T-CloudDisk: *A Tunable Cloud Storage Service for Flexible Batched Synchronization*

Zhenhua Li *, He Xiao, Linsong Cheng *, Zhen Lu, Jian Li, Christo Wilson, Yao Liu, Yunhao Liu, and Yafei Dai

Tsinghua University, Northeastern University, Binghamton University, and Peking University

Abstract

Cloud storage services, such as Dropbox, Microsoft SkyDrive and Google Drive, have quickly gained enormous popularity in recent years. They offer users with convenient and reliable approaches to store and share data from anywhere, on any device, at anytime. However, they are still suffering from the "traffic overuse problem" in the presence of *frequent, short* data updates [1][2]. To address this problem, we are implementing a *tunable* cloud storage service (named "T-CloudDisk" or "ThuCloudDisk") for *flexible* batched synchronization. This poster briefly introduces its characteristics, technical approach, and preliminary implementation.

Motivation

Cloud storage services are very popularity in recent years. Below are some typical examples: **Microsoft SkyDrive**, **Dropbox**, **Google Drive**, **Apple iCloud**, and **Box.com**.

However, we observe that cloud storage services are still suffering from the **traffic overuse problem** in the presence of *frequent, short* data updates [1][2], and thus propose a middleware-based batched synchronization solution (named "**UDS**", *i.e.* **U**pdate-batched **D**elayed **S**ync) to address this problem [1].

Although UDS significantly reduces the traffic overuse, acting as a middleware between the user's local filesystem and the concerned cloud storage application (like Dropbox), it requires considerable additional storage space in the user's local disk, and currently it uses fixed *buffer size* and *timer threshold* for batched sync.

The T-CloudDisk (ThuCloudDisk) System Design and Implementaion

Traffic overuse problem. *Saving network traffic* is a critical goal of cloud storage services, given that billions of files are synchronized to the cloud every day. To minimize the network overhead, cloud storage services (*e.g.* Dropbox [3][4]) employ binary diff, data compression, and other mechanisms when delivering updates among users. However, despite these optimizations, we observe that in the presence of *frequent, short* updates to user data, the network traffic generated by cloud storage services often exhibits pathological inefficiencies.

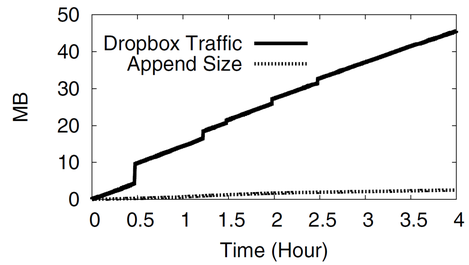

Fig.1: Traffic overuse of Dropbox in the presence of frequent, short updates

Through comprehensive measurements and detailed analysis, we demonstrate that many cloud storage applications generate session maintenance traffic that *far exceeds* the useful update traffic, as shown in **Fig.1**. We refer to this behavior as the "traffic overuse problem" [1][2].

UDS middleware. To address this problem, we propose the UDS middleware. UDS batches updates from clients to significantly reduce the overhead caused by session maintenance traffic, as depicted in **Fig.2**.

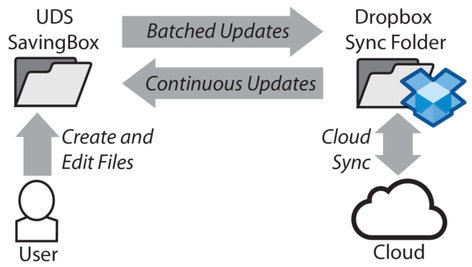

Fig.2: High-level design of the UDS middleware

T-CloudDisk. In order to further overcome the shortcomings of UDS, we are implementing an *independent*, *tunable* cloud storage service (named "T-CloudDisk" or "ThuCloudDisk") for *flexible* batched synchronization:

- *Independent*: T-CloudDisk is a self-contained cloud storage service, rather than a middleware that relies on other services' apps. As depicted in **Fig.3**, it has its own user app and private cloud that maintains the critical meta-data and user liveness information. With such an independent service, we are able to conduct more in-depth, white-box research on cloud storage.
- *Tunable*: To avoid the expensive infrastructure cost and enhance the service scalability, T-CloudDisk outsources all file contents to a public cloud that supports **REST**ful APIs, such as Amazon S3, Aliyun.com OSS, and Openstack Swift (our current implementation).
- *Flexible*: The T-CloudDisk app allows users to customize the buffer size and timer threshold according to their specific requirements, as shown in **Fig.4**. This increases the flexibility of batched sync as it enables users' own tradeoff between traffic usage and user experience.

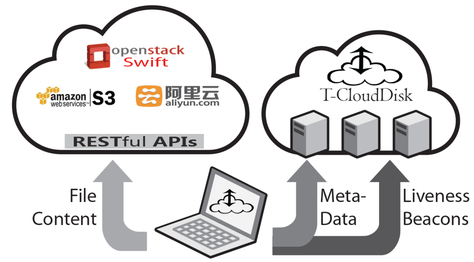

Fig.3: T-CloudDisk system architecture

http://www.thucloud.com

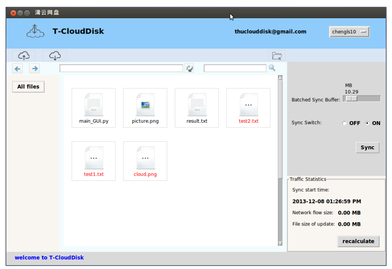

Fig.4: T-CloudDisk client software (Dec. 9, 2013)

Related Work

[1] Zhenhua Li, Christo Wilson, Zhefu Jiang, Yao Liu, Ben Y. Zhao, Cheng Jin, Zhi-Li Zhang, and Yafei Dai. **Efficient Batched Synchronization in Dropbox-like Cloud Storage Services**. The 14th ACM/IFIP/USENIX International Middleware Conference (***Middleware***) Regular Paper, Dec. 9-13, 2013, Beijing, China.

[2] Zhenhua Li, Zhi-Li Zhang, and Yafei Dai. **Coarse-grained Cloud Synchronization Mechanism Design May Lead to Severe Traffic Overuse**. *Journal of Tsinghua Science and Technology* Invited Paper, Vol. 18, No. 3, Jun. 2013.

[3] Idilio Drago, Marco Mellia, Maurizio Munafo, Anna Sperotto, Ramin Sadre, and Aiko Pras. **Inside Dropbox: Understanding Personal Cloud Storage Services**. The 12th ACM Internet Measurement Conference (IMC) Regular Paper, Nov. 14-16, 2012, Boston, MA, US.

[4] Idilio Drago, Enrico Bocchi, Marco Mellia, Herman Slatman, and Aiko Pras. **Benchmarking Personal Cloud Storage**. The 13th ACM Internet Measurement Conference (IMC) Short Paper, Oct. 23-25, 2013, Barcelona, Spain.

Contact

* **Zhenhua Li**, School of Software and TNLIST, Tsinghua University, Beijing, China. lizhenhua1983@tsinghua.edu.cn. Homepage: http://www.greenorbs.org/people/lzh.

* **Linsong Cheng**, School of Software, Tsinghua University, Beijing, China. chengls10@mails.tsinghua.edu.cn.