Avoid Duplicate Entries of Repeating Data in Hybrid Cloud Storage Using Convergent Encryption Techniques

K. Keerthika[1]
Assistant Professor,
Department of CSE, Selvam College of Technology,
Namakkal, India

J. Sagayaraja[3]
UG Student,
Department of CSE, Selvam College of Technology,
Namakkal, India

G.Manikandan[2]
UG Student,
Department of CSE, Selvam College of Technology,
Namakkal, India

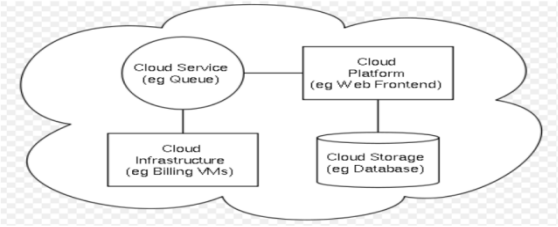
S. Vinoth[4]
UG Student,
Department of CSE, Selvam College of Technology,
Namakkal, India

Abstract:- **In cloud computing the Cloud storage is a model in which data is stored on remote server accessed from the internet or cloud. the cloud storage maintained and manages huge amount of data of multiple users that are built using virtualization techniques stored data includes different file formats such as text, image, video, audio etc. Nowadays the cloud service provider facing a big issue of data duplication on cloud storage. Data deduplication is a technique for reducing the amount of storage space an organization needs to save its data. In most organizations, the storage systems contain duplicate copies of many pieces of data. For example, the same file may be saved in several different places by different users, or two or more files that aren't identical may still include much of the same data. Deduplication eliminates these extra copies by saving just one copy of the data and replacing the other copies with pointers that lead back to the original copy. Companies frequently use deduplication in backup and disaster recovery applications, but it can be used to free up space in primary storage as well. To avoid this duplication of data and to maintain the confidentiality in the cloud we using the concept of Hybrid cloud. To protect the confidentiality of sensitive data while supporting deduplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, this paper makes the first attempt to formally address the problem of authorized data deduplication.**

Keywords: - De-duplication, Cloud Computing, Virtualized and so on.

I. INTRODUCTION

With the potentially infinite storage space offered by cloud providers, users tend to use as much space as they can and vendors constantly look for techniques aimed to minimize redundant data and maximize space savings. A technique which has been widely adopted is cross-user de-duplication. The simple idea behind de-duplication is to store duplicate data (either files or blocks) only once. Therefore, if a user wants to upload a file (block) which is already stored, the cloud provider will add the user to the owner list of that file (block). De-duplication has proved to achieve high space and cost savings and many cloud storage providers are currently adopting it. De-duplication can reduce storage needs by up to 90-95% for backup applications and up to 68% in standard file systems.



Architecture of Cloud Computing

Cloud computing provides seemingly unlimited "virtualized" resources to users as services across the whole Internet, while hiding platform and implementation details. Today's cloud service providers offer both highly available storage and massively parallel computing resources at relatively low costs.

Architecture of Hybrid cloud

As cloud computing becomes prevalent, an increasing amount of data is being stored in the cloud and shared by users with specified privileges, which define the access rights of the stored data. One critical challenge of cloud storage services is the management of the ever-increasing volume of data.

To avoid this duplication of data and to maintain the confidentiality in the cloud we using the concept of Hybrid cloud. It is a combination of public and private cloud. Hybrid cloud storage combines the advantages of scalability, reliability, rapid deployment and potential cost savings of public cloud storage with the security and full control of private cloud storage.

II. LITERATURE REVIEW

A. Fast and Secure Laptop Backups with Encrypted De-duplication

Many people now store large quantities of personal and corporate data on laptops or home computers. These often have poor or intermittent connectivity, and are vulnerable to theft or hardware failure. Conventional backup solutions are not well suited to this environment, and backup regimes are frequently inadequate. This paper describes an algorithm which takes advantage of the data which is common between users to increase the speed of backups, and reduce the storage requirements. This algorithm supports client-end per-user encryption which is necessary for confidential personal data. It also supports a unique feature which allows immediate detection of common subtrees, avoiding the need to query the backup system for every file. We describe a prototype implementation of this algorithm for Apple OS X, and present an analysis of the potential effectiveness, using real data obtained from a set of typical users. Finally, we discuss the use of this prototype in conjunction with remote cloud storage, and present an analysis of the typical cost savings. Data backup has been an important issue ever since computers have been used to store valuable information. There has been a considerable amount of research on this topic, and a plethora of solutions are available which largely satisfy traditional requirements. However, new modes of working, such as the extensive use of personal laptops, present new challenges.

B. Duplets: Server-Aided Encryption for Deduplicated Storage

Cloud storage service providers such as Drop box, Mozy, and others perform de-duplication to save space by only storing one copy of each file uploaded. Should clients conventionally encrypt their files, however, savings are lost. Message-locked encryption (the most prominent manifestation of which is convergent encryption) resolves this tension. However it is inherently subject to brute-force attacks that can recover files falling into a known set. We propose an architecture that provides secure deduplicated storage resisting brute-force attacks, and realize it in a system called Duplets. In Duplets, clients encrypt under message-based keys obtained from a key-server via an oblivious PRF protocol. It enables clients to store encrypted data with an existing service, have the service perform de-duplication on their behalf, and yet achieves strong confidentiality guarantees. We show that encryption for deduplicated storage can achieve performance and space savings close to that of using the storage service with plaintext data. Providers of cloud-based storage such as Dropbox [3], Google Drive [7], and Mozy [63] can save on storage costs via de-duplication: should two clients upload the same file, the service detects this and stores only a single copy.

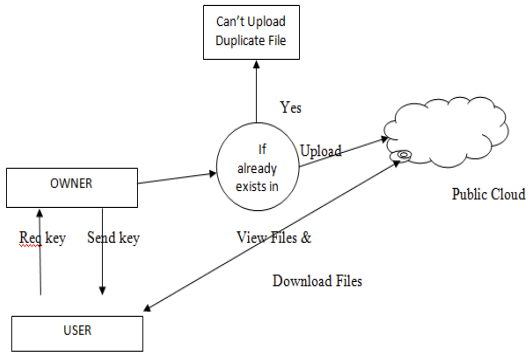
C. Private Data De-duplication Protocols in Cloud Storage

In this paper, a new notion which we call private data de-duplication protocol, a de-duplication technique for private data storage is introduced and formalized. Intuitively, a private data de-duplication protocol allows a client who holds a private data proves to a server who holds a summary string of the data that he/she is the owner of that data without revealing further information to the server. Our notion can be viewed as a complement of the state-of-the-art public data de-duplication protocols of Halevi et al. The security of private data de-duplication protocols is formalized in the simulation-based framework in the context of two-party computations. A construction of private de-duplication protocols based on the standard cryptographic assumptions is then presented and analyzed. We show that the proposed private data de-duplication protocol is provably secure assuming that the underlying hash function is collision-resilient, the discrete logarithm is hard and the erasure coding algorithm can erasure up to a-fraction of the bits in the presence of malicious adversaries in the presence of malicious adversaries. To the best our knowledge this is the first de-duplication protocol for private data storage. In this paper, a new notion which we call private data de-duplication protocol, a de-duplication technique for private data storage is introduced and formalized.

III. METHODOLOGY

In the proposed system we are achieving the data de-duplication by providing the proof of data by the data owner. This proof is used at the time of uploading of the file. Each file uploaded to the cloud is also bounded by a set of privileges to specify which kind of users is allowed to perform the duplicate check and access the files. Before submitting his duplicate check request for some file, the user needs to take this file and his own privileges as inputs. The user is able to find a duplicate for this file if and only if there is a copy of this file and a matched privilege stored in cloud.

ARCHITECTURE:



IV. SYSTEM IMPLEMENTATION

The purpose of **System Implementation** can be summarized as follows:

It making the new system available to a prepared set of users (the deployment), and positioning on-going support and maintenance of the system within the Performing

Organization (the transition). At a finer level of detail, deploying the system consists of executing all steps necessary to educate the Consumers on the use of the new system, placing the newly developed system into production, confirming that all data required at the start of operations is available and accurate, and validating that business functions that interact with the system are functioning properly. Transitioning the system support responsibilities involves changing from a system development to a system support and maintenance mode of operation, with ownership of the new system moving from the Project Team to the Performing Organization.

V. MODULES

A. User Register

Here client has to register their details to log In. It contains unique user name and password. Client is the only authorized person to access this module for security purpose.

B. User Log In

Here client has to log in by using their unique user name and password after registration. Client is the only authorized person to access this module for security purpose. So others don't get rights to access this module.

C. File Upload

In this module user upload a block of files in the cloud with encryption by using his secret key. This ensures the files to be protected from unauthorized user. Encrypt file using triple DES encryption algorithm.

D. Duplicate Detection

Data de-duplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage. In duplicate detection module is to find the duplicate files from the cloud using verification algorithm. Upload files in cloud at that time to check whether the file or already exists in cloud or not. If exists means it displays already exist in file and can't upload in this file. If not exists means file upload in cloud successfully.

E. View Files

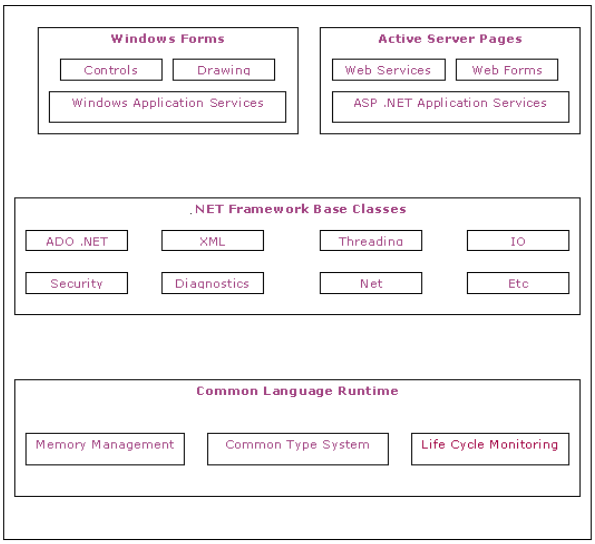
In this module client view their uploaded file from cloud. Client can edit the particular file. The updates will change on cloud. Client is the only authorized person to access this module for security purpose. So others don't get rights to access this module.

F. Download Files

In this module client download their files means give to the correct secret key. The secret key is wrong means the file is not download. Particular date only client download the upload files.

VI. SOFTWARE FEATURES

Many people reckon that it's Microsoft's way of controlling the Internet, which is false. .NET is Microsoft's strategy of software that provides services to people any time, any place, on any device. An accurate definition of .NET is, it's an XML Web Services platform which allows us to build rich .NET applications, which allows users to interact with the Internet using wide range of smart devices (tablet devices, pocket PC's, web phones etc), which allows to build and integrate Web Services and which comes with many rich set of tools like Visual Studio to fully develop and build those applications.



.NET Framework

.NET Framework

.NET is a "Software Platform". It is a language-neutral environment for developing rich .NET experiences and building applications that can easily and securely operate within it. When developed applications are deployed, those applications will target .NET and will execute wherever .NET is implemented instead of targeting a particular Hardware/OS combination. The components that make up the .NET platform are collectively called the .NET Framework.

The .NET Framework is a managed, type-safe environment for developing and executing applications. The .NET Framework manages all aspects of program execution, like, allocation of memory for the storage of data and instructions, granting and denying permissions to the application, managing execution of the application and reallocation of memory for resources that are not needed.

VII. CONCLUSION

In this project, the notion of authorized data de-duplication was proposed to protect the data security by including differential privileges of users in the duplicate check. We also presented several new de-duplication constructions supporting authorized duplicate check in hybrid cloud architecture, in which the duplicate-check tokens of files are generated by the private cloud server with private keys. Security analysis demonstrates that our schemes are secure in terms of insider and outsider attacks specified in the proposed security model. As a proof of concept, we implemented a prototype of our proposed authorized duplicate check scheme and conduct test bed experiments

on our prototype. We showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

VIII. REFERENCES

1. M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller. Secure data de-duplication. In Proc. of StorageSS, 2008.
2. P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.
3. M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Server-aided encryption for deduplicated storage. In USENIX Security Symposium, 2013.
4. M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure de-duplication. In EUROCRYPT, pages 296– 312, 2013.
5. M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 22(1):1–61, 2009.
6. M. Bellare and A. Palacio. Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks. In CRYPTO, pages 162–177, 2002.
7. S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. Twin clouds: An architecture for secure cloud computing. In Workshop on Cryptography and Security in Clouds (WCSC 2011), 2011.
8. J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer. Reclaiming space from duplicate files in a serverless distributed file system. In ICDCS, pages 617–624, 2002.
9. D. Ferraiolo and R. Kuhn. Role-based access controls. In 15th NIST-NCSC National Computer Security Conf., 1992.
10. GNU Libmicrohttpd http:// www.gnu.org/software/libmicrohttpd/.
11. S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, ACM Conference on Computer and Communications Security, pages 491–500. ACM, 2011.

KEERTHIKA.K is professor of computer science and engineering at Selvam College of Technology , Namakkal. She received M.E degree from K..S..Rangasamy college of technology and B.E from Vidhyaa vikas college of engineering and technology,all affiliated to Anna university: Chennai,, in computer science and engineering. Her research interest include Data Analytics And Mining, Computer Programming, Social Networks, Cloud Computing. And published papers in the area of Data mining and Social Networks.

G.MANIKANDAN, currently studying in 4th year Department Of CSE at Selvam College of Technology(SCT), Namakkal, He has developed minor project in the field of Cloud computing, Currently he is doing research on cloud storage. He is expertise in, Computer Networks, Hardware, Software, Web application.

J.SAGAYARAJA, currently studying in 4th year Department Of CSE at Selvam College of Technology(SCT), Namakkal, He has developed minor project in the field of Cloud computing, Currently he is doing research on cloud storage. He is expertise in, Computer Networks, Hardware, Web application.

S.VINOTH, currently studying in 4th year Department Of CSE at Selvam College of Technology(SCT), Namakkal, He has developed minor project in the field of cloud storage, Currently he is doing research on artificial intelligence. He is expertise in IOT, Computer Networks, Hardware, Software, Web application.