



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 10    Issue: V    Month of publication: May 2022**

**DOI: <https://doi.org/10.22214/ijraset.2022.42427>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# StoreSim: Optimizing Information Leakage in Multi-Cloud Storage Services

Prashant Shankar Kapse<sup>1</sup>, Onkar Dhananjaya Swami<sup>2</sup>, Dadaso Keskar<sup>3</sup>, Raturaj Avinash Kadam<sup>4</sup>, Dr. S. N. Gujar<sup>5</sup>  
<sup>1, 2, 3, 4</sup>B.E. Students, <sup>5</sup>Project Guide, Department of Computer Engineering, TSSM's Bhivrabai Sawant College of Engineering & Research Narhe, Savitribai Phule Pune University.

**Abstract:** Many schemes have been recently advanced for storing data on multiple clouds. Distributing data over different cloud storage providers (CSPs) automatically provides users with a certain degree of information leakage control, as no single point of attack can leak all a user's information. However, unplanned distribution of data chunks can lead to high information disclosure even while using multiple clouds. In this paper, to address this problem, we present Store-Sim, an information leakage-aware storage system in multi cloud. Store-Sim aims to store syntactically similar data on the same cloud, thus minimizing the user's information leakage across multiple clouds. We design an approximate algorithm to efficiently generate similarity-preserving signatures for data chunks based on Min-Hash and Bloom filters, and design a function to compute the information leakage based on these signatures. Next, we present an effective storage plan generation algorithm based on clustering for distributing data chunks with minimal information leakage across multiple clouds. Finally, we evaluated our scheme using two real datasets from Wikipedia and GitHub. We show that our scheme can reduce information leakage by up to 60-70 Percent.

**Keywords:** Min-Hash, Bloom-Filter, Multi-Cloud, StoreSim, Data Leakage.

## I. INTRODUCTION

In cloud services, deduplication technology is commonly used to reduce the space and bandwidth requirements of services by eliminating redundant data and storing only a single copy. Deduplication is most effective when multiple users outsource the same data to cloud storage, but it raises issues relating to security and ownership. Proof-of ownership schemes allow any owner of the same data to prove to the cloud storage server that he owns the data in a robust way. However, if encrypted data were outsourced into cloud storage and ownership changes dynamically, deduplication would be hampered. Thus, we propose a secure deduplication scheme that supports dynamic ownership management based on randomized convergent encryption in this study.

## II. LITERATURE SURVEY

Sr. No.	Name	Authors	Disadvantages
1	Design of File Multi-Cloud Secure Storage System Based on Web and Erasure Code-2019.	Wei Shi, Tenglong Liu, and Min Huang.	In this paper the system uses Erasure Code to block the original data and encrypts the data blocks using AES (Advanced Encryption Standard) encryption technology, then stores the encrypted data blocks in different cloud storages terminals.
2	F2MC: Enhancing Data Storage Services with Fog-to Multi Cloud Hybrid Computing-2019.	Wei Shi, Tenglong Liu, and Min Huang.	This paper we introduce F2MC: a Fog-to-Multi Cloud hybrid storage service that combines local fog computing with remote cloud computing to enhance the quality of service (QoS) of data management.
3	Research on multi cloud dynamic secure storage technology-2020.	Jiahao Yao, Xiaoning Jiang	This paper proposes a method of slicing data blocks, encrypting them separately, and finally storing them on multiple clouds. In order to further improve the reliability of data storage, this paper proposes a multi cloud dynamic storage scheduling strategy and local storage optional configuration scheme
4	Optimizing Information Leakage in Multi cloud Storage Services-2015.	Hao Zhuang, Rameez Rahman, Pan Hui, Karl Aberer.	In this paper, to address the problem we present an information leakage aware storage system in multi cloud.

### III. PROBLEM STATEMENT

Dynamic ownership changes in a file-sharing group may occur very frequently in a practical cloud system. However, previous deduplication schemes could not achieve secure access control under a dynamic ownership environment. Therefore, for as long as revoked users keep the encryption key, they can access the corresponding data in cloud storage at any time, which is the problem we attempt to solve in this study. The proposed scheme has the following contributions. First, dynamic ownership management guarantees the backward and forward secrecy of deduplicated data upon any ownership change. Second, the proposed scheme ensures security in the setting of PoW by introducing a re-encryption mechanism that uses an additional group key for dynamic ownership groups. Thus, although the encryption key is revealed, the privacy of the outsourced data is still preserved against outside adversaries, while deduplication is still enabled.

### IV. PROPOSED SYSTEM

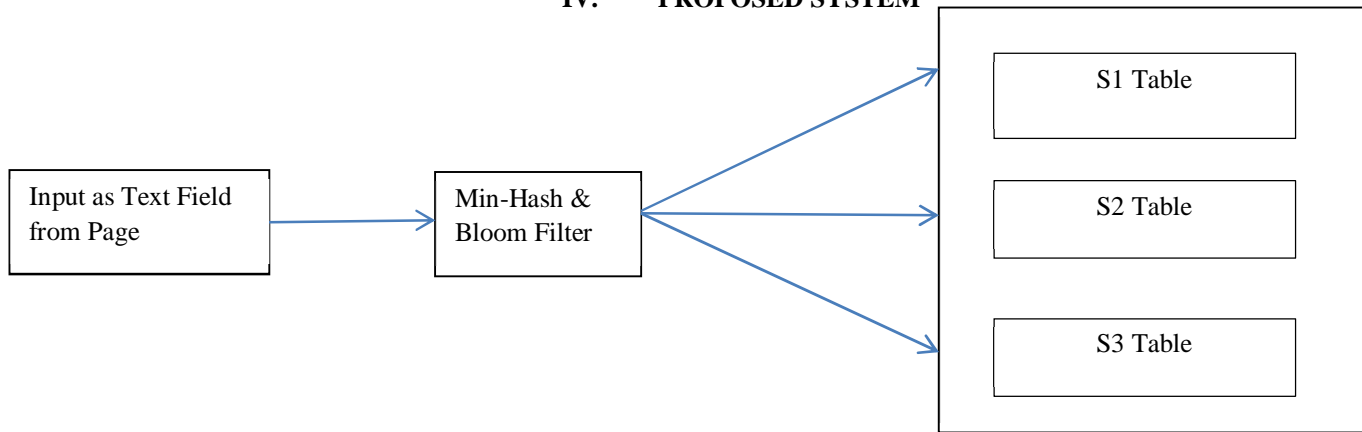


Fig. System Architecture

Whole Database

#### A. Module

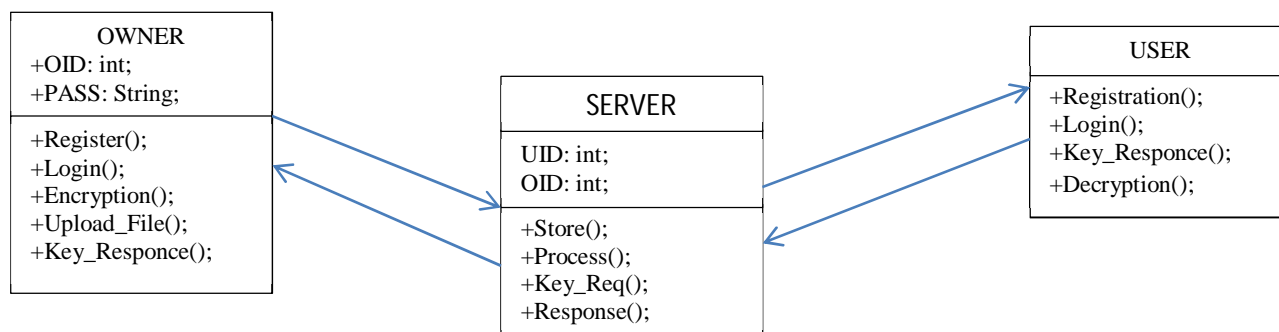
##### Admin

- 1) In this module, the Admin has to log in by using valid user name and password. After login successful he can do some operations such as View All Users and Authorize, View All E-Commerce Website and Authorize, View All Products and Reviews, View All Products Early Reviews, View All Keyword Search Details, View All Products Search Ratio, View All Keyword Search Results, View All Product Review Rank Results.
- 2) View and Authorize Users
- 3) In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, username, email, address and admin authorizes the users.
- 4) View Charts Results

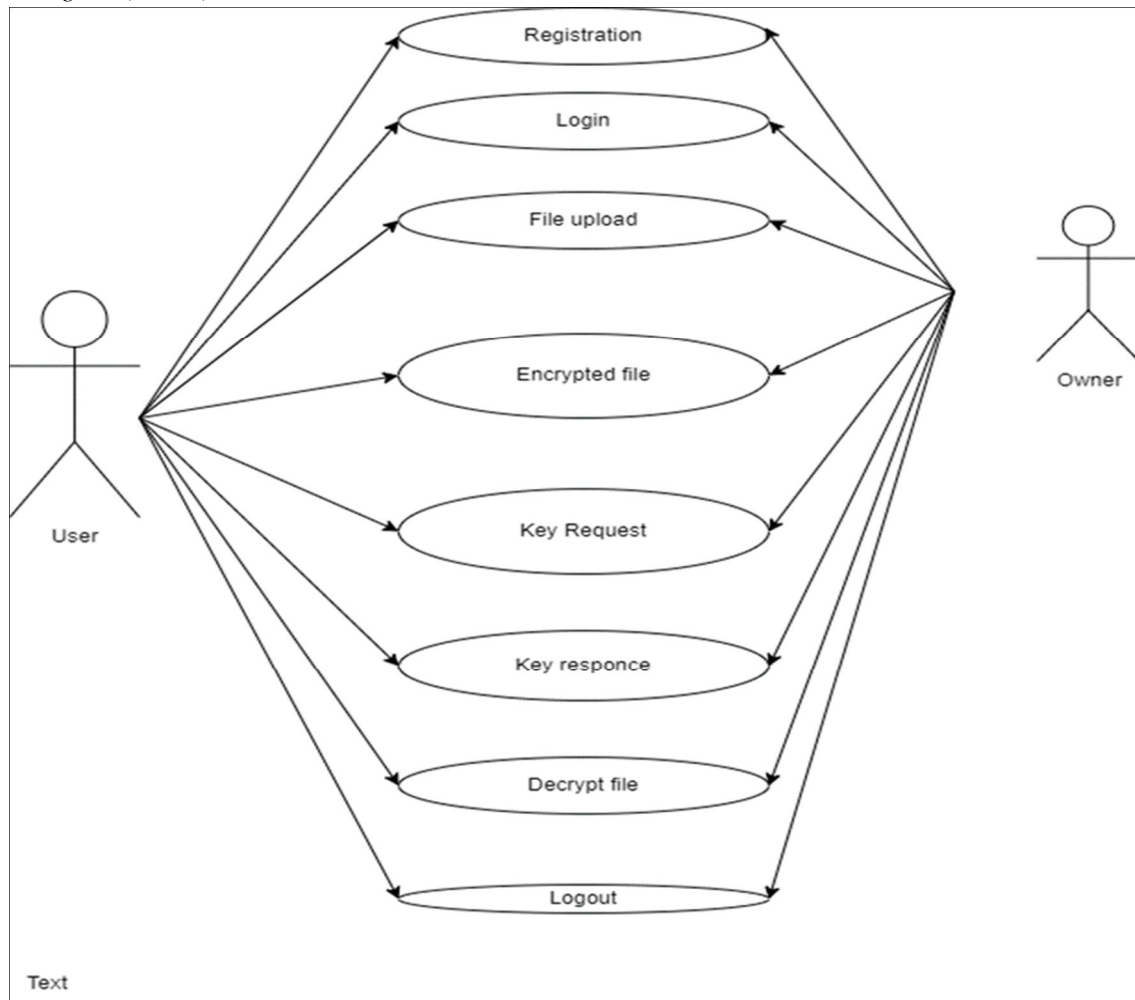
#### B. Data Flow Diagram

In Data Flow Diagram, we Show that flow of data in our system in DFD0 we show that base DFD in which rectangle present input as well as output and circle show our system, In DFD1 we show actual input and actual output of system input of our system is text and output is actual text file in DFD 2 we present operation of user as well as admin.

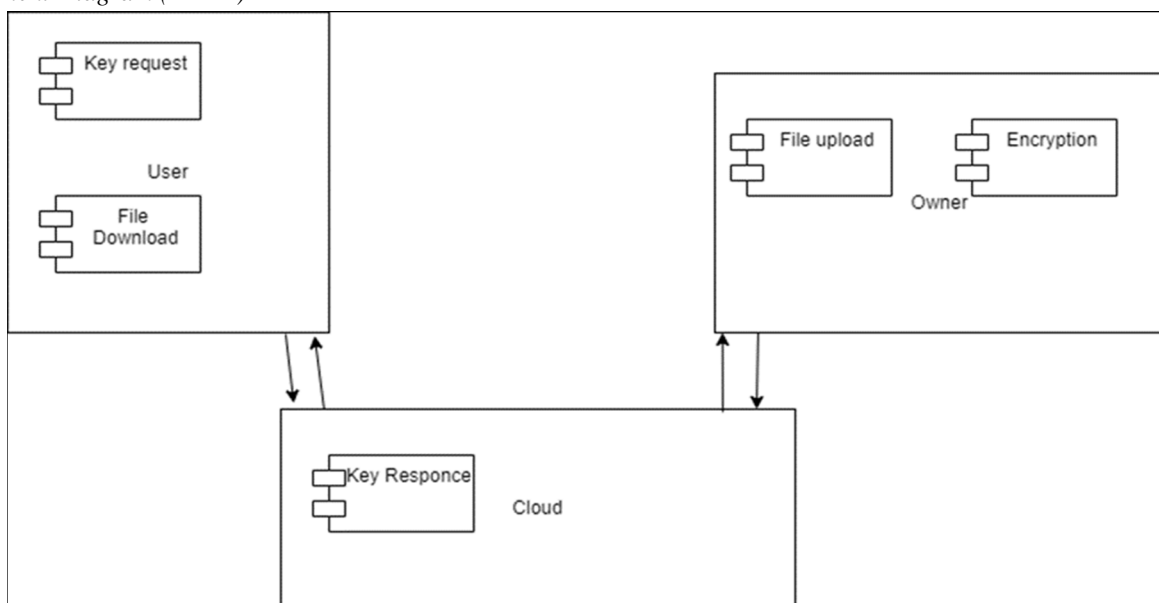
##### 1) Class Diagram (DFD0)



2) Use Case Diagram (DFD1)



3) Component Diagram (DFD2)



## V. SOFTWARE REQUIREMENT

Java can be used to create complete applications that may run on a single computer or be distributed among servers and clients in a network. It can also be used to build a small application module or applet (a simply designed, small application) for use as part of a Web page. There are two kinds of types in the Java programming language: primitive types (§4.2) and reference types (§4.3). There are, correspondingly, two kinds of data values that can be stored in variables, passed as arguments, returned by methods, and operated on: primitive values (§4.2) and reference values (§4.3). What exactly is Java? Java is an object-oriented programming language that produces software for multiple platforms. When a programmer writes a Java application, the compiled code (known as bytecode) runs on most operating systems (OS), including Windows, Linux and Mac OS. Java was developed in the mid-1990s by James Gosling.

The Java™ Programming Language is a general-purpose, concurrent, strongly typed, class-based object-oriented language. It is normally compiled to the byte code instruction set and binary format defined in the Java Virtual Machine Specification. A Java platform is a particular environment in which Java programming language applications run. There are several Java platforms. Many developers, even long time Java programming language developers, do not understand how the different platforms relate to each other.

## VI. CONCLUSIONS

Distributing data on multiple clouds provides users with a certain degree of information leakage control in that no single cloud provider is privy to all the user's data. However, unplanned distribution of data chunks can lead to avoidable information leakage. In this paper, we present StoreSim, an information leakage-aware storage system, to optimize information leakage in the multicloud environment. StoreSim achieves this goal by using novel algorithms, BFS Min-Hash and SP Clustering, which place the data with minimal information leakage (based on similarity) on the same cloud. Through an extensive evaluation based on two real datasets, we demonstrate that StoreSim is both effective and efficient (in terms of time and storage space) in minimizing information leakage during the process of synchronization in a multicloud environment.

## VII. FUTURE SCOPE

In our project, we show only whether the information is modified or not but not about where the information is modified and what information is modified. We get the notification through mail when information is modified. Include user password update option.

## VIII. ACKNOWLEDGMENT

It gives us great pleasure and satisfaction in presenting this paper on "Optimizing Information Leakage in Multicloud Storage Services". We are thankful to and fortunate enough to get constant encouragement, support and guidance from all Teaching staffs of [Computer Department] which helped us in successfully completing our project work. Also, We would like to extend our sincere esteem to all staff in laboratory for their timely support.

## REFERENCES

- [1] J. Crowcroft, "On the duality of resilience and privacy," in Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 471, p. 20140862, The Royal Society, 2015.
- [2] I. Drago, E. Bocchi, M. Mellia, H. Slatman, and A. Pras, "Benchmarking personal cloud storage," in Proceedings of the 2013 conference on Internet measurement conference, pp. 205–212, ACM, 2013.
- [3] Z. Wu, M. Butkiewicz, D. Perkins, E. Katz-Bassett, and H. V. Madhyastha, "Spanstore: Cost-effective geo-replicated storage spanning multiple cloud services," in Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles, pp. 292–308, ACM, 2013.
- [4] T. G. Papaioannou, N. Bonvin, and K. Aberer, "Scalia: an adaptive scheme for efficient multi-cloud storage," in Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, p. 20, IEEE Computer Society Press, 2012.
- [5] H. Chen, Y. Hu, P. Lee, and Y. Tang, "Nccloud: A network-coding-based storage system in a cloud-of-clouds," 2013.
- [6] A. Bessani, M. Correia, B. Quaresma, F. Andre, and P. Sousa, "Depsky: dependable and secure storage in a cloud-of-clouds," ACM Transactions on Storage (TOS), vol. 9, no. 4, p. 12, 2013.
- [7] "Prism surveillance program by nsa" [http://en.wikipedia.org/wiki/Edward\\_Snowden\\_Disclosure](http://en.wikipedia.org/wiki/Edward_Snowden_Disclosure).
- [8] P. Li and C. Konig, "b-bit min wise hashing," in Proceedings of the 19th international conference on World wide web, pp. 671–680, ACM, 2010.
- [9] M. Henzinger, "Finding near-duplicate web pages: a large-scale evaluation of algorithms," in Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 284–291, ACM, 2006.
- [10] T. Suel and N. Memon, "Algorithms for delta compression and remote file synchronization," 2002.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)