Information Sciences 000 (2016) 1–13

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

# Intelligent cryptography approach for secure distributed big data storage in cloud computing

Yibin Li[a], Keke Gai[b,*], Longfei Qiu[c], Meikang Qiu[b,1], Hui Zhao[d]

[a] School of Computer Science and Technology, Shandong University, China
[b] Department of Computer Science, Pace University, New York City, NY 10038, USA
[c] Nanjing Foreign Language School, Jiangsu, China
[d] Software School, Henan University, Kaifeng, Henan, 475000, China

ARTICLE INFO

*Article history:*
Received 27 December 2015
Revised 2 September 2016
Accepted 3 September 2016
Available online xxx

*Keywords:*
Intelligent cryptography
Cybersecurity
Mass distributed storage
Cloud computing
Big data

ABSTRACT

Implementing cloud computing empowers numerous paths for Web-based service offerings to meet diverse needs. However, the data security and privacy has become a critical issue that restricts many cloud applications. One of the major concerns in security and privacy is caused by the fact that cloud operators have chances to reach the sensitive data. This concern dramatically increases users' anxiety and reduces the adoptability of cloud computing in many fields, such as the financial industry and governmental agencies. This paper focuses on this issue and proposes an intelligent cryptography approach, by which the cloud service operators cannot directly reach partial data. The proposed approach divides the file and separately stores the data in the distributed cloud servers. An alternative approach is designed to determine whether the data packets need a split in order to shorten the operation time. The proposed scheme is entitled *Security-Aware Efficient Distributed Storage* (SA-EDS) model, which is mainly supported by our proposed algorithms, including *Alternative Data Distribution (AD2) Algorithm*, *Secure Efficient Data Distributions (SED2) Algorithm* and *Efficient Data Conflation (EDCon) Algorithm*. Our experimental evaluations have assessed both security and efficiency performances and the experimental results depict that our approach can effectively defend main threats from clouds and requires with an acceptable computation time.

## 1. Introduction

As one of the significant technologies used in cloud computing, the distributed storage has enabled the mass remote data storage via *Storage-as-a-Service* (STaaS) service model. This cloud service model has broadly become an acceptable approach in big data along with the development of Web services and networks [9,20]. Many cloud vendors have given attractive storage service offerings that provide giant and scalable cloud-based storage spaces for users, such as Amazon, Dropbox, Google Drive, and Microsoft's OneDrive [14,19,28]. However, the security issue caused by the operations on cloud side is

[*] Corresponding author.
*E-mail addresses:* liyibing@sdu.edu.cn (Y. Li), kg71231w@pace.edu (K. Gai), longfeiqiu2012@gmail.com (L. Qiu), mqiu@pace.edu (M. Qiu), zhh@henu.edu.cn (H. Zhao).

[1] This work was supported in part by the National Science Foundation under Grands CNS-1457506 and NSF CNS-1359557 (Prof. M. Qiu). This work was also supported by the International Science and Technology Cooperation Program of China under Grant 2014DFR70730.

http://dx.doi.org/10.1016/j.ins.2016.09.005
0020-0255/© 2016 Elsevier Inc. All rights reserved.

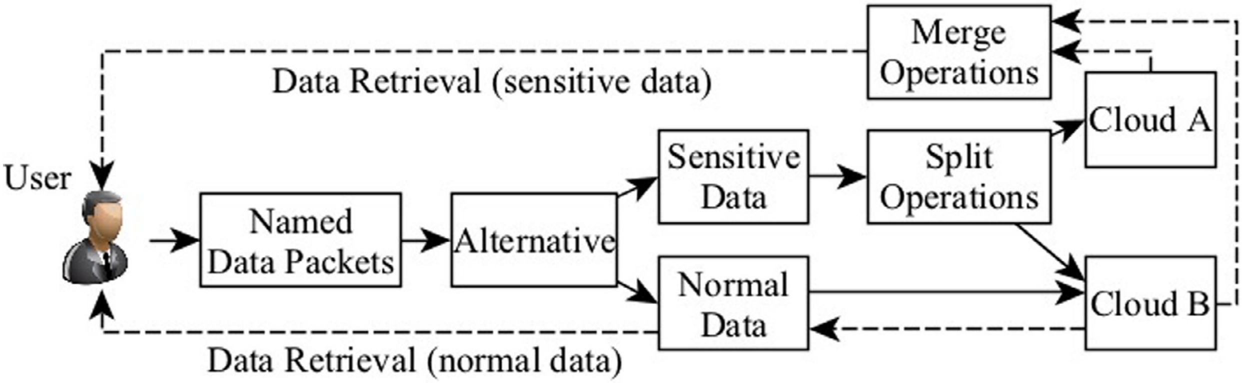

**Fig. 1.** The architecture of the proposed SA-EDS model.

still an obstacle of using STaaS for enterprises [2,11,13,15,55]. Many cloud users concern about their sensitive data to which the cloud operators have the access [17,54]. This matter embarrasses contemporary implementations of STaaS, even though many prior researches have addressed this field [26,32,40,43].

Moreover, *Mass Distributed Storage* (MDS) has been explored to scale up the data storage size in recent years [23,47]. The high level performances of the scalable computation are considered benefits of implementing MDS. One aspect that needs improvements is to secure distributed data storage [4], in which the threats come from a variety of sides. The distributed storage manner can result in more chances of malicious attacks or abuse activities [5,21], such as attack during data transmissions. Currently, the unexpected operations can also occur at the cloud server side, which are mainly constrained by laws and regulations. Meanwhile, it is difficult to balance functionality and security performances due to cost concerns [50]. Therefore, it is a challenging issue to efficiently secure distributed data in cloud systems, since the risks deriving from different network layers are hardly fully addressed [25,44].

This paper concentrates on the problem of cloud operators abuse issues and attempts to avoid cloud users' data release from cloud servers. We propose an intelligent cryptography approach, named *Security-Aware Efficient Distributed Storage* (SA-EDS) model that is designed to obtain an efficient MDS service, as well as high level security protections. Our proposed mechanism aims to encrypt all data and distributively store the data to the different cloud servers without causing big overheads and latency. Fig. 1 illustrates the architecture of SA-EDS model.

As shown in Fig. 1, user's data are assessed by an alternative process in which searchable named-data-packets techniques are applied. The solid arrow lines represent the data splits and storage operations. The broken arrow lines represent the operational directions of the data retrievals. Normal data will be assigned to a single cloud server. Meanwhile, the data with sensitive information are split into two parts that are assigned to two cloud servers, Cloud A and Cloud B. This process is mainly supported by our proposed algorithm, *Alternative Data Distribution* (AD2) algorithm. Moreover, splitting data process is accomplished by the main algorithm, *Secure Efficient Data Distributions (SED2) Algorithm*, which is designed to spilt data in order to prevent sensitive information from leaking on the cloud side using minimum costs. The sensitive data retrieval needs a decryption process that is supported by our proposed algorithm, *Efficient Data Conflation* (EDCon) algorithm.

The significance of the proposed mechanism is that we provide an adaptable approach for those enterprises that intend to use STaaS but require a high level data storage security, such as the financial service industry. The main problem solved by our proposed scheme is preventing cloud providers from directly reaching users' original data. The main contributions of this paper are twofold:

- We propose a novel cryptography approach for delivering mass distributed storage by which users' original data cannot be directly reached by cloud operators. The proposed method is an effectual cryptography means for defending malicious activities occurred on the cloud server.
- We propose an efficient data split mechanism that does not produce big overheads, as well as ensures data retrievability.

The remainder of this paper follows the structure given below. Recent related work is reviewed and summarized in Section 2. In addition, we represent a motivational example to exemplify the execution process in Section 3. Furthermore, the proposed model and the key concepts used in the model are given in Section 4. Next, Section 5 interprets the main algorithms by pseudo codes and algorithm descriptions. Moreover, we evaluate our proposed model in Section 6 via experimental demonstrations. Finally, Section 7 gives our conclusions.

## 2. Related work

This section reviews recent research achievements in cloud security issues, which supports the representation of our research background and the theoretical foundation. We addressed two aspects, including current security issues in cloud