International Journal of Technical Research and Applications e-ISSN: 2320-8163,
www.ijtra.com Volume 4, Issue 2 (March-April, 2016), PP. 291-294

DATA DE-DUPLICATION BY USING HYBRID CLOUD

Akash veer, Charmy Turakhia, Deepali Ause, Piyush Gawali
Department of Information Technology
Pune, India
akashveer7@gmail.com

Abstract— **The cloud computing can make storage outsourcing in which the data having in secure in search literature. Some investigates having problem of efficient and secure public data for shared dynamic data. These are still not secure against the complicity and leakage of cloud storage server from attacker and revoked group users during the user termination in cloud storage system. In this paper, inspecting the integrity of shared data with dynamic group in cloud. According to the paper a new user can added into the group existing group member can be revoked privacy using data backup which is based on the vector commitment and verifier –local revocation group signature. This supports the public validation and efficient user revocation.**

General Terms

De-duplication, Asymmetric Group Key Agreement scheme (ASGKA), Vector commitment.

Index terms- **Dynamic data, cloud computing, Public integrity auditing.**

I. INTRODUCTION

The improvement in cloud computing motivates the organization as well as outsource their data to third party cloud service providers (CSP, s) which will result in improvements the data storage limitation of local devices. In market, already some cloud storage services are available like simple storage service (S3) [1] on-line data backup services of Amazon and software like Google Drive, [2] Dropbox, [3] Mozy, [4] Bitcasa and [5] Memopal built for cloud application. But in some cases the cloud server returns invalid results such as hardware/software failure, attack and maintenance.

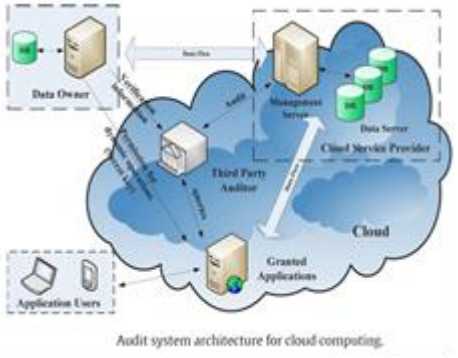
User's data should be protected by data integrity because of security and privacy. To avoid the security issues the cloud storage service, simple replication and protocols sufficient for practical application.

The only and only one the data owner cloud modify the data of the cloud. The development of cloud computing some application where the services are used as a collaboration platform. In this software environments, the one or many users can share source code as well as they need to access, compile, modify and run the code to share by any any user at any time. The new model of cooperation in cloud can provides the data for the remote data, where the data owner can updates its data. The result will be in communication and computation to the data owner which contended the single point of data owner. The data integrity based on ring signature, it does not consider the user revocation problem and the cost of auditing to the data size and group size. The authenticated and private channel exist between the pair of entities and then there is no collusion.Also,the costing of audit is linear to the size of group. Also another attempt to improve the scheme and the scalable and collusion with dynamic public integrity auditing scheme with group user revocation. The schema can supports plain text of data update and integrity auditing, so it's not a cipertext data. So the data owner can share key among the user to update their shared key. Also the owner cannot take part in the user revocation, where the user revocation is work itself as a cloud.

II. PROPOSED SYSTEM

In this paper, we study the problem of public authentication inspection for shared dynamic data with group user revocation. Our contributions are:

1. In cipher text database, we explore on secure and shared data for multi-user operation.
2. An efficient data auditing scheme with new futures such as traceability and countability by vector commitment primitives and group signature.
3. Finally the result shows that our scheme is secure. We provide the security and efficiency of our scheme which the result in back-up and the data storage on cloud.
4. Duplicate check the authorized in the hybrid cloud architecture supported by de-duplication and authorized duplicate check scheme with normal operations.



Audit system architecture for cloud computing.

5. Also in this project we can take backup for another server. This is the advantage of this project because sometimes

the data can crash form main browser then we can take backup from already we can save on the server.

III. SHA1 ALGORITHM

SHA1 algorithm is well defined in RFC 3174 - US Secure Hash Algorithm 1 (SHA1),

SHA1 algorithm consists of 6 tasks:

Tassk 1. Appending Padding Bits. The unique message is "padded" (extended) so that its length (in bits) is corresponding to 448, modulo 512. The padding rules are:

- The unique message is always padded with one bit "1" first.
- Then zero or more bits "0" are expanded to bring the length of the message up to 64 bits less than a multiple of 512.

Task 2. Adding Length. 64 bits are appended to the end of the expanded message to show the length of the original message in bytes. The rules of joining length are:

- The length of the unique message in bytes is transformed to its binary format of 64 bits. If extra happens, only the low-order 64 bits are used.
- Breakdown the 64-bit length into 2 words (32 bits each).
- The low-order word is joined first and tracked by the high-order word.

Task 3. Fixing Processing Functions. SHA1 requires 80 processing functions defined as:

f(t;B,C,D) = (B AND C) OR ((NOT B) AND D) (0 <= t <= 19)

f(t;B,C,D) = B XOR C XOR D (20 <= t <= 39)

f(t;B,C,D) = (B AND C) OR (B AND D) OR (C AND D) (40 <= t <= 59)

f(t;B,C,D) = B XOR C XOR D (60 <= t <= 79)

Task 4. Preparing Processing Constants. SHA1 requires 80 processing endless words defined as:

K(t) = 0x5A827999 (0 <= t <= 19)

K(t) = 0x6ED9EBA1 (20 <= t <= 39)

K(t) = 0x8F1BBCDC (40 <= t <= 59)

K(t) = 0xCA62C1D6 (60 <= t <= 79)

Task 5. Initializing Buffers. SHA1 algorithm requires 5 word buffers with the following initial values:

H0 = 0x67452301

H1 = 0xEFCDAB89

H2 = 0x98BADCFE

H3 = 0x10325476

H4 = 0xC3D2E1F0

Task 6. Processing Message in 512-bit Blocks. This is the main task of SHA1 algorithm, which loops through the expanded and attached message in blocks of 512 bits each. For each input block, a number of procedures are performed. This task can be described in the following pseudo code slightly modified from the RFC 3174's method 1:

Input and predefined functions:

M[1, 2, ..., N]: Blocks of the padded and appended message

f(0;B,C,D), f(1,B,C,D), ..., f(79,B,C,D): Defined as above

K(0), K(1), ..., K(79): Defined as above

H0, H1, H2, H3, H4, H5: Word buffers with initial values

Algorithm:

For loop on k = 1 to N

(W(0),W(1),...,W(15)) = M[k] /* Divide M[k] into 16 words */

For t = 16 to 79 do:

W(t) = (W(t-3) XOR W(t-8) XOR W(t-14) XOR W(t-16)) <<< 1

A = H0, B = H1, C = H2, D = H3, E = H4

For t = 0 to 79 do:

TEMP = A<<<5 + f(t;B,C,D) + E + W(t) + K(t)

E = D, D = C, C = B<<<30, B = A, A = TEMP

End of for loop

H0 = H0 + A, H1 = H1 + B, H2 = H2 + C, H3 = H3 + D, H4 = H4 + E

End of for loop

Output:

H0, H1, H2, H3, H4, H5: Word buffers with final message digest

Step 5. Output. The contents in H0, H1, H2, H3, H4, H5 are returned in sequence the message digest.

A. Group Signature with User Revocation

We define the definition of group signatures with valid user revocation as bellow,

Definition 2. It can consist of authorized group user is a collection of three polynomial-time algorithms, which are VLRKeyGen, VLRSign and VLRVerify as follow:

VLRKeyGen(n). This algorithm takes n parameter as a input where n represent number of group user. The output of the result is in group public key(gpk),an n-element vector of user keys gsk=(gsk(1),gsk(2),....,gsk(n)),the vector of user revocation tokens grt=(grt(1),grt(2)....,grt(n)).

VLRSig(gpk,gsk[i],M). This algorithm takes group of public key(gpk),a private key(gsk[i]) and a message M.

VLRVerify (gpk, RL,M). This algorithm takes group public key gpk, set of revocation tokens RL,M as a input parameter.

B. Supporting Cipher text Database

The outsourced data is usually stored in encrypted database, in previous research. This schema is designed for auditing of both plaintext and cipher text database. This is support for

encrypted database. The group consist of only one user that is data owner, then only need to choose random secrete key

And encrypt the data using encryption. when it needs to support the multiuser data modification, then it is difficult to keep the shared data for encryption, so that the single point can share a secrete key among the number of user. But there is chance of leakage of shared secrete key which break the shared data. So to avoid this problem, we use scheme, which supports multi-user group modification.

IV. PROBLEM FORMULATION

In this section, first we describe cloud storage model of system and second we provide the risk model and also security goals:

A. Cloud Storage Model

Cloud storage model consist of three entities, such as cloud storing server, a Third Party Auditor (TPA) and group users. The group user can contain of data owner and user who authorised to access and the data can be improved by the data owner. The group of user can providing the data storage services by the cloud storage server. The TPA can data integrity of the shared data store in the cloud server. In the remote storage cloud server the data owner could encrypt and upload its data. Also these data can access and modify and share to the number of group user.

B. Threat Model And Security Goals

This model consists of two types of attack:

1. The plain text of the data may be obtain by the attacker outside the group. This attacker can break the security of the group data encryption.

2.The cloud data storage server can revoked group users and then provide the data without being detected.

The cloud can make the data m' and in the user revocation it becomes valid to achieve the following security goals in this paper to overcome the problem as below:

- **a) Security:** It should check the user authenticity by password to verify user identity. By using digital signature it should satisfy privacy certifications.
- **b) Efficiency:** The efficiency for the any data computation as well as storage data issue can facilitated by any group user which is depend on the size of the shared data.
- **c) Countability:** According to improper storage server of the cloud tampered with database.
- **d) Tracebility:**In this the generation algorithm generates the data and the valid group signature, the data owner trace the last user who update the shared data.
- **e) Correctness:** Data updated by valid group user which is supports to encrypted database by correct result.

V. CONCLUSION

In this the database with efficient and secure updates is way to solve the problem of verifiable data storage. We implement a scheme to realize secure and efficient auditing of data for share dynamic data with multiuser modification. In this paper,

The concept of data duplication was proposed to achieve the data security by including privileges of user to check the duplication. This uses vector commitment, Asymmetric Group Key Agreement (AGKA) and group signatures with user revocation are used to achieve the auditing remote data integrity. Beside the public data auditing, the combining of the three primitive enable our scheme to outsource ciphertext database to remote cloud and support secure group users revocation to shared dynamic data.

This paper involve the successful implementation of data backup and efficient storage for maintain the confidentiality of shared data. We provide Security analysis of our scheme, and it prove that our Scheme provide confidentiality shared data for group users, Also, the performance analysis shows that our scheme also efficient in different phases as compare to its relevant schemes.

REFERENCES

[1] T. Jiang, X. Chen, and J. Ma, "Public integrity auditing for shared dynamic cloud data with group user revocation," in Proc. Of IEEE TRANSACTIONS ON COMPUTERS VOL: PP NO: 99 YEAR 2015

[2] Hugo E. Camacho, J. Alfredo Brambila, Alfredo Peña, José M. Vargas, "A cloud environment for backup and data storage," in Engineering Information Technology, Polytechnic University of Altamira, Altamira Tamaulipas, México

[3] Jin Li, Yan Kit Li, X. Chen, Patrick P. C. Lee, Wenjing Lou, "A Hybrid Cloud Approach for Secure Authorize Deduplication," in Proc. of IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEM VOL:PP NO:99 YEAR 2014

[4] Boyang Wang, Baochun Li, Member, IEEE, and Hui Li, Member, IEEE, " Oruta: Privacy-Preserving Public Auditing for Shared Data in the Cloud" in Proc. Of IEEE TRANSACTIONS ON XXXXXX, VOL. X, NO. X, XXXX 201X

[5] C. Wang Student Member, IEEE, Sherman S.-M. Chow, Q. Wang, Student Member, IEEE, K. Ren, Member, IEEE, and W. Lou, Member, IEEE , "Privacy-Preserving Public Auditing for Secure Cloud Storage".

[6] B. Wang, B. Li, Member, IEEE, and H. Li, "Panda: Public Auditing for Shared Data with Efficient User Revocation in the Cloud" in Proc. Of IEEE TRANSACTIONS ON XXXXXX, VOL. X, NO. X, XXXX 201X

[7] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for data storage security in cloud computing," in Proc. of IEEE INFOCOM 2010, CA, USA, Mar. 2010, pp. 525– 533.

[8] D. Catalano and D. Fiore, "Vector commitments and their applications," in Public-Key Cryptography - PKC 2013, Nara, Japan, Mar. 2013, pp. 55–72.

[9] B. Wang, L. Baochun, and L. Hui, "Public auditing for shared data with efficient user revocation in the cloud," in Proc. Of IEEE INFOCOM 2013, Turin, Italy, Apr. 2013, pp. 2904–2912