Productive Client Side Deduplication of Encrypted information in Cloud Storage using Public Auditing

Mr. Kate Ajay Somnath, Miss. Kotmire Nisarga Anil, Miss. Kare Varsha Tukaram, Prof. Nalawade V.S.
SPVP's S.B. Patil College of Engineering, Vangali, Indapur, Pune 413106.

Abstract: Cloud computing offers a replacement approach of service provision by re-arranging various resources over the net. the foremost necessary and well-liked cloud service is information storage. so as to preserve the privacy of information holders, data are often hold on in cloud in associate encrypted type. However, encrypted information introduce new challenges for cloud information deduplication, that becomes crucial for giant information storage and process in cloud. ancient deduplication schemes cannot work on encrypted information. Existing solutions of encrypted information deduplication suffer from security weakness. they can't flexibly support information access management and revocation. Therefore, few of them is promptly deployed in apply. during this paper, we propose a theme to deduplication encrypted information hold on in cloud supported ownership challenge and proxy re-encryption. It integrates cloud information deduplication with access management. we have a tendency to judge its performance supported in depth analysis and pc simulations. The results show the superior potency and effectiveness of the theme for potential sensible readying, particularly for giant information deduplication in cloud storage.

Keywords: Cloud storage, Cryptography, Data security, Public audit, Secure deduplication, Regeneration of data

Introduction: IN cloud storage services, purchasers source information to a remote storage and access the information whenever they have the data. Recently, due to its convenience, cloud storage services became widespread, and there's a rise in the use of cloud storage services. Well-known cloud services such as Dropbox and iCloud area unit utilized by people and businesses for varied applications. A notable modification in information-based services that is going on recently is that the volume of information employed in such services because of the dramatic evolution of network techniques. as an example, in 5G networks, gigabits of information may be transmitted per second, which implies that the dimensions of information that's dealt by cloud storage services will increase because of the performance of the new networking technique. during this viewpoint, we will characterize the amount of data as a main feature of cloud storage services. several service providers have already ready high resolution contents for their service to utilize quicker networks. For secure cloud services within the new era, it's vital to organize appropriate security tools to support this modification. Larger volumes of information need higher price for managing the various aspects of information, since the dimensions of information influences the cost for cloud storage services. the dimensions of storage should be hyperbolic per the number of information to be stored. during this viewpoint, it's fascinating for storage servers to reduce the amount of knowledge, since they will increase their profit by reducing the price for maintaining storage. On the opposite hand, purchasers area unit in the main curious about the integrity of their data keep within the storage maintained by service suppliers. To verify the integrity of keep files, purchasers have to be compelled to perform costly operations, whose complexness will increase in proportion to the scale of knowledge. during this viewpoint, purchasers might want to verify the integrity with a coffee price in spite of the scale of data. due to the strain of storage servers and purchasers, many researches on this subject area unit offered within the literature.

Literature Survey:

Paper 1. Publicly verifiable inner product evaluation over outsourced data streams under multiple keys

Author Name : X. Liu,W. Sun, H. Quan,W. Lou, Y. Zhang and H. Li

Description: Uploading data streams to a resource-rich cloud server for inner product evaluation, an essential building block in many popular stream applications (e.g., statistical monitoring), is appealing to many companies and individuals. On the other hand, verifying the result of the remote computation plays a crucial role in addressing the issue of trust. Since the outsourced data collection likely comes from multiple data sources, it is desired for the system to be able to pinpoint the originator of errors by allotting each data source a unique secret key, which requires the inner product verification to be performed under any two parties' different keys. However, the present solutions either depend on a single key assumption or powerful yet practically inefficient fully homomorphic cryptosystems. In this paper, we focus on the more challenging multi-key scenario where data streams are uploaded by multiple data sources with distinct keys. We first present a novel homomorphic verifiable tag technique to publicly verify the outsourced inner product computation on the dynamic data streams, and then extend it to support the verification of matrix product computation. We prove the security of our scheme in the random oracle model. Moreover, the experimental result also shows the practicability of our design.

Paper 2. Secure and constant cost public cloud storage auditing with deduplication

Author Name : J. Yuan and S. Yu

Description: Data integrity and storage efficiency are two important requirements for cloud storage. Proof of Retrievability (POR) and Proof of Data Possession (PDP) techniques assure data integrity for cloud storage. Proof of Ownership (POW) improves storage efficiency by securely removing unnecessarily duplicated data on the storage server. However, trivial combination of the two techniques, in order to achieve both data integrity and storage efficiency, results in non-trivial duplication of metadata (i.e., authentication tags), which contradicts the objectives of POW. Recent attempts to this problem introduce tremendous computational and communication costs and have also been proven not secure. It calls for a new solution to support efficient and secure data integrity auditing with storage deduplication for cloud storage. In this paper we solve this open problem with a novel scheme based on techniques including polynomial-based authentication tags and homomorphic linear authenticators. Our design allows deduplication of both files and their corresponding authentication tags. Data integrity auditing and storage deduplication are achieved simultaneously. Our proposed scheme is also characterized by constant realtime communication and computational cost on the user side. Public auditing and batch auditing are both supported. Hence, our proposed scheme outperforms existing POR and PDP schemes while providing the additional functionality of deduplication. We prove the security of our proposed scheme based on the Computational Diffie-Hellman problem, the Static Diffie-Hellman problem and the t-Strong Diffie-Hellman problem. Numerical analysis and experimental results on Amazon AWS show that our scheme is efficient and scalable.

Paper 3 Proofs of retrievability via hardness amplification
Author Name: Y. Dodis, S. Vadhan and D. Wichs

Description: For data storage outsourcing services, it is important to allow data owners to efficiently and securely verify that the storage sever stores their data correctly. To address this issue, several proof-of-retrievability (POR) schemes have been proposed wherein a storage sever must prove to a verifier that all of a client's data are stored correctly. While existing POR schemes offer decent solutions addressing various practical issues, they either have a non-trivial (linear or quadratic) communication complexity, or only support private verification, i.e., only the data owner can verify the remotely stored data. It remains open to design a POR scheme that achieves both public verifiability and constant communication cost simultaneously. In this paper, we solve this open problem and propose the first POR scheme with public verifiability and constant communication cost: in our proposed scheme, the message exchanged between the prover and verifier is composed of a constant number of group elements; different from existing private POR constructions, our scheme allows public verification and releases the data owners from the burden of staying online. We achieved these by tailoring and uniquely combining techniques such as constant size polynomial commitment and homomorphic linear authenticators. Thorough analysis shows that our proposed scheme is efficient and practical. We prove the security of our scheme based on the Computational Diffie-Hellman Problem, the Strong Diffie- Hellman assumption and the Bilinear Strong Diffie-Hellman assumption.

s

Paper 4. Short signatures from the Weil pairing

Author Name : D. Boneh, B. Lynn and H. Shacham

Description: We introduce a short signature scheme based on the Computational Diffie-Hellman assumption on certain elliptic and hyper-elliptic curves. The signature length is half the size of a DSA signature for a similar level of security. Our short signature scheme is designed for systems where signatures are typed in by a human or signatures are sent over a low-bandwidth channel.

Paper 5. Dynamic provable data possession

Author Name : C. Erway, A. Küpçü, C. Papamanthou and R. Tamassia

Description: We consider the problem of efficiently proving the integrity of data stored at untrusted servers. In the provable data possession (PDP) model, the client preprocesses the data and then sends it to an untrusted server for storage, while keeping a small amount of metadata. The client later asks the server to prove that the stored data has not been tampered with or deleted (without downloading the actual data). However, the original PDP scheme applies only to static (or append-only) files. We present a definitional framework and efficient constructions for dynamic provable data possession (DPDP), which extends the PDP model to support provable updates to stored data. We use a new version of authenticated dictionaries based on rank information. The price of dynamic updates is a performance change from O(1) toO(log n) (orO(nọ log n)), for a file consisting of n blocks, while maintaining the same (or better, respectively) probability of misbehavior detection. Our experiments show that this slowdown is very low in practice (e.g., 415KB proof size and 30ms computational overhead for a 1GB file). We also show how to apply our DPDP scheme to outsourced file systems and version control systems (e.g., CVS).

Proposed System:

We design a new scheme for secure and efficient cloud storage service. The scheme supports both secure deduplication and integrity auditing in a cloud environment. In particular, the proposed scheme provides secure deduplication of encrypted data. Our scheme performs PoW for secure deduplication and integrity auditing based on the homomorphic linear authenticator (HLA), which is designed using BLS signature. The proposed scheme also supports public auditing using a TPA (Third Party Auditor) to help low-powered clients. The proposed scheme satisfies all fundamental security requirements, and is more efficient than the existing schemes that are designed to support deduplication and public auditing at the same time.

Advantages:

- privacy-preserving, availability and accountability.
- Public Auditing
- Increase Data security

Architecture Diagram:

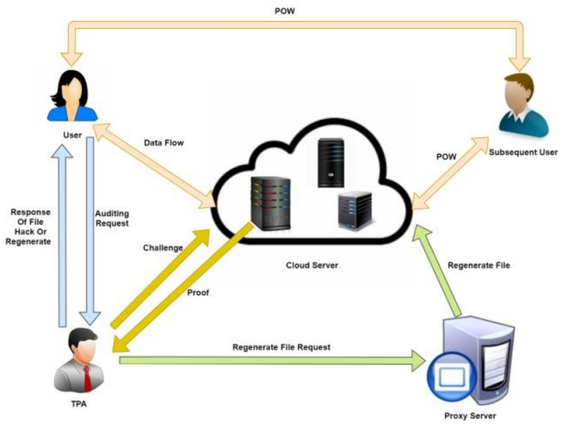

Figure 1 System Architecture

Mathematical Model:

System Description:
System S is defined as
S = {I,P,O,S,F}
where,

- I=Input
- O=Output
- P=Process
- S= Success
- F=Failure

- I= I:Set of outsourced data sets by corresponding data user
- O:store unique file on cloud server
- P: Identify the set of processes as P

P={PRE, TPA, Uo, SE, CSP, Sk}
where,
PRE= proxy re-encryption that store Re-encrypted Files as a backup.
TPA=Third Party Auditor: Perform auditing on users request.
Uo=set of owners that upload data files, if file is duplicate then send POW to the user that means user can access that file.
SE=Symmetric Encryption
CSP=Cloud Service Provider that store all users re-encrypted data
Sk=Symmetric Key used to encryption and decryption of the File

- Identify the initial condition as Ic

Ic= Outsourced data with its privacy privileges to be maintain

- Success Conditions:s

s=check duplicate file that is already store on cloud server If file already exist then duplicate file is not stored on cloud only give reference to new file.

- Failure Conditions:F

F=store duplicate file on cloud server and unable to find file ownership.

ALGORITHM

1. AES Algorithm:

AES 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).

AES performs all its computations on bytes rather than bits. Hence, AES treats the 128 bits of a plaintext block as 16 bytes. These 16 bytes are arranged in four columns and four rows for processing as a matrix.

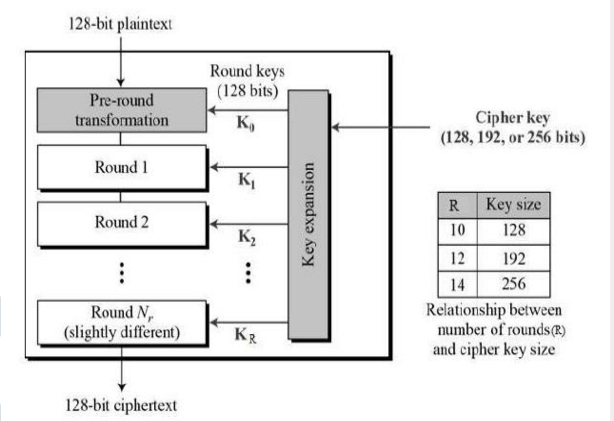

Figure 2 AES Algorithm

2. MD5 Algorithm

MD5 algorithm takes input message of arbitrary length and generates 128-bit long output hash. MD5 hash algorithm consist of 5 steps (described in deatil in Internet RFC 1321 [2]):

Step 1. Append Padding Bits
Step 2. Append Length
Step 3. Initialize MD Buffer
Step 4. Process Message in 16-Word Blocks
Step 5. Output

HARDWARE & SOFTWARE REQUIRMENTS

Software Requirements

Operating system	:	Windows XP/07/08/10.
Coding Language	:	JAVA/J2EE
IDE	:	Eclipse Kepler, XAMP
Database	:	MYSQL

Hardware Requirements

System	:	core i3
Hard Disk	:	40 GB.
Floppy Drive	:	1.44 Mb.
Monitor	:	15 VGA Colour.
Mouse	:	Logitech.
Ram	:	512 Mb.

Conclusion:

When storing information on remote cloud storage, users wish to be assured that their outsourced information square measure maintained accurately within the remote storage while not being corrupted. additionally, cloud servers wish to use their storage tons of expeditiously. To satisfy each of the necessities, we tend to plan a topic to realize each secure deduplication and integrity auditing in exceeding cloud surroundings. to prevent the outpouring of important info about user information, the planned theme supports client-side deduplication of encrypted information, whereas at an equivalent time supporting public auditing of encrypted information. we tend to used BLS signature based homomorphic linear critic to figure out authentication tags for the prisoner of war and integrity auditing. The planned theme glad the security objectives and improved the problems of the prevailing schemes. additionally, it provides higher potency than the prevailing schemes within the viewpoint of client-side machine overhead. Finally, we designed 2 variations for higher security and better performance. the first variance guarantees higher security within the sense that a legitimate user is associated with someone. The second variance provides higher performance from the attitude of the purchasers, by allowing weak purchasers to perform transfer procedure terribly expeditiously by passing on their expensive operations to the CSS.

References:

[1] X. Liu,W. Sun, H. Quan,W. Lou, Y. Zhang and H. Li, “Publicly verifiable inner product evaluation over outsourced data streams under multiple keys,” IEEE Transactions on Services Computing, vol. 10, no. 5, pp. 826-838, Sept.-Oct. 2017.

[2] J. Yuan and S. Yu, “Secure and constant cost public cloud storage auditing with deduplication,” in Communications and Network Security (CNS), 2013 IEEE Conference on, National Harbor, MD, USA, 2013, pp. 145-153.

[3] Y. Dodis, S. Vadhan and D. Wichs, “Proofs of retrievability via hardness amplification,” in Proc. of the 6th Theory of Cryptography Conference on Theory of Cryptography (TCC’09), San Francisco, CA, USA, 2009, pp. 109–127.

[4] D. Boneh, B. Lynn and H. Shacham, “Short signatures from the Weil pairing,” Journal of Cryptology, vol. 17, no. 4, pp. 297–319, Sept. 2004.

[5] C. Erway, A. Küpçü, C. Papamanthou and R. Tamassia, “Dynamic provable data possession,” in Proc. of the 16th ACM conference on Computer and communications security (CCS’09), Chicago, Illinois, USA, 2009, pp. 213–222.

[6] M. Dworkin, “Recommendation for block cipher modes of operation. methods and techniques,” NIST, USA, No. NIST-SP-800-38A., 2001.

[7] J. Gantz and D. Reinsel, “The digital universe decade - are you ready?,” IDC White Paper, 2010.

[8] A. Juels and B.S. Kaliski Jr, “Pors: proofs of retrievability for large files,” in Proc. of the 14th ACM conference on Computer and communications security (CCS’07), Alexandria, Virginia, USA, 2007, pp. 584–597.

[9] S. Keelveedhi and M. Bellare and T. Ristenpart, “DupLESS: serveraided encryption for deduplicated storage,” in Proc. of the 22nd USENIX Security Symposium (USENIX Security 13), Washington, D.C. USA, 2013, pp. 179–194.

[10] J. Li, J. Li, D. Xie and Z. Cai, “Secure auditing and deduplicating data in cloud,” IEEE Transactions on Computers, vol. 65, no. 8, pp. 2386–2396, Aug. 2016.

[11] H. Shacham and B. Waters, “Compact proofs of retrievability,” in Proc. of the 14th International Conference on the Theory and Application of Cryptology and Information Security, Advances in Cryptology - ASIACRYPT 2008, Melbourne, Australia, 2008, pp. 90–107.

[12] O. Goldreich, S. Goldwasser, and S. Micali. How to construct random functions. J. ACM, 33(4):792–807, Aug. 1986.

[13] V. Goyal. Reducing trust in the pkg in identity based cryptosystems. In Proceedings of the 27th annual international cryptology conference on Advances in cryptology, CRYPTO'07, pages 430–447, Berlin, Heidelberg, 2007. Springer-Verlag.

[14] P. Hawkes, M. Paddon, and G. G. Rose. On corrective patterns for the sha-2 family, 2004.

[15] X. Jia and C. Ee-Chien. Towards efficient provable data possession. In Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security, ASIACCS '12, Seoul, Korea, 2012.

[15] Q.Wang, C.Wang, K. Ren,W. Lou and J. Li, "Enabling public auditability and data dynamics for storage security in cloud computing," IEEE Transactions on Parallel and Distributed Systems, vol. 22, no. 5, pp. 847–859, Dec. 2011.

[18] J. Yuan and S. Yu, "Secure and constant cost public cloud storage auditing with deduplication," in Communications and Network Security (CNS), 2013 IEEE Conference on, National Harbor, MD, USA, 2013, pp. 145-153.