Data Sharing Scheme Revisited in Cloud Computing

S.MohanBabuChowdary
Assistant Professor
MallapuBhavani,
M. Tech Student
M.Krishna
Associate Professor
SIR C. R. Reddy College of Engineering, Eluru, West Godavari Dt, AP, India

Abstract-As the cloud computing technology develops during the last decade, outsourcing data to cloud service for storage becomes an attractive trend, which benefits in sparing efforts on heavy data maintenance and management. Nevertheless, since the outsourced cloud storage is not fully trustworthy, it raises security concerns on how to realize data deduplication in cloud while achieving integrity auditing.In this work, we study the problem of integrity auditing and secure deduplication on cloud data. Specifically, aiming at achieving both data integrity and deduplication in cloud, we propose two secure systems, namely SecCloud and SecCloud+. SecCloud introduces an auditing entity with a maintenance of a MapReduce cloud, which helps clients generate data tags before uploading as well as audit the integrity of data having been stored in cloud. Compared with previous work, the computation by user in SecCloud is greatly reduced during the file uploading and auditing phases. SecCloud+ is designed motivated by the fact that customers always want to encrypt their data before uploading,and enables integrity auditing and secure deduplication on encrypted data.

Keywords-Seccloud,Seccloud+, integrity auditing, secure de-duplication, proof of ownership convergent encryption

I. INTRODUCTION

Cloud storage is a model of internet enterprise storage where data is stored in virtualized pools of storage which is hosted by third-party. Cloud storage provide offers for customer which generated more benefit for cloud companies, like popularity, more user. Even though now days cloud storage system has been smart option for work. And also it is affordable, but it has certain limitation .The main problem of client data management and maintenance which is able to Relief by cloud server storage system of cloud is different from another storage System. The first problem is integrity auditing, i.e when we uploaded data it upload various manner like packets tokens which is less secure because if any packet loss while transmitting it's occur problem for client. As well as it's to easy for a professional Attacker to attack. So it's most important that maintain the integrity of data on storage system. The data is transferred via internet and stored in uncertain domain not the under control of client [1]. The uncontrolled cloud server may passively hide the any problem related data for their reputation. It is more important that cloud server might even actively and deliberately discard rearely accessed data files belonging to an ordinary file. The second problem is secure deduplication. In cloud storage among these remote stored files, most of them are already on storage. According to recent survey by EMC, 70% of files are duplicated copies .Because its helps to cloud servers paid more for space from client. That's the one of the reason why many cloud server are store duplicate copies of data. And Its more risky to available duplicate copies of data in storage. Stored data is various manner lie confidential password, banking detail, personal information,it is open invitation for attacker .In cloud server, server store every single file link with the who ask for the file. Cloud server needs to verify whether the user actually owns the file before creating a link for user. In de-duplicate data, when a user wants to upload a data file that already exists in the cloud storage, the cloud server executes a checking algorithm to see whether or not this user actually possesses the whole file i.e. it checks the file attribute. If the user passes the checking, he/she can directly use the file existed on the server without uploading it again. To overcome such problems cloud server uses proofs-of-ownership protocol, which let a client efficiently prove to a server that the client holds a file, rather than short-information about it. In this a file have different ownership which introduce rigorous security definition. For working dynamic data proof-of-retrievability protocol used.Because dynamic data operation can be vital importance to storage outsourcing services.

II. EXISTING SYSTEM

Earlier client upload data file on cloud in plain text format. And wants to maintain the integrity and security on that plain data file. Customer always choose the safest and cheapest method for the data storage and transformation on cloud. but that's not possible to give all feature in such minimum amount .every system has some drawbacks and various problems. Existing system drawback

1. It is very difficult to audit the files huge and large amount of data in cloud using integrity auditing.
2. Lots of Duplicate files in cloud the number of security problems that are faced by cloud computing are Data issues Privacy issues Infected application B. Proposed System: To solve this problem on existing system we propose two secure system. Which generate better And Efficient system for accessing massive data on cloud. In this, firstly encrypted the plain data file and perform integrity auditing on that encrypted file. SecCloudSecCloud system has achieved both integrity

auditing and file deduplication.in this process server doesn't known the contain in the file. So here acceptance confidentially from sever is less secure .In other words, the functionalities of integrity auditing and secure deduplication are only imposed on plain files. SecCloud+ SecCloud+, which is used for maintaining integrity auditing and managing deduplication on encrypted files. In other Word perform the operation on the secure file. i.e encrypted files which encrypted by SecCloud over the plain text file. System Model Compared with SecCloud, our proposed SecCloud+ involves an Additional trusted entity, Namely key server, which is responsible for assigning clients with Secret key (according to the file content) for encrypting files.

Cloud computing minimizes the infrastructure risks in the context malfunction in e.g. equipment, systems, software or services; since in cloud computing a company or a user can access the cloud servers so it is not necessary to purchase the physical servers. This implies that when the work load in the context of processing capacity increases, a large number of servers needed, can be deployed quickly. Even if users are using a private cloud where they have their own servers installed, when their work load increases then this load can be shifted to the public cloud.

2. 3. Low cost of entry

Cloud computing reduces the installation and entry cost in new markets; no need for advanced IT infrastructure. The main reason of the low cost is that the infrastructure installed in the cloud computing is rented, therefore no need to purchase servers, so the initial investment can be zero.

3. Methodology

In the Era Smart Phone technology, Reliability is a big challenge to be considered in the smart Mobile environment. In smart Mobile environment different devices are integrated such as TV, microwave, washing machines, cameras, and telephones etc that integrate with each other and with the environment. Developer must make sure that the devices never crash. In smart Mobile user communicate and interact with the environment in order to perform daily routine functions with the help of sensors and cameras. To some degree the environment is aware of the user and its surroundings. The challenge for the designer is to create a system that ensure the user understands the realistic presence of sensors, other devices, interpretation and machine actions in their Mobiles. The audio visual entertainment requires a smart TV in order to learn the family preferences, access control, record the desired program on given schedule from different channels. This entertainment category is to provide the music according to the user preferences in any location, time and space in the Mobile.

There are four implementation models and three service models for cloud computing that can be implemented depending on the user requirements. Implementation Model which makes cloud as the best of features is listed below.

Low agility and resource intensive management

In the System of automated just-in-time fulfillment without human intervention by the service suppliers most expensive and on demand service. More flexible, responsive and innovative services which were inherently built to be both enduring and flexible to service changing business needs. As of the process of technical up gradation happens in the IT which we sometimes call it as Digital Revolution and if we consider the cloud implementation where we need to upgrade the best of techno local service work at home and revolution will come very soon.

Low utilization of assets

In this aspect we have put forward the approach of the as much as less utilization of asset which leads the optimum level of asset utilization. Greater sharing of resources, improved economies of scale and the ability to closely match allocated resources to demand

Aging technology requiring continual refresh and upgrades

Technically aging and its upgrading is the most changeling terminology where we put to put best for every release. An evergreen model where lifecycle management is predominantly the responsibility of the service provider and quality improvements and cost reductions are driven by highly-competitive market forces

Cost Effective and Technical Managed Modelcal

This is most efficient and typical feature where we like to put forward the feature like Cost effective, i.e. Cloud means cost effective and without this feature cloud has no meaning. Consumption of well-defined, standardized and highly-configurable shared services which continue to evolve and innovate based upon the needs of a large and diverse customer base and is paid for by many customers.Hence, we consider the above implementation features as of to put forward as the mechanism for our research based paper.

III. LITERATURE SURVEY

A. Integrity Auditing

The definition of provable data possession (PDP) was introduced by Ateniese et al. [5][6] for assuring that the cloud servers possess the target files without retrieving or downloading the whole data. Essentially, PDP is a probabilistic proof protocol by sampling a random set of blocks and asking the servers to prove that they exactly possess these blocks, and the verifier only maintaining a small amount of metadata is able to perform the integrity checking. After Ateniese et al.'s proposal [5], several works concerned on how to realize PDP on dynamic scenario: Ateniese et al. [7] proposed a dynamic PDP schema but without insertion operation; Erway et al. [8] improved Ateniese et al.'s work [7] and supported insertion by introducing authenticated flip table; A similar work has also been contributed in [9]. Nevertheless, these proposals [5][7][8][9] suffer from the computational overhead for tag

generation at the client. To fix this issue, Wang et al. [10] proposed proxy PDP in public clouds. Zhu et al. [11] proposed the cooperative PDP in multi-cloud storage. Another line of work supporting integrity auditing is proof of retrievability (POR) [12]. Compared with PDP, POR not merely assures the cloud servers possess the target files, but also guarantees their full recovery. In [12], clients apply erasure codes and generate authenticators for each block for verifiability and retrievability. In order to achieve efficient data dynamics, Wang et al. [13] improved the POR model by manipulating the classic Merkle hash tree construction for block tag authentication. Xu and Chang [14] proposed to improve the POR schema in [12] with polynomial commitment for reducing communication cost. Stefanov et al. [15] proposed aarranged two approaches for recognizing every single jobless tenet Prior work on inter firewall joblessness evacuation requires the data of two firewall techniques and along these lines is just suitable inside one directorial area.POR protocol over authenticated file system subject to frequent changes. Azraoui et al. [16] 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. Li et al. [17] considered a new cloud storage architecture with two independent cloud servers for integrity auditing to reduce the computation load at client side. Recently, Li et al. [18] utilized the key-disperse paradigm to fix the issue of a significant number of convergent keys in convergent encryption.

B. Secure Deduplication

Deduplication is a technique where the server stores only a single copy of each file, regardless of how many clients asked to store that file, such that the disk space of cloud servers as well as network bandwidth are saved. However, trivial client side deduplication leads to the leakage of side channel information. For example, a server telling a client that it need not send the file reveals that some other client has the exact same file, which could be sensitive information in some case. In order to restrict the leakage of side channel information, Halevi et al. [3] introduced the proof of ownership protocol which lets a client efficiently prove to a server that that the client exactly holds this file. Several proof of ownership protocols based on the Merkle hash tree are proposed [3] to enable secure client-side deduplication. Pietro and Sorniotti[19] proposed an efficient proof of ownership scheme bychoosing the projection of a file onto some randomly selectedbit-positions as the file proof.Another line of work for secure deduplication focuses onthe confidentiality of deduplicated data and considers to makededuplication on encrypted data. Ng et al. [20] firstly introducedthe private data deduplication as a complement of publicdatadeduplication protocols of Halevi et al. [3].Convergentencryption [21] is a promising cryptographic primitive forensuring data privacy in deduplication. Bellare et al. [22] Formalized this primitive as message-locked encryption, and explored its application in space-efficient secure outsourced storage. Abadi et al. [23] further strengthened Bellare et al's security definitions [22] by considering plaintext distributions that may depend on the public parameters of the schemas. Regarding the practical implementation of convergent encryption for securing deduplication, Keelveedhi et al. [4] designed the DupLESS system in which clients encrypt under file-based keys derived from a key server via an oblivious pseudorandom function protocol. As stated before, all the works illustrated above considers either integrity auditing or deduplication, while in this paper, we attempt to solve both problems simultaneously. In addition, it is worthwhile noting that our work is also distinguished with [2] which audits cloud data with deduplication, because we also consider 1) outsource the computation of tag generation, 2) audit and deduplicating encrypted data in the proposed protocols.We additionally consider an adaptable clash determination technique to empower a fine-grained strife determination with the assistance of a few compelling determination methodologies as for the danger evaluation of ensured systems and the aim of approach definition. In our system strife location and determination, clashing portions are distinguished in the initial step. Each clashing fragment partners with a strategy strife and an arrangement of clashing standards. Likewise, the connection connections among clashing sections are distinguished and strife relationship gatherings are determined. Strategy clashes having a place with various clash connection gatherings can be determined independently, therefore the scanning space for determining clashes is lessened by the connection procedure.

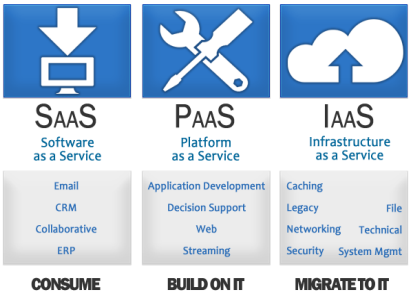


Fig. 3.1 Illustration of the Social TV Connecting the HD of Cloud

Today we have TV, Digital Box, smart phones, security alarms, digital locks, computers, game consoles, and smart Mobile appliances in our Mobiles. Usually, these devices are not connected to each other since these are often from different vendors and run on separate platforms.

Cloud Services

Cloud computing provides services for all the needs ranging from hardware to end user applications. It allows for rental access to hardware resources such as servers, storages, routers, switches and also provides required application for

end users on demand. It proposed three categories of services SaaS, IaaS, and PaaS. These categories are described as following.

Software as a Service (SaaS): is on demand service that provides a complete software application. The software is installed on a single computer in the organization and multiple users access it over the cloud within the organization.

Infrastructure as a service (IaaS): provides the all solution required to build an information technology (IT) infrastructure that usually consists of equipment, systems, software, and services. It provides storage and computing features as service on network. For hardware concerns it provides servers, switches, storage solutions, and routers, etc, and for computing purpose it provides all kind of applications from simple to high performance applications.

Platform as a service (PaaS): provides platform as a service to create high level services. The platform is equipped with all the resources required such as operating systems, application software, security, middleware, storage, programming language and development environment.

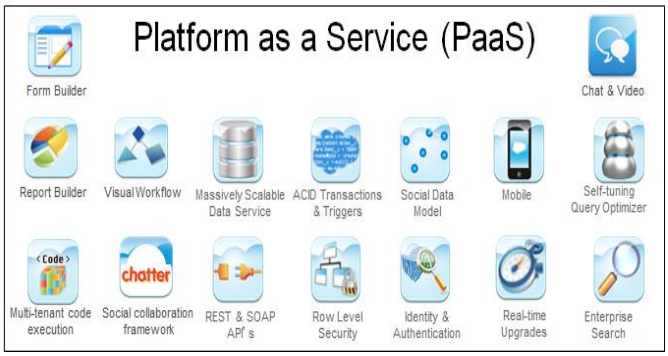


Fig. 3.2 Illustration of Platform as a Service

RELATED WORK

IV. PROTOCOL

We used different protocol to define the working of system. In this we used 3 different protocols on data file. File Uploading Protocol In this protocol client upload data files with the help of Auditor. For uploading data file it fulfil the requirement of uploading protocol. Specifically, the file uploading protocol includes three Steps:

• Phase 1 (cloud client → cloud server): in this phase when client upload the data file, firstly checks the duplication of file. If file is already present then perform POW (Proof of Ownership)on that file.

• Phase 2 (cloud client → auditor): After performing Client side operation client send data file to auditor, and receives a receipt from auditor. • Phase 3 (auditor → cloud server): after receive data file from client auditor preform some task like-helps generate a set of tags for the uploading file, and send them along with this file on cloud server Integrity Auditing Protocol This protocol work on the maintaining integration of data file. i.e perform the verification on data file This protocol includes two Phase:

• Phase 1 (cloud user/auditor → cloud server): verifier (i.e., client or auditor) generates a set of challenges and sends them to the prover (i.e., cloud server). • Phase 2 (cloud server → cloud user/auditor): on the basis of stored file, prover (i.e., cloud server) tries to prove that it exactly owns the target file by sending the proof back to verifier (i.e., cloud client or auditor). At the end of this protocol, integrity verification is done on file, if the verifier output become true. Proof of Ownership Protocol It is an interactive protocol which run on cloud server to verify the client, in this client play the role of prover to cloud server for its own claimed file. This protocol also includes two phase.

• phase 1 (cloud server → user): cloud server generates a set of challenges and sends them to the client. Server paly important role here, because the basic principle is file can access only authorized user .if any unauthorized get the file access it become dangerous for client as well as cloud server also.

• phase 2 (user → cloud server): ones the client responds with the proof for file ownership, and cloud server finally verifies the validity of proof. The access granted for data file without any challenges.

V. WORKING OF SYSTEM MODEL MODULES DESCRIPTON

Cloud Clients Cloud Client is fixed hardware or software which perform accessing and storing data on virtual server. It is important because cloud services are useless if client not used. It may be any device like computer, mobile, browser etc. Cloud Servers Cloud Server is nothing but virtual pool server which provide different services for different client with the help of internet. It is service – oriented architecture which provides high-capacity network, low-cost, hardware-virtualization. It is work platform independent. Its support various devices with their compatibility. Auditor Auditor is the system or manual software which helps clients upload and audit their out-sourced data maintains a Map Reduce cloud and acts like a certificate authority. This assumption presumes that the auditor is associated with a pair of public and private keys. Its public key is made available to the other entities in the system.

ADVANTAGE OF SYSTEM:

1. It provides the Integrity auditing by clustering the files with removing the duplicate
2. The duplicate files are mapped with a single copy of the file by mapping with the existing file in the cloud files.

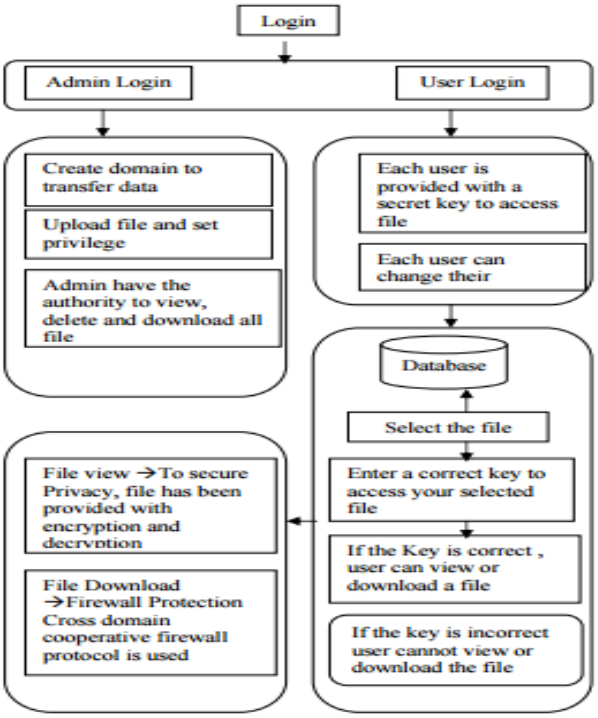


Fig. 5. 1 Architecture of Data Sharing Scheme

VI. ALGORITHM

A. Bilinear Map and Computational Assumption Definition 1 (Bilinear Map): Let G and GT be two cyclic multiplicative groups of large prime order p. A bilinear pairing is a map e : G × G → GT with the following properties:

B. • Bilinear: e(ga 1 ; gb2) = e(g1; g2)ab for all g1; g2 ∈R G and a; b ∈R Zp; • Non-degenerate: There exists g1; g2 ∈ G such that e(g1; g2)≠ 1;

C. • Computable: There exists efficient algorithm to compute e(g1; g2) for all g1; g2 ∈R G. The examples of such groups can be found in supersingular elliptic curves or hyperelliptic curves over finite fields, and the bilinear pairings can be derived from the Weil or Tate pairings. For more details, We then describe the Computational Diffie-Hellman problem, the hardness of which will be the basis of the security of our proposed schemes. Definition 2 (CDH Problem): The Computational Diffie-Hellman problem is that, given g; gx; gy∈ G1 for unknown x; y ∈ Z∗ p, to compute gxy. B. Convergent Encryption Convergent encryption provides data confidentiality in deduplication. A user (or data owner) derives a convergent key from the data content and encrypts the data copy with the convergent key. In addition, the user derives a tag for the data copy, such that the tag will be used to detect duplicates. Here, we assume that the tag correctness property holds, i.e., if two data copies are the same, then their tags are the same. Formally, a convergent encryption scheme can be defined with four primitive functions:

D. • KeyGen(F) : The key generation algorithm takes a file content F as input and outputs the convergent key ckF of F;

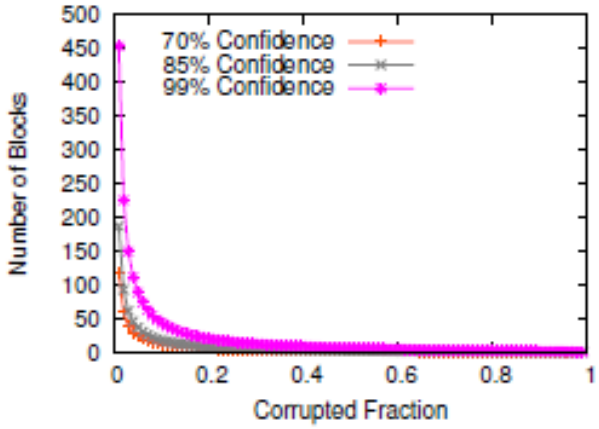
E. • Encrypt(ckF;F) : The encryption algorithm takes the convergent key ckF and file content F as input and outputs the ciphertextctF;

F. • Decrypt(ckF; ctF) : The decryption algorithm takes the convergent key ckF and ciphertextctF as input and outputs the plain file F;

G. • TagGen(F) : The tag generation algorithm takes a file content F as input and outputs the tag tagF of F. Notice that in this paper, we also allow TagGen(·)

VII. PERFORMANCE ANALYSIS

In this section, we will provide a thorough experimental evaluation of our proposed schemes. We build our testbed 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 G and G*T* are respectively chosen as 160 and 512 bits in length. We also set the block size as 4 KB and each block includes 25 sectors.



VIII. CONCLUSION

We examine that different algorithm that helps to secure transmitting data on cloud server. Like uploading, downloading data using key ,word search algorithm. Which ensure the cloud Storage security. And to overcome all existing system proposed SecCloud and SecCloud+. SecCloud introduces an auditing entity with maintenance which helps client to tag their file/data before uploading on server as well as maintain the integrity of that data. SecCloud uses proof-of-ownership protocol for secure data de-duplication also prevent from data leakage on internet. SecCloud+ is an advanced method for SecCloud that encrypt clients data before uploading , and Allow secure integrity auditing and data de-duplication on that encrypted data.

IX. REFERENCES

[1]. Dr. Anwar M. Mousa, ―Prospective of Fifth Generation Mobile Communications‖International Journal of Next-Generation Networks (IJNGN) Vol.4, No.3, September 2012

[2]. Ahmed and Moore. Sparql / odatainterop.Technical report, W3C,2013.
[3]. Z. Huang, C. Mei, L. E. Li, and T. Woo, "Cloudstream: Delivering high-quality streaming videos through a cloud-based svc proxy," in INFOCOM'11, 2011, pp. 201–205.
[4]. T. Coppens, L. Trappeniners, and M. Godon, "AmigoTV: towards a social TV experience," in Proc. of EuroITV, 2004.
[5]. N. Ducheneaut, R. J. Moore, L. Oehlberg, J. D. Thornton, and E. Nickell, "Social TV: Designing for Distributed, Sociable Television Viewing," International Journal of Human-Computer Interaction, vol. 24, no. 2, pp. 136–154, 2008.
[6]. A. Carroll and G. Heiser, "An analysis of power consumption in as smartphone," in Proc. of USENIXATC, 2010.
[7]. What is 100% Pure Java, http://www.javacoffeebreak.com/faq/faq0006.html.
[8]. J. Santos, D. Gomes, S. Sargento, R. L. Aguiar, N. Baker, M. Zafar, and A. Ikram, "Multicast/broadcast network convergence in next generation mobile networks," Comput. Netw., vol. 52, pp. 228–247, January 2008.
[9]. DVB-H, http://www.dvb-h.org/.[10] K. Chorianopoulos and G. Lekakos, "Introduction to social tv: Enhancing the shared experience with interactive tv," International Journal of Human- Computer Interaction, vol. 24, no. 2, pp. 113–120, 2008.
[10]. M. Chuah, "Reality instant messaging: injecting a dose of reality into online chat," in CHI '03 extended abstracts on Human factors in computing systems, ser. CHI EA '03, 2003, pp. 926–927.