An Approach of Extending Net Disks’ Collaborative Document Editing Functionality

Yunmei Shi ¹ ², Ning Li ¹ ², Qi Liang ¹ ², Zhichao Yu ¹ ² and Yina Qiao ¹ ²

¹ Beijing Key Laboratory of Internet Culture and Digital Dissemination Research, Beijing, China
² School of Computing of Beijing Information Science & Technology University, Beijing China
sym@bistu.edu.cn, lining@bistu.edu.cn

Abstract. Net disks are playing an increasingly important role in our daily works. Currently, net disks have given us an ideal ecosystem to store and share files; however, they generally cannot support collaborative document editing. To overcome the shortage, after analyzing the model of cloud storage and net disks, this paper proposes a solution to extend net disks’ collaborative editing functionality via API. Based on the proposed method, a net disk system coupled with collaborative document editing functionality is built. The experiment result shows that the extending approach is feasible and effective. This approach can not only make net disks support document collaborative editing, but also be easy to be implemented.

1. Introduction

With the development of the cloud computing technologies, cloud office has become a new office pattern that is looked for and sought for by people. In a narrow sense, cloud office [1] means document-centric, and provides companies and institutions with the services of document editing, co-authoring, mobile office, workflow, and etc. This office pattern can greatly improve people’s working efficiency and reduce the running costs of the companies and institutions, besides, facilitate information sharing. Collaborative document editing is necessary for the cloud office, as it is an important basis for co-authoring and cooperating office documents.

From the perspective of cloud office, a net disk is not a part of office platform, as it is mainly used for resource storage, rather than editing or collaborative editing demanded by the cloud office. Some famous domestic net disks, such as Baidu Netdisk, Kingsoft fast-disk, Lenovo Disk etc., only have resource storage and sharing functionality, and some more powerful products can support online preview of the documents, however none of which provide online document editing, especially the collaborative editing.

Besides the cloud storage services, both Google Drive and Microsoft OneDrive provide online document editing, so they bring us a great ecosystem to work [2,3,4]. However, For OneDrive, it supports the document sharing and co-authoring with the help of Office Online or local Office software of Microsoft, which means that collaborative editing depends on Microsoft products. Similarly, Google Drive provides users with collaborative document editing via Google Docs [5].

The document storage and online editing products of Microsoft and Google have driven us to think about a problem: how to find a way to extend net disks’ collaborative editing functionality. In the following, the paper will discuss the solution of the problem. Section 1 analyses the models of the cloud storage and net disks. Section 2 discusses the interface design between the net disk and collaborative editing system, and the model design of net disks coupled with collaborative editing.
functionality. Section 3 and 4 demonstrates a web-based application based on the proposed method, and verify the experimental results. Finally, the paper discusses the implications of author’s study and consider future work in the area.

2. Model Analysis
In literature [20], a five-layer cloud storage model is described. Each layer and its functions are as following:

Network and storage infrastructure layer provides storage infrastructure interconnected by network. Storage management layer organizes distributed storage resources. Metadata management layer is for load balancing purposes. Storage overlay layer is for virtualizing, service retrieving and redirecting. Service interface layer provides clients with a uniform interface to access the cloud storage resources.

Literatures [6, 7] proposes a four-layer net disk model, including storage layer, basic management layer, application interface layer, and access layer.

Storage layer: consists of various storage devices interconnected by a wide area network, Internet or a fiber channel network. This layer is the basic component of net disks. Virtualization technologies, such as KVM, Xen, VMWare, etc., are often used in this layer for the purpose of virtualizing and monitoring hardware devices, and maintaining faults [12]. This layer corresponds to the IaaS (Infrastructure as a Service) of cloud computing core services [8, 9].

Management layer: Various technologies, such as cluster, distributed file system, are usually utilized together to provide uniform external services for the upper layer. Generally, this layer uses different middleware, such as Google's GFS, Amazon's S3, open source software Hadoop etc. The middleware can coordinate the multiple storage devices in storage layer to provide better and stronger data access performance [13, 14]. Owing to open source characteristics, hadoop and HDFS are frequently used in this layer. This layer corresponds to the PaaS (Platform as a Service) of cloud computing core services [8, 9].

Application interface layer: This layer corresponds to SaaS (Software as a Service) of cloud computing core service [8, 9]. This layer provides users with extra services through new application interfaces implemented by developers [10, 11], that means developers can design and implement uniform interfaces to extend the functions of net disks. So the layer is the most flexible compared with the other layers.

Access layer: is located on top of the net disk model. Users can operate the documents from this layer, for example, uploading a document to the net disk or downloading a document stored in the net disk. In general, users can be provided two ways to access the documents: client-side or web-based mode.

A net disk is essentially a kind of cloud storage, so their models have great similarities: each model contains several layers; every layer is responsible for relatively independent functions; each model always contains interface layer. Based on product research and survey, we found that most net disks have similar layered models, and especially, almost all the models have an interface layer.

3. Proposed Solution
3.1. Design Consideration
Considering the layered models discussed above, following considerations should be kept in mind while extending collaborative editing functionality of the net disks: 1) extended net disks should maintain the same layered model as the original net disks. 2) extended net disks should support all functionalities of collaborative document editing system and the net disks. 3) attention should be paid to the functionality separation between the net disk and collaborative editing system. 4) extending approach should be easy to be carried out.

Obviously, based on the layered model, appropriate API design at the interface layer can satisfy the above considerations to a great degree.

3.2. Interface Design
Functionally, the net disks mainly offer three type of services: user management (user creation and
logout), folder management (folder creation and deletion) and file management (file upload, download, etc.). Correspondingly, three types of interfaces should be designed: user management interface, folder management interface, and document management interface, as shown in Fig. 1.

The user management interface manages all the users of the net disks, such as the users’ registering or logging out. The folder management interface mainly manages the users’ folders, such as creating or deleting folders, moving or renaming folders, and file management interface mainly provides the documents’ uploading, downloading and deleting.

The users of the net disks can be divided into two categories: document creators and normal users. For a document, the creator and privileged normal users are the legal ones of the document, which means these users can access the document uploaded by the creator. Users’ legality and their access privilege to the document are checked by the collaborative editing system. Generally, the document creator has the highest operation privilege, while the normal users’ privileges are controlled by the document creator.

In Fig. 1, the users directly interact with the collaborative editing system, so users’ requests on documents are firstly sent to it, which manages user’s operation authorities, reading and writing operations. Here, taking opening a document as an example, the steps under the collaboration of a net disk and collaborative documents editing system are as following:

1) A logged-in user sends a request of opening a document to the collaborative document editing system.
2) The system judges the user's access privilege.
3) If the request is permitted, and the document is already downloaded from the net disk into the editing system, the system will transfer the document to the user via a user-defined protocol.
4) If the request is permitted, while the document is still not downloaded into the collaborative editing system, the system will get the document from the net disk through a proper interface.

After obtaining the document, the user can perform all the privileged operations to it, such as inserting, deleting, modifying, etc. under the control of the collaborative editing system. Besides the common operations, the collaborative editing system also provides conflict detection, consistency maintenance and document synchronization to ensure the correctness of shared documents. When all users close the shared document from their respective client side, the collaborative editing system will upload the completed document to the net disk.
3.3. Model Design
Based on the interface design principles discussed above, the model of the net disks with collaborative editing functionalities is divided into four layers: infrastructure layer, platform layer, document collaboration service layer and user access layer, as shown in Fig. 2.

![Figure 2. Model of extended net disks with collaborative document editing](image)

Comparing Fig. 2 with the layered models discussed in section 2, a mapping relationship can be found, as shown in Table 1. From the table, we can find that the extended net disk’s model has similar layers with the other models.

| Net disks             | Extended net disks          | Cloud storage                      |
|-----------------------|------------------------------|------------------------------------|
| storage layer         | infrastructure layer         | network and storage infrastructure layer |
| management layer      | platform layer               | storage management layer            |
| application interface layer | collaborative service | metadata management layer           |
| access layer          | interface layer              | storage overlay layer               |

The details of each layer of the extended net disk are as follows.
1) Infrastructure layer: located at the bottom of the model, is the basis of distributed storage in the cloud computing environment, providing users with hardware infrastructure services [15]. This layer can deliver physical computing, storage, and networking resources etc., and also provide highly reliable and customized, scalable infrastructure services via virtualization techniques. Users’ documents are stored in this layer.

2) Platform layer: provides users with application deployment and management services. This layer uses hadoop, a basic framework to build distributed systems for large-scale data processing and analysis, to build a basic platform. Hadoop can be easily used for developing distributed applications, and besides, it takes full advantages of the cluster for coordinative computing and storage to achieve a horizontal extension without paying attention to the details of distributed infrastructure [17, 18]. HDFS manages users’ files stored in cloud storage. As a distributed file management system, HDFS has highly fault tolerance. More importantly, HDFS enables to provide high-throughput data access for processing large-scale data applications [16]. Users’ documents in net disks will be increasing with the extension of application scale, so the number of users and documents will become larger and larger. HDFS is greatly suitable for the large-scale document management. Hbase in platform layer is used to manage user information.

3) Collaborative service interface layer: provides users with collaborative editing service through proper interface, including collaborative shared documents’ editing, access control, version control, consistency maintenance and synchronization control.
4) Document access layer: provides users with user interface, such as web pages, to access the documents stored in a net disk. By way of the user interface, users can upload or download documents, create, delete or update documents, share his own documents with others.

The main task of service management on the right in Fig. 2 is to ensure the availability, reliability, and security of core service layers [8]. The block mainly provides access control and security management to protect the security of all the users and documents, such as the management of user access privileges, encryption and decryption of documents’ storage and transmission, and the management of user requests from different domains.

4. Implementation of System

Based on the interface and model design described above, this paper implemented a system, in which a net disk and collaborative editing system are connected through an interface, named InteractionAPI. Fig.3 demonstrates the schematic block diagram of the system.

Figure 3. Schematic block diagram of a net disk extending collaborative editing functionalities

In the block of collaborative document editing system in Fig. 3, there are two servers: file and control server. The file server puts or gets documents in HDFS cluster via InteractionAPI, and temporarily stores the documents in its memory. The file server supports three common document formats: OOXML, UOF and ODF, in which OOXML is an office document format developed by Microsoft, UOF is a self-developed office document format of China, and ODF is one released by OASIS. Each document format is supplied with corresponding APIs to directly operate documents without using office software, for example, .docx documents can be directly accessed by way of OOXML API. Control server is responsible for all the control functions, including revision, synchronization and edit control. Revision control is mainly used to backtrack documents to any historical state. Synchronization control maintains the shared documents in the uniform state among the file server and all co-authoring users. Editing control provides document-editing functionality, in which concurrency control can avoid conflicts and maintain documents consistency.

The bottom block in Fig. 3, mapping to the infrastructure layer and platform layer of the extended net disks, is for managing the users’ information stored in Hbase and documents in HDFS cluster. Cluster state management in this block manages cluster's running logs to ensure the cluster to work...
successively.

When a user wants to view or edit a document, the net disk first gets the document from HDFS through HDFS API, and then deliver it to a temporary folder of the file server via InteractionAPI. Thereafter, the document is managed by the document collaborative editing system. When all the co-authoring users close the same shared document from their own webpages, the collaborative editing system uploads the completed document to HDFS cluster. The web-based user interface of the system is shown in Fig. 4.

![Image of web-based user interface](image)

**Figure 4.** Web-based user interface implemented

5. Experimental Validation

To validate our approach, we performed function comparison tests between two systems: DAVoffice, a collaborative editing system developed by our research group; an extended net disk combined with DAVoffice via InteractionAPI. The tests are divided into two groups: conventional function tests for single user as shown in Table 2, and collaborative editing function tests in Table 3. System 1 in Tables 2 and 3 refers to DAVoffice, and System 2 refers to the extended net disk.

**Table 2.** Test results of conventional document operations (single user)

| User management | Folder management | File management |
|-----------------|-------------------|-----------------|
| User creation   | Folder creation   | Document creation|
| User log-off    | Folder deletion   | Document deletion|

| System 1 | System 2 |
|----------|----------|
| √        | √        |
| √        | √        |
| √        | √        |
| √        | √        |

The test items in Table 2 mainly test the system user management, folder management, and file management functions. √ in the table indicates that the test result is correct and consistent under the same test conditions.

The tests in Table 3 contains the following test items:

1) Tests of editing control: multiple users insert, delete, and update the shared documents from their client at the same time.

2) Tests of synchronous control: the test covered inserting, deleting, and updating the shared documents.

3) Tests of version control: check if the shared documents can be backtracked to any historical
state. The test results in Table 3 demonstrate that both systems maintain the same results under the same test scenario.

| Editing control | Synchronous control | Version control |
|-----------------|---------------------|-----------------|
| System 1        | √                   | √               |
| System 2        | √                   | √               |

The comparative analysis indicates that the extending approach for net disks is feasible and effective. Performance test will not be discussed here, and it was discussed in literature [19].

6. Conclusions

This article presents a solution to extend collaborative document editing functionalities for net disks, and then builds a system based on the solution. The system test show that the solution can effectively and conveniently make a net disk extend collaborative editing functionalities, rather than change its original model. More importantly, this approach can keep independence between the net disk and collaborative editing system. Obviously, combining a net disk with a collaborative editing application will bring great conveniences for people’s office work, however may raise certain new concerns, such as document security and privacy protection in cloud storage[4], transmission security from the cloud platform to users, fine-grained document access control and cross-domain document access [15, 16]. These issues will be solved in our future work.

This work was supported by National Key R&D Program of China (Grant No. 2018YFB1004100) and Opening Project of Beijing Key Laboratory of Internet Culture and Digital Dissemination Research (Grant No. ICDDXN006).

7. References

[1] Jiang Tang, Yun-Mei Shi, Ning Li, C.: Inspirations for office document formats from web office. In: International Conference on Cyberspace Technology (CCT 2013), pp: 242-246. Beijing (2013)
[2] Work together on Office documents in OneDrive, https://support.office.com/en-us/article/Work-together-on-Office-documents-in-OneDrive-ea3807bc-2B73-406f-8c9-a493de18258b?CorrelationId=32fe03d2-e42b-4a08-871f-b115f9cf1c6e&ui=en-US&rs=en-US&ad=US
[3] Office Live Workspace vs Google Docs: Feature-by-Feature Comparison, http://technology. feedfury.com/content/13089748-office-live-workspace-vs-google-docs-feature-by-feature-comp arison.html
[4] Shashank Arora, Gaurav Varshney, Pradeep K. Atrey, et al., C.: SecureCEdit: An Approach for Secure Cloud-based Document Editing. In: The 2nd IEEE Workshop on Security and Privacy in the Cloud (SPC 2016).
[5] Google Stores, Synchs, Edits in the Cloud, http://allthingsd.com/20120424/google-stores-synchs- edits-in-the-cloud
[6] Yang Yuexiang, Deng Weiping, Deng Jinheng, et al: Research on Cloud-based Storage System Architecture and Key Technologies. J. Telecommunications Science, 28 (10), 68-72(2012)
[7] Xie Jinxing: Netdisk System Design and Implementation Based on Cloud Storage. Hunan University (2016)
[8] Luo Junzhou et al.: Cloud computing: Architecture and Key Technologies. J. Journal of Communication. 32 (7), 3-20(2011)
[9] Jiang Yongsheng et al.: Implementation Standards for Cloud Computing and Cloud Computing: Review and Exploration. J. Journal of Shanghai University.19 (1), 5-13(2013)
[10] Yu Qi, Ling Jie: Research on Cloud Storage Security Technology based on HDFS. J. Computer
[11] Fu Yingxun, Luo Shengmei, Shu Jiwu: Safety Netdisk System under Cloud Storage Environment. J. Journal of Software. 8, 1831-1843(2014)
[12] Zhang Xing: Hadoop-based Cloud Storage Platform Research and Implementation. University of Electronic Science and Technology(2013)
[13] Huang Xiaoyun: Research on Cloud Storage Service System Based on HDFS. Dalian Maritime University(2010)
[14] Wang Yijie, Sun Weidong, Zhou Song, et al.: Key Technologies of Distributed Storage in Cloud Computing Environment. J. Journal of Software. 23 (4), 962-986 (2012).
[15] Feng Dengguo, Zhang Min, Zhang Yan, et al.: Research on Cloud Computing Security. J. Journal of Software. 22 (1), 71-83(2011)
[16] Ma Qiang, Ai Zhongliang: Access Control Model Oriented to Cloud Computing Environment. J. Computer Engineering and Design. 33 (12), 4487-4492(2012)
[17] Shuji Ji, Fu Yingxun: A Method for Implementing Secure Netdisk systems in the Cloud Storage Environment. CN102685148A.P. (2012)
[18] Xue Mao, Xue Wei, Shu Jiwu, et al.: A Secure Storage System in the Cloud Storage Environment. J. Chinese Journal of Computers. 38 (5), 987-998(2015)
[19] Yi-Na Qiao, Yun-Mei Shi, C.: A Method of Extending Document Collaborative Editing Function of Netdisks in Cloud Environment. In: Proceedings of the 3rd International Conference on Wireless Communication and Sensor Networks (WCSN 2016), pp: 537-543. ATLANTIS PRESS, (2017)
[20] Isaac Odun-Ayo, Olasupo Ajayi, Ravin Ahuja, Boladele Akanle, C.: An Overview of Data Storage in Cloud Computing. In: International Conference on Next-Generation Computing and Information Systems (ICNGCIS 2017), pp: 11-12(2017)