Cloud Based System for Streamlined Final Year Project Allocation and Tracking with Deduplication

G. Abinaya[1], K. Meenatchi[2]

[1]M.C.A Scholar, A.R.J College of Engineering and Technology, Mannargudi
[2]Assistant professor, A.R.J College of Engineering and Technology, Mannargudi

Abstract

Projects are the only thing that matters on final graduation. Project work demonstrates the depth of knowledge and some soft skills, such as creativity and problem-solving. The final year Projects will also improve your interview prospects. Therefore, it is necessary and mandatory for students to complete a project in their final year of graduation. Managing and controlling the final year projects of students using manual or traditional process is a very tedious job. This project is aimed at developing a web-based system, which manages the activity of Student Project Allocation and Tracking system. The application helps students, project advisors, and faculty administrators manage the project from the beginning to the end. The main features of this system are project title recommendation, project and team prioritization, project and team matching, project management and scoring, and report generation. This deduplication functionality ensures the elimination of redundant or overlapping project proposals, streamlining the project selection and allocation process. The system is therefore useful to help Project Internal Guide to arrange project selection and allocation procedure, as well as helping the students to submit their preferences. HOD will also be able to employ the system for keep tracking the progress of the projects. This application will be used by the Project Co-ordinator, Internal Guide, HoD and FYP students to increase the efficiency and quality of the collaboration. This application will make supervision process formal and professional.

Introduction

Cloud computing has become a popular information technology service by providing huge amount of resources (e.g., storage and computing) to end users based on their demands. Among all cloud computing services, cloud storage is the most popular. Since the volume of data in the world is increasing rapidly, saving cloud storage becomes essential. One of the key reasons that causes storage waste is duplicate data storage. Multiple users may save same files or different files containing same pieces of data blocks at the cloud. Obviously, duplicate data storage at the cloud introduces a big waste of storage resources. Data deduplication provides a promising solution to this issue. In a deduplication scheme, the CSP can cooperate with the cloud user to first check whether a pending uploaded file has been saved already or not, and then provide the user whose pieces of file data are checked duplicate a way to access the file without storing another copy at the cloud. However, since the CSP cannot be fully trusted, the cloud users may suffer from some security and privacy issues. Notably, a semi-trusted CSP may modify, tamper or delete the uploaded data driven by some profits. The damage of deduplicated data could cause huge loss to all related users (e.g., data owners and holders). Thus, the integrity of the data stored at the cloud should be verified, especially for duplicate data storage with deduplication. Duplicate data storage with deduplication. Several Proof of Retrievability (PoR) schemes have been proposed to address the issue of integrity check on cloud data storage in recent decade. In such schemes, a user adds verification tags along with a file. During the verification, the user creates a random challenge and sends it to the CSP, the CSP has to use all the data in user's corresponding files it stored as inputs to compute a response back to the user. The user then checks the integrity of the stored file by verifying the response. However, existing PoR solutions mainly aim to improve the performance at the user side and assume that the CSP has infinite computation and storage resources. Duplicated at the CSP as shown in Fig. 1a. Message-locked PoR, provides a promising solution to check data integrity when performing deduplication. It derives a same file into a same verification tag based on message-locked encryption technique as shown in Fig. 1b. However, such design restricts the users from creating their own individual tags with their private keys. Practically, we expect an effective method that can check data integrity with the support of deduplication where each user can generate its own individual verification tags from its private key against brute-force attacks

System Model

VeriDedup offers grarantee on the correctness of duplication check and supports the integrity check of deduplicated encrypted data in cloud storage. Our target system contains three types of entities: 1) Data holder who owns data and saves its data that consists of multiple blocks at CSP. It is possible that a number of eligible data holders share the same encrypted data blocks in the CSP. In particular, the data holder that first uploads the data blocks to the CSP is denoted as a data owner with regard to the same blocks. CSP who provides a data storage service with deduplication to data holders. Only one data copy is stored at the CSP, which can be accessed by all data holders with authority. Authenticated auditor (AA) who serves as a third party to check data ownership, authorize data access and cooperate with other two types of entities aiming to audit the whole procedure of data duplication check. Our research based on the following assumptions. We assume that the data holder is honest. We assume the CSP is semi-trusted. It may raise the following three security threats: 1) Snooping the private data of the data holders Cheating the data holders by providing a wrong duplication check result in order to ask a higher storage fee Causing data loss due to carelessness of data maintenance. VeriDedup is a verifiable cloud data deduplication storage scheme with integrity and duplication Proof. It holds the following design goals: Independent integrity check when deduplication: VeriDedup allows the data holder to check the integrity of its files stored at the CSP without downloading the whole files and interacting with the corresponding data owner. Flexible tag generation: VeriDedup allows each data holder to create its own individual verification tags while still can perform data deduplication over those tags. Correctness guarantee of duplication check: VeriDedup can assure the correctness of duplication check. Thus, a semi-trusted CSP can never cheat the data holders to upload any data that have already been stored by the CSP.

Research Methodology

We recognize the fact that the CSP is likely to increase its income with massive amounts of computation/storage from deduplication. In this case, confirming deduplication happened already at the CSP to get an offer of low storage charge becomes essential, our paper aims to solve this issue. For motivating the adoption of our scheme, in another line of our work, we study how to make all related stakeholders to accept and use deduplication schemes by applying game theory to design proper incentive or punishment mechanisms in three cases: client-controlled deduplication server-controlled deduplication and hybrid deduplication. Since our scheme design is built upon the one in, belonging to server-controlled deduplication, the incentive mechanism suitable for the server-controlled deduplication schemes can be applied to motivate scheme adoption. Moreover, linking a trust value to each CSP can help the users to choose a trustworthy CSP. We applied five metrics in our simulation studies to evaluate TDICP, including the data owner's computational complexity for creating and inserting the note set the data holder's storage overhead for extra data storage in integrity check the data holder's computational complexity for challenging CSP and retrieving the inserted note set for verification CSP computational complexity for responding the challenge from the data holder Data holder-CSP communication cost for transferring extra data in integrity check. The communication cost of AA to broadcast the hidden function f is omitted, since it is a one-time cost regardless with integrity check interactions. Meanwhile, we used six metrics in our simulation studies to evaluate UDDCP, including the data holder's computational complexity for initializing duplication check CSP's computational complexity for pre-processing its tag set and responding the challenge from data holders AA's computational complexity for verifying CSP computation and setting up the cuckoo filter the data holder's computational complexity for confirming duplicate blocks the communication cost from CSP to AA for constructing the cuckoo filter the communication cost between the data holder and CSP for transferring extra data in duplication check. The communication cost from AA to the data holder for transferred the cuckoo filter is omitted, since it depends on the concrete type of the cuckoo filter.

Conclusion

In this paper, we introduced VeriDedup to check the integrity of an outsourced encrypted file and guarantee the correctness of duplication check in an integrated way. The integrity check protocol TDICP of VeriDedup allows multiple data holders to verify the integrity of their outsourced file with their own individual verification tags without interacting with the data owner. On the other hand, we employed a novel challenge and response mechanism in the duplication check protocol UDDCP of VeriDedup to let the data holder instead of the CSP first tell whether a file is duplicate in order to guarantee the correctness of duplication check. Security and performance analysis show that VeriDedup is secure and efficient under the described security model. The result of our computer simulation further shows its efficiency compared with highly related prior arts,

References

1. Z. Yan, L. Zhang, W. Ding, and Q. Zheng, "Heterogeneous data storage management with deduplication in cloud computing," IEEE Trans. Big Data, vol. 5, no. 3, pp. 393–407, Sep. 2019.
2. Z. Yan, W. X. Ding, and H. Q. Zhu, "A scheme to manage encrypted data storage with deduplication in cloud," in Proc. Int. Conf. Algorithms Archit. Parallel Process., 2015, pp. 547–561.
3. Z. Yan, M. Wang, Y. Li, and A. V. Vasilakos, "Encrypted data management with deduplication in cloud computing," IEEE Cloud Comput., vol. 3, no. 2, pp. 28–35, Apr. 2016.
4. W. Shen, Y. Su, and R. Hao, "Lightweight cloud storage auditing with deduplication supporting strong privacy protection," IEEE Access, vol. 8, pp. 44 359–44 372, 2020.
5. Q. Zheng and S. Xu, "Secure and efficient proof of storage with deduplication," in Proc. 2nd ACM Conf. Data Appl. Secur. Privacy, 2012, pp. 1–12.
6. Giuseppe, R. Burns, and C. Reza, "Provable data possession at untrusted stores," in Proc. 14th ACM Conf. Comput. Commun. Secur., 2007, pp. 598–609.
7. G. Ateniese et al., "Remote data checking using provable data possession," ACM Trans. Inf. Syst. Secur., vol. 14, pp. 1–34, 2011.
8. Z. Wen, J. Luo, H. Chen, J. Meng, X. Li, and J. Li, "A verifiable data deduplication scheme in cloud computing," in Proc. Int. Conf. Intell. Netw. Collaborative Syst., 2014, pp. 85–90.
9. P. Meye, P. Raı¨pin, F. Tronel, and E. Anceaume, "A secure twophase data deduplication scheme," in Proc. IEEE Int. Conf. High Perform. Comput. Commun., IEEE 6th Int. Symp. Cyberspace Saf. Secur., IEEE 11th Int. Conf. Embedded Softw. Syst., 2014, pp. 802–809.