OPTIMIZING INFORMATION LEAKAGE IN MULTICLOUD SERVICES

K.Rambabu [1], B Naga Venkata Jyothi,

[1]**Assistant professor(HOD) , PG DEPT,** Dantuluri Narayana Raju College**, Bhimavaram, Andhrapradesh**
Email:- kattarambabudnr@gmail.com
[2]**PG Student of PG, Dantuluri** Narayana Raju College**, Bhimavaram, Andhrapradesh**
Email:- jyothiballa912@gmail.com

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, for no single point of attack can leak all the information. However, unplanned distribution of data chunks can lead to high information disclosure even while using multiple clouds. In this paper, we study an important information leakage problem caused by unplanned data distribution in multicloud storage services. Then, we present StoreSim, an information leakage aware storage system in multicloud. StoreSim 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 MinHash and Bloom filter, and also 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.

1 INTRODUCTION

With the increasingly rapid uptake of devices such as laptops, cell phones and tablets, users require ubiquitous and massive network storage to handle their ever-growing digital lives. To meet these demands, many cloud-based storage and file sharing services such as Dropbox, Google Drive and Amazon S3, have gained popularity due to the easy-to-use interface and low storage cost. However, these centralized cloud storage services are criticized for grabbing the control of users' data, which allows storage providers to run analytics for marketing and advertising . Also, the information in users' data can be leaked e.g., by means of malicious insiders, backdoors, bribe and coercion. One possible solution to reduce the risk of information leakage is to employ multicloud storage systems in which no single point of attack can leak all the information. A malicious entity, such as the one revealed in recent attacks on privacy, would be required to coerce all the different CSPs on which a user might place her data, in order to get a complete picture of her data. Put simply, as the saying goes, do not put all the eggs in one basket.

Literature Survey

Depsky: dependable and secure storage in a cloud-of-clouds

The increasing popularity of cloud storage services has lead companies that handle critical data to think about using these services for their storage needs. Medical record databases, large biomedical datasets, historical information about power systems and financial data are some examples of critical data that could be moved to the cloud. However, the reliability and security of data stored in the cloud still remain major concerns. In this work we present DepSky, a system that improves the availability, integrity, and confidentiality of information stored in the cloud through the encryption, encoding, and replication of the data on diverse clouds that form a cloud-of-clouds. We deployed our system using four commercial clouds and used PlanetLab to run clients accessing the service from different countries. We observed that our protocols improved the perceived availability, and in most cases, the access latency, when compared with cloud providers individually. Moreover, the monetary costs of using DepSky in this scenario is at most twice the cost of using a single cloud, which is optimal and seems to be a reasonable cost, given the benefits.

3 IMPLEMENTATION STUDY

EXISTING SYSTEM:

In fact, the data deduplication technique, which is widely adopted by current cloud storage services in existing clouds, is one example of exploiting the similarities among different data chunks to save disk space and avoid data retransmission . It identifies the same data chunks by their fingerprints which are generated by fingerprinting algorithms such as SHA-1, MD5. Any change to the data will produce a very different fingerprint with high probability . However, these fingerprints can only detect whether or not the data nodes are duplicate, which is only good for exact equality testing. Determining identical chunks is relatively straightforward but efficiently determining similarity between chunks is an intricate task due to the lack of similarity preserving fingerprints (or signatures).

Disadvantages:

- Unplanned distribution of data chunks can lead to high information disclosure even while using multiple clouds.
- Frequent modifications of files by users result in large amount of similar chunks1;

Proposed System & alogirtham

We present StoreSim, an information leakage aware multicloud storage system which incorporates three important distributed entities and we also formulate information leakage optimization problem in multicloud.

4.1 Advantages:

- However, previous works employed only a single cloud which has both compute and storage capacity. Our work is different since we consider a mutli cloud in which each storage cloud is only served as storage without the ability to compute.
- Our work is not alone in storing data with the adoption of multiple CSPs these work focused on different issues such as cost optimization , data consistency and availability.

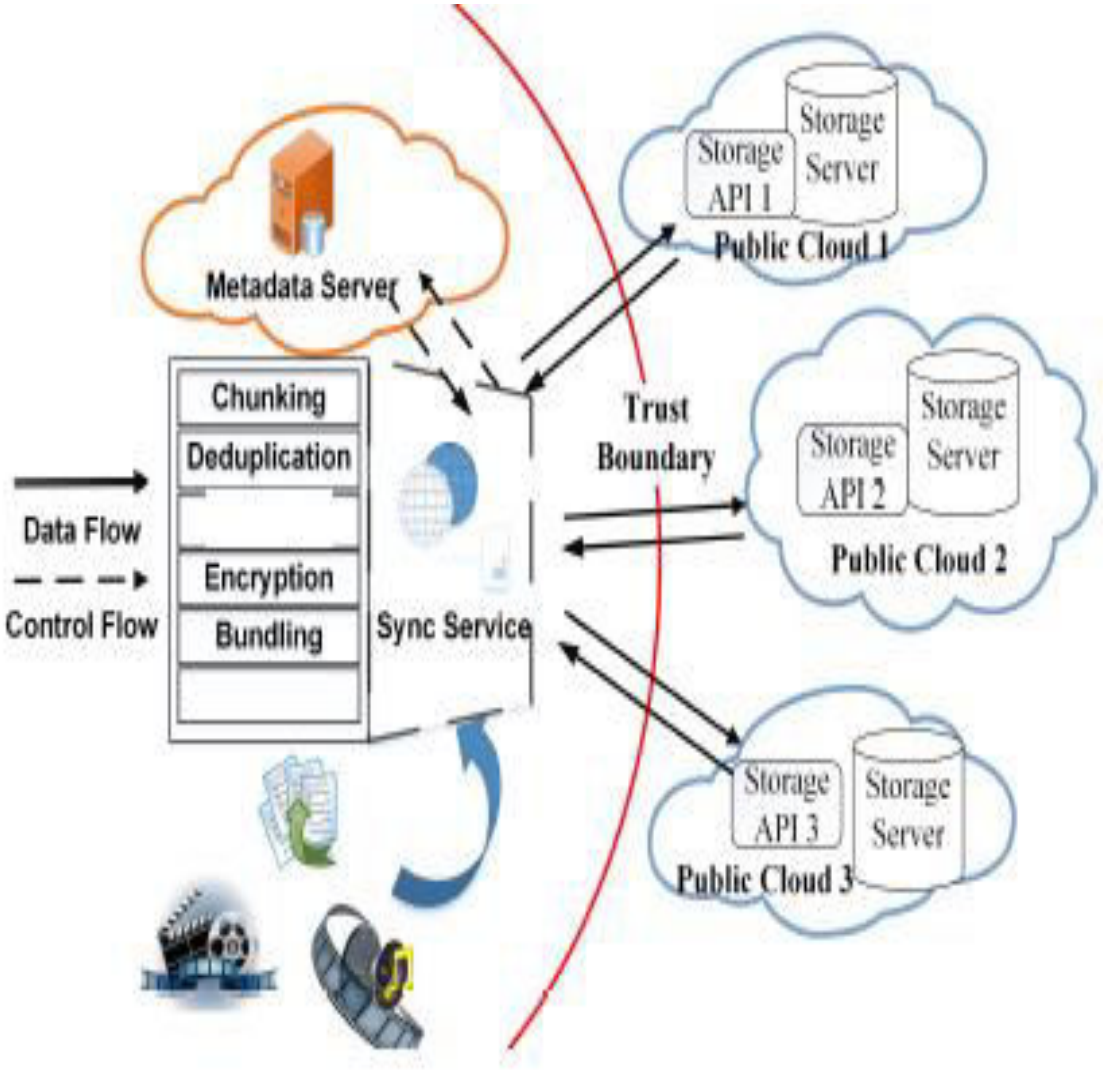


Fig:3.1 System Architecture

IMPLEMENTATION

MODULES

1. DATA OWNER
2. METADATA SERVER
3. CLOUD SERVICE PROVIDER
4. DATA USER

1.DATA OWNER:

- In this module, we develop the Customer features functionalities. Customer first register his/her details and login. Customer can outsource sensitive and valuable data to cloud by encrypting data and splitting data in to multiple parts.
- Data owner has option to modify data which is uploaded to cloud. In this process when user updates data stored in cloud1 with data which is already available in cloud2 then total data will be visible in cloud1 only. In order to solve this problem owner will check data similarity using minhash and data matching percentage is calculated and refer to user where to upload data.

2.METADATA SERVER:

Metadata servers are used to store the metadata database about the information of files, CSPs and users,which usually are structured data representing the whole cloud file system.

3.CLOUD SERVICE PROVIDER:

- In this module, we design the Cloud functionalities. The Cloud entity can view all customer details, file upload details and customer file download details. In this module, we use the DriveHQ Cloud Service API for the Cloud Integration and develop the project.

5 RESULTS AND DISCUSSION

HOME PAGE

Fig 5.1: home page

CLIENT REGISTRARION

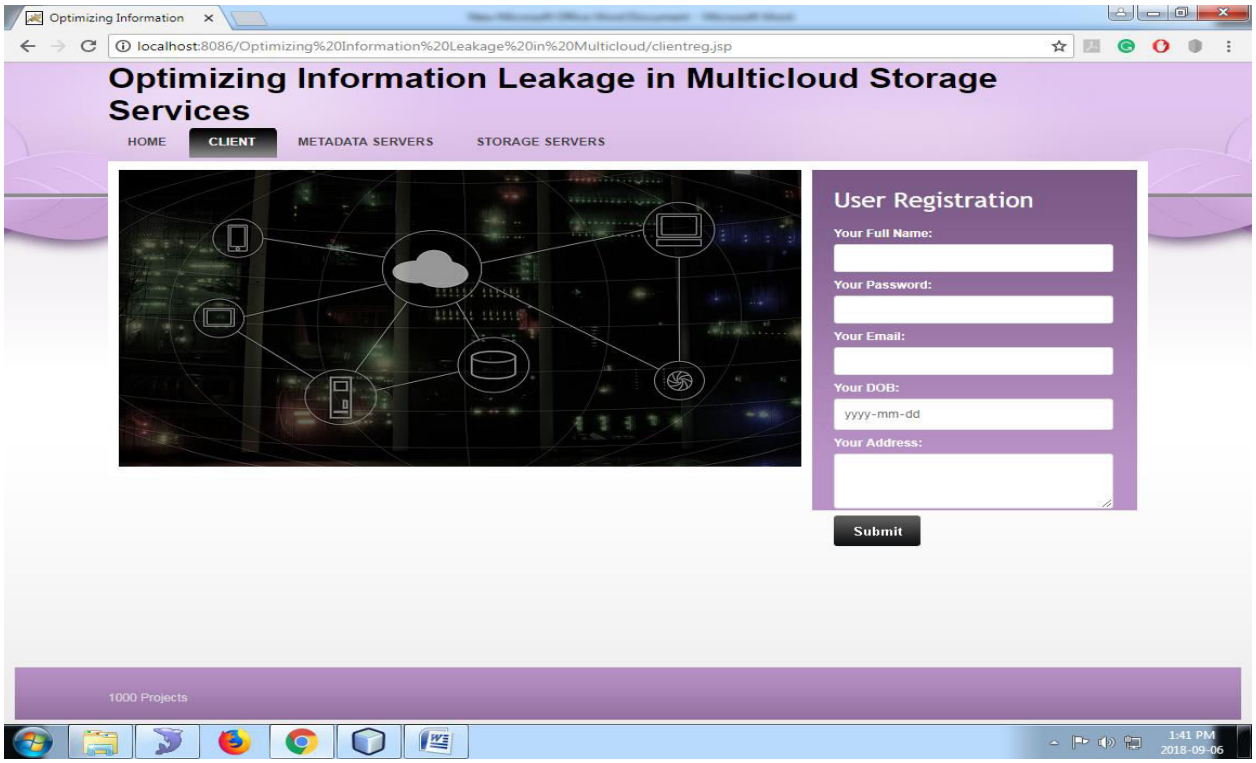


Fig 5.2: Client Registration

CLIENT LOGIN

Fig : client login

CLIENT HOME

Fig : client home

UPLOAD

FILES

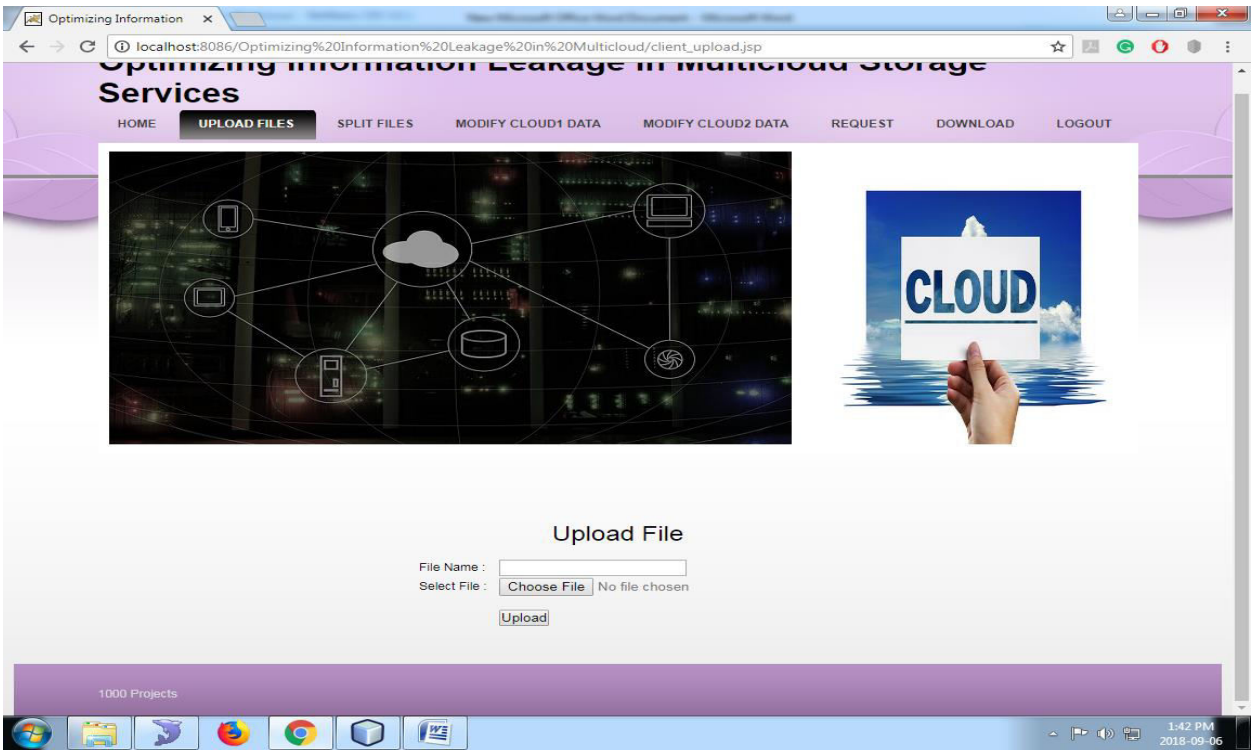


Fig5.3 : Upload files

ENCRYPT DATA

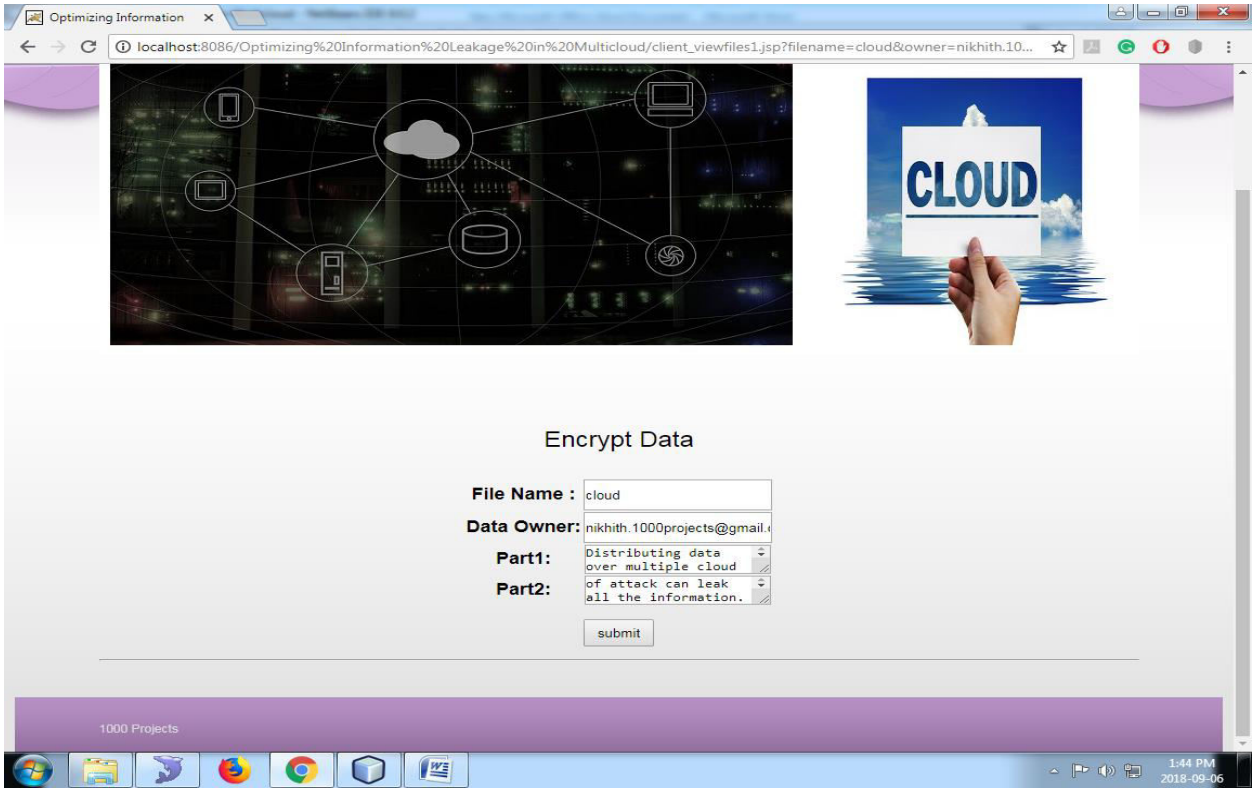
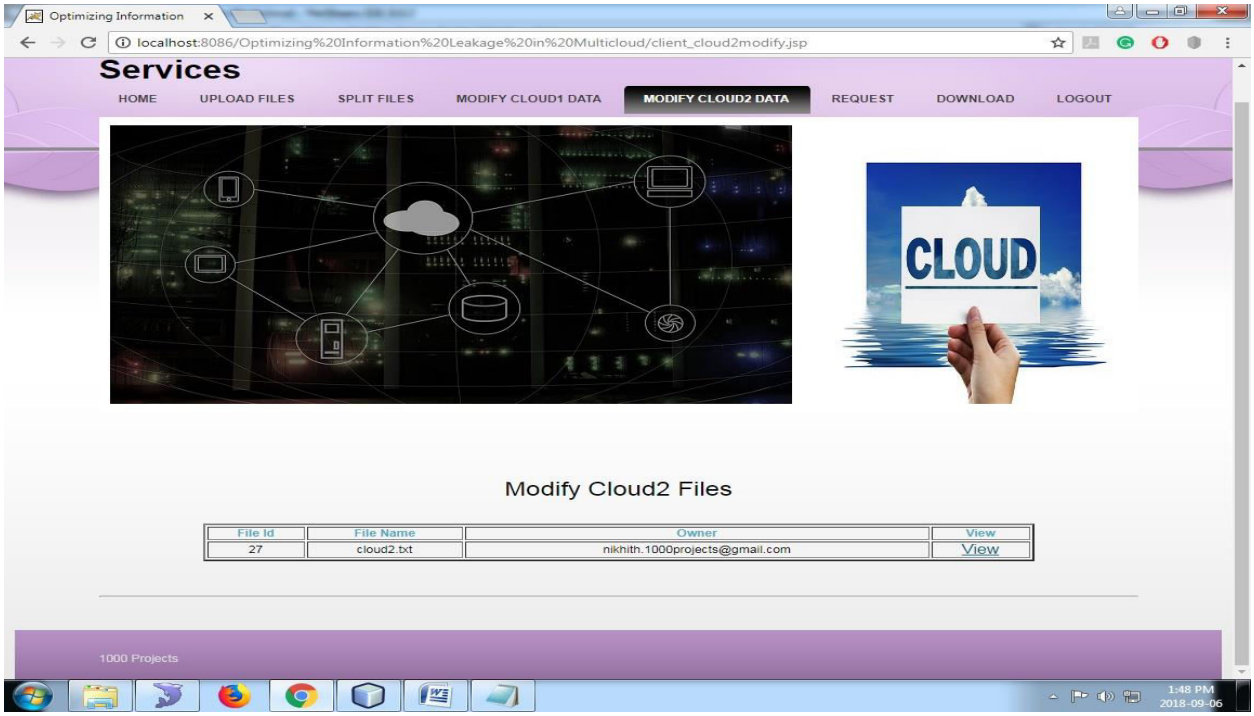


Fig5.4: encrypt data

MODIFY FILES



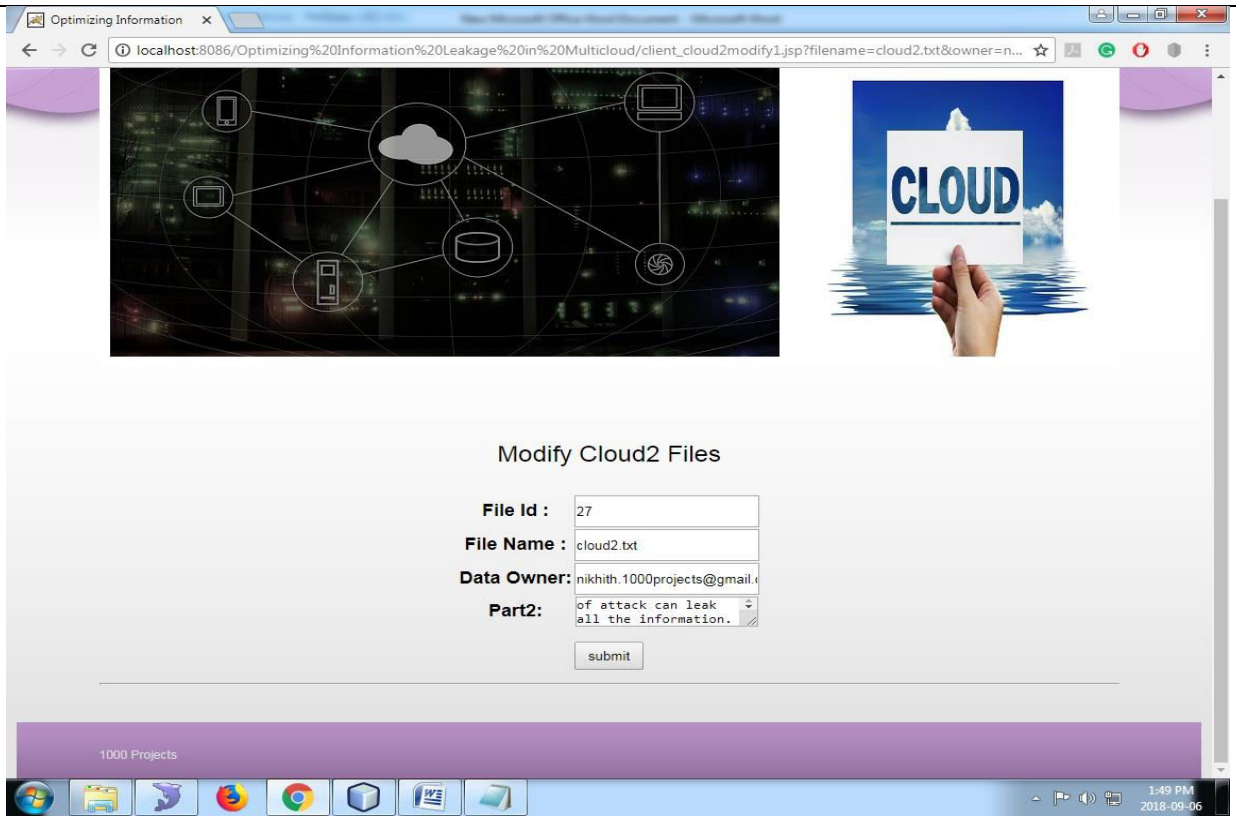


Fig 5.5 modify files

VIEW FILES & REQUESTS

Fig5.6: view files & requests

DOWNLOAD FILES

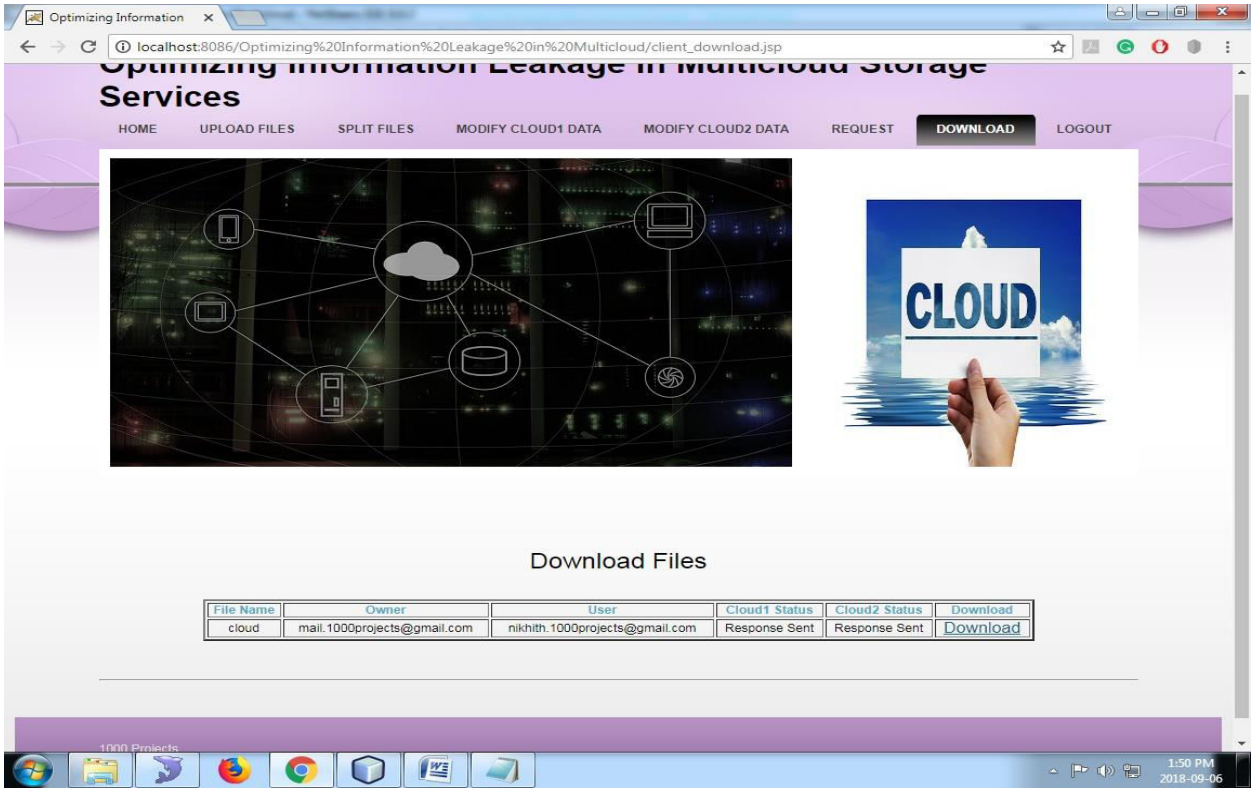


Fig: download files

VIEW SECURITY KEY

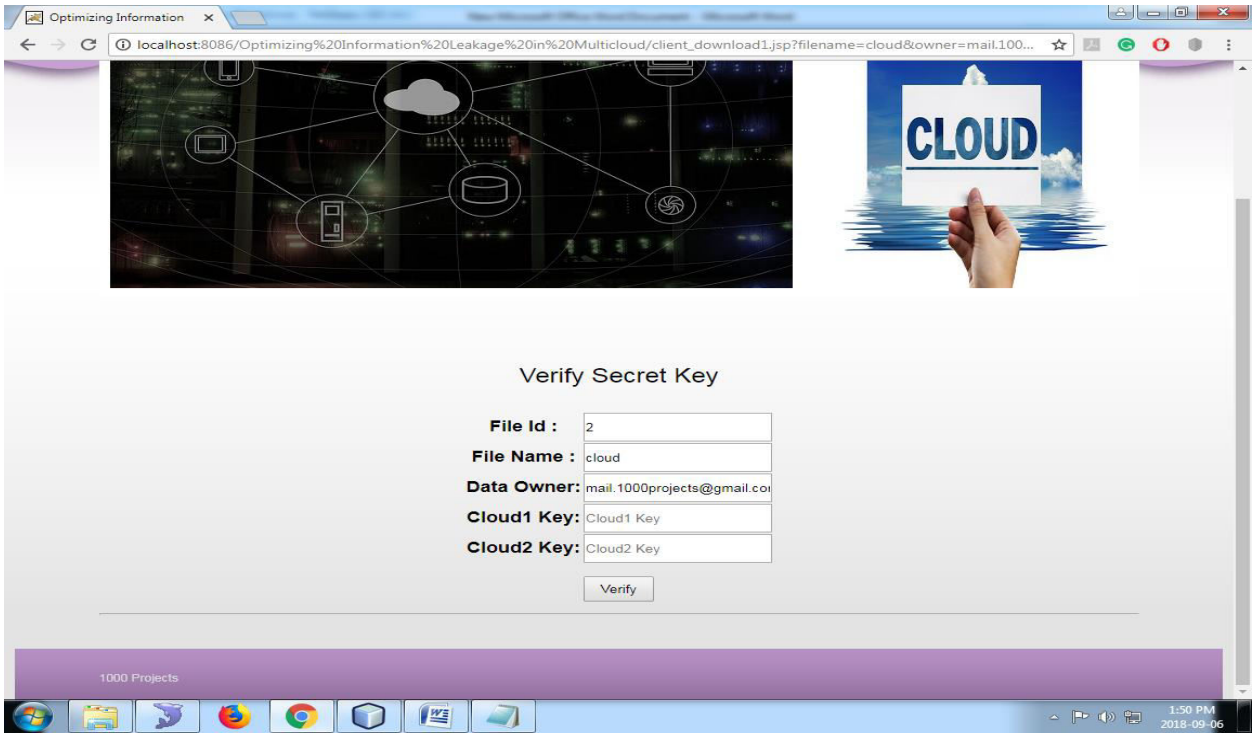


Fig5.7: view security key

DOWNLOAD CLOUD1 & CLOUD2 DATA

Fig5.8 : download cloud1 & cloud2 data

METADATA SERVERS LOGIN

Fig5.9 : metadata servers login

VIEW FILES

Fig5.10:view files

VIEW CLIENT REQUESTS

Fig5.11: view client requests

6. CONCLUSION AND FUTURE WORK

CONCLUSION

Distributing data on multiple clouds provides users with a certain degree of information leakage control in that no single cloud provider is privy to the entire user's data. However, unplanned distribution of data chunks can lead to avoidable information leakage. We show that distributing data chunks in a round robin way can leak user's data as high as 80% of the total information with the increase in the number of data synchronization. To optimize the information leakage, we presented the StoreSim, an information leakage aware storage system in the multicloud. Store Sim achieves this goal by using novel algorithms, BFSMinHash and SPClustering, 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 multicloud. We show that our StoreSim can achieve near-optimal performance and reduce information leakage up to 60% compared to unplanned placement. Finally, through our attackability analysis, we further demonstrate that StoreSim not only reduces the risk of wholesale information leakage but also makes attacks on retail information much more complex.

7. REFRENCES

1. . Crowcroft, “On the duality of resilience and privacy,” in Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 471, no. 2175. The Royal Society, 2015, p. 20140862.
2. Bessani, M. Correia, B. Quaresma, F. Andr´e, 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.
3. H. Chen, Y. Hu, P. Lee, and Y. Tang, “Nccloud: A network-coding-based storage system in a cloud-of-clouds,” 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. IEEE Computer Society Press, 2012, p. 20.
5. 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. ACM, 2013, pp. 292–308.
6. G. Greenwald and E. MacAskill, “Nsa prism program taps in to user data of apple, google and others,” The Guardian, vol. 7, no. 6, pp. 1–43, 2013.
7. T. Suel and N. Memon, “Algorithms for delta compression and remote file synchronization,” 2002.
8. [8] I. Drago, E. Bocchi, M. Mellia, H. Slatman, and A. Pras, “Benchmarking personal cloud storage,” in Proceedings of the 2013 conference on Internet
9. measurement conference. ACM, 2013, pp. 205–212.
10. [9] I. Drago, M. Mellia, M.MMunafo, A. Sperotto, R. Sadre, and A. Pras, “Inside dropbox: understanding personal cloud storage services,” in Proceedings of the 2012 ACM conference on Internet measurement conference. ACM, 2012, pp. 481–494.
11. [10] U. Manber et al., “Finding similar files in a large file system.” in Usenix Winter, vol. 94, 1994, pp. 1–10.
12. [11] P. Mahajan, S. Setty, S. Lee, A. Clement, L. Alvisi, M. Dahlin, and M.Walfish, “Depot: Cloud storage with minimal trust,” ACM Transactions on Computer Systems (TOCS), vol. 29, no. 4, p. 12, 2011.
13. [12] A. J. Feldman, W. P. Zeller, M. J. Freedman, and E. W. Felten, “Sporc: Group collaboration using untrusted cloud resources.” in OSDI, vol. 10, 2010, pp. 337–350.