Computer Communications 35 (2012) 1516–1526

Contents lists available at SciVerse ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Off-line incentive mechanism for long-term P2P backup storage

Marco Gramaglia [a,b,*], Manuel Urueña [b], Isaias Martinez-Yelmo [c]

[a] Institute IMDEA Networks, Av. Mar Mediterraneo, 22, E-28912 Leganés, Madrid, Spain
[b] Universidad Carlos III de Madrid, Av. Universidad, 30, E-28911 Leganés, Madrid, Spain
[c] Universidad de Alcalá, Escuela Politécnica Superior, Campus Universitario, N-II Km 33, 6. E-28871 Alcalá de Henares, Madrid, Spain

ARTICLE INFO

Article history:
Received 25 May 2011
Received in revised form 2 March 2012
Accepted 19 April 2012
Available online 27 April 2012

Keywords:
Peer-to-peer (P2P)
Long-term storage
Incentive mechanism
P2P backup
Digital cheque

ABSTRACT

This paper presents a micro-payment-based incentive mechanism for long-term peer-to-peer storage systems. The main novelty of the proposed incentive mechanism is to allow users to be off-line for extended periods of time without updating or renewing their information by themselves. This feature is enabled through a digital cheque, issued by the user, which is later employed by the peers to get a gratification for storing the user's information when the user is off-line. The proposed P2P backup system also includes a secure and lightweight data verification mechanism. Moreover, the proposed incentive also contributes to improve the availability of the stored information and the scalability of the whole system. The paper details the verification and cheque-based incentive mechanisms in the context of a P2P backup service and analyzes its scalability and security properties. The system is furthermore validated by means of simulation, proving the effectiveness of the proposed incentive.

1. Introduction

The increasing number of consumer devices that can generate all kinds of digital media (e.g. audio, video, photos) has worsened the old problem of safely storing all these space-consuming data. To relieve users from the laborious and eventually expensive task of maintaining their own dedicated storage hardware, in the past few years many on-line storage services made their appearance on the market, ranging from the most basic ones like *Dropbox*[1] to more complex and professional-oriented ones like *Amazon S3*.[2] The growing success of new networking paradigms such as peer-to-peer or, more recently, cloud computing, is offering more tools to tackle this storage problem.

To use a peer-to-peer (P2P) paradigm radically changes the nature of the solutions to this problem, offering some advantages (i.e. utilization of unexploited space in users' hard disks, tunable level of reliability, low cost, etc.), but posing other challenges regarding security, privacy and a fair use of the network. Some P2P-based solutions are already present in the market: LaCie's Wuala[3] or Fiabee[4] exploit this paradigm to offer on-line storage services at a reduced price.

One of the storage services that may benefit most from a distributed P2P architecture is backup, because data is replicated and stored in the hard disks of many different users, distributed worldwide. Thus, it can be hardly affected by a single failure or even a set of failures that may otherwise wipe out a local backup or even a whole data center. Of course, any existing P2P distributed file system could be employed as the basis of a P2P backup service. However, there are two specific characteristics that play a major role in P2P backup: the presence of a local copy, and looser access-time constraints. In a distributed file system, usually, the information is just saved in remote hard disks in order to offload the local one and to better balance their utilization. In a network-based backup solution this assumption is no longer true: the user always has a local copy of the data in order to continue working and updating it. The network backup will only be used in case that some failure happens to the local one. The second key difference of a P2P backup system are the access-time requirements. P2P-based file systems impose hard time constraints (in order to guarantee the performance of input/output operations), while in a P2P backup system these timing constraints are much less strict. A user could tolerate some extra time as long as the backup is completely restored in a reasonable period of time.

In this paper we present an incentive mechanism based on micro-payments and digital cheques for long-term P2P storage systems, such as a P2P backup service. In our proposal, a user pays other peers to store its backup data, whereas charges other (possibly different) peers for using their local hard disk. This kind of monetary incentive approach has already been proposed for different P2P applications [1–6] and, even if it is a hidden market,

* Corresponding author at: Institute IMDEA Networks, Av. Mar Mediterraneo, 22, E-28912 Leganés, Madrid, Spain. Tel.: +34 91 4816210; fax: +34 91 4816965.
E-mail addresses: marco.gramaglia@imdea.org (M. Gramaglia), muruenya@it.uc3m.es (M. Urueña), isaias.martinezy@uah.es (I. Martinez-Yelmo).

[1] http://www.dropbox.com.
[2] http://aws.amazon.com/s3/.
[3] http://www.wuala.com.
[4] http://www.fiabee.com.

0140-3664/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.comcom.2012.04.017

for a P2P backup service [7]. We extend this micro-payment incentive framework by means of digital cheques to motivate peers to keep storing backup data even when the owner (user) is off-line for an extended period of time. Albeit the presented schemes also mentioned the problem of long-term availability, they were just focused on employing redundancy techniques to minimize the impact of a lost chunk due to a failure of an, otherwise well-behaved, peer. To the best of our knowledge, this is the first proposal that tackles the problem of long-term availability in a P2P backup system with selfish peers. These selfish peers can deliberately erase a chunk when the owner (user) goes off-line to free their local resources. Our proposal introduces the possibility to keep on charging users even if they are off-line (e.g. due to a hardware problem), providing an incentive to not erase their backup when it is even more necessary. This last point is crucial for any P2P backup service.

The paper is structured as follows: after studying the related works in Section 2, Section 3 introduces an overview of the proposed P2P backup system and defines the incentive mechanism that governs it. Later, Section 4 presents a detailed discussion about the design of our proposal, emphasizing on the behavior of the system when a user is on-line or off-line. The different incentive mechanism is evaluated by means of simulation in Section 5. In Section 6, we analyze the possible threats that could affect our proposal and some security mechanisms to prevent them. Finally, we summarize the main conclusions of this paper in Section 7.

2. Related work

The idea of taking advantage of unused space in remote hard disks that are part of a P2P network was first studied at the beginning of the past decade. One of the earliest proposals was OceanStore by Kubiatowicz et al. [8], which provided solutions to many of the issues caused by relying on an untrusted infrastructure for data storage. Also Farsite [9] by Adya et al. was focused on the problems concerning the fault-tolerance and reliability of the stored data. However, these initial proposals considered a distributed file system in a heterogeneous, yet cooperative and trusted scenario. The first proposal of a backup-oriented solution was done by Batten et al. in [10]. It features file encryption, version control, and provides reliability in case of multiple nodes failures. In this first stage, the research was more focused on how to achieve scalability, reliability and fault-tolerance in such a system [11,12].

More recently, researchers have started considering additional aspects of P2P storage and backup systems. In such a distributed environments, peers are service users and providers at the same time: they want to store their data in the system, but to do so, they should also share part of their unused capacity with other peers. This peculiarity raises obvious fairness issues, especially when peers can behave selfishly. The so-called *free-riding* problem of P2P systems is well-known since Adar and Huberman showed in [13] that the 70% of peers were not sharing any files in the Gnutella network. Since then, a plethora of works [14–17] have been proposed to try and mitigate this problem in P2P networks. They usually exploit some physical constraint of the system or they are based on fundamental principles taken from economics or game theory, modeling the problem as the "Tragedy of the commons" dilemma. In fact, most of the P2P systems currently deployed implement some kind of incentive mechanism like BitTorrent's "Tit-For-Tat" [18].

However, the *free-riding* problem in a P2P backup (or file storage) system is quite different from the one that can be found in a file-sharing system. A first difference is the potential "audience" of the resources being stored by the peers. For example, let us consider one file being uploaded with some file-sharing software. This file is "public" and could be potentially downloaded by any user of the network. In fact, in most P2P file-sharing systems, peers store a file because they are actually interested in it. On the other hand, in long-term storage services and especially in P2P backup services, each piece of information is usually encrypted and, thus, belongs to a single user (or to a restricted group of users). This consideration makes the problem even more complicated since there is not any implicit incentive for peers to store (useless) file chunks from other peers. Thus, how to reward a peer that is sharing an amount of disk space much greater than the one it is asking for its data? Or the opposite case, how to incentive a peer to be more generous with the system, if it is using more space than the one is sharing? Furthermore, how can such storage quotas be enforced in a fully distributed system? A possible solution was proposed by Cox et al. in [19,20], using a framework for limiting the amount of data that a user is allowed to store into the network. This leads to a symmetrical behavior that, although keeps the system in a stable state, limits its flexibility. More recently, other studies have tried to tackle this problem without forcing the users to share a fixed amount of disk space. In Seuken et al. [7] have proposed to solve the problem by introducing a virtual market where a central system computes the exact amount of resources that users have to share (including uplink and downlink bandwidth) by following a trade mechanism that, under certain conditions, leads the system towards an equilibrium. However, this proposal still implies that peers follow a fair share, and it does not define any verification or penalty mechanism for selfish peers. Using monetary incentives in P2P systems was also studied in [1–6], but these works are mostly focused on the security aspects about coining digital currency.

Furthermore, the incentive mechanisms for P2P file systems cannot be directly applied to P2P backup systems due to their specific characteristics. Whereas in a standard file system the saved data is frequently read and written, this is usually not true in a backup system. Instead, the data are commonly stored only once. Read operations are not frequent at all and, hopefully, null unless the user's data is lost from the local storage. Moreover, whereas in a distributed file-system the most important feature is access performance, the long-term durability of stored data is paramount in a backup service. The terms data durability and availability are often used in the literature regarding P2P storage systems [21]. The durability is the property that guarantees the fact that data stored in a peer will last for a time ideally infinite. This property is valid even when the peer is off-line. The availability property is more restrictive: it is valid when the data stored in a peer is correctly saved and available for downloading. Therefore, if we consider that the users of a P2P network behave selfishly, we cannot just rely on replication mechanisms and consider that the data stored in a working peer is "safe" without a secure control mechanism that continuously verifies this. If both peers are on-line, the problem could easily be solved by performing periodical checks of the data availability as proposed by Toka et al. in [22,23]. Michiardi and Toka in [24] presented an analytical model based on game theory for the detection of selfish peer and a similar solution was provided previously by Pamies-Juarez et al. in [25]. They propose a proactive monitoring system that checks the availability of the peers and assigns different quotas of the system to the users according to their obtained score. This architecture is further researched by the same authors in [26] where they found a relation between the system health (in terms of data availability) and the peer selection algorithm. Selecting the best peers increases the efficiency of the P2P system and provides better results when trying to retrieve the stored chunks. The solution proposed by Oualha and Roudier in [27] also identifies the data reliability as one of the biggest issues in the field of P2P storage. They propose a distributed system to find out malicious peers, either passive (which do not