Data Deduplication with Dynamic Management in Cloud Storage

Nethra M.V.O.

Professor,
Department of CSE,
R. R. Institute of Technology,
Bengaluru, Karnataka, India

Yashaswini R.

UG Student,
R. R. Institute of Technology,
Visvesvaraya Technological University,
Bangalore, India

ABSTRACT

In cloud storage 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 of them. Deduplication is most effective when multiple users outsource the same data to the 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, many users are likely to encrypt their data before outsourcing them to the cloud storage to preserve privacy, but this hampers deduplication because of the randomization property of encryption. Recently, several deduplication schemes have been proposed to solve this problem by allowing each owner to share the same encryption key for the same data. However, most of the schemes suffer from security flaws, since they do not consider the dynamic changes in the ownership of outsourced data that occur frequently in a practical cloud storage service. In this paper, we propose a novel server-side deduplication scheme for encrypted data. It allows the cloud server to control access to outsourced data even when the ownership changes dynamically by exploiting randomized convergent encryption and secure ownership group key distribution.

Keywords: Deduplication, Proof-of-ownership, Encryption

INTRODUCTION:

Cloud computing provides scalable, low-cost, and location-independent online services ranging from simple backup services to cloud storage infrastructures. The fast growth of data volumes stored in the cloud storage has led to an increased demand for techniques for saving disk space and network bandwidth. To reduce resource consumption, many cloud storage services, such as Dropbox, Wuala, Mozy, and Google Drive, employ a deduplication technique, where the cloud server stores only a single copy of redundant data and provides links to the copy instead of storing other actual copies of that data, regardless of how many clients ask to store the data. The savings are significant, and reportedly, business applications can achieve disk and bandwidth savings of more than 90%. However, from a security perspective, the shared usage of users 'data raises a new challenge. As customers are concerned about their private data, they may encrypt their data before outsourcing in order to protect data privacy from unauthorized outside adversaries, as well as from the cloud service provider. This is justified by current security trends and numerous industry regulations such as PCIDSS. However, conventional encryption makes deduplication impossible for the following reason. Deduplication techniques take advantage of data similarity to identify the same data and reduce the storage space. In contrast, encryption algorithms randomize the encrypted files in order to make cipher text indistinguishable from theoretically random data. Encryptions of the same data by different users with different encryption keys results in different cipher texts, which make it difficult for the cloud server to determine whether the plain data are the same and deduplicate them. The majority of cloud computing infrastructure currently consists of reliable services delivered through data-centers that are built on computer and storage virtualization technologies. The services are accessible anywhere in the world, with The Cloud appearing as a single point of access for all the computing needs of

consumers. Commercial offerings need to meet the quality of service requirements of customers and typically offer service level agreements.

RELATED WORK:

Cloud can offer you the possibility of storing your files and accessing, storing and retrieving them from any web-enabled interface. The web services interfaces are usually simple. At any time and place you have high availability, speed, scalability and security for your environment. In this scenario, organizations are only paying for the amount of storage they are actually consuming, and do so without the worries of overseeing the daily maintenance of the storage infrastructure. There is also the possibility to store the data either on or off premises depending on the regulatory compliance requirements. Data is stored in virtualized pools of storage hosted by a third party based on the customer specification requirements. Backing up data has always been a complex and time-consuming operation. This included maintaining a set of tapes or drives, manually collecting them and dispatching them to a backup facility with all the inherent problems that might happen in between the originating and the backup site. This way of ensuring a backup is performed is not immune to problems such as running out of backup media , and there is also time to load the backup devices for a restore operation, which takes time and is prone to malfunctions and human errors. Cloud-based backup, while not being the panacea, is certainly a far cry from what it used to be. You can now automatically dispatch data to any location across the wire with the assurance that neither security, availability nor capacity are issues. While the list of the above uses of cloud computing is not exhaustive, it certainly give an incentive to use the cloud when comparing to more traditional alternatives to increase IT infrastructure flexibility , as well as leverage on big data analytics and mobile computing.

PROPOSED SYSTEM:

We propose an efficient group key management protocol in distributed group communication. This protocol is based on Elliptic Curve Cryptography and decreases the key length while providing securities at the same level as that of other cryptosystems provides. We provide the high level security and avoid the replication of file in the cloud service provider. In proposed system, we are using hash function to generate key for the file.By using hash function to avoid the duplication in cloud. After that we apply cryptographic technique for security purpose. We using ECC algorithm for encryption and decryption process. Some advantages of proposed system are: Avoid duplication in cloud, Increase the security level, High efficient.

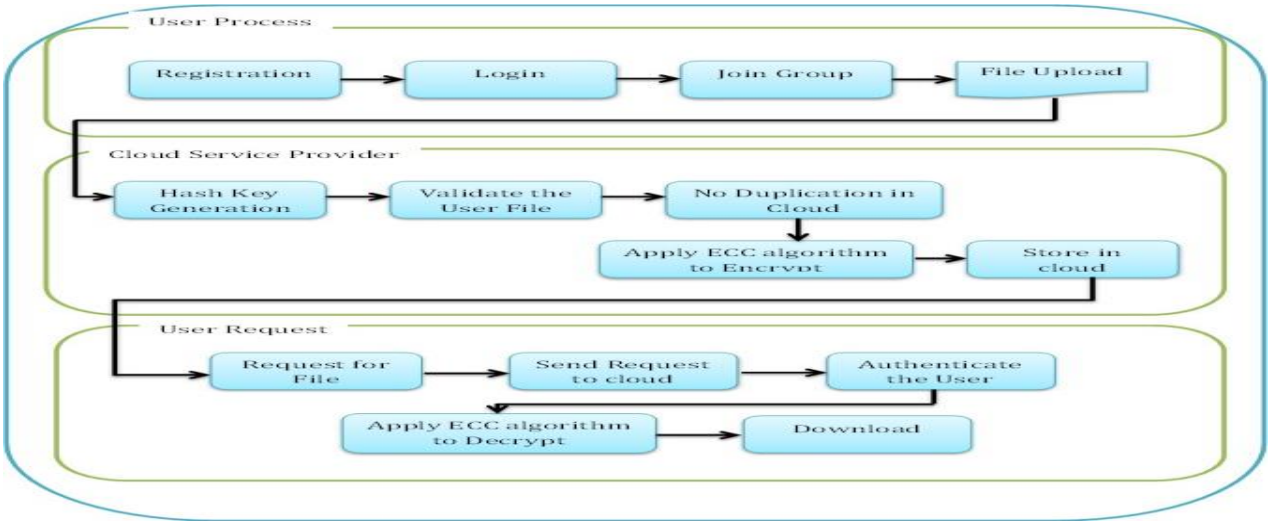
EXISTING SYSTEM:

In existing system, Cryptographic techniques were applied to access control for remote storage systems. The data owners encrypt files by using the symmetric encryption approach with content keys and then use every user's public key to encrypt the content keys. It requires each data owner to be online all the time. Some methods deliver the key management and distribution from the data owners to the remote server under the assumption that the server is trusted or semi-trusted. The key management is very complicated when there are a large number of data owners and users in the system. The key distribution is not convenient in the situation of user dynamically system. The server is cannot be trusted by the data owners in cloud storage systems. It cannot be applied to access control for cloud storage systems.

SYSTEM DESIGN:

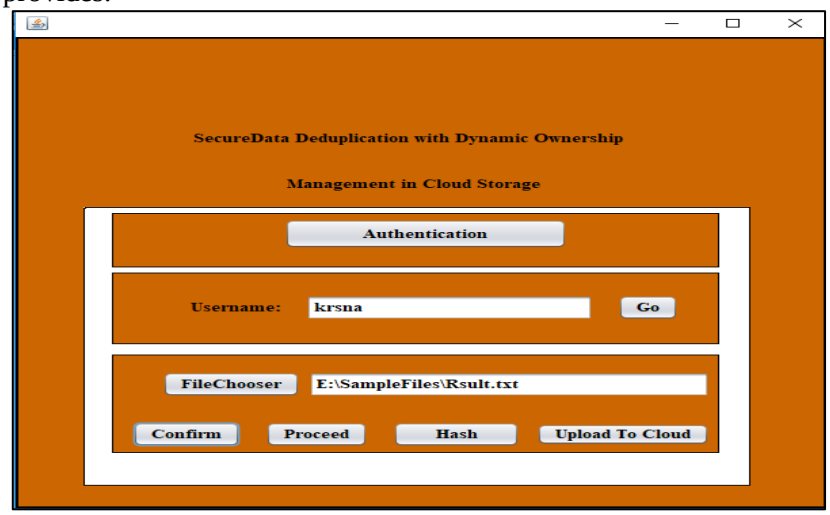
In our process, new user registers the details and gets the username and password for further process. Using Username and Password, user login into Group. Group generate key for the valid user and process inside the group under the valid key. In file upload process, user chooses the file from the system and generate hash key for each file. Hash key generation is provided to avoid duplication of file to the cloud. If the file is already in cloud, user should upload another file to cloud. User send request to the cloud, cloud service provider decrypt the file .For cryptographic technique, we using Elliptic Curve Cryptography (ECC) algorithm for decrypting the file. Send the requested file to the user after validate the user. Then file will be downloaded in user location.

System Architecture

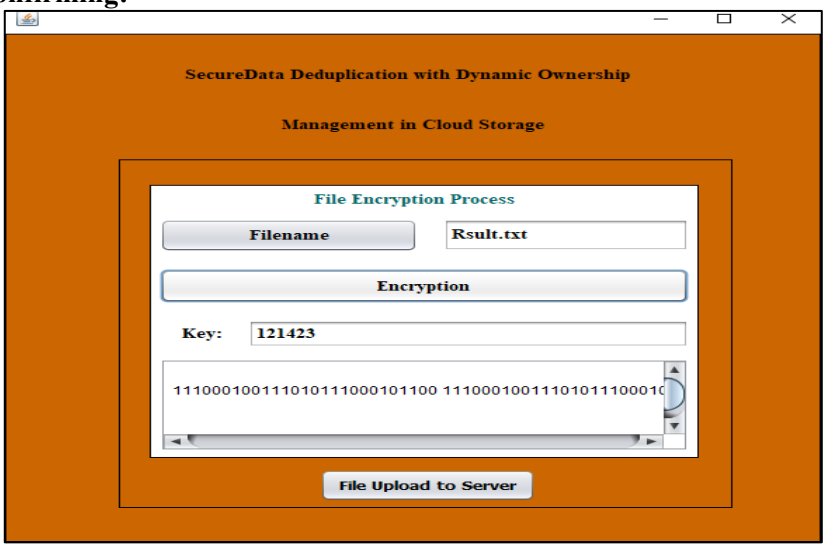


RESULTS:

An efficient group asymmetric key management protocol in distributed group communication which is based on Elliptic Curve Cryptography and decreases the key length while providing securities at the same level as that of other cryptosystems provides.



File choosing and confirming:



File encryption and Upload:

```
mysql> select * from reg;
+---------+-----------+----------+------------+--------+---------+
| name    | uname     | password | dob        | gender | country |
+---------+-----------+----------+------------+--------+---------+
| lata    | lata123   | 1234567  | 2000-02-01 | F      | India   |
| Bhusal  | gaurav123 | asdfgh   | 1990-02-01 | M      | India   |
| krsna   | krsna123  | qwerty   | 1995-02-01 | M      | Nepal   |
| meghana | meghana123| zxcvbnm  | 1995-02-01 | F      | India   |
+---------+-----------+----------+------------+--------+---------+
4 rows in set (0.00 sec)
```

Backend database of registered User

CONCLUSION AND FUTURE WORK:

We provide the high level security and avoid the replication of file in the cloud service provider. We have used hash function to generate key for the file. We have used hash function to avoid the duplication in cloud. After that we apply cryptographic technique for security purpose. ECC algorithm for encryption and decryption process to decrease probability of attacking the file is used. Validate the File for confidential in cloud and high efficient. For future work. Iris scanner and Fingerprint scanner can be used for Registration and login to provide higher security and QR scanner can be used instead of key which is used to download the file.

REFERENCES:

C. Wang, Z. Qin, J. Peng, and J. Wang, "A novel encryption scheme for data deduplication system," Proc. International Conference on Communications, Circuits and Ssytems (ICC-CAS), pp. 265–269, 2010.

C. Wang, Z. Qin, J. Peng, and J. Wang, "A novel encryption scheme for data deduplication system," Proc. International Conference on Communications, Circuits and Ssytems (ICC-CAS), pp. 265–269, 2010.

D. Harnik, B. Pinkas, and A. Shulman Peleg, "Side channels in cloud services, the case of deduplication in cloud storage, "IEEE Security & Privacy, vol. 8, no. 6, pp. 40–47, 2010.

D. T. Meyer, and W. J. Bolosky, "A study of practical deduplication," Proc. USENIX Conference on File and Storage Technologies 2011, 2011.

M. Dutch, "Understanding data deduplication ratios," SNIA Data Management Forum, 2008.

M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, "Secure data deduplication," Proc. StorageSS‟08, 2008.

Malicious insider attacks to rise, http://news.bbc.co.uk/2/hi/7875904.stm

N. Baracaldo, E. Androulaki, J. Glider, A. Sorniotti, "Reconciling end-to-end confidentiality and data reduction in cloud storage," Proc. ACM Workshop on Cloud Computing Security, pp. 21–32,2014.

P. S. S. Council, "PCI SSC data security standards overview,"2013.

W. K. Ng, W. Wen, and H. Zhu, "Private data deduplication protocols in cloud storage," Proc. ACM SAC‟12, 2012.
