Secure Data Deduplication with Dynamic Ownership Management in Cloud Storage

Authors:Archana Patil, Anil Dangi

Abstract—In this paper, we propose a novel server-side deduplication scheme for encrypted data. It allows the cloud server to control access to outsourced data even when the ownership changes dynamically by exploiting randomized convergent encryption and secure ownership group key distribution. This prevents data leakage not only to revoked users even though they previously owned that data, but also to an honest-but-curious cloud storage server. In addition, the proposed scheme guarantees data integrity against any tag inconsistency attack. Thus, security is enhanced in the proposed scheme. The efficiency analysis results demonstrate that the proposed scheme is almost as efficient as the previous schemes, while the additional computational overhead is negligible.

Keywords: Identification of Deduplication, Password authentication.

I. INTRODUCTION

In cloud storage services, deduplication technology is commonly used to reduce the space and bandwidth requirements of services by eliminating redundant data and storing only a single copy of them. Deduplication is most effective when multiple users outsource the same data to the cloud storage, but it raises issues relating to security and ownership. Proof of-ownership schemes allow any owner of the same data to prove to the cloud storage server that he owns the data in a robust way. However, many users are likely to encrypt their data before outsourcing them to the cloud storage to preserve privacy, but this hampers deduplication because of the randomization property of encryption. Recently, several deduplication schemes have been proposed to solve this problem by allowing each owner to share the same encryption key for the same data. However, most of the schemes suffer from security flaws, since they do not consider the dynamic changes in the ownership of outsourced data that occur frequently in a practical cloud storage service.

Cloud computing provides scalable, low-cost, and location-independent online services ranging from simple backup services to cloud storage infrastructures. The fast growth of data volumes stored in the cloud storage has led to an increased demand for techniques for saving disk space and network bandwidth. To reduce resource consumption, many cloud storage services, such as Dropbox, Wuala, Mozy, and Google Drive, employ a deduplication technique, where the cloud server stores only a single copy of redundant data and provides links to the copy instead of storing other actual copies of that data, regardless of how many clients ask to store the data. The savings are significant, and reportedly, business applications can achieve disk and bandwidth savings of more than 90%. However, from a security perspective, the shared usage of users' data raises a new challenge.

II. LITERATURE SURVEY

A Dynamic Layering Scheme of Multicast Key Management Group key management is a difficult task in implementing large and dynamic secure multicast. In this paper, a new scheme is proposed in the basis of in-depth analysis of the requirements of the secure multicast and group key management. The scheme is based on the multicast group security architecture and multicast security group key management architecture proposed by IETF. This scheme constructs group key based on pairings and distributes the group key using HSAH function polynomial, and manages group key making use of the dynamic layering GCKS. The scheme is better in security, lower in computation cost and communication cost. The analysis comparison proves that the scheme has strong scalability and efficiency.

Tree-based Group Key Agreement: Fault-tolerant, scalable, and reliable communication services ha v e become critical in modern computing. An important and popular trend is to con v ert traditional centralized services (e.g., file sharing, authentication, web, and mail) into distributed services spread across multiple systems and networks. Man y of these newly distributed and other inherently collaboration applications (e.g., conferencing, white-boards, shared instruments, and command-and-control systems) need secure communication. However , experience shows that security mechanisms for collaboration and dynamic peer groups tend to be both expensive and unexpectedly complex. In that regard, dynamic peer groups are very different from non-collaboration v e, centrally managed, one-to-man y (or few-to-man y) broadcast groups such as those encountered in Internet multicast. Dynamic Peer Groups (DPGs) are common in man y layers of the network protocol stack and man y application areas of modern computing. Examples of DPGs include replicated servers (such as database, web, time), audio and video conferencing and, more generally, applications supporting collaboration work. In contrast to large multicast groups, DPGs tend to be relatively small in size, on the order of hundred members. Larger groups are harder to control on a peer basis and are often organized in a hierarchy. DPGs typically assume a man y-to-man y (or , equivalently , an y-to-an y) communication pattern rather than one-to-man y pattern common of larger hierarchical groups. Despite their relatively small number, group members in a DPG may be spread throughout the Internet and must be able to deal with arbitrary partitions due to network failures, congestion, and hostile attacks. In essence, a group can be split into a number of disconnected partitions each of which must persist and function as an independent peer group. Security requirements in collaboration eDPGs present several interesting research challenges. In this paper, we focus on services.

Energy and Communication Efficient Group Key Management Protocol for Hierarchical Sensor Networks: The security of sensor networks has become one of the most pressing issues in further development of these networks. Compared to the traditional wireless network, Wireless Sensor Network (WSN) provides a different computation and communication infrastructure. These differences originate not only from their physical characteristics, but also from their typical application s. For example, the physical characteristics include the large scale of deployment, limited computing capability, and constraints on power consumption. As a result, the requirements for the key management of a WSN are noticeably different from those for traditional networks.

The major requirements for the key management in a WSN are as follows. First, sensor s' communication involves a key distribution procedure between the communication parties, in which the key may be transmitted through some insecure channels. Therefore, key confidentiality, integrity and ownership must be enforced in the whole procedure. Second , it must be power aware, such that the power limitations need to be taken into consideration in the design of a key management protocol for WSN. Third, the key management scheme must be scalable, whereby it must be able to support larger networks and flexible against substantial increase in the size of network even after deployment. Forth, in the event of sensor node compromisation, security credentials which are stored in a sensor node or exchanged over the radio links should not reveal any useful information about any other links in the WSN. This is essential in upholding the resiliency of the WSN. Last but not least, the key management design for WSN needs to be balanced in terms of computation and storage overhead among the participated entities. In fact, this balanced property appears to be more important than the performance of an individual entity as an unbalanced design may results in performance degradation in some entities within a sensor network. Despite the challenges and requirements described above, we had proposed key management schemes for hierarchical self-organizing sensor networks in a formal and systematic manner. The first proposal, described in section 4, is a hybrid group key management scheme that uses high and middle powered nodes perform an asymmetric key agreement protocol to compute a group key. The group key will later be used for clustered low powered nodes' communication. During the group key transport phase, mutual authentication is performed between the low-powered sensors and the middle- powered nodes, and subsequently allow the establishment of secure group- wise local links. The second proposal, described in section 5, is a group key establishment scheme with initial shared keys . With initial trust built from a shared key, low - cost symmetric protocols enable the low- powered sensors to authenticate and establish secure group- wise local links. Once secure group keys are established, other security services such as group key refresh can be provided. The key management scheme enables a sensor network to set up cryptographic keys in an autonomous fashion, without relying on expensive cryptography for low level nodes and requires only two shared key independent of the network size.

III. EXISTING SYSTEM

In existing system, Cryptographic techniques were applied to access control for remote storage systems. The data owners encrypt files by using the symmetric encryption approach with content keys and then use every user's public key to encrypt the content keys. It requires each data owner to be online all the time. Some methods deliver the key management and distribution from the data owners to the remote server under the assumption that the server is trusted or semi-trusted.

Disadvantages:

- The key management is very complicated when there are a large number of data owners and users in the system.
- The key distribution is not convenient in the situation of user dynamically system.
- The server is cannot be trusted by the data owners in cloud storage systems.
- It cannot be applied to access control for cloud storage systems.

IV PROPOSED SYSTEM:

We propose an efficient group key management protocol in distributed group communication. This protocol is based on Elliptic Curve Cryptography and decreases the key length while providing securities at the same level as that of other cryptosystems provides. We provide the high level security and avoid the replication of file in the cloud service provider. In proposed system, we use hash function to generate key for the file .By using hash function to avoid the duplication in cloud. After that we are applying cryptographic technique for security purpose. We using ECC algorithm for encryption and decryption process.

Advantages:

- Avoid duplication in cloud.
- Increase the security level.
- High efficient.
- ECC algorithm provides high end security.

V SYSTEM ARCHITECTURE

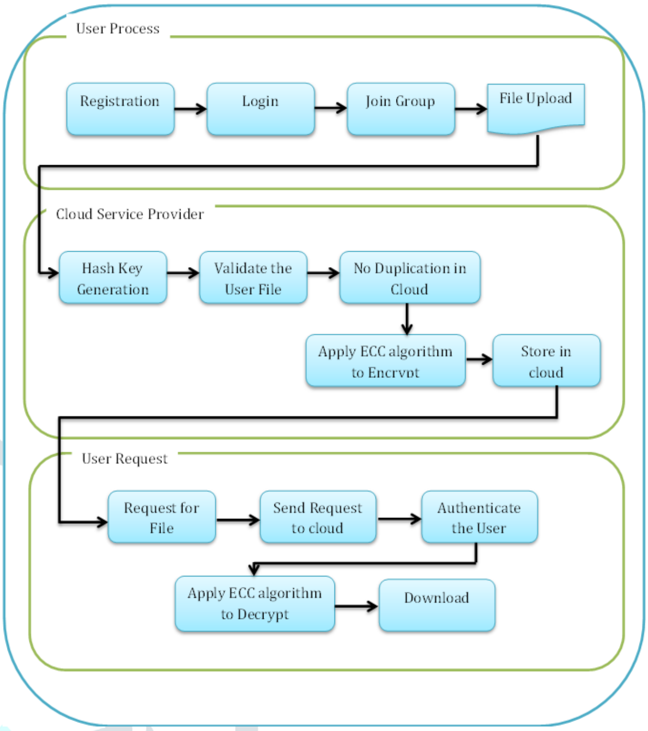

Figure 1 System Architecture

MODULES EXPLANATION

1. Registration and Login
2. Join Group and File Upload
3. File encrypt and store into Cloud
4. User request and Download

- **Registration and Login:** In our process , new user register the details and get the username and password for further process.Using Username and Password , user login into Group.Group generate key for the valid user and process inside the group under the valid key .
- **Join Group and File Upload:** In file upload process , user choose the file from the system and generate hash key for each file.Hash key generation is provided to avoid duplication of file to the cloud.If the file is already in cloud ,user should upload another file to cloud.
- After the validation of file from the user with cloud , we apply cryptographic technique to improve the security level in cloud.For cryptographic technique , we using Elliptic Curve Cryptography(ECC) algorithm for encrypting the file. In Elliptic Curve Cryptography(ECC),it convert the file into binary format and store it in cloud.

User request and Download: User send request to the cloud, cloud service provider decrypt the file. For cryptographic technique, we using Elliptic Curve Cryptography (ECC) algorithm for decrypting the file. Send the requested file to the user after validate the user. Then file will be downloaded in user location.

VI. DATA FLOW DIAGRAM

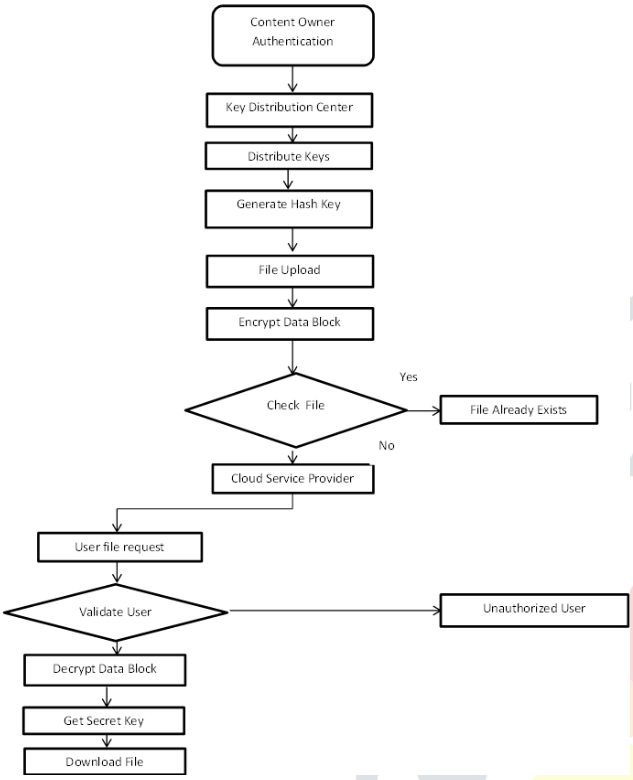

Figure 2 Data Flow Diagram

CONCLUSION

Dynamic ownership management is an important and challenging issue in secure deduplication over encrypted data in cloud storage. In this study, we proposed a novel secure data deduplication scheme to enhance a fine-grained ownership management by exploiting the characteristic of the cloud data management system. The proposed scheme features a reencryption technique that enables dynamic updates upon any ownership changes in the cloud storage. Whenever an ownership change occurs in the ownership group of outsourced data, the data are reencrypted with an immediately updated ownership group key, which is securely delivered only to the valid owners. Thus, the proposed scheme enhances data privacy and confidentiality in cloud storage against any users who do not have valid ownership of the data, as well as against an honest-but-curious cloud server. Tag consistency is also guaranteed, while the scheme allows full advantage to be taken of efficient data deduplication over encrypted data. In terms of the communication cost, the proposed scheme is more efficient than the previous schemes, while in terms of the computation cost, taking additional 0.*1*-0:2 ms compared to the RCE scheme, which is negligible in practice. Therefore, the proposed scheme achieves more secure and fine-grained ownership management in cloud storage for secure and efficient data deduplication.

REFERENCES

[1] M. Bellare, S. Keelveedhi, "Interactive message-locked encryp-tion and secure deduplication," Proc. PKC 2015, pp. 516–538, 2015. N. Baracaldo, E. Androulaki, J. Glider, A. Sorniotti, **"Reconciling end-to-end confidentiality and data reduction in cloud storage,"** roc. ACM Workshop on Cloud Computing Security, pp. 21–32, 2014.

[2] D. T. Meyer, and W. J. Bolosky, **"A study of practical deduplica-tion,"** Proc. USENIX Conference on File and Storage Technolo-gies 2011, 2011.

[3] A. Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Lee, J. C. S. Lui, **"A secure cloud backup system with assured deletion and version control,"** Proc. International Workshop on Security in Cloud Computing, 2011.

[4] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, **"Secure dedupli-cation with efficient and reliable convergent key management,"** IEEE Transactions on Parallel and Distributed Sytems, Vol. 25, No. 6, 2014.

J. Li, Y. K. Li, X. Chen, P. Lee, and W. Lou, **"A hybrid cloud approach for secure authorized deduplication,"** IEEE Transactions on Parallel and Distributed Systems, Vol. 26, No. 5, pp. 1206–1216, 2015.