A Review on Cloud Storage Architectures

Pallavi D. Dudhe[1], P. L. Ramteke[2]

[1]M.E. First Year CSE, HVPM C.O.E.T. Amravati, India

[2]Professor, IT Department, HVPM C.O.E.T. Amravati, India

Abstract: Cloud Computing has recognized as the next-generation architecture of IT companies. Cloud computing provides elastic, scalable, reliable services for individuals as well as organizations. In the cloud computing, everything is considered as the service and the service provided by Cloud is dynamic, diverse, and context-related. cloud computing moves the application software and databases to the centralized large data centers, where the management of the data and services may not be fully trustworthy. Storage services based on public clouds which provide customers with scalable and dynamic storage. By moving their data to the cloud customers can avoid the costs of building and maintaining a private storage infrastructure, opting instead to pay a service provider as a function of its needs. For most customers, this provides several benefits including availability (i.e., being able to access data from anywhere) reliability (i.e., not having to worry about backups) and confidentiality (i.e., accessed only by authorized parties) at a relatively low cost.

Keywords: Cloud storage, CP-ABE, Data Security, cloud storage servers, Fine grained access, attribute-based encryption, Cloud Storage Server (CSS), Cloud Service Provider (CSP).

1. Introduction

The model of networked online storage is cloud storage where data is stored in public clouds which are generally hosted by third parties. Hosting companies operate large data centers, and people who require their data to be hosted buy or lease storage capacity from them. However, cloud computing brings not only reliability and efficiency, but also security issues, especially data security and privacy protection. In the cloud computing, cloud servers and clients often placed in different security domains and the provide services with a dynamic, fast, continuous, reliable issues [1]. The data center operators, in the background, consider the real effect of the resources according to the requirements of the customer and lay open them as storage pools as database, which the customers can themselves use to store files or data objects. Physically, the resource may spread across multiple servers [2].

By outsourcing, organizations can concentrate on their basic tasks and operate other business applications via the Internet, rather than incurring substantial hardware, software and personnel costs involved in maintaining applications in house [3]. As data owners no longer physically possess the storage of their data, traditional cryptographic primitives for the purpose of data security protection cannot be directly adopted. In particular, simply downloading the data for its integrity verification is not a practical solution due to the high cost of input/output (I/O) and transmission across the network. Improper use of the data by the storage server or unauthorized access by outside users could be potential threats to their data. People would like to make their sensitive or private data only accessible to the authorized people with credentials they specified [4].

The introduction of ABE in implementing ingrained access control systems, a lot of works have been proposed to design flexible ABE schemes. There are two methods to realize the fine-grained access control based on ABE: KP-ABE and CP-ABE.

This paper presents a comprehensive survey on cloud storage architecture. Section II provides a brief overview of cloud data storage architecture, and. Section III discusses public cloud infrastructure Then, Section IV presents architecture of cryptographic storage service. Section V presents CP-ABE based secured cloud storage architecture. Finally we conclude the cloud storage architectures in Section VI.

2. Cloud Data Storage Architecture

In this model, spot-checking and error-correcting codes are used to ensure both “possession” and “retrievability” of data files on archive service systems. Specifically, some special blocks called “sentinels” are randomly embedded into the data file F for detection purpose and F is further encrypted to protect the positions of these special blocks. However, the number of queries a client can perform is also a fixed priori and the introduction of pre-computed “sentinels” prevents the development of realizing dynamic data updates. In addition, public verifiability is not supported in their scheme [2], [3]. Like the construction they use publicly verifiable homomorphism authenticators and provably secure in the random oracle model. Public retrievability is achieved and the proofs can be aggregated into a small authenticator value.

In particular, to support updates, especially for block insertion, they try to eliminate the index information in the “tag” computation. To achieve this before verification, they employ authenticated skip list data structure to authenticate the tag information of challenged or updated blocks first. In the cloud paradigm, by putting the large data files on the remote servers, the clients can be relieved of the burden of storage and computation. As clients no longer possess their data locally, it is of critical importance for the clients to ensure that their data are being correctly stored and maintained. That is, clients should be with security means that they can periodically verify the correctness of the remote data even without the existence of local copies [5].

Three different network entities can be identified as follows: Client : an entity, which has large data files to be stored in the cloud and relies on the cloud for data maintenance and computation, that can be either individual consumers or organizations; Cloud Storage Server (CSS): an entity, which is managed by Cloud Service Provider (CSP), has allows storage space and computation resource to maintain clients' data; Third Party Auditor (TPA): a TPA, which has capabilities that clients do not have, is trusted to assess and expose risk of cloud storage services on behalf of the clients upon request.

3. Public Cloud Infrastructure

Public cloud infrastructure introduces significant security and privacy risks. In fact, it seems that the biggest hurdle to the adoption of cloud storage (and cloud computing in general) is concern over the confidentiality and integrity of data. While, so far, consumers have been willing to trade privacy for the convenience of software services (e.g., for web-based email, calendars, pictures etc…), this is not the case for enterprises and government organizations. This reluctance can be attributed to several factors that range from a desire to protect mission-critical data to regulatory obligations to preserve the confidentiality and integrity of data [6]. The latter can occur when the customer is responsible for keeping personally identifiable information (PII), or medical and financial records. More precisely, such a service should provide.

- **Confidentiality: -** the cloud storage provider does not learn any information about customer data.
- **Integrity: -** any unauthorized modification of customer data by the cloud storage provider can be detected by the customer.
- **Non-repudiation:-** any access to customer data is logged, while retaining the main benefits of a public storage service.
- **Availability: -** customer data is accessible from any machine and at all times.
- **Reliability: -** customer data is reliably backed up [7].
- **Efficient retrieval: -** data retrieval times are comparable to a public cloud storage service.
- **Data sharing: -** customers can share their data with trusted parties.

4. Architecture of Cryptographic Storage Service

We now describe possible architecture for cryptographic storage service. The architecture consists of three components: a data processor (DP), that processes data before it is sent to the cloud; a data verifier (DV), that checks whether the data in the cloud has been tampered with; and a token generator (TG), that generates tokens which enable the cloud storage provider to retrieve segments of customer data.

4.1 Consumer Architecture

Consider three parties: a user Alice that stores her data in the cloud; a user Bob with whom Alice wants to share data; and a cloud storage provider that stores Alice's data. To use the service, Alice and Bobby begin by downloading a client application that consists of a data processor, a data verifier and a token generator. Upon its first execution, Alice's application generates a cryptographic key [9], [10]. We will consider this key as a master key and assume it is stored locally on Alice's system and that it is kept secret from the cloud storage provider. Whenever Alice wishes to upload data to the cloud, the data processor is invoked. It attaches some metadata (e.g., current time, size, keywords etc…) and encrypts and encodes the data and metadata with a variety of cryptographic primitives Data sharing between Alice and Bobby proceeds in a similar fashion. Whenever she wishes to share data with Bobby, the token generator is invoked to create a token and a decryption key which are both sent to Bob. He then sends the token to the provider who uses it to retrieve and return the appropriate encrypted documents. Bob then uses the decryption key to recover the files.

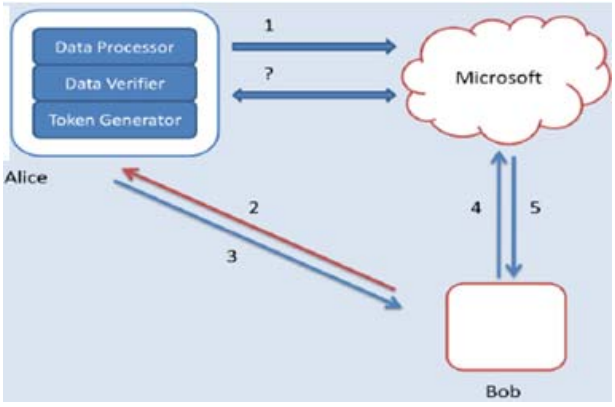

Figure 1: (1) Alice's data processor prepares the data before sending it to the cloud; (2) Bob asks Alice for permission to search for a keyword; (3) Alice's token generator sends a token for the keyword and a decryption key back to Bob; (4) Bob sends the token to the cloud; (5) the cloud uses the token to find the appropriate encrypted documents and returns them to Bob. At any point in time, Alice's data verifier can verify the integrity of the data.

4.2 Enterprise Architecture

In the enterprise architecture we consider an enterprise MegaCorp that stores its data in the cloud; a business partner PartnerCorp with whom MegaCorp wants to share data; and a cloud storage provider that stores MegaCorp's data. To handle enterprise customers, we introduce an extra component: a credential generator. The credential generator implements an access control policy by issuing credentials to parties inside and outside MegaCorp [8]. To use the service, MegaCorp deploys dedicated machines within its network. Depending on the particular architecture, these dedicated machines will run various core components. Since these components make use of a master secret key, it is important that they be adequately protected and, in particular, that the master key be kept secret from the cloud storage provider and PartnerCorp. If this is too costly in terms of resources or expertise, management of the dedicated machines (or specific components) can alternatively be outsourced to a trusted entity. Whenever a MegaCorp employee generates data that needs to be stored in the cloud, it sends the data together with an associated decryption policy to the dedicated machine for processing. The decryption policy specifies the type of credentials necessary to decrypt the data

(e.g., only members of a particular team). To retrieve data from the cloud (e.g., all files generated by a particular employee), an employee requests an appropriate token from the dedicated machine [12]. The employee then sends the token to the cloud provider who uses it to find and return the appropriate encrypted files which the employee decrypts using his credentials. Whenever MegaCorp wants to verify the integrity of its data, the dedicated machine's data verifier is invoked [11]. The latter uses the master secret key to interact with the storage provider and ascertain the integrity of the data. In the case of a medium-sized enterprise with enough resources and expertise, the dedicated machines include a data processor, a data verifier, a token generator and a credential generator. To begin, each MegaCorp and PartnerCorp employee receives a credential from the credential generator.

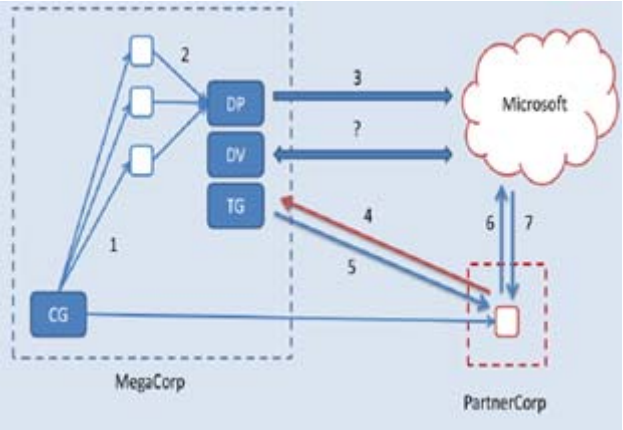

Figure 2: (1) Each MegaCorp and PartnerCorp employee receives a credential; (2) MegaCorp employees send their data to the dedicated machine; (3) the latter processes the data using the data processor before sending it to the cloud; (4) the PartnerCorp employee sends a keyword to MegaCorp's dedicated machine ; (5) the dedicated machine returns a token; (6) the PartnerCorp employee sends the token to the cloud; (7) the cloud uses the token to find the appropriate encrypted documents and returns them to the employee. At any point in time, MegaCorp's data verifier can verify the integrity of MegaCorp's data.

4.3 Benefits of Cryptographic Storage Service

The properties of a cryptographic storage service are that control of the data is maintained by the Customer and the security properties are derived from cryptography.

- **Regulatory compliance:** A cryptographic storage service, the data is encrypted on-premise by the data processor(s). This way, customers can be assured that the confidentiality of their data is preserved irrespective of the actions of the cloud storage provider. This greatly reduces any legal exposure for both the customer and the provider.
- **Geographic restrictions:** In a cryptographic storage service data is only stored in encrypted form so any law that pertains stored data has little to no effect on the customer. This reduces legal exposure for the customer and allows the cloud storage provider to make optimal use of its storage infrastructure, thereby reducing costs [13].
- **Subpoenas:** If an organization becomes the subject of an investigation, law enforcement agencies may request access to its data. If the data is stored in a public cloud, the request may be made to the cloud provider and the latter could even be prevented from notifying the customer.
- **Electronic discovery:** Digital information plays an important role in legal proceedings and often organizations are required to preserve and produce records for litigation.
- **Data retention and destruction:** A cryptographic storage service alleviates these concerns since data integrity can be verified and since the information necessary to decrypt data (i.e., the master key) is kept on-premise. Secure data erasure can be effectively achieved by just erasing the master key.

4.4 Cloud Services

- **Secure Extranet: -**In addition to simple storage, many enterprise customers will have a need for some associated services. These services can include any number of business processes including sharing of data among trusted partners, litigation support, monitoring and compliance, back-up, archive and audit logs [14]. We refer to a cryptographic storage service together with an appropriate set of enterprise services as a secure extranet.
- **Electronic Health Records: -**This move towards electronic health records promises to reduce medical errors, save lives and decrease the cost of healthcare. Given the importance and sensitivity of health-related data, it is clear that any storage platform for health records will need to provide strong confidentiality and integrity guarantees to patients.
- **Interactive Scientific Publishing:-** As scientists continue to produce large data sets which have broad value for the scientific community, demand will increase for a storage infrastructure to make such data accessible and sharable. To incent scientists to share their data, scientific societies such as the Optical Society of America are considering establishing a publication forum for data sets in partnership with industry.

5. CP-ABE Based Secured Cloud Storage Architecture

The CP-ABE based secured cloud storage architecture includes the following parameters:

Key Generation Center: It is a key authority that generates public and secret parameters for CP-ABE. It is in charge of issuing, revoking, and updating attribute keys for users [15]. It grants differential access rights to be honest but curious. That is, it will honestly execute the assigned tasks in the system; however, it would like to learn information of encrypted contents as much as possible. Thus, it should be prevented from accessing plaintext of the encrypted data even if it is honest.

Data Storing Center: It is an entity that provides a data sharing service. It is in charge of controlling the accesses from outside users to the storing data and providing corresponding contents services. The data storing center is another key authority that generates personalized user key with the KGC, and issues and revokes attribute group keys to valid users per each attribute, which are used to enforce a fine-grained user access control.

Data Owner: It is a client who owns data, and wishes to upload it into the external data storing center for ease of sharing or for cost saving [16]. A data owner is responsible for defining (attribute- based) access policy, and enforcing it on its own data by encrypting the data under the policy before distributing it.

User: It is an entity who wants to access the data. If a user possesses a set of attributes satisfying the access policy of the encrypted data, and is not revoked in any of the valid attribute groups, then he will be able to decrypt the cipher text and obtain the data.

6. Conclusion

In this paper, Cloud storage architectures will provide many exciting future with new opportunities and enable innovative applications and support different businesses also. Design and architecture of cloud storage system plays a vital role in cloud computing infrastructure in order to improve the storage capacity as well as cost effectiveness. Usually cloud storage system provides users to efficient storage space with elasticity feature. One of the challenges of cloud storage system is difficult to balance the providing huge elastic capacity of storage and investment of expensive cost for it.

References

[1] Sahai, A., Waters, B.: Fuzzy identity-based encryption. EUROCRYPT 2005.
[2] Goyal, V., Pandey, O., Sahai, A., Waters, B.: Attribute-based encryption for finegrained access control of encrypted data. ACM CCS 2006.
[3] Bethencourt, J., Sahai, A., Waters, B.: Ciphertextpolicy attribute-based encryption. IEEE Symposium on Security and Privacy, 2007
[4] H. Shacham and B. Waters, “Compact proofs of retrievability,” in Proc. of ASI- ACRYPT’08. Springer-Verlag, 2008, pp. 90–107.
[5] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, “Provable data possession at untrusted stores,” in Proc. of CCS’07. NewYork, NY, USA: ACM, 2007, pp. 598–609.
[6] Juels and B. S. Kaliski, Jr., “Pors: proofs of retrievability for large files,” in Proc. of CCS’07. New York, NY, USA: ACM, 2007, pp. 584–597.
[7] Abdalla, Michel, et al. «Searchable Encryption Revisited: Consistency Properties, Relation to Anonymous IBE, and Extensions. » Advances in Cryptology -- *CRYPTO '05.* Springer, 2005. 205-222. Ateniese, Giuseppe, et al. «Provable data possession at untrusted stores. » ACM Conference on Computer and Communications Security. ACM Press, 2007. 598-609.
[8] Ateniese, Giuseppe, Seny Kamara, et Jonathan Katz. «Proofs of Storage from Homomorphic Identification Protocols.» Advances in Cryptology -- Asiacrypt '09. Springer, 2009.
[9] Cloud Security Alliance. «Security Guidance for Critical Areas of Focus in Cloud Computing.» April2009.http://www.cloudsecurityalliance.org/guidance/csaguide.pdf.
[10] Curtmola, Reza, Juan Garay, Seny Kamara, et Rafail Ostrovsky. «Searchable symmetric encryption: improved definitions and efficient constructions.» 13th ACM Conference on Computer and Communications Security (CCS). ACM Press., 2006. 79-88.
[11] Erway, Chris, Alptekin Kupcu, Charalampos Papamanthou, et Roberto Tamassia. «Dynamic Provable Data Possession.» ACM Conference on Computer and Communications Security. 2009.
[12] Goh, E.-J. *Secure Indexes.* IACR ePrint, 2003.
[13] Golle, Phillipe, Brent Waters, et Jessica Staddon. «Secure Conjunctive Keyword Search over Encrypted Data.» Applied Cryptography and Network Security (ACNS *'04).* 2004. 31-45.
[14] Goyal, Vipul, Omkant Pandey, Amit Sahai, et Brent Waters. «Attribute-Based Encryption for Fine-Grained Access Control of Encrypted.

Author Profile

Miss Pallavi D. Dudhe has Completed Degree in B. Tech (Computer Science and Engineering) from Government College of Engineering, Amravati, Maharashtra, India. She is pursuing M.E. First Year Computer Science and Engineering in HVPM COET, Amravati, Maharashtra, India.

Prof. P. L. Ramteke is working as Associate Professor, Department of Information Technology, HVPM COET, Amravati, Maharashtra, India.