Proposal on Project 5: Assessment of Cloud Storage Service

He Xiao∗, Cihang Liu∗, Jingchao Feng∗  
∗School of Software, Tsinghua University, Beijing, China

I. Introduction

Cloud storage services, such as Dropbox, Google Drive, and Baidu Yun, etc, provide users with a convenient and reliable way to store and share data from anywhere, on any device, and at any time. The cornerstone of these services is the data synchronization operation which automatically maps the changes in users’ local filesystems to the cloud via a series of network communication. If no designed properly, however, the tremendous cost can potentially cause(financial) pains to both service providers and users. In this project, we will perform a profound evaluation on COTS(Commercial-off-the-shelf) cloud storage services, from different aspects on both MACs and mobile devices. We can compare the QoE(Quality of Experience, including computation&memory cost, sync traffic, sync time and the power consumption in mobile client) of different services by synchronizing the same file. Moreover, after editing the file, the sync performance of different services will show how differencing synchronization works to improve the QoE.

II. Objectives

We propose to measure and compare the QoE of several personal cloud storage services. In this measurement, we will achieve to following three goals:

1. The QoE including sync traffic, sync time CPU usage and memory cost among the sync operation when we synchronizing raw files between the client and cloud servers. We will compare the QoE of different cloud services on different devices.
2. The power consumption of the sync operation on mobile phones on different cloud services.
3. The effect of differencing synchronization algorithm on compressed files between different cloud services.

III. Background

1. General benchmark of personal cloud storage

Personal cloud storage services are data-intensive applications already producing a significant share of Internet traffic. Several solutions offered by different companies attract more and more people. However, little is known about each service capabilities, architecture and most of all performance implications of design choices. Excellent and general works have been done in the area of Benchmarking personal cloud storage[2]. In this article

2. Measure cloud storage with passive measurement

Dropbox is the leading solution in personal cloud storage because dropbox has been focusing on new futures to enhance user experience. Previous study about dropbox present how to characterize a closed-source COTS cloud storage with passive measurement[4].

3. Network traffic is concerned

Of all the QoE features of personal cloud storage services, network traffic are specially concerned by both cloud providers and users because network traffic is money-related. Excellent passive measurement and a ingenious solutions are presented by an earlier study[1].

4. Differencing sync is an easy but efficiency-arguable solution

Speaking of network traffic problems, Differencing sync[5] between clients and servers are recommended when users try to frequently modify the same files and sync to the cloud servers.

However, due to the complexity of file structure and modification by the users. Synchronization strategy can be very challenging and complicated for the cloud storage. Some sync algorithms are cpu-usage-high and slow which can harm user experience than more network traffic when the files synced are not very large[6]. So, how to trade off the balance between computing source and network traffic is hard.

IV. Plan of Action

1. Choose and install several representative cloud storage

We choose dropbox, iCloud and Baidu Yun as our experimental Platform because dropbox is the professional leader of cloud storage , iCloud is one of the thousands of software of Apple Inc. while Baidu Yun is popular and high-speed in China.

1. Generate the proper files and design reasonable modification on them to measure the QoE of cloud storage.

1) New files of different size

2) Big files with tiny modification

3) Compressed files with tiny modification

1. Use Wireshark to analyze the network traffic among the sync operations.
2. Use System Manager Tools to analyze the cpu usage , memory cost and power consumption among the sync operations.

V. APPENDIX

REFERENCES

[1]  Li Z, Jin C, Xu T, et al. Towards network-level efficiency for cloud storage services[C]//Proceedings of the 2014 Conference on Internet Measurement Conference. ACM, 2014: 115-128.

[2]   Drago I, Bocchi E, Mellia M, et al. Benchmarking personal cloud storage[C]//Proceedings of the 2013 conference on Internet measurement conference. ACM, 2013: 205-212.

[3]   M. Islam, “Spectrum Survey in Singapore: Occupancy Measurements and Analyses,” Proc of IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN, Chicago, IL, USA, 2008

[4]   Drago I, Mellia M, M Munafo M, et al. Inside dropbox: understanding personal cloud storage services[C]//Proceedings of the 2012 ACM conference on Internet measurement conference. ACM, 2012: 481-494.

[5]   Tridgell A, Mackerras P. The rsync algorithm[J]. 1996.

[6]  Zhang H, Xiao N, Liu F, et al. S-rsync: An efficient differencing algorithm with locally chunk digests generating for file synchronization   
services[C]//Proceedings of the International Conference on Human-centric Computing 2011 and Embedded and Multimedia Computing 2011. Springer Netherlands, 2011: 173-185.