INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 6)
Available online at: www.ijariit.com

Encrypted data management with deduplication in cloud computing

Pokala Phanitej
pphanitej@gmail.com
Loyola Institute of Technology and Management, Guntur, Andhra Pradesh

Y. Suresh
sureshyallamati@gmail.com
Loyola Institute of Technology and Management, Guntur, Andhra Pradesh

ABSTRACT

Cloud Computing is an information technology concept which plays a vital role in data processing and data storing. It also plays a crucial role in the Internet of Things (IoT). The data stored in the cloud should be secured to prevent the unauthorized access. There comes a data security concept known as Encryption. In order to maintain the users, Privacy and the security of the data is stored in the cloud in the encrypted or cipher-text format. By this, only the encrypted data is going to be stored in the cloud which reduces the usage of the storage devices up to a great extent. This is mainly used for storing very large size datasets like for the big data. We have a lot of deduplication schemes which avoids the duplicate data, but the main problem with those schemes are lack of security and lack of tractability for the secure data access control. Due to these two problems, very few of them are taken into practice. In this, we used a scheme known as Attribute-based Encryption to deduplicated the encrypted data to provide secure data access control.

Keywords— *Cloud computing, Encryption, Access control, Internet of Things (IoT), Attribute-based encryption, Data storage, Decryption, Deduplication management, Cipher-text*

1. INTRODUCTION

Data storage [1] service is the most important cloud service. The personal data is uploaded in the cloud service provider (CSP) and that is maintained by the cloud service provider (CSP). In existing research to maintain data privacy only outsource encrypted data are proposed to the cloud. Especially for shared data, the duplicated data in an encrypted form are stored by the same or different users which lead to wastage of networking resources. Due to this, the objective of deduplication is to achieve high speed and cost saving. As existing systems suffer from brute-force attacks, they cannot support data access control and revocation flexibly. The present solutions cannot ensure security, relia-bility, and privacy [2]. In the existing system for cloud data supports data deduplication [3] and few can ensure security [4] and flexibility of sound performance for cloud in which data deduplication is directly controlled by data owners.

To deduplicate encrypted data stored in the cloud [5] and support secure data access control at the same time and efficient a scheme based on attribute-based encryption (ABE) [6] is proposed. Attribute-based encryption is one type of public-key [7] encryption [8] in which the hidden key of a user and the ciphertext key depends upon the attributes. The decryption of the encrypted data is possible only if the set or group of attributes of the user key matches with the attributes of the encrypted data. It follows an adversary that multiple keys should only be able to access data if at least one individual key grants permission to access the data.

The new system mainly depends upon the attribute-based encryption (ABE) algorithm in which the public/private key depends upon the attributes. We use this algorithm to deduplicate the encrypted data stored in the cloud which provides the user secure [9] data access control. This encrypted data is handled by the Cloud Service Providers. The same or different user can save the duplicated data, but it is stored only once, in the cloud by the CSPs. This mainly reduces the storage space, reduces the resources and managing the data becomes very easy. Two types of deduplication are used: Intra user deduplication and Inter-user deduplication. This will not work for multiple users after deduplication.

2. EXISTING SYSTEM

The existing system outsources the encrypted data only to the Cloud Service Providers. In this existing system, the same or different users have the chance to save the duplicate data into the cloud. There are various deduplication schemes, but those are not safe and easily susceptible to brute-force attacks and there is a lack of tractability also in this existing system.

2.1 Disadvantages of the Existing System

Deduplication technology became very important for storing the data into the cloud to avoid the duplicate data. But the scheme used in one data center may not be suitable for storing the data in another data center. This requires a separate hardware, but it degrades the performance and also there will be a lack of data integrity.

3. PROPOSED SYSTEM

The new system mainly depends upon the attribute-based encryption (ABE) algorithm in which the public/private key depends upon the attributes. We use this algorithm to deduplicate the encrypted data stored in the cloud which provides the user with secure data access control. This encrypted data is handled by the Cloud Service Providers. The same or different user can save the duplicated data, but it is stored only once in the cloud by the CSPs. This mainly reduces the storage space, reduces the resources and managing the data becomes very easy. Two types of deduplication are used: Intra user deduplication and Inter-user deduplication. This will not work for multiple users after deduplication.

3.1 Advantages of the proposed system

This proposed system has a lot of improvements than the existing system. This system mainly saves the storage space and works effectively in real-time applications at the same or similar type of data is stored once in the cloud. This provides a secure way of storing and can be accessed for other users based on the data owner's policies. Data backups which are done to prevent the data loss and where a lot of storage space is required. In that case, our proposed system will be more helpful.

4. MODULES

Three modules used in this project are:

4.1. Deduplication

Deduplication means that it will not allow the duplicate data to be stored in the cloud. The data which is encrypted and is stored in the cloud can be accessed by the other users if and only if the original copyright 24owner gives permission to use the data. There are many deduplication schemes to handle the encrypted data, but some of those are not secure and some of those ensure the data security and data flexibility in the cloud. Deduplication saves the amount of huge space and cost which is very useful in the maintenance of the data effectively in several applications.

4.2. Cloud Computing

There are various cloud service providers (CSPs) who are responsible for the maintenance of the data. By maintaining the data, they ensure some properties like scalability, flexibility, and elasticity. CSP allows different users to store different types of data such as pictures, videos, and personal files etc. and also allows retrieving the data. This feature became important in IoT [10] applications.

4.3. Inter and Intra User

Data owners want to protect their data or information from unauthorized access. So they are given this responsibility to the CSPs to control and protect the data. Based on the data owner expectations, CSPs perform the access control. Data owners want CSPs not only to control the data access but also a way of storing the data and usage of the data. CSP restricts the other users to duplicate the stored data in the cloud and the data in the cloud can be accessed by the other users based on the data owner rules and regulations. Data owner will be 1 i.e. M=1 and the data users can be many i.e. u=1, 2, 3…. N.

5. ALGORITHMS

Cipher-text Policy ABE (CP-ABE) or Key Policy ABE (KP-ABE)

Attribute-based encryption [11] is one type of public-key encryption [12] in which the hidden key of a user and the ciphertext [13] key depends upon the attributes. The decryption [14] of the encrypted data is possible only if the set or group of attributes of the user key matches with the attributes of the encrypted data. It follows an adversary that multiple keys should only be able to access data if at least one individual key grants permission to access the data.

Encryption Algorithm

Encryption is the scheme or process in which the data are converted to a specific format which is understood or read by the specified user. Before encrypting the data or the information is referred to as the plain text and after the encryption the data are referred to as the ciphertext. Encrypted data are only read if decrypted otherwise we cannot read the data. An encryption algorithm is used for the encryption which is very important to secure the data in the cloud.

Decryption Algorithm

After the encryption of the data, if we need to use the data we have to decrypt the data. We cannot use the encrypted data directly, so we use the decryption algorithm to convert the encrypted data into the decrypted data. There are two types of cryptography to do this: symmetric key cryptography [15] and asymmetric key cryptography. In asymmetric key cryptography, different (public/private) keys are used and in symmetric key cryptography, the same key is used for both encryption and decryption.

6. DESIGN

6.1 Activity Design

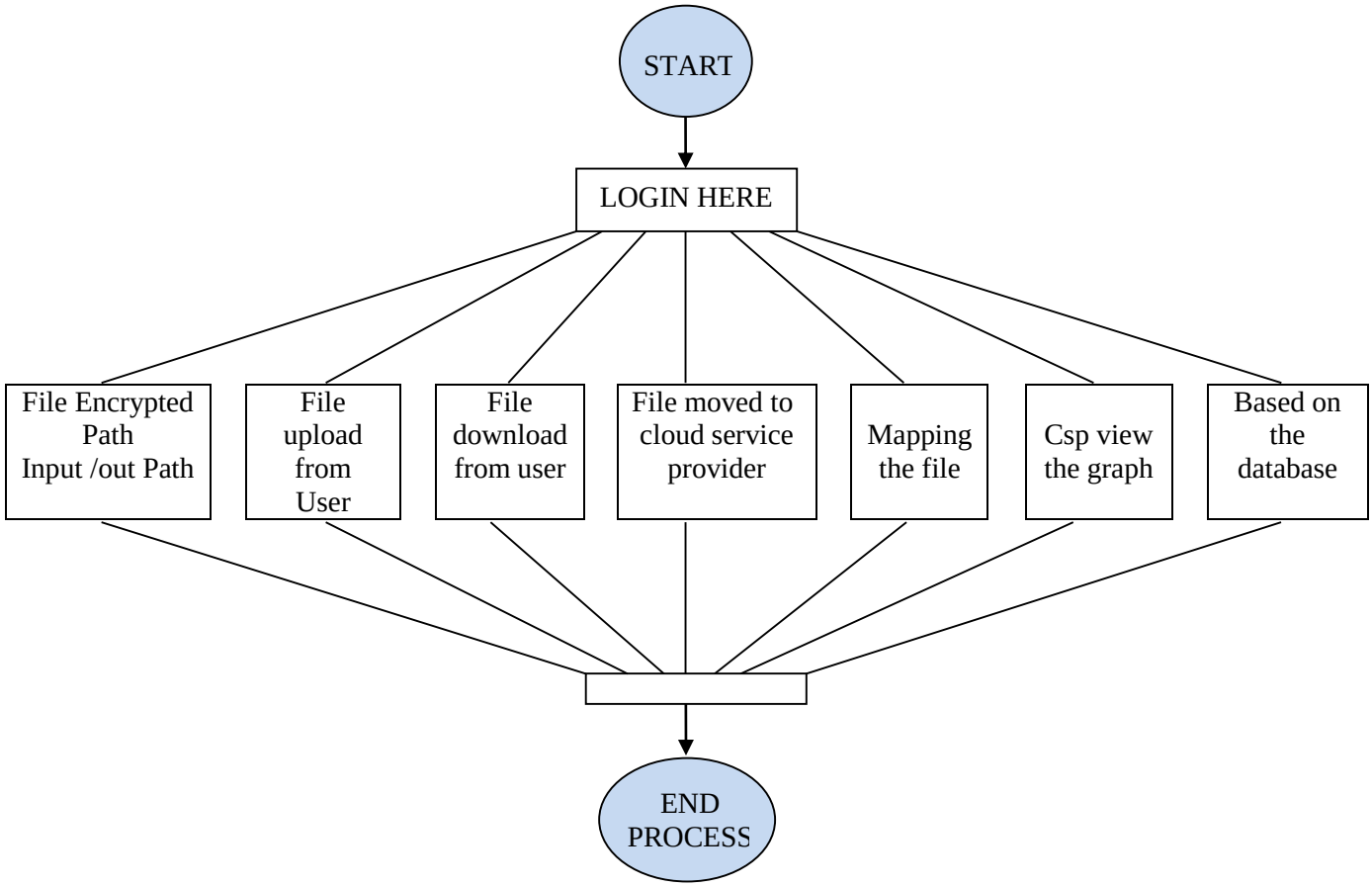

Fig. 1: Activity design

6.2 System Design

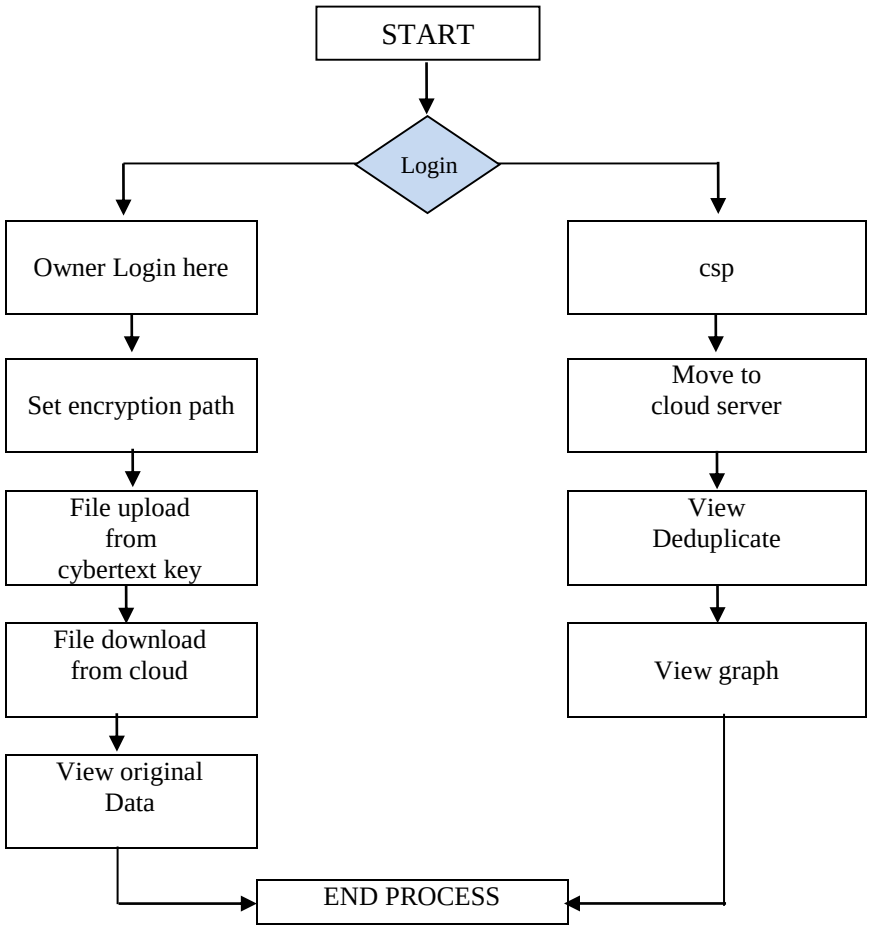

Fig. 2: System design

6.3 ER Diagram

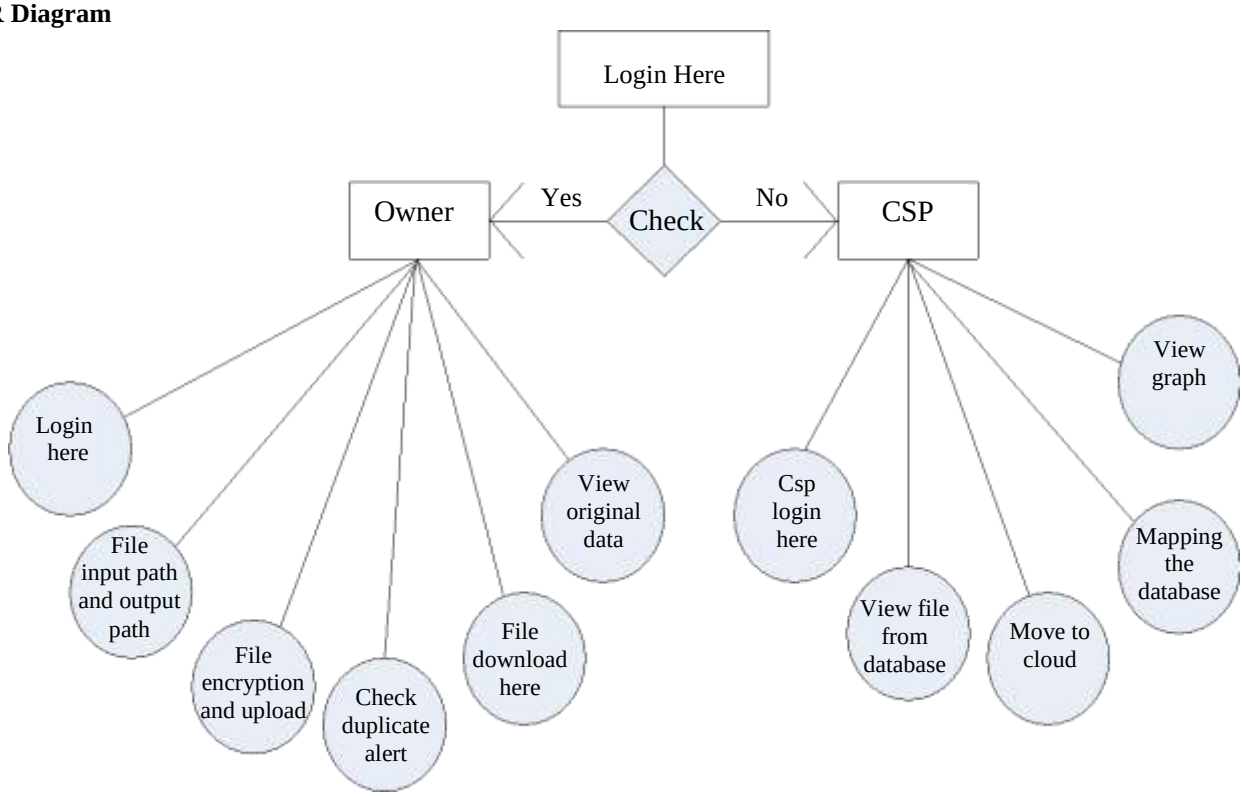

Fig. 3: ER Diagram

7. IMPLEMENTATION

This is the most critical stage where the theoretical design will be implemented in a real system. At this stage, we need to have a lot of careful planning and idea about the existing system so that the newly proposed system will work effectively. In this, we may face so many errors which will stop the implementation of the project and according to them, we have to take the steps to reduce the errors.

8. OUTPUT SCREENSHOTS

Figure 4 is the home page of the Encrypted Data Management with Deduplication in Cloud Computing. On this page, we can select the option to Login as CSP or Data Owner or Logout. By choosing the appropriate option further will be proceeded. Fig. 5 is the Owner login page. The owner has to enter an owner name and password. If the login details are matched, then the owner gets a login and the public key is sent to the owner's email id. The public key is used at the time of downloading files from the cloud. If the owner is new, he/she can sign up by clicking the signup option which will be later used to sign in.

Fig. 6 is the User-Registration page where the user can register which will be used to sign in. This Page contains username, email id, password, and mobile number and gender fields. Figure 7 is the page showing the message that the user successfully registered after user registration. Figure 8 is the page appears after CSP move files to the cloud, showing the file names with the size of the file and a download link to download the file. Any user can download the files by signing into their accounts. Figure 9 is the CSP login page. The CSP owner has to enter the CSP name and password to log in. CSP owner moves files to the cloud. After moving the files only, any user can download the files. Figure 10 is the page of moving files to the cloud. Here CSP owner enters a file name and Decrypt key to move files to the cloud. Whatever file CSP enters, only that file will be moved to the cloud.

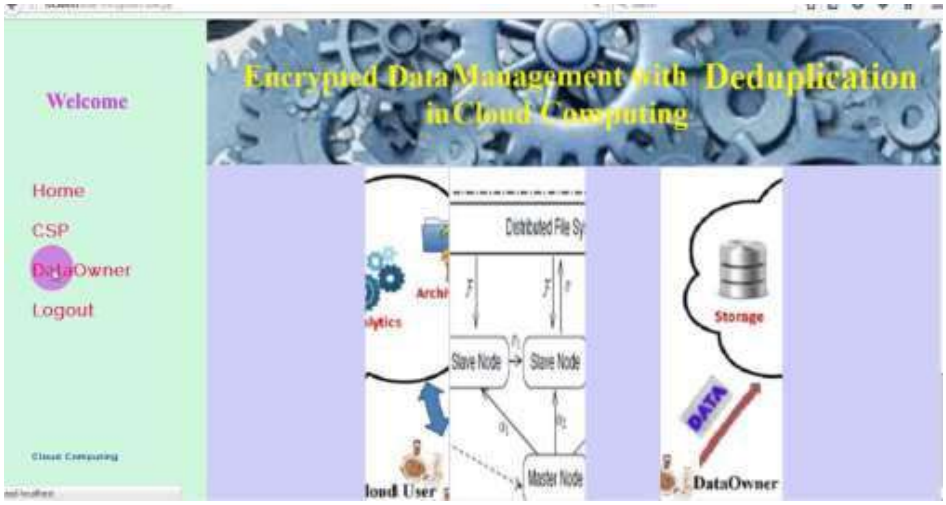

Fig. 4: Homepage

Fig. 5: Data Owner login page

Fig. 6: Data User Registration page

Fig. 7: Page after registration

Fig. 8: Files available in CSP

Fig. 9: CSP login page

Fig. 10: To move the file to the cloud

9. CONCLUSION

Encrypted data with data deduplication is very useful in cloud computing in storing the data securely and ensures the consistency and flexibility of the data. This is mainly used in the data analysis where we use a large amount of data which is secured by using this scheme. This system adds the data claiming to the original copyright owner, optimization of the schemes with the hardware and improves the solutions to assist the user to make his/her data secure and the access controlled by the original owner.

10. REFERENCES

[1] A Parak and Kak, Online Data Storage using Implicit Security, *Information Sciences*, 2009, 179(19), 3323-3331.
[2] Wayne Jansen, Timothy Grance, Guidelines on Security and Privacy in Public Cloud Computing, NIST, *Draft Special Publication, 800-144*, 2011.
[3] Open Dedup, Global Inline Deduplication for Block Storage and Files, *Available from http://opendedup.org/index.php*, 2010.
[4] MQ Zhou, R Zhang, W Xie, WN Qian and A Zhou, Security, and Privacy in Cloud Computing: A Survey, *Sixth International Conference on Semantics, Knowledge and Grids(SKG)*, 2010, 105-112.
[5] T Dillon, C Wu and E Chang, Cloud Computing: Issues and Challenges, *24th IEEE International Conference on Advanced Information Networking and Applications (AINA)*, 2010, 27-33.
[6] J Bethencourt, A Sahai and B Waters, Cipher Text-Policy Attribute-Based Encryption, *IEEE Symposium on Security and Privacy (SP'07)*, 2007, 321–334.
[7] Li H and Dai Y, Identity-Based Authentication for Cloud Computing, *MG Jaatun, G Zhao and C Rong (Eds): Cloud Computing Lecture Notes in Computer Science*, 2009, 5931, 157-166.
[8] Agrawal, S Gorbunov, S Vaikuntanathan and V Wee, Functional Encryption: New Perspectives and Lower Bounds. *In Canetti, R., Garay, J.A. (eds.) CRYPTO 2013, Part II. LNCS*, 8043, 2013, 500–518.
[9] D Catteddu, Cloud Computing: Benefits, Risks, and Recommendations For Information Security Web Application Security, Communications in Computer and Information Science book series, (CCIS, volume 72), 2010.
[10] Rafiullah Khan, Sarmad Ullah Khan, Rifaqat Zaheer and Shahid Khan, Future Internet: The Internet of Things Architecture and Possible Applications and Key Challenges, *In Proceedings of Frontiers of Information Technology (FIT)*, 2012, 257-260.
[11] J Bethencourt, A Sahai and B Waters, Cipher Text-Policy Attribute-Based Encryption, *In IEEE Symposium on Security and Privacy*, 2007, 321–334.
[12] K Kurosawa and Y. Desmedt, A New Paradigm of Hybrid Encryption Scheme, *In Proceedings of Crypto 2004 of LNCS*, 2004, 3152, 426–442.

[13] Y Dodis and J Katz, Chosen-Cipher Text Security of Multiple Encryption, *In Proceedings of TCC 2005, LNCS. Springer-Verlag,* 2005

[14] Obaida Mohammad Awad Al-Hazaimeh, A New Approach for Complex Encrypting and Decrypting data," International Journal of Computer Networks & Communications, 2013, 5(2), 95-103.

[15] D Chatterjee, J Nath, S Dasgupta, and A Nath, A New Symmetric Key Cryptography Algorithm Using Ex-tended MSA Method: DJSA Symmetric Key Algorithm, International Conference on Communication Systems and Network Technologies (CSNT), DOI: 10.1109/CSNT.2011.25. 2011.

BIOGRAPHY

Pokala Phanitej
Master of Technology in Computer Science Engineering
Loyola Institute of Technology and Management, Guntur, Andhra Pradesh

Bachelor of Technology in Electronics and Communication Engineering
Amara Institute of Engineering and Technology, Guntur, Andhra Pradesh

Y. Suresh
Associate Professor
Loyola Institute of Technology and Management, Guntur, Andhra Pradesh

