Journal of Network and Computer Applications 34 (2011) 562–574

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

Cooperative caching versus proactive replication for location dependent request patterns

Niels Sluijs *, Frédéric Iterbeke, Tim Wauters, Filip De Turck, Bart Dhoedt, Piet Demeester

Department of Information Technology (INTEC), Ghent University—IBBT, Gaston Crommenlaan 8, Bus 201, 9050 Ghent, Belgium

ARTICLE INFO

Article history:
Received 24 June 2010
Received in revised form
28 October 2010
Accepted 5 December 2010
Available online 19 December 2010

Keywords:
Caching strategy
Distributed Hash Table
Performance analysis
Personal content
Scalability

ABSTRACT

Today's trend to create and share personal content, such as digital photos and digital movies, results in an explosive growth of a user's personal content archive. Managing such an often distributed collection becomes a complex and time consuming task, indicating the need for a personal content management system that provides storage space transparently, is quality-aware, and is available at any time and at any place to end-users. A key feature of such a Personal Content Storage Service (PCSS) is the ability to search worldwide through the dataset of personal files. Due to the extremely large size of the dataset of personal content, a centralized solution is no longer feasible and an interesting approach for an efficient distributed PCSS implementation is to use a structured Peer-to-Peer network, and more in particular a Distributed Hash Table (DHT), providing a logarithmic lookup performance in the number of network nodes. In order to further increase the lookup performance, a caching layer is typically used between the application layer and the DHT. These caching strategies are location neutral, and usually do not exploit location dependence of request patterns. In this article we present our cooperative caching framework and introduce the cooperative Request Times Distance (RTD) caching algorithm. Since, lookup patterns in a PCSS typically have a power law popularity distribution and exhibit location dependent requests patterns, the proposed caching solution takes into account popularity and distance metrics. To enable cooperation between caches we introduce an update protocol, which only occasionally introduces one hop delay for a lookup operation. We present a systematic analysis of the caching framework and compare the cooperative caching algorithm to the state-of-the-art Beehive replication strategy. The cooperative RTD caching solution shows that when request patterns are more localized, the increase in lookup performance through cooperation is significantly better than Beehive.

1. Introduction

The interaction with digital information plays an important role in our daily life. Different websites, such as YouTube[1] and Flickr,[2] offer platforms to store and share personal content (e.g. text documents, music files, digital photos and personal movies). Due to the explosive growth of the user's personal content collections, managing those archives becomes a complex and time consuming task. Nevertheless, end-users expect that they can access and share their personal content from any device, anywhere and at any time. Current systems that offer storage space for personal content fail to achieve this in a scalable and quality-aware way, constraints (e.g. on files sizes and formats) need to be set on the content in order to cope with the workload. To be able to deal with the future workload, a centralized approach is no longer feasible. A Personal Content Storage Service (PCSS) is a networked solution that offers storage space to end-users in a transparent manner, which can be accessed from different types of devices independent of place and time. Fig. 1 presents an architectural view on such a distributed content management platform, where users (i.e. clients) are connected to super nodes in the PCSS overlay network.

The PCSS uses a (hybrid) Peer-to-Peer (P2P) architecture to support all necessary operations and the architecture is split-up in two high-level components: super nodes and clients. The key functions of the super node component (as schematically shown in Fig. 1) concern user and content management (including replica management and indexing), query handling, presence management, security provisioning, monitoring of the underlying P2P network. The client component is responsible for advertising shared content as well as retrieving and uploading personal content items. For end-users the PCSS acts as a virtual hard disk, as if personal content were accessed using their local file system. Additionally, end-users are relieved from cumbersome back-up issues, since the PCSS provides data integrity through replication.

To efficiently lookup personal content references (i.e. through optimal content indexing), this paper presents a novel cooperative caching strategy that is able to react on location dependent request patterns and making use of an underlying Distributed Hash

* Corresponding author.
E-mail address: Niels.Sluijs@intec.ugent.be (N. Sluijs).
[1] http://www.youtube.com/.
[2] http://www.flickr.com/.

1084-8045/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jnca.2010.12.007