ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

Deduplication and Secure Auditing Of Data in Cloud with Convergent Key Management

Kunal S. Bhutwani, Prof. S. P. Kosbatwar

Department of Computer Engineering, SKNCOE, Savitribai Phule Pune University, Maharashtra, India

ABSTRACT: Data de-duplication is a strategy for wiping out copy duplicates of information, and has been generally utilized as a part of distributed storage to decrease storage room and transfer transmission capacity. Be that as it may, there is just a single duplicate for every document put away in cloud regardless of the possibility that such a record is claimed by countless. Therefore, de-duplication framework enhances stockpiling use while diminishing dependability. Moreover, the test of security for touchy information likewise emerges when they are outsourced by clients to cloud. Intending to address the above security challenges, this paper makes the primary endeavor to formalize the thought of conveyed dependable de duplication framework. In this paper new circulated de duplication frameworks with higher dependability in which the information pieces are dispersed over different cloud servers is being proposed.

KEYWORDS: Sec cloud, sec cloud+, integrity auditing, secure de-duplication, proof of ownership, convergent encryption.

I. INTRODUCTION

The distributed storage benefit (CSS) eases the weight for capacity administration and upkeep. In any case, if such an essential administration is helpless against assaults or disappointments, it would convey hopeless misfortunes to the customers in light of the fact that their information or documents are put away in a dubious stockpiling pool outside the ventures. Third-party auditor (TPA) who has expertise and capable to audit the outsourced data when needed. Public audit ability allows an external party, in addition to the user himself, to verify the correctness of remotely stored data In this way, it is fundamental for CSP to offer a productive review administration to check the respectability and accessibility of put away data. It is attractive that cloud just engages confirmation ask for from a solitary assigned gathering. To completely guarantee the information respectability and spare the cloud client's calculation assets and in addition online weight, it is of basic significance to empower open examining administration for cloud information stockpiling, with the goal that clients may depend on an autonomous outsider inspector (TPA) who has skill and proficient to review the outsourced information when required. Open review capacity permits an outer gathering, notwithstanding the client himself, to confirm the accuracy of remotely put away information This extreme disadvantage extraordinarily influences the security of these conventions in distributed computing. It is an endeavor to demonstrate the security by applying different systems and legitimize the execution of proposed plans through solid trials and examinations. It is our endeavor to give security to the cloud by just basically utilizing Kerberos frameworks for open review capacity. In particular, proposed plot accomplishes group examining where various assigned inspecting undertakings from various clients can be performed at the same time by the TPA in a protection safeguarding way.

II. RELATED WORK

Distributed computing framework has been anticipated as the cutting edge design of IT Enterprise. It moves the application programming and databases to the concentrated with vast server farms, where the administration of the information and administrations may not be completely dependable. This one of a kind ensample realizes numerous new security challenges, which have not been surely knew. Our examination work analyze the issue of guaranteeing the honesty of information stockpiling in Cloud Computing. Specifically, we consider the undertaking of permitting an outsider reviewer (TPA), on worry of the cloud customer, to check the uprightness of the dynamic information put

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

away in the cloud. While managing take a shot at guarantee remote information trustworthiness frequently do not have the backings of either open unquestionable status or element information operation. [1] Distributed storage frameworks are turning out to be increasingly prevalent. In this work we recognize assaults that endeavor customer side deduplication, conceding an aggressor to access discretionary size documents of different clients in light of a little hash mark of these records. All the more particularly, an aggressor who knows the hash mark of a document can guarantee the capacity benefit that it claims that record, consequently the server gives the assailant a chance to download the whole record. [2] Distributed storage specialist organizations, for example, Drop box, Mozy, and others perform deduplication to spare space by just putting away one duplicate of every record transferred. We propose an engineering that gives secure deduplicated stockpiling contradicting animal constrain assaults, and acknowledge it in a framework called DupLESS. In DupLESS, customers encode the under message-based keys got from a key-server by means of a careless PRF convention. It empowers customers to store encoded information with a present administration, have the administration perform deduplication for their sake, but accomplishes solid secrecy ensures. We demonstrate that encryption for deduplicated stockpiling can accomplish execution and space investment funds close to that of utilizing the capacity benefit with plaintext information.[3] We present a model for provable information ownership (PDP) that permits a customer that has put away information at an untrusted server to confirm that the server has the first information without recovering it. In this way, the PDP display for remote information checking bolsters huge information sets in generally - dispersed capacity frameworks.[4] We recommend a model for provable information ownership (PDP) that can be utilized for remote information checking: A customer that has put away information at an untrusted server can confirm that the server has the first information without recovering it. The model creates probabilistic evidences of ownership by examining irregular arrangements of pieces from the server, which radically decreases I/O costs. The customer keeps up a steady measure of metadata to confirm the verification. The test/reaction convention transmits a little, consistent measure of information, which minimizes organize correspondence. Therefore, the PDP display for remote information checking is lightweight and backings substantial information sets in dispersed stockpiling frameworks. The model is additionally hearty in that it fuses instruments for moderating subjective measures of information defilement.[5] The author Qian wang researched the new paradigm for security challenges in cloud where the management of data is not trustworthy. For overcome problem of integrity author introduces TPA (third party auditor) model, on behalf of the cloud client ,to verify the integrity of the dynamic data stored in the cloud. TPA Eliminates the involvement of client through the auditing of whether his data is stored in secure manner[11]. The popularity of cloud increasing now days, but problem is deduplication of data .Which is helpful for attacker that exploit client-side deduplication. Allowing an attacker to gain access to arbitrary-size files of other user based on very small hash signature of a file can convince the storage of these files.[12] The author jiawei yuan and shucheng yu improves storage security using proof-of-retrievability(POR) and proof-of-data possession(PDP) by removing unnecessarily duplicated data on storage server[14]. The author Karyn Benson, Hovav Shacham,BrentWaters propose Bilinear Map cryptographic algoritham. Which is Identity-Based Encryption. System based on. The author build bilinear map syatem that depend on weaker assumptions than the decisional-BHD assumption[15].The author Karyn Benson, Hovav Shacham,San Diego,Brent Waters proposed convergent key management for data deduplication. Using public key and private key decide the access for file and the secure transmission using keys on multiple server[16].

III. EXISTING SYSTEM APPROACH

Proposed a dynamic PDP schema but without insertion operation. Improved and supported insertion by introducing authenticated flip table. proxy PDP in public clouds. thecooperative PDP in multi-cloud storage. Improved the POR model by manipulating the classic Merkle hash tree construction for block tag authentication. To improve the POR schema with polynomial commitment for reducing communication cost. proposed a POR protocol over authenticated file system subject to frequent changes. combined the privacy-preserving word search algorithm with the insertion in data segments of randomly generated short bit sequences, and developed a new POR protocol. considered a new cloud storage architecture with two independent cloud servers for integrity auditing to reduce the computation load at client side.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

DRAWBACKS OF EXISITING SYSTEM

1. The first issue is trustworthiness reviewing. The cloud server can mitigate customers from the substantial weight of capacity administration and support. The most distinction of distributed storage from customary in-house stockpiling is that the information is exchanged by means of Internet and put away in a dubious area, not under control of the customers by any means, which unavoidably raises customers awesome worries on the trustworthiness of their information.
2. The second issue is secure deduplication. The fast appropriation of cloud administrations is joined by expanding volumes of information put away at remote cloud servers. Among these remote put away documents, the greater part of them are copied: by late study by EMC, 75% of late computerized information is copied duplicates.
3. Unfortunately, this activity of deduplication would prompt to various dangers conceivably influencing the capacity framework, for instance, a server telling a customer that it (i.e., the customer) does not have to send the record uncovers that some other customer has precisely the same, which could be touchy once in a while. These assaults start from the reason that the verification that the customer claims a given record (or piece of information) is exclusively in light of a static, short esteem (much of the time the hash of the document).

IV. PROPOSED SYSTEM APPROACH

1. In this paper, going for accomplishing information honesty and De-duplication in cloud, we propose two secure frameworks in particular SecCloud and SecCloud+.
2. SecCloud presents an inspecting element with upkeep of a MapReduce cloud, which helps customers create information labels before transferring and also review the trustworthiness of information having been put away in cloud.

Figure 1: Proposed System Model

3. Besides supporting honesty examining and secure deduplication, SecCloud+ empowers the certification of record secrecy.
4. We propose a strategy for straightforwardly examining respectability on scrambled information.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

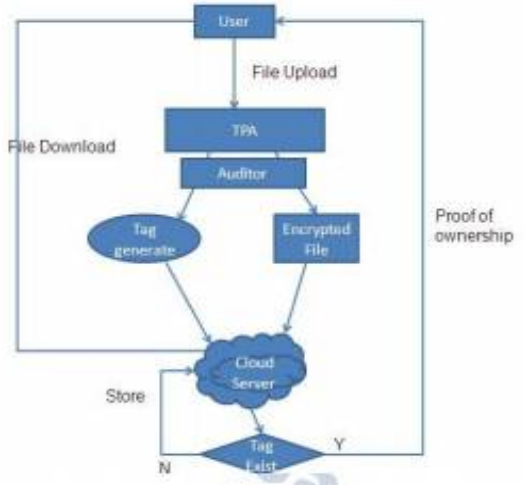

Figure 2: Proposed System Architecture

V. BENEFITS OF PROPOSED SYSTEM

[1] This plan settles the issue of past work that the computational load at client or inspector is excessively enormous for label era. For culmination of fine-grained, the usefulness of evaluating planned in SecCoud is bolstered on both piece level and division level. What's more, Sec cloud likewise empowers secure deduplication.
[2] The test of deduplication on encoded is the aversion of word reference assault.
[3] Our proposed SecCloud framework has accomplished both trustworthiness inspecting and record deduplication.

VI. MATHEMATICAL MODEL

Let us consider S as a system for CONCEPT BASED USER PROFILE.
S= {……
INPUT:

- Identify the inputs

F= {f_1, f_2, f_3 ……,f_n| 'F' as set of functions to execute commands.}
I= {i_1, i_2, i_3…|'I' sets of inputs to the function set}
O= {o_1, o_2, o_3….|'O' Set of outputs from the function sets}
S= {I, F, O}
I = {Data submitted by the user, ...}
O = {Output of desired data,...}
F = {Functions implemented to get the output, ASA Algorithm, RSA Algorithm.

VII. ALGORITHMS USED

1) AES Algorithm :

This symmetric encryption Algorithm which are AES is an iterative rather than Festal cipher. It is based on 'substitution–permutation network'. It comprises of a series of linked operations, some of which involve replacing inputs by specific outputs (substitutions) and others involve shuffling bits around (permutations).

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

2) Secure Key Distribution Algorithm DSS(Digital Signature Standard And Mail method)

Key are generated and given to particular user which are given to requested to key with mechanism of mail or sms to that user.

3)Key Generation RSA (Ron Rivest, Adi Shamir and Leonard Adelman) :

RSA is the algorithm used by modern computers to encrypt and decrypt messages. It is an asymmetric cryptographic algorithm. Asymmetric means that there are two different keys. This is also called public key cryptography, because one of them can be given to everyone.

Key generation has two phases. The first phase is a choice of algorithm parameters which may be shared between different users of the system, while the second phase computes public and private keys for a single user.

VIII. SOFTWARE AND HARDWARE REQUIREMENTS

Functional Requirement:
The non-functional requirements of the system are explained below as performance requirements and design constraints.

Performance requirements:

1. Accuracy
Since we will give the priority to the accuracy of the software, the performance of the Music Recommender will be based on its accuracy on recommendations.
2. Failure handling
System components may fail independently of others. Therefore, system components must be built so they can handle failure of other components they depend on.
3. Openness
The system should be extensible to guarantee that it is useful for a reasonable period of time.
Security requirements:

Sensitive information should be kept in safe.

Software quality attributes:

1. Usability
The software will be embedded in a website. It should be scalable designed to be easily adopted by a system.
2. Reliability
The system should have accurate results and fast responses to user's changing habits.

3. Security
User profile information will be used, so data security is one of the most important concern of the system.

EXTERNAL INTERFACE REQUIREMENT

Hardware Interfaces:

1. System: Pentium IV 2.4 GHz.
2. Hard Disk:40 GB
3. Floppy Drive: 44 Mb.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

4. Monitor : 15 VGA Color

Software Interfaces:

1. Operating system: Window 7,8,10.
2. Coding Language : Jdk 1.7
3. Database :MYSQL 5
4. Server apache tomcat 7
5. Services Web Based
6. IDE : Eclipse Luna
7. Front End Jsp
8. Back End Servlet/Data Base(MYSQL)

IX. EXPECTED RESULT

In this section, we will provide a thorough experimental evaluation of our proposed schemes. We build our test bed by using 64-bit t2.Micro Linux servers in Amazon EC2 platform as the auditing server and storage server. In order to achieve_ = 80 bit security, the prime order p of the bilinear group Gand GT are respectively chosen as 160 and 512 bits in length. We also set the block size as 4 KB and each block includes25 sectors.

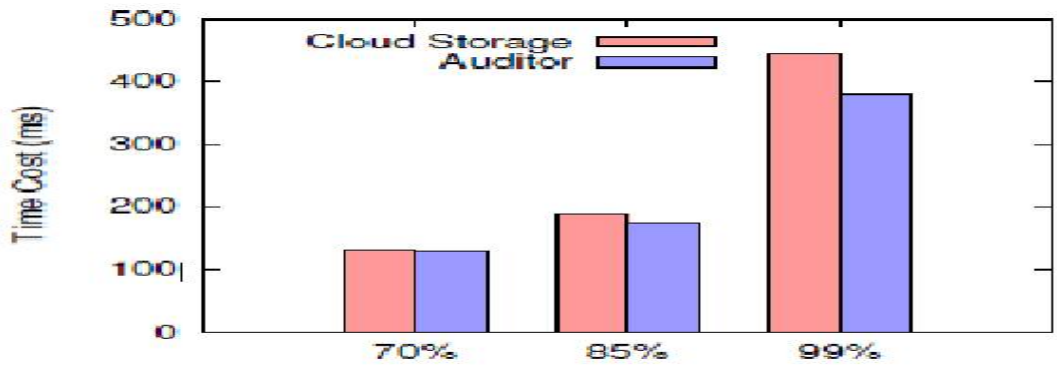

Figure 3: File Auditing

Now, we come back to evaluate the time cost of file auditing in Fig. 7, which shows the time cost of auditing for detecting the misbehavior of cloud storage respectively with 70%; 85% and 99% confidence. Obviously, as the growth of the number
of blocks for challenge (to guarantee higher confidence), the time cost for response from cloud storage server is increasing. This is because it needs to compute all the exponentiations for each challenge block as well as the coefficient for each column of S. Correspondingly, the time cost at auditor grows with the number of challenge blocks as well. But compared with cloud storage, the rate is slightly lower, because auditor only needs to aggregate the homomorphic signature of the challenged blocks.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

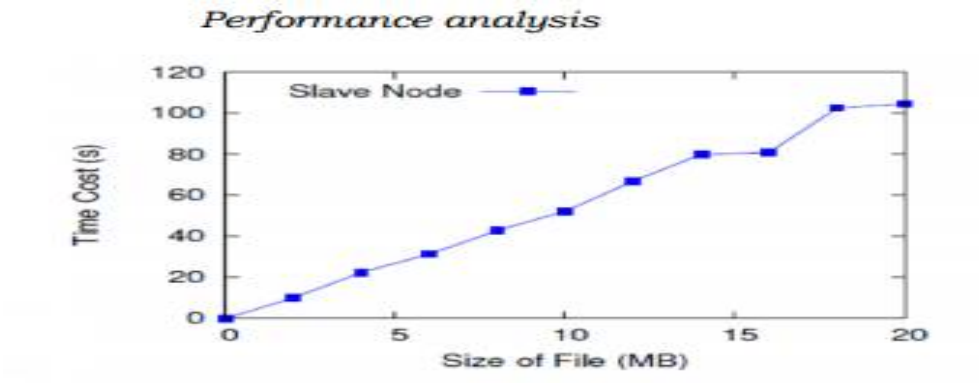

Figure 4: Tag Generation

Fig.4 shows the time cost of slave node in MapReduce for generating file tags. It is clear the time cost of slave node is growing with the size of file. This is because the more blocks in file, the more homomorphic signatures are needed to be computed by slave node for file uploading.

Results Table:
Table 1 Analysis of file uploading and downloading time of de duplicate files

File Size in Kb	Time for uploading in ms	Time For Downloading File
1	1.5	1
2	2	1.5
3	2.5	2
4	3	2.5

X. IMPLEMENTATION

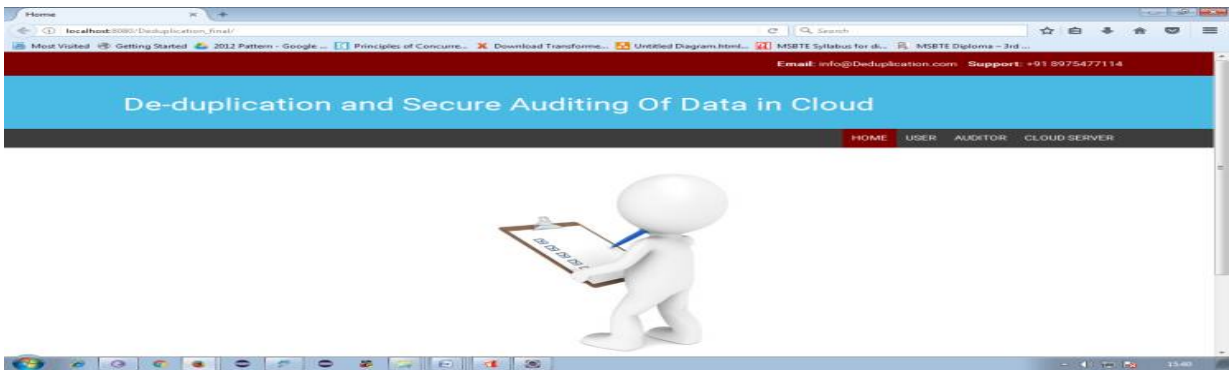

Figure 5 :Home Page

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

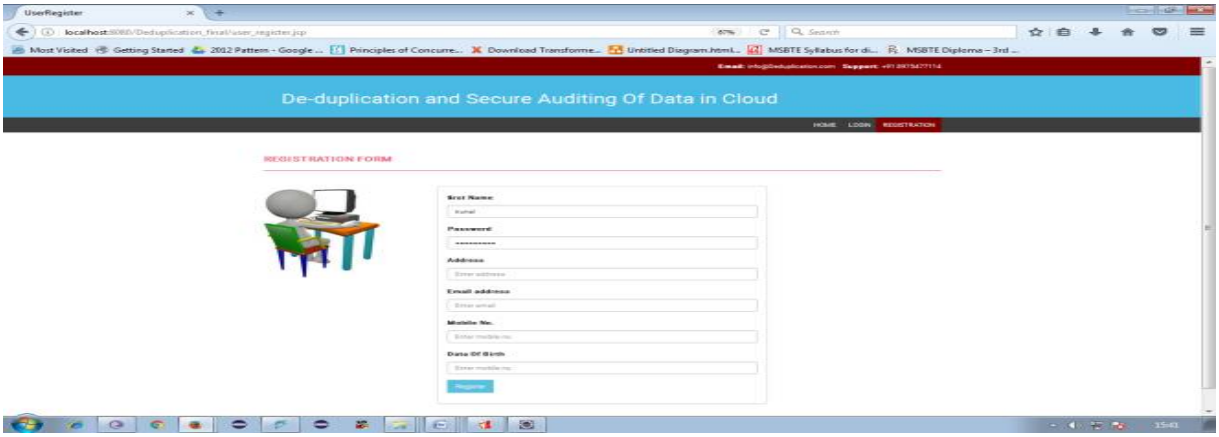

Figure 6 : Registration Page

Figure 7: Admin Login

X. FUTURE WORK

In addition, SecCoud enables secure deduplication through introducing a Proof of Ownership protocol and preventing the leakage of side channel information in data deduplication. Compared with previous work, the computation by user in SecCloud is greatly reduced during the file uploading and auditing phases. SecCloud+ is an advanced construction motivated by the fact that customers always want to encrypt their data before uploading, and allows for integrity auditing and secure deduplication directly on encrypted data.

XI. CONCLUSION

In this paper, we research the minimization of capacity taken a toll when the client stores its information in different untrustful what's more, temperamental mists. We give a lower bound on the expense what's more; present a coding plan that can accomplish this bound. This ideal plan can be comprehended through two-dimensional seek, which has low computational stockpiling frameworks.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

REFERENCES

[1] Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou, "Enabling public verifiability and data dynamics for storage security in cloud computing," in Computer Security – ESORICS 2009, M. Backes and P. Ning, Eds.,vol. 5789. Springer Berlin Heidelberg, 2009, pp. 355–370.
[2] S. Keelveedhi, M. Bellare, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in Proceedings of the 22Nd USENIX Conference on Security, ser. SEC'13. Washington, D.C.: USENIX Association, 2013, pp. 179–194.
[3] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems," in Proceedings of the 18th ACM Conference on Computer and Communications Security. ACM, 2011,pp. 491–500.
[4] H. Wang, "Proxy provable data possession in public clouds," IEEE Transactions on Services Computing, vol. 6, no. 4, pp. 551–559, 2013.
[5] Joseph K. Liu, Member, IEEE, Man Ho Au, Member, IEEE, Xinyi Huang, Rongxing Lu, Senior Member, IEEE, and Jin Li, "Fine-Grained Two-Factor Access Control for Web-Based Cloud Computing Services", IEEE transactions on information forensics and security, vol. 11, no. 3, march 2016
[6] Y. Zhu, H. Hu, G.-J. Ahn, and M. Yu, "Cooperative provable data possession for integrity verification in multicloud storage," IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 12, pp. 2231– 2244, 2012.
[7] Zhongma Zhu and Rui Jiang "A Secure Anti-Collusion Data Sharing Scheme for Dynamic Groups in the Cloud ," IEEE transactions on parallel and distributed systems, vol. 27, no. 1, january 2016
[8] Mr.Satish Shelar, Prof.S.Y.Raut,"Review On Deduplicating Data and Secure Auditing in Cloud", Volume: 02 Issue: 09 | Dec-2015.
[9] jie xu, qiaoyan wen, wenmin li, and zhengping jin, "circuit ciphertext-policy attribute-based hybrid encryption with verifiable delegation in cloud computing", IEEE transactions on parallel and distributed systems, vol. 27, no. 1, january 2016
[10] J. Xu and E.-C. Chang, "Towards efficient proofs of retrievability," in Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security, ser. ASIACCS '12. New York, NY, USA: ACM, 2012, pp. 79–80.
[11] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," Communication of the ACM, vol. 53, no. 4, pp.50–58, 2010.
[12] J. Yuan and S. Yu, "Secure and constant cost public cloud storage auditing with deduplication," in IEEE Conference on Communications and Network Security (CNS), 2013, pp. 145–153.
[13] S. Keelveedhi, M. Bellare, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in Proceedings of the 22Nd USENIX Conference on Security, ser. SEC'13. Washington, D.C.: USENIX Association, 2013, pp. 179–194. [Online].Available:https://www.usenix.org/confere nce/usenixsecurity13/technicalsessions/presentati on/bellare
[14] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable data possession at untrusted stores," in Proceedings of the 14th ACM Conference on Computer and Communications Security, ser. CCS '07. New York, NY, USA: ACM, 2007, pp. 598–609.
[15] G. Ateniese, R. Burns, R. Curtmola, J. Herring, O. Khan, L. Kissner, Z. Peterson, and D. Song, "Remote data checking using provable data possession," ACM Trans. Inf. Syst. Secur., vol. 14, no. 1, pp. 12:1–12:34, 2011