Computer Communications 52 (2014) 71–81

Contents lists available at ScienceDirect

Computer Communications

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

Performance evaluation of a peer-to-peer backup system using buffering at the edge

Anne-Marie Kermarrec [b], Erwan Le Merrer [a,*], Nicolas Le Scouarnec [b], Romaric Ludinard [a], Patrick Maillé [c], Gilles Straub [a], Alexandre Van Kempen [a]

[a] Technicolor, ZAC des Champs Blancs, 35576 Cesson Sévigné, France
[b] INRIA-Rennes, Campus Universitaire de Beaulieu, 35042 Rennes, France
[c] Telecom Bretagne, rue de la Châtaigneraie, 35576 Cesson Sévigné, France

ARTICLE INFO

Article history:
Received 9 October 2013
Received in revised form 6 June 2014
Accepted 10 June 2014
Available online 27 June 2014

Keywords:
Storage
Backup
Availability
Peer-to-peer

ABSTRACT

The availability of end devices of peer-to-peer storage and backup systems has been shown to be critical for usability and for system reliability in practice. This has led to the adoption of hybrid architectures composed of both peers and servers. Such architectures mask the instability of peers thus approaching the performances of client–server systems while providing scalability at a low cost. In this paper, we advocate the replacement of such servers by a cloud of residential gateways, as they are already present in users' homes, thus pushing the required stable components at the edge of the network. In our gateway-assisted system, gateways act as buffers between peers, compensating for their intrinsic instability. We model such a system, for quick dimensioning and estimation of gains. We then evaluate our proposal using statistical distributions based on real world traces, as well as a trace of residential gateways for availability (that we have collected and now make available). Results show that the time required to backup data in the network is substantially improved, as it drops from days to a few hours. As gateways are becoming increasingly powerful in order to enable new services, we expect such a proposal to be leveraged on a short term basis.

1. Introduction

While digital data clearly dominates, backup is of the utmost importance. More specifically, online (*i.e.* off-site) backup is often preferred over simple backup on external devices as it ensures data persistence regardless of the damage cause (*e.g.* failures, burglars or even fires). To enable their deployment, online backup systems should run in the background and provide reasonable performances so that archives can be stored safely in reasonable time. While cloud backup systems are increasingly adopted by users (*e.g.,* justcloud, SugarSync, Egnyte HybridCloud, Amazon S3 or DropBox), their peer-to-peer alternatives, potentially offering *virtually unlimited storage* for backup [1,2], are still not appealing enough performance-wise, as *e.g.* retrieval times for saved data can be an order of magnitude higher that the time required for direct download [3]. A particularly illustrative example is the Wuala case: the Wuala company gained fame by proposing a peer-assisted (advertised as fully peer-to-peer) and practical storage service; nevertheless, this technical choice was abandoned to move to a centralized architecture [4], probably for cost/performance matters. Beside this initial example and academic systems, we are not aware of a peer-to-peer storage system deployed at large scale for common needs.

Indeed, peer-to-peer backup systems are limited by the low to medium availabilities of participating peers and by the slow up-links of peers' network connections. This limits the amount of data that peers can transfer and places peer-to-peer systems way behind datacenter-based systems [5]. Not only this may impact the reliability of the stored content but also this does not provide a convenient system for users. We focus on this performance problem and investigate a new way of performing efficient backup on commodity hardware in a fully peer-to-peer way. Other specific issues with peer-to-peer solutions include security or QoS [6], but are out of the scope of this article.

In this article, we propose a new architecture for peer-to-peer backup, where residential gateways are turned into a *stable buffering layer* between the peers and the Internet. The residential gateways are ideal to act as stable buffers: they lay at the edge of the network between the home network and the Internet, and are highly available since they remain powered-on most of the time [7]. Our idea is to temporarily store data on gateways to

* Corresponding author.
E-mail address: erwan.lemerrer@technicolor.com (E. Le Merrer).

http://dx.doi.org/10.1016/j.comcom.2014.06.002
0140-3664/

compensate for peers transient availability. In this article, we advocate the use of gateways as buffers and not storage; this choice is motivated by the increasing number of devices embedding storage, within the home and attached to a gateway. Dimensioning the storage of the gateway accordingly would be costly and would break the peer-to-peer paradigm by creating a central point in charge of hosting resources of attached devices durably: the contributed resources would no longer scale with the number of clients. In our buffer model, each device is required to provide a portion of its available space [1,2], to participate to the global backup system. Such a system enhances the backup system's performance along two lines:

- The network connection can be used more efficiently: if devices upload data continuously while they are up, the available bandwidth can be exploited typically 21 h/day instead of only 6–12 h/day on average, based on actual measured availabilities. This leads to significant enhancements. For example, we observe that the time to backup a 1 GB archive is reduced from few weeks in a pure peer-to-peer system to around one day in our system.
- Additionally, the gateways, offering a high availability (86% on average, according to our measurements), can act as rendezvous to allow any two peers to communicate efficiently, even if they are not up at the same time. In our application, this enhancement mainly has an impact on the time to restore.

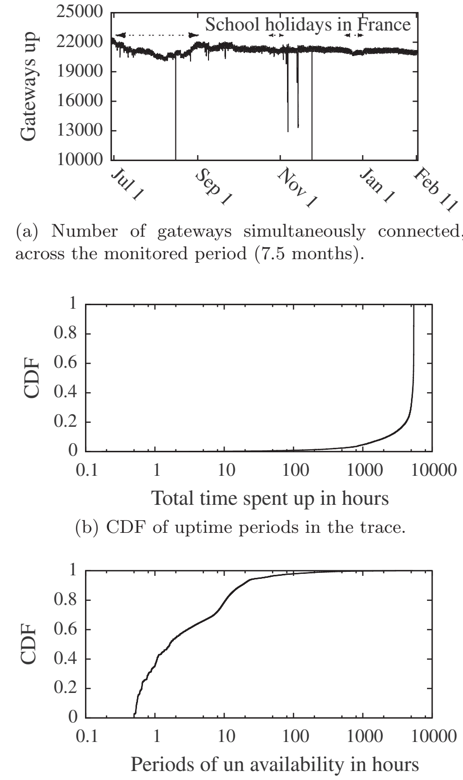
Our proposal differs from existing approaches [4,5,8–13] by taking into account the low-level structure of the network. Indeed, most peer-to-peer applications ignore the presence of a gateway in between each peer and the Internet. We believe that leveraging the gateway storage space may render peer-to-peer systems viable alternatives for backup. This should provide a reasonable solution even when peers experience a low availability as long as they connect frequently enough to the system. Using those gateways as buffers between peers participating in a backup or restore operation, enables to implement a stable rendezvous point between transient peers.

The remainder of this article is structured as follows. In Section 2, we briefly review some pieces of work that motivated our proposal. In Section 3, we detail our architecture and sketch the storage system. We then propose in Section 4 a model for this system, in order to estimate its performance. Section 5 introduces a framework for comparison of our proposal to that of competitors, and presents our evaluation study. We discuss respectively some specific points and related work in Sections 6 and 7. Finally, we conclude the article in Section 8.

2. Background

Peer-to-peer storage systems initially relied on the set of all participating peers, typically constituted of users' desktop PCs, without any further infrastructure [8,9]. However, it has been acknowledged since then [14,15] that those pure peer-to-peer architectures may fail to deliver reliable storage by exploiting the resources of peers, mainly due to the low availability of peers and the slow up-link of their network connections. One straightforward solution is to exclude peers with a low availability or a slow network connection to access the service [16]; this nevertheless excludes many participants and significant amounts of exploitable resources [1,2].

Hybrid architectures, where both servers and peers coexist, have been proposed in various contexts, in order to move towards practical system deployment while still leveraging users' resources [17]. The problem of sharing files while mitigating the load of central servers is addressed in [18]. This article proposes a



(a) Number of gateways simultaneously connected, across the monitored period (7.5 months).

(b) CDF of uptime periods in the trace.

(c) CDF of dowtime periods in the trace.

Fig. 1. Availability of residential gateways mesured on a French ISP. The dataset has been acquired sending pings to a random sample of gateway IPs.

BitTorrent like server-assisted architecture where central servers act as permanently available seeders. Lastly, a server assisted peer-to-peer backup system is described in [5]. In this system, which can be referred to as *CDN-assisted*, the CDN enables to reduce the time needed to backup data, while the use of peers guarantees that the burden of storage and communication on the data center remains low. In this last approach, a peer uploads data to a set of other peers if they are available, and falls back on the datacenter otherwise, thus using the datacenter as a stable storage provider.

Finally, another aspect of interest is the network setting of home networks. Residential gateways connect home local area networks (LAN) to the Internet. They act as routers between their WAN interface (Cable or DSL) and their LAN interfaces (Ethernet and WiFi). They started to be deployed in homes to share Internet access among different devices and got additional functions as services (VoIP, IPTV) were delivered over the Internet. It is now fairly common to have home gateways embedding a hard drive, acting as Network Attached Storage to provide storage services to other home devices and offering some other ones to the outside world [7,19–21].

3. A gateway assisted system

3.1. Stability of residential gateways

As residential gateways provide not only Internet connectivity, but also often VoIP, IPTV and other services to the home, the intuition tells us that they remain permanently powered on. To confirm this assumption, we extracted a trace of residential gateways of the French ISP Free, using active measurements.[1] We

[1] This trace and additional information can be found at the following URL http://www.thlab.net/lemerrere/trace_gateways/.