Performance Evaluation of Online Backup Cloud Storage

Xiao Zhang *
School of Computer Science
Northwestern Polytechnical
University 127 Youyi xi Road,
Xi'an Shaanxi China
zhangxiao@nwpu.edu.cn

Wan Guo
School of Computer Science
Northwestern Polytechnical
University 127 Youyi xi Road,
Xi'an Shaanxi China
xilouyouki@163.com

Xiao Qin
Department of Computer
Science and Software
Engineering Auburn
University,Auburn AL,USA
xqin@auburn.edu

ABSTRACT
Cloud storage provide storage service to user through internet. There are many different access interface for different applications. Online backup is the most developed application. Most comparison of online backup service focus on functional characters. There is not too much consider about performance evaluation. In this paper, we present a method to evaluate performance of different online backup services with client. We evaluate the performance from view of end-user, compare the upload and download speeds of different service. Because these clients do not prompt when the download/upload process start and finish, we capture the packets transferred between user and cloud storage provider. This method can help user find the right provider under their situation.

Categories and Subject Descriptors
C.4 [**Performance of Systems**]: General—*complexity measures, performance measures*; C.2.4 [**Computer-Communication Networks**]: Distributed Systems—*distributed applications*

General Terms
Measurement, Performance

Keywords
Cloud storage, Performance evaluation, network monitor

1. INTRODUCTION
Cloud storage become an important aspect in IT industry. Cloud storage is an application of cloud computing. It's the most developed part in cloud application. It depends on the cluster application, grid technology and distributed file system, and provides storage service to user through internet. In most conditions, Cloud storage can provide high reliability and security storage service at competitive price.

*Corresponding author. Tel.: 86-159-0297-8022; E-mail address: zhangxiao@nwpu.edu.cn

There are more than 50 cloud storage service providers, such as Amazon, Google, Microsoft and Dropbox. The services can vary greatly when it comes to features, accessibility and support options. Most past researches aimed at comparison on functional features. There are many software working together through disk array and clusters in cloud storage system. Compared with traditional storage devices, cloud storage is not only a certain hardware, it includes network, storage, server and application software. It provides different public access interfaces, including File system interface, database access interface and API used for applications running in the cloud.

The most developed cloud storage application is online backup or file-syncing. Disaster may happen at any time: fire, flood, tornado, hard driver failure. These disasters can destroy all local-stored data. User can use remote backup to protect their data from disasters. The other choice is to store copies of files in the cloud storage. Online backup is an Internet-based system that is set to automatically back up all selected files. These files are stored online, and can be accessed anywhere. Which is especially useful in case of local computer or server gets lost or damaged. The benefit of using online storage services does not limited in protecting data. Cloud storage services make it easy to share files from different machines and mobiles.

There are many comparison between different online backup services. But most of them aimed at function and price. Performance is an important feature for storage system. But it is difficult to evaluation because of the complexity of system and network conditions. In this paper, we present an evaluation method for cloud storage from user view. We tested three of different cloud storage service and analyzed the results.

The rest of this paper is organized as following: Section 2 introduce the previous studies on cloud storage performance measurement. Section 3 introduce the evaluation environments. Section 4 present the method to test performance of storage service. Section 5 shows the test results of three different cloud storage service. At last, we point out the future work to improve the method.

2. RELATED WORKS
The cloud storage market is now crowded with different service providers. There are plenty of services to choose. Many articles introduced how to choose new cloud storage

services. Vince compared the most well known cloud storage providers, including Dropbox, SkyDrive, SugarSync and Box.net [4]. He compared the space and functionality of different providers. And there is a comparison of 51 different online backup services [11]. Most of them compare the free and paid space, support platform, convenience and security.

Too many think of cloud storage as just another or the next type of storage. As usual with this view, it is associated with a view that the "next" storage type is bigger, faster and cheaper. Because each generation of storage is always bigger, faster and cheaper. As such, proponents of this view generally believe that access via traditional approaches, like WebDAV, NFS, CIFS and others, is a critical capability [3]. Performance is an important factor for storage device. Cloud storage provide abundant interface for different usage. However, the novel interfaces (usually neither SQL nor POSIX), elasticity, and new use cases of cloud serving systems motivate a new benchmark.

Most cloud storage performance evaluation researches are focus on the database operation in the cloud. CloudCMP compare the public cloud provider[15]. They also compared the performance of persistent storage in cloud. They selected three types of storage services: table, blob and queue. They use operation response time, time to consistency and cost per operation to compare quality of cloud storage. YCSB provide the performance compare on database view [13]. Some scholars do research on web based cloud storage [12, 17]. Cumulus implement a system for efficiently implementing filesystem backups over the Internet[16]. this can be used to test performance of online bakcup service support S3 access interface.

Most online backup service implement clients on multiple platforms. The client maps the space in the cloud as a local disk devices or a folder. Many test tools were developed to test local disk performance, such as IOmeter [14], IOZone[7] and Hdtune[8]. But these tools can not be used on cloud storage devices because it is a virtual device which has longer delay. Some clients save the data on local disks and upload data background. In these condition, the test tool will only get the time of saving file on local disks or folders. For the online backup user, they want to know when the data will stored in the cloud indeed. Customers need a method to test which cloud storage platform is more suitable, but there is rare cloud storage testing method and tools.

To improve the performance of online backup, a network appliance or server which resides at the customer premises and translates data through internet. These appliance or server are called as cloud storage gateway. Unlike the cloud storage services which they complement, cloud storage gateways use standard network protocols which provide a seamless integration with existing applications. Cloud storage gateways can also serve as intermediaries to multiple cloud storage providers. Some cloud storage gateways also include additional storage features such as backup and recovery, caching, compression, encryption, de-duplication and provisioning. Some vendors announce that they can accelerate 10-30 times of transmission. Amazon provide storage gateway as a virtual machine (VM), user download it and choose gateway work method, gateway-cache or gateway-stored. The AWS storage gateway only uploads data that has changed, minimizing the amount of data sent over the internet[2]. CloudArray is available as a virtual, physical or in-cloud appliance that takes just minutes to configure and integrates with more than 20 public and private cloud options, as well as local or remote storage devices [5].

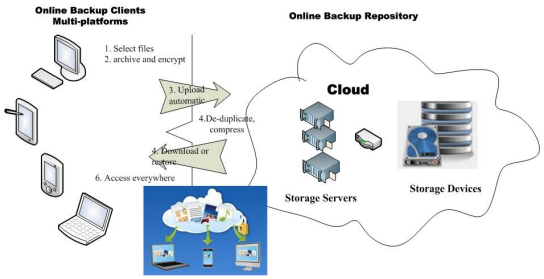

Figure 1: Process of online backup and restore

3. EVALUTION ENVIRONMENT

Online backup service is an easy and reliable way to store valuable information online. The process of backup and restore with online backup services is showed in Figure 1. Most online backup service provider provide clients on multiple platforms. User select which files or folders should backup to cloud storage. Then clients will do some archive or encrypt job before upload. There are many things the client can do to reduce the transmit file size. It can compress files to decrease the size of transmit, which needs more local compute resource. Another method is check same blocks in files and transmit only once. The last is it can compute the fingerprint of each file and check if there is same file stored in the cloud. If the file was stored by other user, it just send control information to create a symbol link to the file. The client upload files to online backup cloud storage system through internet. The cloud storage system will check same blocks among the whole system. The data should be partitioned in distribute file system with single name-space. And the clients on other machine or platform can view and download files anywhere.

Online backup client is the most important part in the system. It provides a mapping mechanism, the data stored in remote side is mapped to a local disk devices. It makes remote space acts as a regular computer drive, but is hosted remotely on a file server. Because the mapping protocol is not a standard protocol, user have to install many different clients to use different cloud storage service. The clients communicate with cloud storage service with different protocols. These can be simply divided into three categories: transfer protocol based on HTTP, File transfer protocols such as FTP and private protocols. Amazon Simple Storage Service provides a simple web services interface that can be used to store and retrieve any amount of data, at any time, from anywhere on the web[1]. It gives any developer access to the same highly scalable, reliable, secure, fast, inexpensive infrastructure that Amazon uses to run its own global network of web sites. Files in Amazon S3 provide a are organized in buckets and objects. Using protocols based on HTTP or FTP make it easy to implement on different platform. It provides cloud storage APIs such as SOAP

or REST. But there are more convenience to implement clients based on private protocols. By using a customized client, cloud storage service provider can present a similar GUI for user to manage their files just like local file system. Clients also can compress files before upload to cloud storage, which get better upload performance. De-duplicate technology also can be implemented in clients to decrease transfer bytes. The clients can upload files or directories background automatically.

Cloud storage service provider present storage service at an incredible low price. There are many technologies used to decrease the cost per GB. Thin provisioning is the act of using visualization technology to give the appearance of having more physical resources than is actually available. They will not allocate the space until the user store files in it. Data de-duplication is a specialized data compression technique for eliminating duplicate copies of repeating data. Related and somewhat synonymous terms are intelligent (data) compression and single-instance (data) storage. The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. Different with de-duplication applied in client side, de-duplicate on server side will search same blocks in whole system instead of one user. Some people save music and other public documents in their own space. In these condition, clients can send the digital fingerprint of a file to check if the file was saved before. If the file was uploaded by other user before, it just adds a link to indicate that the file is shared by multiple users.

We use a fixed PC terminal install several clients one by one, and test the download/upload speed of different file types. After upload files to the cloud storage, we test the download speed from another client. The test environment is shown in figure 2. In our test, we used same PC install clients, and tested under same network condition. We selected three most popular cloud storage provider: Data Bank(DBank) from huawei company[6], Kingsoft fast disk [9] and surfing disk from China Telcomm [10].

- Model: HP Pavilion p6-1236cx
- Hardware: Intel Celeron CPU G540, 2GB memory.500G SATA disk(7200 rpm)
- Software: Windows XP sp2
- Network: 1Gbps intranet, education network to internet

4. EVALUTION METHODOLOGY

Client mode is the most flexible mode. It can support several data compress functions to accelerate the upload/download speed. There are three kinds of cloud storage clients. First use local disk as cache to save files in cloud, second access files on remote side directly, and the third store same files in local devices. The first type includes DBank data bank, Kingsoft fast disk and Surfing disk. After installation of DBank and Kingsoft fast disk, the client will create a local hard disk cache to synchronous with cloud storage. The local disk cache is a folder or virtual disk partition. Files to be uploading will save to this folder until upload process finished. If upload process was interrupted by network disruption, the file will upload next time. When the synchronization finished, the file will remove from local folders. By using local disk cache, it presents a very fast upload speed, but it did not finish upload process at the point it reports.

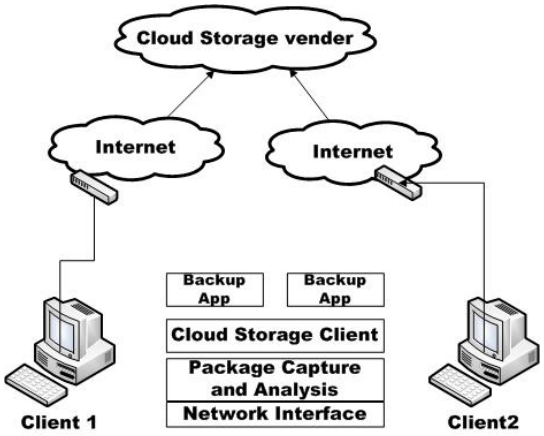

Figure 2: Test environment

Another type is Surfing disk. After user log into cloud storage, client directly displays the storage space in the remote side. And the client will download/upload files directly without local cache. When the network interruption, synchronous folder disappears, the local hard disks not provide the file mapping.

Our goal is to provide a fair comparison among various online backup cloud storage providers. This will help customer to choose which provider is best under his own condition. The test results of certain service provider may vary under different network condition.We use a fixed PC terminal install several clients one by one, and test the download/upload speed of different file types. To evaluate the performance of different cloud storage service, we must solve following problems:

- Measure the time of file transfer background. Most clients upload/download files background, and do not prompt when it finishes. We must find an effective way to measure the transfer time.
- Make the test process reproducible. Although the test results may change under different network condition and workloads of cloud storage. The test results should be stable under described environment. We need to find out the factors taking effect on the final test results.
- Measure the acceleration effect of different data preprocessing. Data compress and de-duplicate can speed up the transfer rate. We need design different test scenario to measure these effect.

Cloud storage client transfer files background and does not prompt when it begin and stop. It is difficult to measure the transmission time. We solved this problem by capture packets transferred between user and cloud storage vendors. Transmission is most time cost part of backup and restore, So we mainly focus on how many time used in transmission. We used a network sniffer etheral to capture all pack-

ets transmitted between client and cloud storage vendors. Ethereal is high performance sniffer software. It can capture packets of given IP or port. We can get port number which client used by "netstat -ano". We use the time between first and last packet send to cloud storage services as the transmission time. This method can apply to all kinds of cloud storage clients.

There are many routers and hubs between end-user and cloud storage service provider, transmit speed varies over time. Which is an important factor to cloud storage performance. Because cloud storage nodes do not response ping packet, we can not identify the network condition from end-user to cloud storage provider. Network performance is descibed in upload/download speed and packet transfer delay and packet loss rate.

There are several data pre-processing technology can accelerate the transmission. We choose different test scenario to measure these effect. At first, we test the transfer speed of raw file. We create files full of random number to make it difficult to compress or do any de-duplicate process. Then we compare the transmit speed of a normal file and the file after compressed by winzip. If the client compress the file before transmission, the speed of normal file should be much higher than compressed file. At last, we test upload same files using different users and clients. If the client check the digital fingerprint of files before it upload, the speed of upload should be very quickly.

5. EVALUATION RESULTS AND ANALYSIS

In this section, we illustrate the test results on three different cloud storage services. We have done following three different tests:

- Upload/download speed for different size files. The file was generated by random numbers. This test can measure the performance varies range when handle different files.
- Upload files and compressed version. This test can measure if the cloud storage client optimize the transfer process by compress files before it upload/download file.
- Upload same files through different client. This test can measure if the cloud storage vendor uses de-duplicate technologies to optimize transfer and space needed.

5.1 upload/download speed of different file size

Figure 3 shows that according with file size increasing, the average upload speed increases too. And we can find that the speed reach the max when file size is 40M. In the current network condition, DBank is the fastest one and Kingsoft fast disk is the slowest.

Figure 4 shows that download speed changes during file size change. Surfing is the best one in the condition of file size less than 20M, when file size greater than 20M, DBank is the best one. Another conclusion is that download/upload speed varies rapidly when file size change. It shows that the performance is not stable. All of these three cloud storage provider should improve their performance stable. Kingsoft fast disk and Surfing disk meets the small file upload/download condition. DBank disk can do better in the condition of big file upload/download. But this test is under certain network condition, user can do these tests again to choose suitable provider.

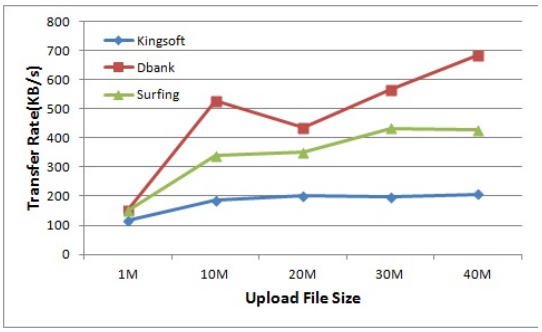

Figure 3: Upload speed of different files

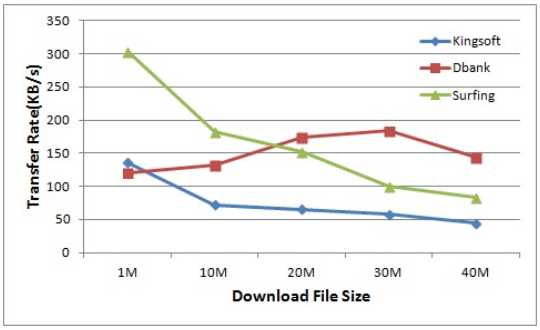

Figure 4: download speed of different files

5.2 upload raw file and compressed version

Figure 5 show upload speed of compress and uncompressed files. Because limit of network speed, some client compress files before upload. In this case, upload compress file should slow than uncompressed one. According to the test result, we can find that Dbank maybe compress files before transfer. It's strange that Surfing transfer compressed files faster than normal files.

The results show that three clients have different pre-upload

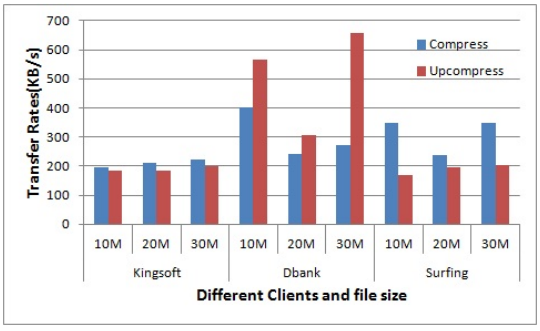

Figure 5: upload speed of raw file and compressed file

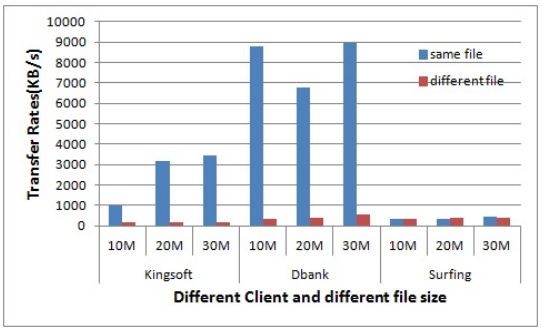

Figure 6: Upload same file from different clients

process method. Compress before transfer may save time in some condition, such as text file or documents. But for compressed file format, such as JPEG, MPEG, RM, compress will not decrease the size too much. In these cases, the performance may worse than normal clients. The clients can judge the format of files to be transferred, and then choose compress or not. It may be the best way to solve this problem.

5.3 upload same file through different client

Deduplicate is a specialized data compression technique for eliminating duplicate copies of repeating data. Figure 6 shows the upload speed of duplicate files. There are huge differences between three clients. Kingsoft and DBank accelerate the upload speed by de-duplicate or other similar technologies.

In cloud storage environments, many users may save same pop songs or pictures in their space. This technology can reduce the amount of transfer data and space needed in the cloud storage devices. Surfing does not check the content of file, so it uploads each file at same speed. Kingsoft and DBank checked the fingerprint of files and transfer second same file very quickly. It require time to check if the file same with existing files, but in our test, we did not take these time into consider because of file size is too small. In worst conditions, cloud storage client calculate the character of file to be upload, but it is different with any existing one. In this condition, upload process cost calculate time and transfer time. Another argument about this technology is security. Cloud storage services provider compare all users data to decrease the space needed, which means all user data was scanned.

6. CONCLUSIONS

In this paper, we introduce a method to evaluate the performance of cloud storage clients. From the end-user perspective, the method test network disk transfer rate. By using the sniffer capture packets, this method can obtain the accurate transmission time. And we use this method test performance of three cloud storage provider. The results show that as the file size changes, the speed of transmission is different. File size take effect on performance of cloud storage. It's important to improve big file transfers. Compressed file transfer rate is slow than normal file, maybe cloud storage vendor compressed the file before transfer. De-duplication technology also can improve the transmission performance. The method needs analysis packet transferred, it's difficult for normal user. We need develop a test tool to analysis these packet automatic. Another discourage of the method is that it does not take concurrency into consider. When many user use same service at same time, the performance maybe varies. It also need more work to measure the performance of cloud storage from view of service provider.

Acknowledgments

This work was made possible thanks to the NPU Fundamental Research Foundation under Grant No.JC20110227, the Industry Innovation Foundation under Grant No.2011BAH04B05 and the National Natural Science Foundation of China under Grant No.61033007. And this work was supported by China Scholarship Council.

7. REFERENCES

[1] Amazon simple storage service(s3). `http://aws.amazon.com/s3/`.
[2] Amazon storage gateway. `http://aws.amazon.com/storagegateway/`.
[3] Cloud storage and the innovator's dilemma. `http://cloudstoragestrategy.com/2009/07/cloud-storage-and-the-innovators-dilemma.html`.
[4] Cloud storage comparison. `http://blog.gsmarena.com/cloud-storage-comparison-google-drive-vs-dropbox-vs-sug`
[5] Cloudarray storage gateway. `http://www.twinstrata.com/CloudArray-overview`.
[6] Data bank. `http://www.dbank.com/`.
[7] Filesystem benchmark tool. `http://www.iozone.org/`.
[8] Hard disk utility. `http://www.hdtune.com/`.
[9] Kingsoft fast disk(kuaipan). `http://www.kuaipan.cn/`.
[10] Surfing disk. `http://webmail30.189.cn`.
[11] Wikipedia,compare of online backup storage service. `http://en.wikipedia.org/wiki/Comparison_of_online_backup_services`.
[12] D. Chiu and G. Agrawal. Evaluating caching and storage options on the amazon web services cloud. In *Grid Computing (GRID), 2010 11th IEEE/ACM International Conference on*, pages 17–24. IEEE, 2010.
[13] B. Cooper, A. Silberstein, E. Tam, R. Ramakrishnan, and R. Sears. Benchmarking cloud serving systems with ycsb. In *Proceedings of the 1st ACM symposium on Cloud computing*, pages 143–154. ACM, 2010.
[14] I. Corporation. I/o subsystem measurement tool. `http://www.iometer.org/`.
[15] A. Li, X. Yang, S. Kandula, and M. Zhang. Cloudcmp: comparing public cloud providers. In *Proceedings of the 10th annual conference on Internet measurement*, pages 1–14. ACM, 2010.
[16] M. Vrable, S. Savage, and G. Voelker. Cumulus: Filesystem backup to the cloud. *ACM Transactions on Storage (TOS)*, 5(4):14, 2009.
[17] L. Zhao, A. Liu, and J. Keung. Evaluating cloud platform architecture with the care framework. In *Software Engineering Conference (APSEC), 2010 17th Asia Pacific*, pages 60–69. IEEE, 2010.