IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

Data Secure De-Duplication in Cloud Environment

L. Kavitha[1] and Sumalatha. V[2]
PG Student, Department of Computer Applications[1]
Associate Professor, Department of Computer Applications[2]
Vels Institute of Science Technology and Advanced Studies, Pallavaram, Chennai, India
lkavitha2794@gmail.com and sumalathav.research@gmail.com

Abstract: *In the current area of information explosion, users' demand for data storage is increasing, and data on the cloud has become the first choice of users and enterprises. Cloud storage facilitates users to backup and share data, effectively reducing users' storage expenses. As the duplicate data of different users are stored multiple times, leading to a sudden decrease in storage utilization of cloud servers. Data stored in plaintext form can directly remove duplicate data, while cloud servers are semi-trusted and usually need to store data after encryption to protect user privacy. In this paper, we focus on how to achieve secure re-duplication and recover data in ciphertext for different users, and determine whether the indexes of public key searchable encryption and the matching relationship of trapdoor are equal in cipher text to achieve secure de-duplication. For the duplicate file, the data user's re-encryption key about the file is appended to the ciphertext chain table of the stored copy. The cloud server uses the re-encryption key to generate the specified transformed ciphertext, and the data user decrypts the transformed ciphertext by its private key to recover the file. The proposed scheme is secure and efficient through security analysis and experimental simulation analysis.*

Keywords: PEKS, Secure de-duplication, Proxy Re-encryption, Data recovery

I. INTRODUCTION

As a major service provided by cloud computing technology, cloud storage enables users to back up and share data easily and quickly, which can efficiently reduce users' storage expenses and improve work efficiency. With the increasing maturity of cloud computing technology. There are many Cloud Service Providers (CSP) in the market, such as Baidu Cloud, Amazon Cloud, and other famous CSPs. Users will upload and store their confidential data to the data storage centre of the cloud server, which is managed and maintained by the CSP, but with this comes the frequent occurrence of cloud computing security issues. For enterprises or individual users will be personal files, business contracts, user transaction records, environmental geographic data, and other susceptible data stored in the cloud server. However, user privacy leaks and sensitive data leaks have emerged, and even more, there are CSP through the sale of user data to achieve corporate profits. The issue of data security in cloud storage deserves widespread attention. Big data and cloud computing are developing rapidly, with an explosion of data from users around the world, resulting in a dramatic increase in demand for cloud servers. An effective solution to the need for storage of massive amounts of data will be deduplicate data. For plaintext data, the equality test can be achieved by direct comparison, while user data involves the user's personal privacy, and uploading or storing it in plaintext form to cloud servers can cause user privacy leakage. Encrypting data can protect user privacy effectively. In practical scenarios, different users use different keys for encrypting files, and there are random parameters in the encryption, then the ciphertext generated from the same file is different.

Therefore, there is an urgent need to design a secure de-duplication scheme for encrypted data with different keys in multi-user scenarios. Currently, convergent encryption [1] is widely used to construct secure data de-duplication systems, but convergent encryption also faces various dangers such as data leakage, faking attacks and chosen-plaintext attacks [2], [3], [4]. Since the encryption key used in convergent encryption is generated by the hash value of the user's data file, multiple files of the same user will generate multiple different keys, thus causing a key management problem [5]. The operations of encryption and de-duplication of data affect each other. Encrypting data with the same key by different users will generate the same ciphertext. Secure data de-duplication is achieved by directly comparing cipher

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

texts with each other, but this will cause the problem of key management. If different users use different keys to encrypt data, the key management problem can be effectively reduced, but it is difficult to achieve an equality test. Therefore, how different users can encrypt data with the same key without communicating with each other, thus producing the same ciphertext after encrypting the same data, and how users can recover their data are the main research directions of this paper.

II. LITERATURE SURVEY

Secure Cloud Storage: Review existing literature on secure cloud storage architectures, focusing on techniques for ensuring data confidentiality, integrity, and availability.

Explore different encryption schemes used in cloud storage systems to protect data at rest and in transit.

Examine methods for access control and authentication in cloud environments to prevent unauthorized access to stored data.

Data Deduplication:

Investigate previous research on data deduplication techniques, particularly in the context of cloud storage.

Explore approaches for identifying and eliminating duplicate data across multiple users to optimize storage utilization.

Review challenges and solutions associated with deduplicating data in encrypted form while preserving data privacy.

Searchable Encryption:

Survey literature on searchable encryption schemes that enable efficient searching over encrypted data.

Explore different types of searchable encryption, such as symmetric searchable encryption (SSE) and public-key searchable encryption (PEKS).

Examine the trade-offs between search efficiency and security in searchable encryption schemes.

Efficient Data Recovery:

Review existing methods for data recovery in cloud storage systems, particularly in scenarios involving encrypted data.

Explore techniques for securely recovering data from backups or redundant copies stored in the cloud.

Investigate approaches for efficient key management and access control to facilitate data recovery while maintaining security.

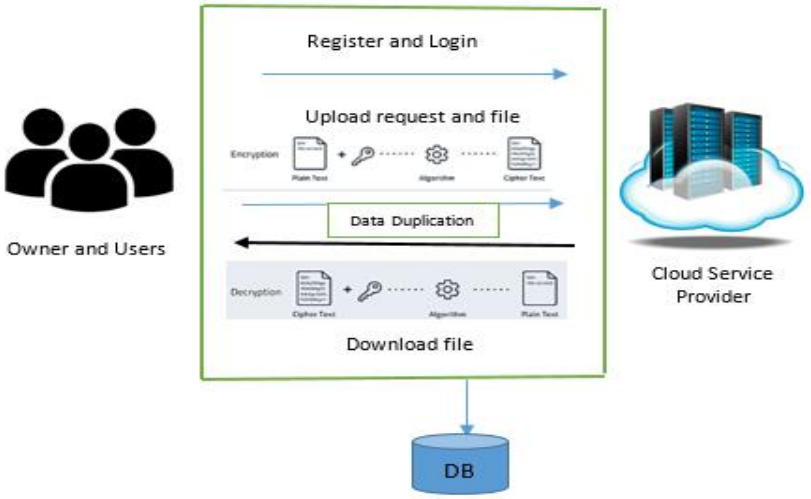
Security Analysis and Experimental Evaluation:

Examine previous studies that have conducted security analyses of encryption schemes, deduplication techniques, and data recovery methods.

Explore experimental simulation studies that evaluate the performance and effectiveness of proposed solutions in realistic cloud storage environments.

Identify gaps in existing research and areas for improvement or further investigation.

III. SYSTEM ARCHITECTURE



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

IV. EXPERIMENTAL RESULT

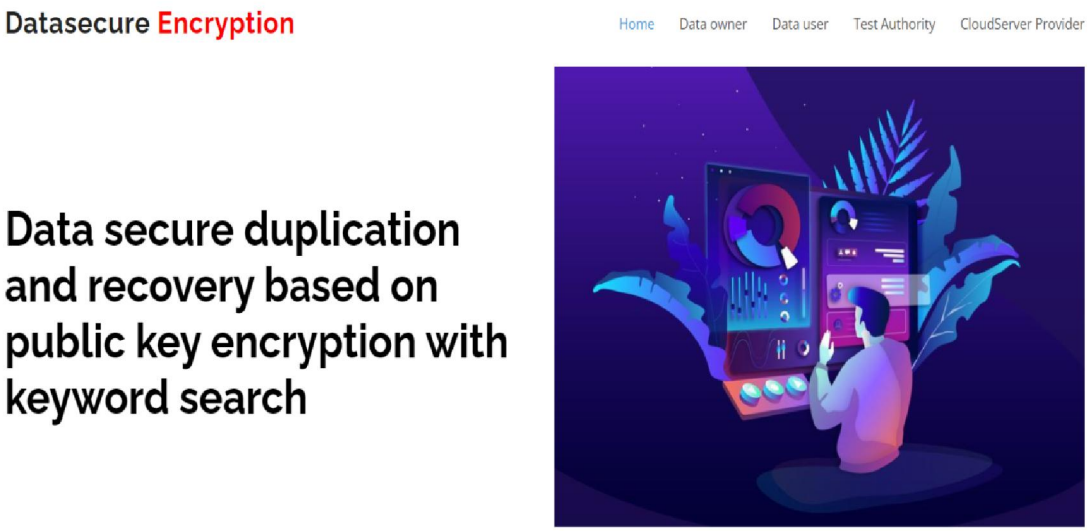


Fig 1. Home Page

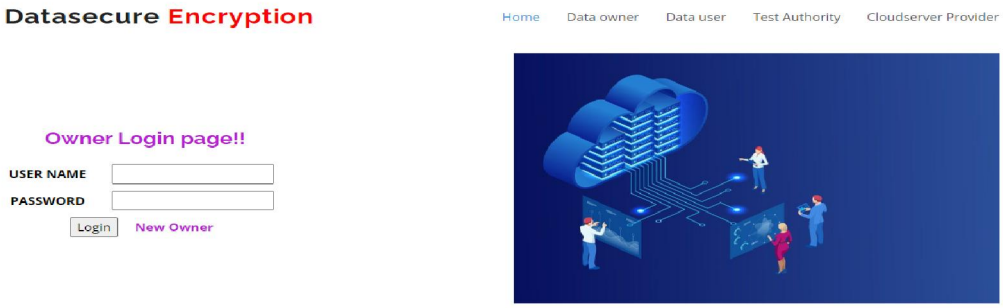


Fig 2. Owner Login Page

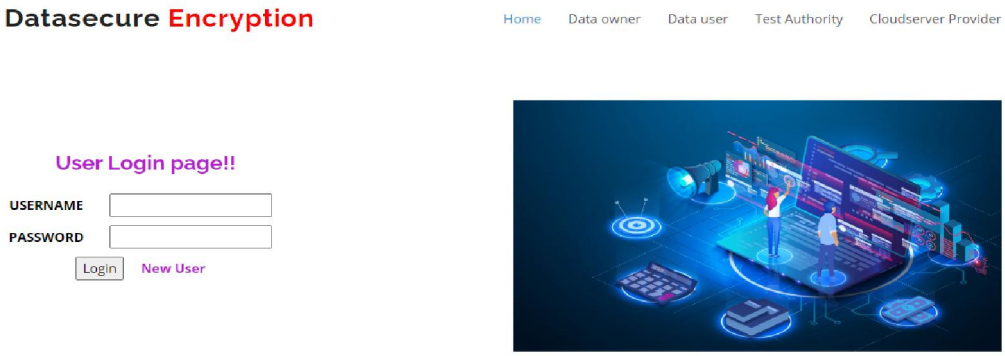


Fig 3. User Login Page

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

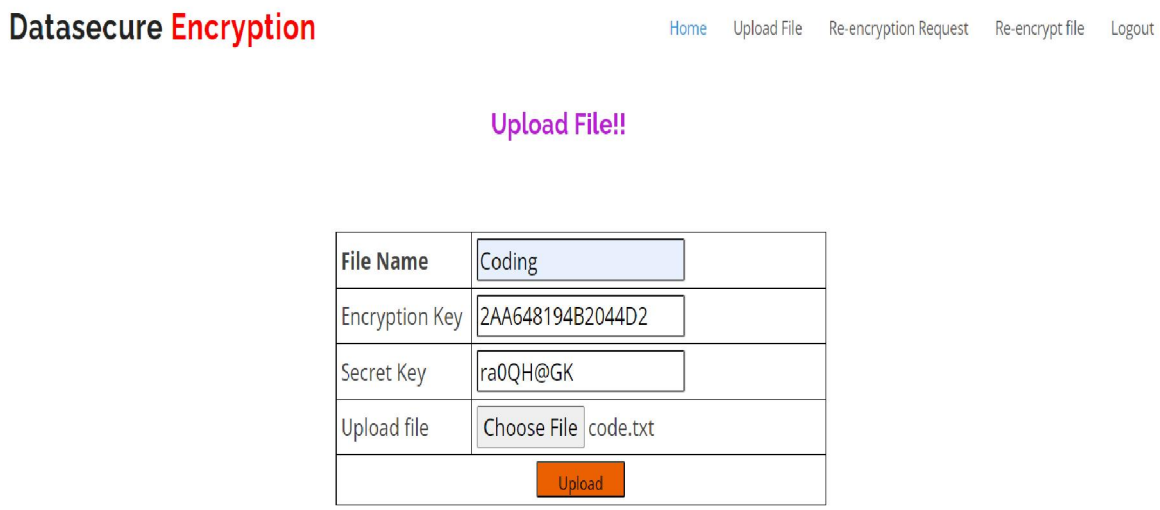


Fig 4. User Upload file details

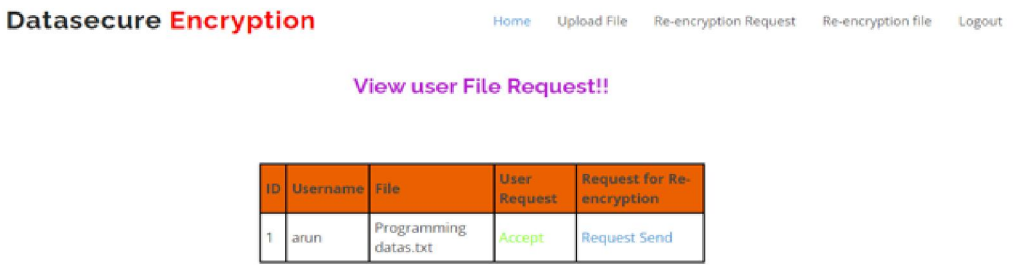


Fig 5. User File Request

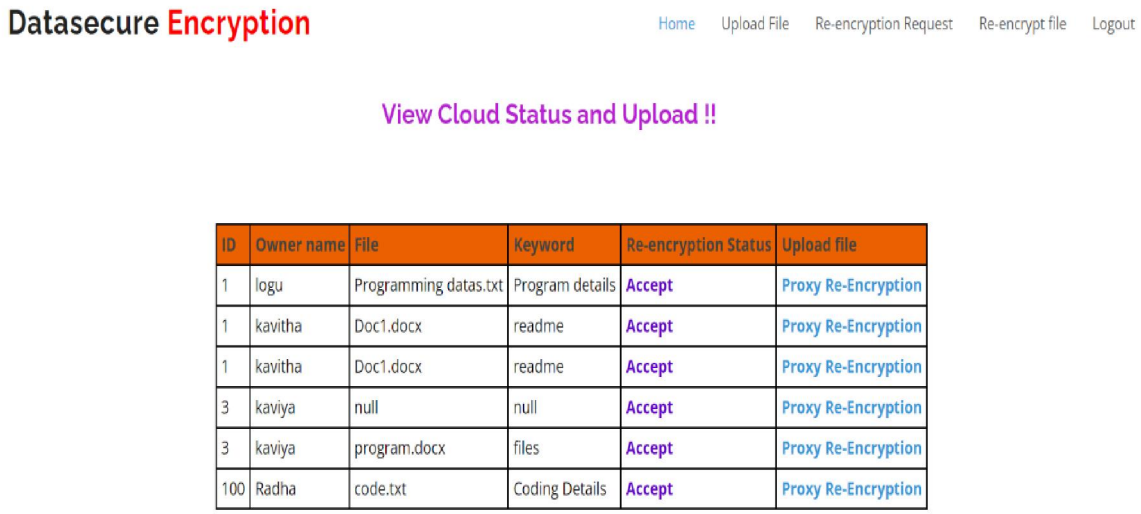
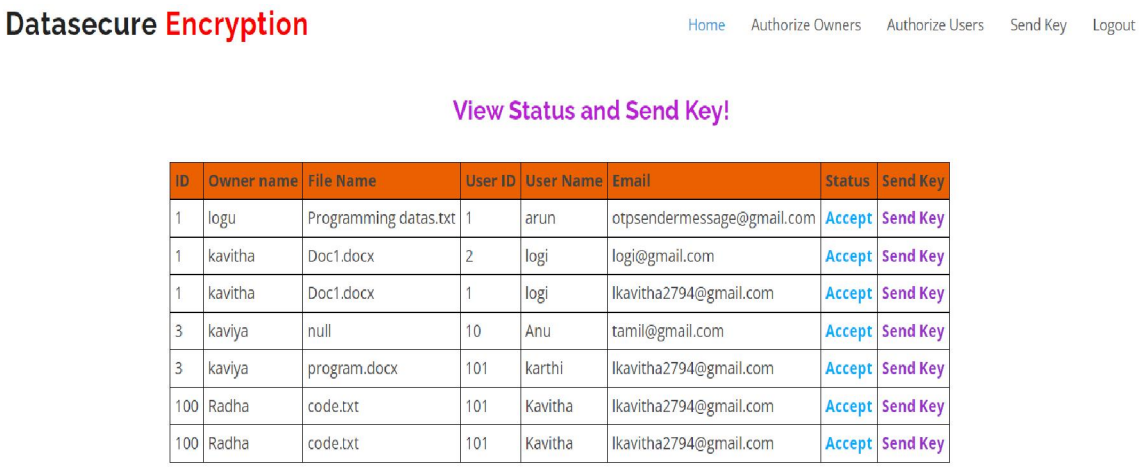


Fig 6. File Re-Encryption Request & Accept Details

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

ID	Username	Email	Keyword	File	Status
1	logu	otpsendermessage@gmail.com	Program details	Programming datas.txt	Accept
1	kavitha	logi@gmail.com	readme	Doc1.docx	Accept
1	kavitha	lkavitha2794@gmail.com	readme	Doc1.docx	Accept
3	kaviya	tamil@gmail.com	null	null	Accept
3	kaviya	lkavitha2794@gmail.com	files	program.docx	Accept
100	Radha	lkavitha2794@gmail.com	Coding Details	code.txt	Accept

Fig 7. View Re-Encryption File Request



ID	Owner name	File Name	User ID	User Name	Email	Status	Send Key
1	logu	Programming datas.txt	1	arun	otpsendermessage@gmail.com	Accept	Send Key
1	kavitha	Doc1.docx	2	logi	logi@gmail.com	Accept	Send Key
1	kavitha	Doc1.docx	1	logi	lkavitha2794@gmail.com	Accept	Send Key
3	kaviya	null	10	Anu	tamil@gmail.com	Accept	Send Key
3	kaviya	program.docx	101	karthi	lkavitha2794@gmail.com	Accept	Send Key
100	Radha	code.txt	101	Kavitha	lkavitha2794@gmail.com	Accept	Send Key
100	Radha	code.txt	101	Kavitha	lkavitha2794@gmail.com	Accept	Send Key

Fig 8. View File Status & Send Key

Fig 9. Enter Key & Download File

V. CONCLUSION

Secure data de-duplication is of great value in cloud storage, and it can effectively improve the space utilization of cloud storage systems. In this paper, a secure data de-duplication and recovery scheme based on PEKS is constructed

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

by using the matching relationship between keyword and trapdoor of searchable encryption to achieve file equality test in ciphertext state and using proxy re-encryption to achieve data recovery. Since the de-duplication process of a single file requires the execution of multiple equality test algorithms depending on the size of the database, this scheme is designed to avoid the computational overhead of this algorithm as much as possible. Through experimental simulation, the results show that the scheme in this paper has good performance in a cloud storage system. At present, scholars have made some achievements in the study of secure data de-duplication and have applied it to practical scenarios. This paper conducts in-depth research based on the previous work, but there are still some shortcomings, such as the current scheme of this paper only supports the equality test at the file level. In the future, the main consideration is the de-duplication rate. When the data user has two files with only minor differences, this paper will determine them as different files, which will reduce the de-duplication rate.

REFERENCES

[1] J. R. Douceur, A. Adya, W. J. Bolosky, P. Simon, and M. Theimer, ‘‘Reclaiming space from duplicate files in a serverless distributed file system,’’ in Proc. 22nd Int. Conf. Distrib. Comput. Syst., Vienna, Austria, 2002, pp. 617–624.
[2] A. Agarwala, P. Singh, and P. K. Atrey, ‘‘DICE: A dual integrity convergent encryption protocol for client side secure data deduplication,’’ in Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC), Banff, AB, Canada, Oct. 2017, pp. 2176–2181.
[3] P. Anderson and L. Zhang, ‘‘Fast and secure laptop backups with encrypted deduplication,’’ in Proc. 24th Large Installation Syst. Admin. Conf., San jose, CA, USA, 2010, pp. 1–12.
[4] D. Harnik, B. Pinkas, and A. Shulman-Peleg, ‘‘Side channels in cloud services: Deduplication in cloud storage,’’ IEEE Security Privacy, vol. 8, no. 6, pp. 40–47, Nov./Dec. 2010.
[5] J. Li, X. Chen, M. Li, J. Li, P. P. C. Lee, and W. Lou, ‘‘Secure deduplication with efficient and reliable convergent key management,’’ IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 6, pp. 1615–1625, Jun. 2014.
[6] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, ‘‘Public key encryption with keyword search,’’ in Proc. Int. Conf. Theory Appl. Cryptograph. Techn., C. Cachin and J. Camenisch, Eds. Interlaken, Switzerland, 2004, pp. 506–522.
[7] M. H. Au, J. Chen, J. K. Liu, Y. Mu, D. S. Wong, and G. Yang, ‘‘Malicious KGC attacks in certificateless cryptography,’’ in Proc. 2nd ACM Symp. Inf., Comput. Commun. Secur., New York, NY, USA, 2007, pp. 302–311.
[8] M. Bellare, S. Keelveedhi, and T. Ristenpart, ‘‘Message-locked encryption and secure deduplication,’’ in Proc. 32nd Annu. Int. Conf. Theory Appl. Cryptograph. Techn., T. Johansson and P. Q. Nguyen, Eds. Athens, Greece, 2013, pp. 296–312.
[9] S. Keelveedhi, M. Bellare, and T. Ristenpart, ‘‘DupLESS: Server-aided encryption for deduplicated storage,’’ in Proc. 22th USENIX Secur. Symp., S. T. King, Ed. Washington, DC, USA, 2013, pp. 179–194. [10] M. Abadi, D. Boneh, I. R. A. Mironov, and G. Segev, ‘‘Message-locked encryption for lock-dependent messages,’’ in Proc. 33rd Annu. Cryptol. Conf., Santa barbara, CA, USA, 2013, pp. 374–391.
[11] M. Abadi, D. Boneh, I. R. A. Mironov, and G. Segev, ‘‘Message-locked encryption for lock-dependent messages,’’ in Proc. 33rd Annu. Cryptol. Conf., Santa barbara, CA, USA, 2013, pp. 374–391.
[12] J. Liu, N. Asokan, and B. Pinkas, ‘‘Secure deduplication of encrypted data without additional independent servers,’’ in Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur., Denver, CO, USA, Oct. 2015, pp. 874–885.
[13] P. Puzio, R. Molva, M. Önen, and S. Loureiro, ‘‘PerfectDedup: Secure data deduplication,’’ in Proc. 10th Int. Workshop, 4th Int. Workshop, Vienna, Austria, 2015, pp. 150–166.
[14] J. Li, C. Qin, P. P. C. Lee, and J. Li, ‘‘Rekeying for encrypted deduplication storage,’’ in Proc. 46th Annu. IEEE/IFIP Int. Conf. Dependable Syst. Netw., Toulouse, France, Jun. 2016, pp. 618–629.
[15] M. Li, C. Qin, J. Li, and P. P. C. Lee, ‘‘CDStore: Toward reliable, secure, and cost-efficient cloud storage via convergent dispersal,’’ IEEE Internet Comput., vol. 20, no. 3, pp. 45–53, May 2016.