses on game theory applied to networks: resource pricing, routing games, consequences of user selfishness on network performance.



László Toka graduated in 2007 at Budapest University of Technology and Economics and received his MSc degree in Telecommunications. He also obtained the engineer diploma of Telecom Bretagne and Eurecom in France. After graduating he enrolled as a PhD candidate at Telecom Paris and at Budapest University of Technology and Economics. His research interests are on economic modeling of distributed systems, with special focus on incentives and mechanism design.