A Method of Extending Document Collaborative Editing Function of Netdisks in Cloud Environment

Yi-Na Qiao, Yun-Mei Shi
Beijing Key Laboratory of Internet Culture and Digital Dissemination Research of Beijing Information Science and Technology University
School of Computer, Beijing Information Science and Technology University
Beijing, China
Email: sym@bistu.edu.cn

***Abstract*-Nowadays,netdisks have played an important role in our life. The main problem of those netdisks is that they generally provide functions of file storage and retrieve, but fail to support document collaborative editing. In this paper, we analyse architectures of the general netdisks and propose a method to expand the function of document collaborative editing for netdisks. Based on the extended method, we further design, implement a document collaborative editing system named DAVOffice, and combine it to a netdisks interface. Extensive experiments are conducted and the results verify that the proposed method is reasonable, easy to combine netdisks and extend collaborative editing functions for netdisks. Specifically, the proposed method can effectively expand applications of netdisks and this connection has achieved lower communication traffic compared with several similar well-known products.**

Keywords-cloud platform; netdisks; collaborative editing; online storage system; interface.

I INTRODUCTION

Google proposed the concept of cloud computing in 2007.Since then, more and more Internet business giants focus on cloud computing, so cloud computing has a rapid development. As a new type of service delivery and usage model, cloud computing is a calculated method that is based on the Internet, which can provide users with the software and hardware platform and network resources. It is more and more widely used because of its good expansibility, stability, generality and cheap cost.

With the development of office products, cloud office and online office have become a new type of office platform. In a narrow sense, cloud office focuses on “office documents”, providing document editing, storage, collaboration, communication, mobile office, workflow and other SaaS cloud service, which can improve work efficiency, reduce operating costs and help users to share information and communication [1].

There is no doubt that netdisks office is an important field of cloud office. More and more companies provide netdisks service for the public, such as Baidu, 360, Kingsoft, and etc. At present, the domestic netdisks provide online storage as main function, also support documents management which including upload and download files. In addition, documents can be shared with each other by links. Moreover, online preview also can be supported. However, almost all the domestic netdisks cannot provide document content collaborative editing which bring inconvenience to people’s daily work.

OneDrive and Google Drive are very famous foreign netdisks. Compared with domestic netdisks, they can support more functions. OneDrive can share documents with other users through getting documents links. Only in this way can users edit the documents collaboratively. However, users can only use Microsoft’s Office Online or local Office to edit the documents. Obviously, OneDrive has poor platform versatility [2-3]. Moreover, OneDrive doesn’t support version recovery. Google Drive can support document collaborative editing and can also recover documents’ changes of within 30 days. But the disadvantage of Google Drive is that it can only use GoogleDoc to edit documents [4-7]. So Google Drive also has a strong dependence on platform.

From the analysis mentioned above, we know that netdisks either cannot provide document collaborative editing, or support this type of function but has the poor platform independence. To solve this problem, this paper analyses the framework of common cloud storage netdisks. Based on the framework features, a method named DAVOffice is proposed in this paper to extend document collaborative editing for netdisks. This method can make netdisks easily call collaborative editing system, and let netdisks have the ability of collaborative editing. More importantly, the communication traffic of DAVOffice is lower and the system performance is better than other solutions. Therefore, users’ experiences have been improved as the reduced communication traffic is able to shorten the transmission delay [8-10] and increase the quality of experience.

The reminder of the paper is organized as follows. In section 2, we analyse the common netdisk’s architecture. In section 3, we present a method of extending the function of document collaborative editing for netdisks. A prototype system is realized and system architecture is described in section 4. To prove the correct of presented method, a set of intensive experiments is shown in section 5, and finally, the conclusion and future works in section 6.

II ANALYSIS OF COMMON NETDISKS’ ARCHITECTURE

Some literatures [11-14] analysed netdisks framework from different perspectives, in which, [11] and [12] divided the netdisks architecture into four layers. This four-layered cloud storage architecture can improve system performance and scalability. Literature [13] also divided cloud storage into four layers in order to solve data-intensive applications. Compared with other cloud storage architecture, this architecture is more efficient when dealing with the data-intensive applications. Five-layered netdisks architecture is proposed in literature [14], including network and storage infrastructure layer, storage management layer, metadata management, storage overlay layer and service interface layer. This architecture is based on the four-layered architecture. The difference between them is that five-layered architecture adds the metadata management layer to facilitate information management and load balancing among different domains.

At present, four-layered cloud storage architecture is a common one used by domestic vendors,for example Baidu, 360, Kingsoft and so forth. This kind of architecture has many significant advantages, such as good system performance, strong scalability and high security.

Therefore, this paper will analyse four-layered netdisks architecture, shown as Figure 1. Based on the architecture, document co-editing function is added and implemented over netdisks.

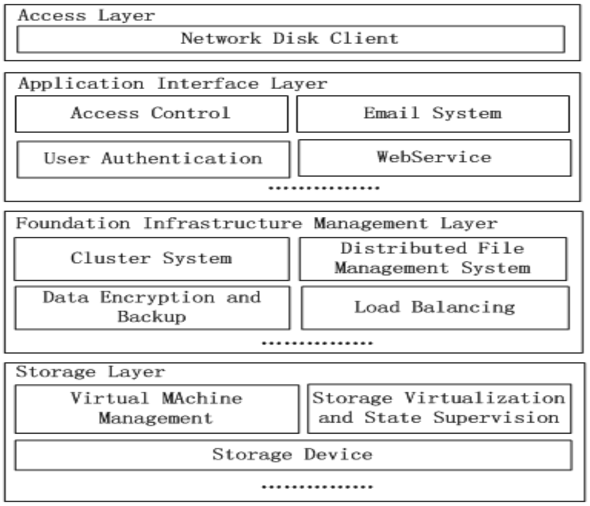

Access Layer
Network Disk Client
Application Interface Layer
Access Control
Email System
User Authentication
WebService
Foundation Infrastructure Management Layer
Cluster System
Distributed File Management System
Data Encryption and Backup
Load Balancing
Storage Layer
Virtual MAchine Management
Storage Virtualization and State Supervision
Storage Device

Figure 1. Common architecture of netdisks.

Storage layer, providing virtual resources for users, focuses on virtual machine management, storage virtualization and state supervision. In order to facilitate the management of hardware resources, VMware Sphere, a management software of hardware resources, is usually used to monitor and maintain hardware devices, and manage device virtualization and multi-link redundancy.

Cluster system, distributed file management system, data encryption and backup and load balancing belong to the infrastructure management layer. In this layer, hadoop, HDFS (distributed file management system) and data encryption technology are widely used as the middleware, which can help multiple storage devices work together to provide services to outside. In this way, we can get stronger and better data access performance.

Because of interface for interacting with outside, application interface layer becomes the most flexible part of cloud storage. By using interface, netdisks can combine third-party software with them to extend their functions. For example, netdisks can provide mail service by intergrating e-mail system into netdisks. In short, application interface layer plays an important role in extending netdisks functions.

Access layer, facing to users directly, is located on top of netdisks architecture. Generally, netdisks provide two access methods: by client or by webpage. All the functions of uploading, downloading sharing files stored in netdisks are implemented in this layer. Accordingly, access layer can make users communicate with netdisks.

IIISOLUTION OF EXTENDING NETDISKS' FUNCTION

A. *Design Philosophy*

According to the analysis of previous section, common architecture of network disks can be divided into four layers, in which interface layer provides a channel of extending netdisks' functions. Accordingly, when extending document co-editing function over netdisks, application interface layer should be used. Based on the idea, we rebuild the netdisks' architecture, shown as Figure 2.

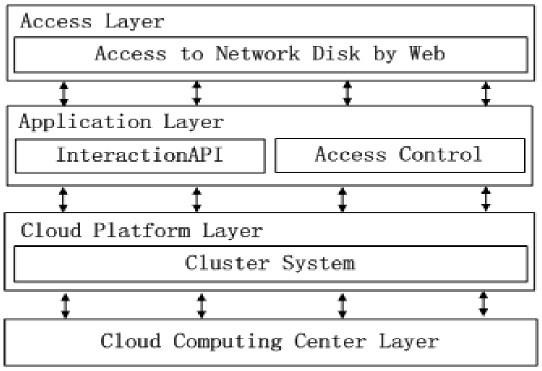


Figure 2. The architecture of netdisks with document collaborative editing function.

To keep consistent with the common architecture of network disks, the architecture that will integrate collaborative editing system still keeps four layers, shown as figure 2. The differences between these two architectures are top two layers. Application layer, corresponding to the interface layer in figure 1, adds collaborative editing InteractionAPI and access control. The former can make netdisks realize document collaborative editing function. Users can collaboratively operate documents stored in netdisks by InteractionAPI. InteractionAPI is an interface that can let netdisks interact with document co-editing system. Actually, it is an important issue for netdisks to ensure the secure of the documents stored in netdisks, so access control is essential, such as controlling read or write document paragraphs. Access layer corresponds to the top layer in Figure 1. Web access way is used in this layer.

Compared with other access ways, web access way can guarantee users to access netdisks conveniently without downloading any client application.

Cloud computing center layer corresponds to the storage layer in Figure 1. This layer is the foundation of cloud storage, and it manages physical devices and connects cloud storage devices together by WAN, Internet or dedicated network. This layer provides convenient service for the upper layers of netdisks by scheduling storage resources.

B. *Interface Design*

Interface design aims at free data exchange and share among different systems. Good interface design can reduce the system interdependence and improve system compatibility and portability. In the background of netdisks, the interface design relevant to collaborative editing not only needs to meet the common architecture requirements of netdisks but also need to shield the implementation details of co-editing function. With the help of interface, netdisks only need to transmit documents and user information to co-editing system instead of considering collaborative control.

InteractionAPI defined in Figure 2 is an interface which help netdisks invoke the document collaborative editing system. Necessary information needs be transferred from by InteractionAPI . Apart from document collaborative editing function, user and security management over netdisks can also be realized. Based on the design idea of interface, four basic methods are defined in InteractionAPI:

```
interface InteractionAPI {
openFile(FileName);//open file method
getUserInfo(UserName);//get user informations
downloadFileToDir(FileName,Target Path,User Information);
//download document to the target directory the user specified
uploadFileToDIr(FileName,Target Path,User Information);
//upload document to the target directory the user specified
```

1) openFile: This method can open the documents stored in cloud platform layer. Documents need to be loaded to local browser when users edit documents firstly. Then document collaborative editing system will display the contents of documents.

2) getUserInfo: As the same as documents, user information is also stored in the cloud platform layer. The current user information (including user name, document update time and the colour of document content) is obtained by getUserInfo method which can provide information for the initialization of document collaborative editing system.

3) downloadFileToDir: This method can download documents to the user-specified directory, and then open the file. This method must be used with openFile method simultaneously.

4) uploadFileToDir: This method can upload documents to the user-specified directory. When users finish editing, document should be saved. Then, uploadFileToDir method is called by system and upload finished editing documents to the distributed file management system. Save method need to work with this method together.

Document collaborative editing function for netdisks is supported by InteractionAPI, and the process of open operation is shown in Figure 3.

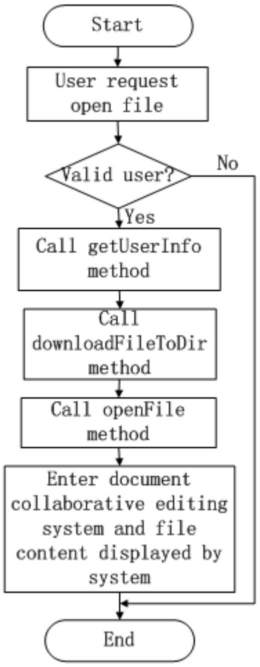

Start
User request open file
Valid user?
No
Yes
Call getUserInfo method
Call downloadFileToDir method
Call openFile method
Enter document collaborative editing system and file content displayed by system
End

Figure 3. Netdisks and document collaborative editing system to open file by calling the interface.

When user requests to open file, system will first call the method which is in access control to check the user's legality. If invalid, system will refuse the request directly, otherwise, netdisks call getUserInfo method, which is in InteractionAPI, to initialize document collaborative editing system. The initialization of document co-editing system includes two parts: 1) show user interface in web page. 2) trigger the monitor of document co-editing system to capture user's behaviour.

After completing initialization, the system will invoke downloadFileToDir method to download a document to current user directory, and then invoke openFile method to open the document. Since then, all the document operations are totally controlled by co-editing system, which is responsible for the multi-person collaboration editing, synchronous and concurrent control to ensure the document's content consistency.

IV IMPLEMENTATION OF COLLABORATIVE EDITING SYSTEM OVER NETDISKS

According to the method discussed above, we implemented a document collaborative editing system over a netdisk. The system principle is illustrated in Figure 4. Interaction interface is one that interconnects netdisks with document co-editing system. By using , netdisks can provide powerful collaboration editing function.

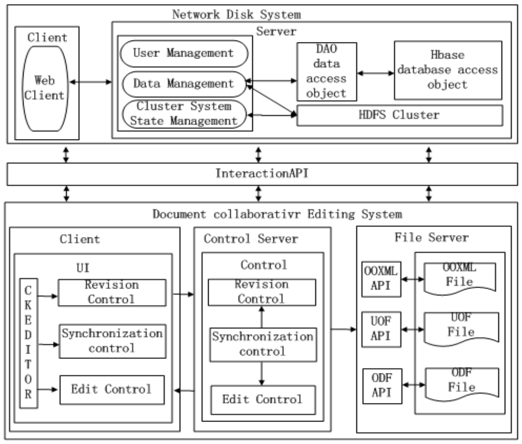


Figure 4. The principle of document collaborative editing system over netdisk.

Based on the common architecture of cloud storage, we build a prototype using hadoop, HDFS and Hbase. The functions of the prototype are listed as follows:

1) User Management: including user information management, user registration, user login and user information updating.

2) Folder Management: including add, delete and rename folder.

3) Documents Management: including upload, download, move, and delete document, and document sharing.

Download method in this section has different definition with section 2.2 mentioned. The former aims at user and user can download document by using download method directly. The latter can help document collaborative editing system to work. Document is downloaded in the temporary directory specified by user. User cannot get document directly.

User management, data management and cluster state management is supported by netdisks. User can access netdisks by Web page.

The object of user management is user information. System can exchange information with Hbase database by data access object.

Data management aims at managing files. Using HDFS store document and operate file by HDFS API.

Cluster operation log is managed by cluster state management which can ensure the cluster work methodically.

Document collaborative editing system is mainly divided into three parts: client, control server and file server.

Client is oriented for the user, providing document preview and editing interface. Control server is responsible for the revision control, synchronization control and editing control. In file server, system accesses OOXML document, UOF document and ODF document by different document API. OOXML[15] is a technical specification for Office 2007 and is the international standard of document format. UOF[16] is the China’s domestic document standard and is short for “standard text”. ODF[17] is a standard based on the XML file format specification, which is developed by Sun Microsystem and is an international standard issued in 2006.

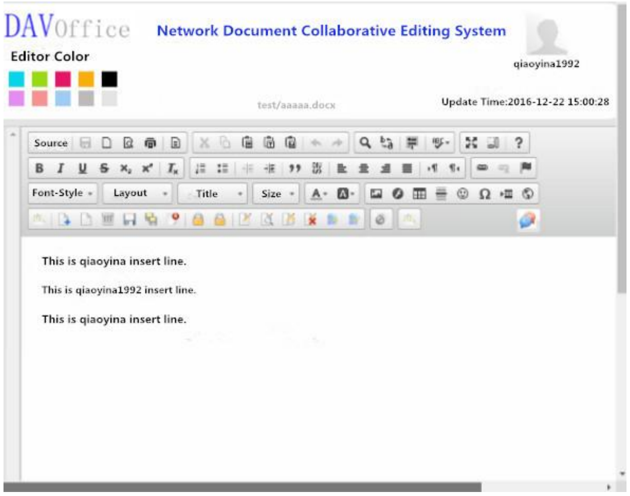
Revision control means to produce and manage modification information according to the procedure of editing documents when processing online documents, in order to backtrack documents to any state before. Viewing document revision information, document editing trace, content reply and revision revocation are realized in this module. Client displays revision interface and submit edited content to server. Control server, according to submitting time, line up the document operations from different users to form diverse document versions. Then, control server copes conflicts generated in the control processing and returns result to client. Finally, client will analyse protocol and display result.

Synchronous control provides the function of collaborative editing user sharing the latest status of the Web document in anytime. This function can avoid the conflicts occurring in the process of collaborative work which is the foundation of revision control and editing control. In synchronous control, client is responsible for the encapsulation of request protocol, the analysis of the response protocol and the display of document content. Control server is in charge of analysing the protocol sent by client, adding concurrency control and synchronous display control according to the result, afterwards sending the operation request to server. After that, dealing requests backed from server. Finally, packing operation response protocol and sending to client which can achieve synchronous display in every client.

Editing control is used to control the editing function of the document collaborative system. Document collaborative editing system can support the function of many people to edit the same document together. For example, when a user edits a paragraph, it is necessary to lock the editing paragraph to avoid conflicts during the editing process. The function provided by client is different from server supported. Client provides editing function and the server is responsible for concurrency control in multiple user document editing.

The most important function provided by netdisks is document storage. Documents are stored in HDFS. When user wants to access document stored in HDFS, document collaborative editing system will download document to the temporary folder from HDFS by InteractionAPI. After that, the document collaborative editing system is responsible for document collaborative editing process. When all users about this document finish document editing, document is closed and uploaded to the HDFS through InteractionAPI.

The main interface of system is shown in Figure 5. DAVOffice is the name of system.


DAVOffice Network Document Collaborative Editing System
Editor Color
qiaoyina1992
test/aaaaa.docx
Update Time:2016-12-22 15:00:28

This is qiaoyina insert line.

This is qiaoyina1992 insert line.

This is qiaoyina insert line.

Figure 5. Document collaborative editing system interface.

V EXPERIMENT ANALYSIS

This paper verifies the proposed method from two aspects, including function test and performance test. For the former, experiment focuses on traditional netdisks and DAVOffice, and the latter, experiment focuses on well-known document collaborative system and DAVOffice.

A. *Function Test*

The following tests are performed by using the traditional netdisks and DAVOffice.

1) Add, delete and rename folder.

2) Add, delete and rename document.

3) Test and verify if documents can be shared with other users in netdisks.

4) Insert, delete and update document content.

5) Test and verify synchronization: if a document can be correctly edited by three different users synchronously.

6) Test and verify if documents can be correctly recovered to any previous version by different users.

The result of function test is shown in Table 1, which demonstrates that traditional netdisks can only provide file or folder management and document sharing. The function-expanding netdisk, namely DAVOffice, can provide not only the functions that traditional netdisks have, but also document co-editing services that traditional netdisks don’t have. Apparently, DAVOffice has more powerful functions compared with traditional netdisks, therefore, can improve users’ work efficiency and be convenient for daily work.

TABLE I. COMPARISON OF FUNCTION TEST.

Test Items	Traditional Netdisks	DAVOffice
File Management	√	√
Folder Management	√	√
Document Sharing	√	√
Editing Control	×	√
Synchronization Control	×	√
Revision Control	×	√

B. *Performance Test*

In document collaborative editing system, communication traffic and the response time are the important indicator of system performance evaluation. The lower communication traffic is, the better performance system will be. Similarly, the shorter response time is, the better performance system will be. However, collaborative editing is severely affected by network environment, therefore, it’s hard to make a meaningful comparison for response time between other collaborative editing system and DAVOffice. Communication traffic, by contrast, is suitable for system performance evaluation for its independence of transmission latencies. So finally, we decided to test and compare communication traffic of four different cooperative editing systems, including DAVOffice, OneDrive, Google Drive and Baihui Write.

For the four systems mentioned above, we executed insert, delete and update operations respectively in different size of OOXML documents from 10KB, 50KB, 100KB, 150KB, 200KB, 500KB to 1M. The tests show that the communication traffic has no concern with document size, but relevant to the number of operated characters. In this paper, for simplicity, we only take an OOXML document of 10KB size as an example to illustrate the relationship between communication traffic and the number of operated characters.

In four document collaboration systems mentioned above, Insert operation, delete operation and update operation are added in 10KB OOXML document. The character number of each operation is 2, 10, 30, 50, 70, 100, 130, 150, 170 and 200, respectively. The communication traffic statistics is illustrated in Figure 6, 7 and 8.

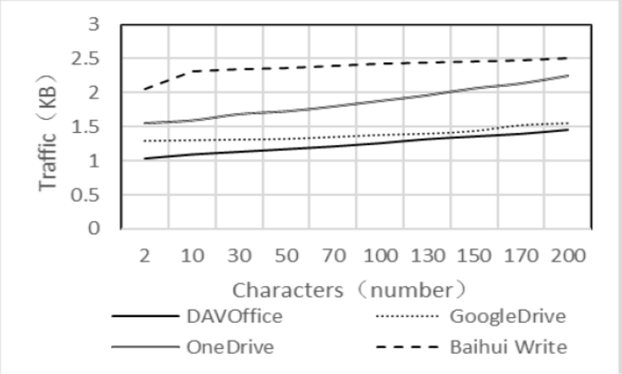


Figure 6. Insert operation statistic.

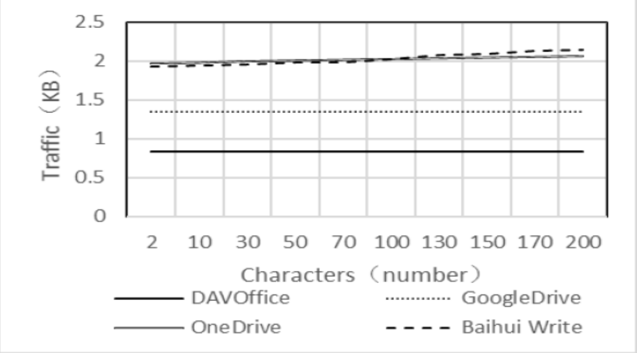


Figure 7. Delete operation statistic.

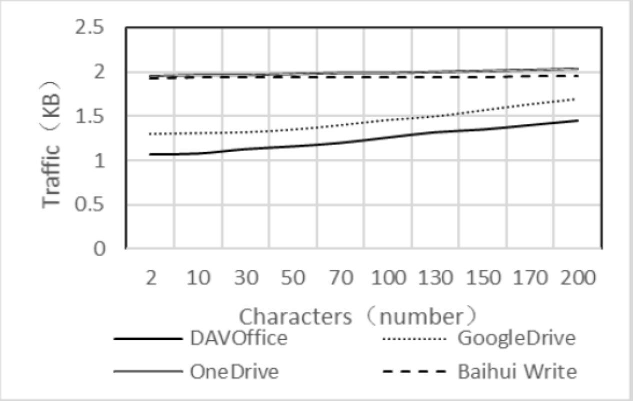


Figure 8. Update operation statistic.

Apparently, from the above statistical results, the communication traffic of DAVOffice is lower than other document collaboration systems. The result shows that DAVOffice's communication cost is low, while the system performance is better. This is especially significant to increase users' quality of experience, as more communication traffic definitely result in more time delay to finish those transmissions [8-11].

The result shows that the presented method of extending document collaborative editing function over netdisks is reasonable, and can easily extend the document collaborative editing function for netdisks and system performance is better than other well-known document collaborative editing system.

VI CONCLUSION AND FUTURE WORK

In this paper, we analyse the development of netdisks and point out problems in the current netdisks system. Specifically, most of the current netdisks cannot provide document collaborative editing. Although some systems support this type of functions, they have very strong platform independence. Based on the universal four layer netdisks architecture, in this paper, we propose a method to expand document collaborative editing functions for netdisks. Leveraged by this method, a prototype is designed and implemented in this paper. Extensive experiments have been conducted and those results verify that the proposed method is effective and can easily combine netdisks with document collaborative editing. Most importantly, users' quality of experience has been improved as the communication traffic has been greatly decreased.

Combining netdisks with document collaborative editing system can bring convenience but security concerns are also raised. Those concerns include the security of streaming documents over netdisks, the security of paragraph content and the request from different domain [18-19]. All those concerns are going to be addressed in our future work. Specifically, one of our future works is to design a streaming document access control model in cloud environment aiming to address the streaming document security problem under cloud environment.

ACKNOWLEDGMENTS

This work originated in the Beijing Municipal Education Commission, under the grant (KM201511232012). We would like to thank the funding agency for supporting our research.

REFERENCES

[1] J. Tang,Y.M.Shi,N.Li, International Conference on Cyberspace Technology, Inspirations for office document formats from web office,242-246(2013)

[2] Microsoft. Work together on Office documents in OneDrive,(2014) https://support.office.com/en-us/article/Work-together-on-Office-documents-in-OneDrive-ea3807bc-2b73-406f-a8c9-a493de18258b?CorrelationId=32fe03d2-e42b-4a08-871f-b115f9cfd1c6&ui=en-US&rs=en-US&ad=US.

[3] P.Sarah, Office Live Workspace vs Google Docs: Feature-by-Feature Comparison,(2008) http://technology.feedfury.com/content/13089748-office-live-workspace-vs-google-docs-feature-by-feature-comparison.html.

[4] The Wall Street Journal. Google Stores, Syncs, Edits in the Cloud,(2012) http://allthingsd.com/20120424/google-stores-syncs-edits-in-the-cloud.

[5] PB. Moody,DA.Shrum, Method and apparatus for consolidating edits made by multiple editors working on multiple document copies,(1999)

[6] VJ.Vagell,A.Pavese,Collaborative electronic document editing,(2012)

[7] K.Dutta,Document collaboration system,(2006)

[8] G. Wang, Y. Ren, and J. Li, An effective approach to alleviating the challenges of transmission control protocol, IET Communications, vol. 8, no. 6, pp. 860–869, (2014)

[9] G. Wang, Y. Ren, K. Dou, and J. Li, IDTCP: An effective approach to mitigating the TCP Incast problem in data center networks, Information Systems Frontiers, vol. 16, pp. 35–44, (2014)

[10] G. Wang, Y. Wu, K. Dou, Y. Ren, and J. Li, AppTCP: The design and evaluation of application-based TCP for e-VLBI in fast long distance networks, Future Generation Computer Systems, vol. 39, pp. 67–74, (2014)

[11] Y.X.Yang,W.P.Deng,J.S.Deng,Y.Li,Telecommunications Science,System Architecture and Key Technology of Network Disk Based on Cloud Storage, 28,68-72,(2012)

[12] X.Zhang, Computer Science and Engineering,Research and Implementation of Cloud Storage Platform Based on Hadoop,(2013)

[13] Y.Huo,H.Wang,L.Hu,et al, International Conference on Computer and Management, A Cloud Storage Architecture Model for Data-Intensive Applications,1-4,(2011)

[14] W.Zeng,Y.Zhao,K.Ou,et al, International Conference on Interaction Sciences: Information Technology,Research on cloud storage architecture and key technologies,1044-1048,(2009)

[15] ISO JTC 1/SC 34,ISO/IEC 29500:2008 Information technology -- Office Open XML file formats, Geneva: ISO/IEC,(2008)

[16] National Information Technology Standardization Technical Committee,GB/T20916-2007, Chinese office software document format specification,(2007)

[17] ISO JTC 1/SC 34,ISO/IEC 26300:2006 Information technology -- Open Document Format for Office Applications (Open Document),(2006)

[18] F.D.Guo,M.Zhang,Y.Zhang,Z.Xu,Journal of Software,Study on Cloud Computing Security,22,71-83,(2011)

[19] M.Qiang,Z.L.Ai, Computer Engineering and Design,Access control models for cloud computing,33,4487-44,(2012)