Survey on Scrambled Data Control with Dedupulication in Cloud Computing

J. Velmurugan*, A. Vishnukumar, S.K. Manigandan, V. Anupriya, S. Minnu Priyanga
Department of Information technology, Vel tech high tech Dr.rangarajan Dr. Sakunthala Engg College, Avadi, chennai-62
***Corresponding author: E-Mail: vel.jme@gmail.com**

ABSTRACT

Cloud computing is a type of internet based computing that provide shared computer processing resources and data to computer and other devices on demand. It is a model for enabling ubiquitous, on-demand access to shared pool of configurable computing resources which can be rapidly provisioned and released with minimal management effort. In this paper, we study the problem of auditing the integrity and providing security to de-duplication on cloud. Specifically, in order to achieve the integrity of data and eliminating the duplicate data in cloud, we provide the systems called as Security Cloud and Security Cloud+. Security Cloud is nothing but entity which is to be audited with the help of Map Reduce, usually provides the customer to generate the data which is to be tagged before uploading as well as auditing the data integrity in cloud. Security Cloud is greatly decreased before uploading and auditing the file. SecCloud+ is the fact that is to be motivated with the customer wants to encrypt the data before uploading, and enables the auditing and providing security to data which is to be encrypted.

KEY WORDS: Cloud Computing, Network, CSP, Data Holder, Data Owner, Privacy.

1. INTRODUCTION

Cloud computing is the practice of using a network of remote servers hosted on internet to store, manage and process the data rather than a local server. The word Cloud is commonly used in science to describe a large agglomeration of objects that visually appear from a distance as a cloud and describe any set of things whose details are not further inspected and a cluster for sever in network. Cloud computing provides resource pool that usually includes storage data space, internet, processing power of computer and user applications.
Cloud services example: file storage, networking sites, web mails.

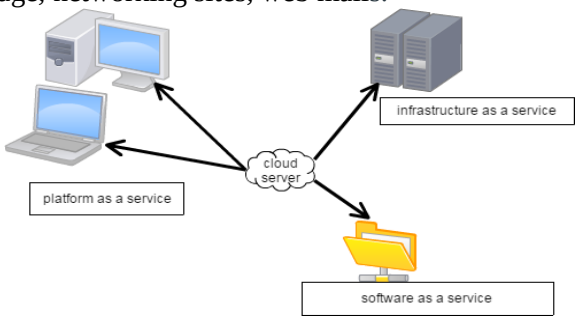


Figure.1. Cloud model

Existing System: An attribute-based storage system which employs CP-ABE and supports secure deduplication. The main contributions can be described as follows. a. Firstly, the system is the first that achieves the standard notion of semantic security for data confidentiality in attribute-based deduplication systems by resorting to the mixed character of cloud b. Secondly, method to identify the cipher text over the policy which is to be accessed in same cipher text over the plaintext under any other access policies without revealing the underlying plaintext. This technique might be free in adding the application in the proposed storage system. c. Thirdly, an approach based on two cryptographic primitives, including a zeroth proof of knowledge and a commitment scheme to achieve data consistency in the system. In storage system of typical data with secure deduplication, to store a file in the cloud, a data provider generates a tag and a cipher text.

The data provider uploads the tag and the cipher text to the cloud. Upon receiving an outsourcing request from a data provider for uploading a cipher text and an associated tag, the cloud runs a so-called equality checking algorithm, which checks if the tag in request is identical to any tags in the storage system.

If there is a match, then the underlying plaintext of this incoming cipher text has already been stored and the new cipher text is discarded. It is similar that the system with a tag appended to the cipher text does not provide the standard notion of semantic security for data confidentiality, because if the plaintexts can be predicated from their tags can always make a correct guess by computing the tag of a plaintext and then testing it against the tag in the challenge phase in the semantic security game.

Existing way to providing duplication of data are by avoiding the attack of brute force from data support which are flexible of control revocation (see the “Related Work in Data Deduplication” is the discussion of some other work in the area).Few schemes of existing access data control is to support data deduplication simultaneously, and few can ensure flexibility and security with sound performance that data owners control directly.

A scheme based on ABE to encrypt the data which is stored in cloud and to support the accessing of data control at the particular time. Analysis and implementation demonstrate is secure, effective and efficient. It includes the data's hash code M (H(M)) being applied as its indicator, which is used to check the duplication of data during data storage. We accept the holder of the data who signs the code hash function for right ownership for verification at the code. This code hash is given privacy and it makes difficult to attackers. We further guess that the owner of data has higher priority for data storage management. A data holder should provide valid proof to request special treatment. The CSP, data owners, and data holders make discussion with other through a secure channel. The scheme consists of several fundamental algorithms. We can adopt either: Algorithm1: CP-ABE or Algorithm2: KP-ABE.

Proposed System: In using advanced deduplication system supporting authorized duplicate check. In this new deduplication system, a hybrid cloud architecture is introduced to slove the problem. The private key for privileges will not be issued to user directly, which will be kept and managed by the private cloud server instead. In this way, the users cannot share these private keys of privileges in this proposed construction, which means that it can prevent the privilege key sharing among users in the above straight forward construction. To get the file token, the user needs to send a request to a private cloud server. The private cloud server will also check the users identity before issuing the corresponding file token to the user.

The authorized duplicate check for this file can be performed by the user with the public cloud before uploading this file. Based on the results of duplicate check, the user either uploads the file or runs POW. We use MD5 to check the integrity of the data. MD5 files use an algorithm that is based on the number of bits that a file should contain, so that a public verifier is able to audit shared data integrity without retrieving the entire data, yet it cannot distinguish who is the signer on each block. To improve the efficiency of verifying multiple auditing tasks have implemented.

The Registration involves in getting the information about the users who wants to use this application. Detect Deduplication and data encryption provides a confidentiality in deduplication.

Some of the advantages in proposed system is security issue will not be there, Privacy issues are minimized, Reducing the space required to store data in cloud.

To detect duplicates, the user first sends the server side of the tag to identical copy which have been stored already. The key of the convergent form and the tags are derived and cannot be used to deduce the key and compromise data. In data utilization, every user perform the operation on the block, and perform the ring formation on the block.

This scheme enables the auditor to audit the cloud data used by the user without retrieving the entire data from cloud. Both shared data and its verification data are stored in cloud.

A TPA of the public verifier gives the data expert of services of auditing form and group of data user to utilize data share and the ability to public verification of data share in cloud server.

Architecture Diagram: At the user side, each user can download an item, and the cipher text to decrypt with the key of private attribute generated by the user's attribute set satisfies the access structure. Each user checks decrypted message using the label, and accepts the message if it is consistent with the label.

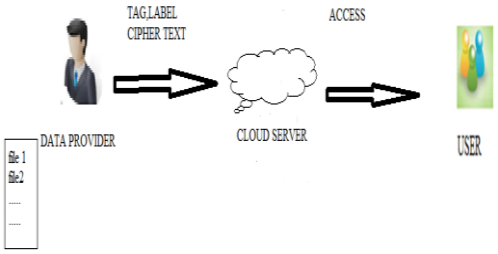


Figure.2. System architecture of scrambled data control

Concerning the model of our storage system, we guess the private cloud is "curious-but-honest" such that it will attempt to obtain the encrypted messages but it will follow the protocols. In addition to trying to obtain plaintext from cloud, malicious outsiders may also commit duplicate faking attacks. The architecture of our attribute-based system to secure deduplication includes data provider and user (fig.2). A data provider wants to outsource his/her cloud data it with users possessing certain credentials. At the user side, each user can download an item, and the cipher text to decrypt with the key generated.

Algorithm: The algorithm used for enhancing technique to detect deduplication is a.MD5 b.AES.

Algorithm1: MD5 algorithm can be used as a signature mechanism.

The message digest algorithm is that it takes input as arbitrary length and gives the output as 128 bit of input ram. These transformation function must fulfil these requirements:

- No one should produce the different inputs for which the transformation function returns the same output.
- No one should produce input for given pre specified output. Message-Digest algorithm applications for guaranteeing consistency (integrity) of data.

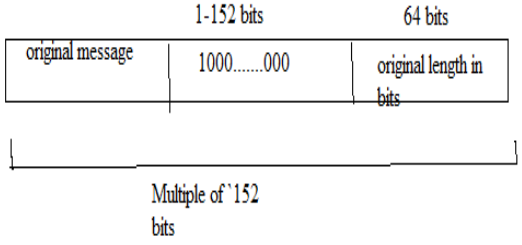
Commonly used model is as follows (message-digest in cooperation with asymmetric cryptography): Input message is created by the sender and digested message (sMD). The customer uses the private key and message digest is to encrypted (esMD) is attached to the whole message (M-esMD) is send to receiver.

Receiver gets the message (M-esMD) and the encrypted message digest (esMD). Then he uses his own message digest (rMD) of the received MD5. It takes the input message and generates the bit of 128 long output hash. MD5 hash algorithm consists of 5steps

Step 1.Append stepwise padding
Step 2.Append Length
Step 3.Initialize the step of MD buffer
Step 4.Process Message in 16-Word Blocks
Step 5.Output MD5 is commonly used hash algorithm.

It makes many implementations (available on some Unix-based system as utility md5 class MD5Crypto provider in Microsoft's. NET Framework example implementation in Visual C++. It is used sometimes as file CRC cipher in authentication operations (for storing user password hash).

MD5 was also used as cryptographic methods in DS applications or in SSL and others. The bit order within octet is most significant to least significant, the octet order is least significant to most significant.



Padding for MD5

Algorithm 2: The process of encryption was used as a set of specially derived keys called as round keys. The data that holds exactly one block of data is encrypted and are applied along with other operation. This type of array is called as state array.

Step:1. Derive the round keys from cipher key
Step:2. Initialize the plain text.
Step:3. Add the keys to the state array.
Step:4. Perform state manipulation
Step:5. Copy the encrypted data as a final state array.

In the encryption process each round requires a series of step to alter state array.

These steps involve four types of operations called: Sub Bytes Shift Rows Mix Columns Xor Round Key AES cipher which is used to replace AES commercially. And uses block size of 128 bit and key size of 128,192& 256 bits. AES does not use a Feistel structure. Also consist of four separate functions

- Byte substitution
- Permutation
- Arithmetic operation
- XOR operation with keys.

The AES structure and cipher is complex therefore it cannot be explained easily. The cipher takes a size of PT block as 128 bits or 16 bytes. The length of the key can be 16, 24 or 32 bytes. The algorithm used here is AES-428, AES-192 or AES-256. The encryption and decryption algorithm takes the input as 128 bit which is 4*4 matrix. Each word is 4 bytes, and the schedule key will be 44 words for 128 bit. The N rounds will be consists for the cipher which depends on the key length,10-rounds,12-rounds for 924 byte key and 14 rounds for 32 byte key.8 bit bytes will be operated by AES. In addition there is two bytes referred as bitwise XOR operation.

Storage system of cipher text-policy consists of the following algorithms:

- Setup algorithm
- Encryption algorithm Encrypt,
- Re-encryption algorithm Re-encrypt and
- Decryption algorithm Decrypt.

Setup (pars, msk): The input will be security parameter and it setup the output algorithm as public parameter and master private key for the system. KeyGen (pars, msk, A) skA: Public parameter and master private key is taken and A as the input, to generate an attribute based private key.

Encrypt(pars, M, A) (skT , CT): An access structure A over the universe of attributes is taken as a input , output algorithm as a encryption trapdoor key skT and a tuple CT = (T, L, ct, pf), where T and L are the tag and the label associated with M respectively, ct is the ciphertext which includes the encryption of M and it access the structure

A, and pf is a proof on the relationship of tag T, label L and ciphertext ct. d.

Re-encrypt (pars, skT, (L, ct), A') (L, ct'). Public parameter is taken, trapdoor key skT , a tag and cipher text pair (L, ct) and input as a access structure reencryption the outputs new cipher textct' associated with A' sharing the same label L of the cipher textct'. e.

Decrypt(pars, (L; ct), A, skA) Public parameter pars is taken, a label and cipher text pair (L; ct) and this decryption algorithm outputs either the message M when the private key skA satisfies the access structure of the cipher textct and the label L is consistent with M (to be defined later), or a symbol ? Indicating the failure of the decryption.

Formally, for all messages M, and all attribute sets A and access structures A with authorized A satisfying A, a. if (pars, msk)-> Setup, b.sKA<- KeyGen(pars, msk, A), c.(skT , CT) <-Encrypt(pars, M, A), then d. Decrypt(pars,(L, ct), A, skA) = M. (a)Setup (b)Key Gen (c)encrypt graph (d)Re-encrypt graph (e)decrypt graph .

3. CONCLUSION

Aiming at obtaining the integrity of data and security of duplication in cloud, we study security cloud and security cloud+. SecCloud provides and entity for auditing the maintenance of cloud by Map Reduce, which helps clients to data logs which are generated before auditing the integrity as well as uploading is usually stored in cloud. In adding to the above data, secuirty Cloud enables deduplication through ownership of protocol by the proof and the leakage of side channel information in data deduplication. SecCloud+ is the construction of the fact that the encrypted data is used by the customer before uploading the data and for auditing the integrity and the deduplication on data.

REFERENCES

Armbrust M, Above the clouds, A berkeley view of cloud computing. Tech. Rep. UCB/EECS-2009-28, EECS Department, U.C. Berkeley, 2009.

Ateniese G, Burns R, Curtmola R, Herring J, Kissner L, Peterson Z and Song D, Provable data possession at untrusted stores, In ACM CCS '07, 2007, 598–609.

Bellare M, Namprempre C and Neven G, Security proofs for identity-based identification and signature schemes, 2004.

Jin Li, Xiaofeng Chen, Mingqiang Li, Jingwei Li, Patrick Lee P.C and Wenjing Lou, Secure Deduplication with Efficient and Reliable Convergent Key Management, IEEE Transactions on Parallel And Distributed Systems, 25 (6), 2014.

Johnson R, Molnar D, Song D and Wagner D, Homomorphic signature schemes, In Proc. of CT-RSA, of LNCS, 2271, 2002, 244–262.

Li J, A Twin Cloud Approach for Secure Authorized Deduplication, IEEE Trans. Parallel Distributed Systems, 26 (5), 2015, 1206–1216.