Revamping Optimal Cloud Storage System

Shraddha Ghogare, Ambika Pawar and Ajay Dani

Abstract With the prosperous growth of cloud computing, it is being used widely in business as well as in researches today, considering its numerous advantages over traditional approaches. However, security and privacy concerns are ascending day by days. Cloud is being utilized for not only to use software and platform over the Internet, but also for storing confidential data. This paper presents a novel approach wherein multi-cloud environment is used to mitigate data privacy apprehensions. The viability and design of this method are described in the proposed paper.

Keywords Cloud computing · Cloud storage · Flat files · Privacy · Multi-cloud environment

1 Introduction

In an always changing economic and business world, businesses strive to achieve more profit and enhance future growth. This can be effectuated by using cloud computing, wherein businesses can perform computing on the cloud. The cloud is essentially a pool of hardware, network, software resources, storage, interfaces and services that leads to delivery of computing as a service. Figure 1 shows services offered by cloud.

S. Ghogare (✉) · A. Pawar
CSE Department, SIT, Symbiosis International University (SIU), Pune, India
e-mail: shraddha.ghogare@sitpune.edu.in

A. Pawar
e-mail: ambikap@sitpune.edu.in

A. Dani
Indian Business School, Hadapsar, Pune, India
e-mail: ardani_123@rediffmail.com

A. Nagar et al. (eds.), *Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics*, Smart Innovation, Systems and Technologies 44, DOI 10.1007/978-81-322-2529-4_48

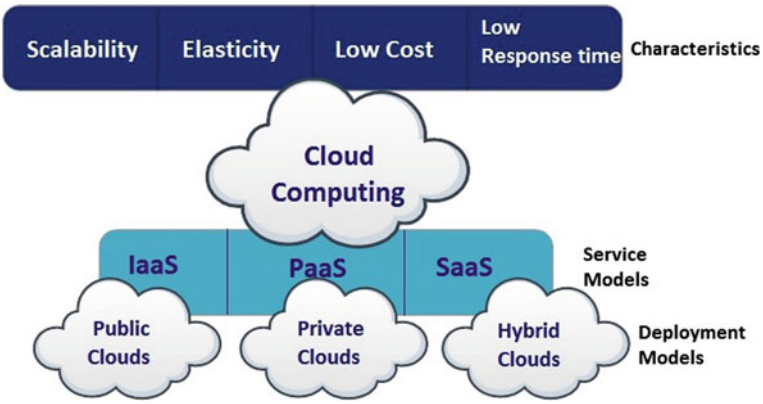

Fig. 1 Cloud computing: an overview

Cloud computing has several advantages over traditional computing systems such as on-demand service, elasticity, scalability and reduced response time. Even so, there exist certain concerns related to security of data. Many researchers and professionals have focused on security of the data deployed on cloud. However, data privacy has not gained much attention. These data privacy issues on service provider's sides are handled using SLAs that are agreed by end users and service provider in the negotiation phase. Furthermore, monitors are used to keep a check up on insiders. Privacy issues on a client's application server's side still persist.

Users are skeptical about their data stored on cloud data storage being mined or used for advertising and marketing purpose. Existing cryptographic techniques can be used for data security; but protecting data privacy needs significant focus. Who should have how much access right is another challenge to data security. In Google's approach to IT Security [1], Google's security strategy that provides control at multiple levels of data storage, transfer and access have been addressed. Security concerns have been addressed precisely; nevertheless, data privacy concerns still need to be stressed. Another approach to solve data privacy issues is to have a role-based access control wherein permissions and users are assigned to roles. Any authenticated user may misuse the access right assigned thereby compromising data privacy.

Moreover, in a cloud based environment, the data are stacked away at several geographic locations. Laws and Regulations vary for each country that defines the way confidential information is collected and stored and the access rights for that data. Therefore, protecting private data on public cloud is a major concern. Use of several clouds solves this problem as files are divided into parts and are uploaded on multiple clouds. This paper extends this approach. It provides a method in that at files are divided into chunks and are uploaded on different cloud service providers (CSPs).

2 Related Work

Maintaining privacy of data on public clouds is a core concern. Once a user deploys his data-processing applications on the cloud, the cloud provider is granted full control over the processes. Hence, trust and governance are two major issues. [2, 3] suggested an approach wherein the multi-cloud concept had been proposed, which includes three patterns: replication of application, partition of an application system into layers and small partitions of the application tier into fragments. All aforementioned approaches have been implemented in a multi-cloud environment. Moreover, assessment and comparison of this approach in terms of security, feasibility and regulation have been presented in [4]. However, switching time from one cloud to another and cost of each cloud are two important factors to be addressed to [5].

Many organizations now have moved to cloud storage to store data. These providers either use it to store archive data (passive data) or to store data on the social network that is sensitive. Although there are many advantages of using cloud storage, data privacy and losing control over data are major issues [6]. [7] proposes that the user would trust small part of memory on their system wherein they would encrypt the data and then store on the cloud. The pitfall is that not all users are aware of how to encrypt or secure this little part. However, this would be beneficial rather than uploading plain text.

[8] have proposed a mechanism which addresses Privacy concerns in a cloud environment by suggesting an optimal cloud storage management system wherein multiple cloud storages are used to store the data. The application developed offers the user to select cloud storage(s) out of 67 providers who have been analysed. The feature of this application is that the user can customize sign-up privacy, throughput, response time, capacity. Furthermore, the user has an option to select storage strategy as UseAllInParallel or RoundRobin. However, if the user is not aware of what these techniques are, he might keep the default one as selected. Also, the file splitter does not run concurrently as of now, which may hamper application's overall performance and file splitting operation does not consider a file type so this may incur extra processing overhead on the application server. Users have to sign-up manually to various cloud storage providers at their website; no single sign-on in presence as yet.

It has also been depicted how cloud storage services take no liability of any loss of data or corruption and declare rights to read, write, modify or even delete data. Furthermore, they can sell the data for analysis purpose. [9] proposes a technique by introducing an additional layer of security has been suggested termed as overlay structure here. Cost and switching times are two issues user may face with an increased use of such models. Moreover, performance may degrade due to introduction of an additional layer (overlay structure). Key management issues are dealt well in this paper.

Some governing bodies still continue to use flat-files to store their information [10]. In addition, users upload their data to the cloud in the form of files. Even

though files are being used widely with growing usage of cloud storage, there has been little work done to provide privacy to files. This scheme can enhance the privacy of such files.

In order to upload files to multiple clouds, file splitting is called for. One of the methods used to develop this scheme is: split the file based on its size in bytes. IDA another algorithm which is inculcated for splitting the file in this model [11, 12]. This system uses IDA algorithm to generate erasure codes by applying either Reed-Solomon or Rabin's IDA wherein simple matrix multiplication of bytes of file and coefficient of a matrix which are computed by using a prime number or Fermat's number system is done [13]. In future, file splitting will be done based on file type, various methods as discussed in [14] will be utilized.

3 Our Approach

This paper considers all the pitfalls of existing methods and brings out a proficiency to enhance privacy protection of cloud storage by using file splitter, encryption and multiple cloud storage.

3.1 Proposed Design

There have been tremendous efforts taken to protect data uploaded on a cloud. The approaches discussed and implemented earlier were primarily engineered for exclusive cloud environments wherein data on singular cloud was provided with enhanced privacy protection by applying certain models. Figure 2 shows how users can upload their private files to the cloud by using client application server with the help of privacy algorithms.

Although this approach may solve some of the concerns related to data privacy, it may fall through in cases where cloud itself goes wrong or is attacked. Moreover, threat of insiders still needs to be addressed effectually.

Considering aforementioned issues, this model makes use of multiple cloud environments for storing the data. Figure 2 shows the flow of operations.

The detailed description of the system is as given below:

As illustrated in Fig. 3, when the user selects a file to be uploaded, it first is divided into m chunks by using a simple file splitter. After this, IDA (Information Dispersal Algorithm) is applied on every chuck of the original file to ensure availability and integrity. Moreover, to achieve confidentiality, encryption algorithm is applied on these n fragments of chunks. And then, these encrypted fragments of chunks are uploaded on multiple clouds. To begin with, the project is using Dropbox accounts to store these fragments. On the other hand, to download a file user has uploaded via this system, first decryption is done, after that out of n

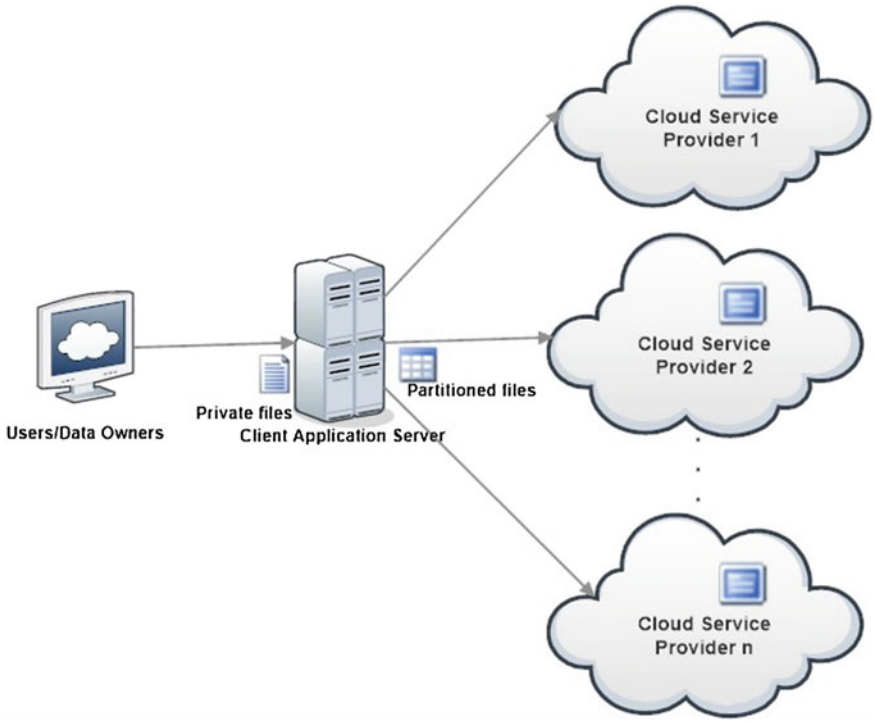

Fig. 2 Privacy protection with multiple clouds

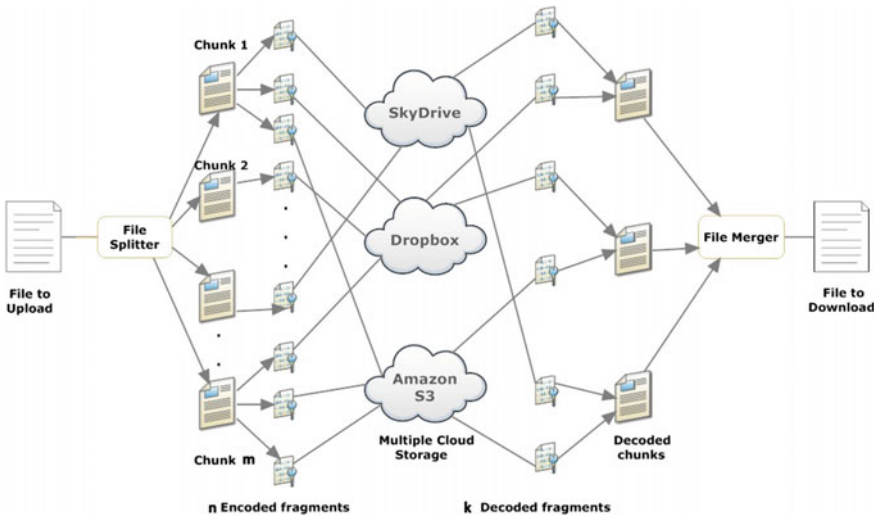

Fig. 3 Working principle of the system

fragments of m chunks, k fragments are retrieved from multiple clouds, and file is subsequently merged by using a simple merger program.

This entire process has to be done on byte stream to make the system independent of the file format. Furthermore, to improve response time, parallel processing of applying IDA on chunks can be applied. In future, the system shall be developed for accomplishing these things.

3.2 Results

As of now, the aforementioned system has been developed for uploading text files wherein, a simple file splitter-merger, erasure codes' generator, and database are used. AES algorithm is applied on the partitioned fragments of the chunks optionally. Moreover, users are not asked to sign up on various cloud storage providers as this would add an extra overhead on them. So, this model utilizes admin accounts for storing users' data on clouds.

The results delineated in Fig. 4 indicate that as the file size grows, the time taken to upload and download a file grows subsequently. However, as the application server does splitting, encoding and encryption; security, privacy, availability and integrity of the files being uploaded are substantially higher. In addition, as of now the system is developed for only text files (as splitting is done on a character array). So, in time to come, the splitting will be done on bytes' array; therefore, time taken will be less and the whole process will be file format independent.

There is no single-point failure here, as IDA takes care of it [12, 13]. Also, more than one user can use this system simultaneously.

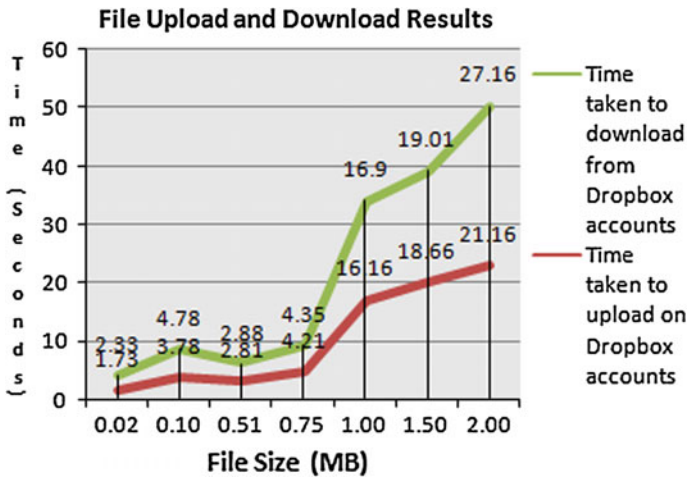

Fig. 4 Results for text files

3.3 Conclusion

This paper addressed core concerns regarding privacy of data stored on the cloud by applying multi-cloud environment and encryption. This approach would reduce privacy issues in cloud. Furthermore, as this model would be developed as a web-based application, number of users can use it simultaneously for storing the data in cloud storage. In the future, the utilization of parallel algorithms or Hadoop map-reduce has been planned in order to reduce processing overhead from the application server and time needed to store/retrieve data on the cloud. Moreover, for partitioning files into pieces, dynamic selection of the algorithm according to file format will be done.

References

1. Kincaid, J.: Google privacy blunder shares your docs without permission. In: TechCrunch (2009)
2. Bohli, J.M., Jensen, M., Gruschka, N., Schwenk, J., Iacono, L.L.L.: Security prospects through cloud computing by adopting multiple clouds. In: Proceedings of IEEE Fourth International Conference on Cloud Computing (CLOUD) (2011)
3. Iacono, L.L., Marnau, N.: Security and privacy-enhancing multicloud architectures
4. Hubbard, D., Sutton, M.: Top threats to cloud computing v1. 0. Cloud Security Alliance (2010)
5. Kumar, G., Shrivastava, N.: A survey on cost effective multi-cloud storage in cloud computing
6. Yao, A.C.: Protocols for secure computations. In: Proceedings of the 23rd Annual IEEE Symposium on Foundations of Computer Science, FOCS82, p. 160164 (1982)
7. Cachin, C., Keidar, I., Shraer, A.: Trusting the cloud. ACM Sigact. News **40**(2), 81–86 (2009)
8. Muller, J., Spillner, J., Schill, A.: Creating optimal cloud storage systems. In: Future Generation Computer Systems, p. 10621072 (2013)
9. Svenn, M.: Secure data management for cloud-based storage solutions. Cloud-Based Softw. Eng. 59, 2013
10. Chickowski, E.: Flat-file databases often overlooked in security schemes (2010). http://www.darkreading.com/risk/flat-file-databases-often-overlooked-in-security-schemes/d/d-id/1133363?
11. Béguin, P., Cresti, A.: General information dispersal algorithms. Theoret. Comput. Sci. **209** (1), 87–105 (1998)
12. Rabin, M.O.: Efficient dispersal of information for security, load balancing, and fault tolerance. J. ACM (JACM) 36(2), 335–348 (1989)
13. Lin, S.-J., Chung, W.-H.: An efficient (n, k) information dispersal algorithm for high code rate system over fermat fields. IEEE Commun. Lett. 16(12), 2036–2039 (2012)
14. Bian, J.: JIGDFS: A secure distributed file system and its applications. PhD thesis, University of Arkansas (2007)