A Novel Cloud Data De Duplication Backup using ALG De Dupe

Voppalanchu Santhoshi[1] Bhukya Ramji[2]
[1]M.Tech. Student [2]Assistant Professor
[1,2]Department of Computer Science & Engineering
[1,2]Sri Venkateshwara Engineering College, Suryapet

Abstract— The cloud backup is utilized for the personal storage of the people in terms of reducing the mainlining process and managing the structure and storage space managing process. The challenging process is the reduplication process in both the local and ecumenical backup de-duplications. In the prior work they only provide the local storage de-duplication or vice versa ecumenical storage de-duplication in terms of amending the storage capacity and the processing time. In this paper, the proposed system is called as the ALG- De dupe. It signifies the Application cognizant Local Ecumenical Source De-duplication proposed system to provide the efficient de-duplication process. It can provide the efficient de duplication process with the low system load, minimized backup window, and incremented power efficiency in the user's personal storage. In the proposed system the sizably voluminous data is partitioned into more minute part which is called as chunks of data. Here the data may contain the redundancy it will be eschewed afore storing into the storage area.

Key words: Data De duplication, Backup Window, ALG De dupe, Cloud Backup

I. INTRODUCTION

To make data management scalable in cloud computing, de duplication has been a well-kenned technique and has magnetized more and more attention recently. Data de duplication is a specialized data compression technique for eliminating duplicate replicas of reiterating data in storage. The technique is utilized to ameliorate storage utilization and can withal be applied to network data transfers to reduce the number of bytes that must be sent. In lieu of keeping multiple data copies with the same content, de duplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. De duplication can take place at either the file level or the block level. For file level de duplication, it eliminates duplicate facsimiles of the same file. De duplication can additionally take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files.

Albeit data de duplication brings a plethora of benefits, security and privacy concerns arise as users' sensitive data are susceptible to both inside and outside attacks .Traditional encryption, while providing data confidentiality is incompatible with data de duplication. Concretely, traditional encryption requires different users to encrypt their data with their own keys. Thus, identical data replicas of different users will lead to different cipher texts, making de duplication infeasible. Convergent encryption has been proposed to enforce data confidentiality while making de duplication feasible. It encrypts/decrypts a data copy with a convergent key, which is obtained by computing the cryptographic hash value of the content of the data copy. After key generation and data encryption, users retain the keys and send the cipher text to the cloud. Since the encryption operation is deterministic and is derived from the data content, identical data copies will engender the same convergent key and hence the same cipher text . To avert unauthorized access, a secure proof of ownership protocol is withal needed to provide the proof that the utilizer indeed owns the same file when a duplicate is found. After the proof, subsequent users with the same file will be provided a pointer from the server without needing to upload the same file. A utilizer can download the encrypted file with the pointer from the server, which can only be decrypted by the corresponding data owners with their convergent keys. Thus, convergent encryption sanctions the cloud to perform de duplication on the cipher texts and the proof of ownership obviates the unauthorized utilizer to access the file.

II. RELATED WORK

A. *Existing System:*

Data de duplication is one of paramount data compression techniques for eliminating duplicate replicas of reiterating data, and has been widely utilized in cloud storage to reduce the amount of storage space and preserve bandwidth. To for fend the confidentiality of sensitive data while fortifying de duplication, Cloud computing provides ostensibly illimitable "virtualized" resources to users as accommodations across the whole Internet, while obnubilating platform and implementation details. Today's cloud accommodation providers offer both highly available storage and massively parallel computing resources at relatively low costs. As cloud computing becomes prevalent, an incrementing amount of data is being stored in the cloud and shared by users with designated privileges, which define the access rights of the stored data.

1) *Disadvantages of Existing System:*

One critical challenge of cloud storage services is the management of the ever-increasing volume of data.

B. *Proposed System:*

The convergent encryption technique has been proposed to encrypt the data afore outsourcing. To better bulwark data security, this paper makes the first endeavor to formally address the quandary of sanctioned data de duplication. Different from traditional de duplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. We additionally present several incipient de duplication constructions fortifying sanctioned duplicate check in a hybrid cloud architecture.

Security analysis demonstrates that our scheme is secure in terms of the definitions designated in the proposed security model. As a proof of concept, we implement a prototype of our proposed sanctioned duplicate check scheme and conduct testbed experiments utilizing our prototype. We show that our proposed sanctioned duplicate

check scheme incurs minimal overhead compared to mundane operations.

C. Hybrid Highlights:

Hybrid cloud can be built utilizing any technology it varies according to different vendors. Key components In many of the situations, implementation of the hybrid cloud has a controller that will keep track of all locations of private and public clouds, IP address, servers and other resources that can run systems efficiently.

Some of the key components include:

- Orchestration manager and cloud provisioning for storage, public cloud resources which includes virtual machines and networks, the private and public clouds, which are not obligatorily compatible or identical.
- Synchronization element and Data transfer efficiently exchange information between private and public clouds.
- Transmuting configurations of storage, network and some other resources are being tracked by configuration monitor.[1]

In the Fig 1, the simplest view of hybrid cloud is provided, a single off-premises public cloud and on-premises private cloud is within the Enterprise Datacenter is shown and public cloud establishes the safe connection to store data on to the cloud is denoted by the arrow:

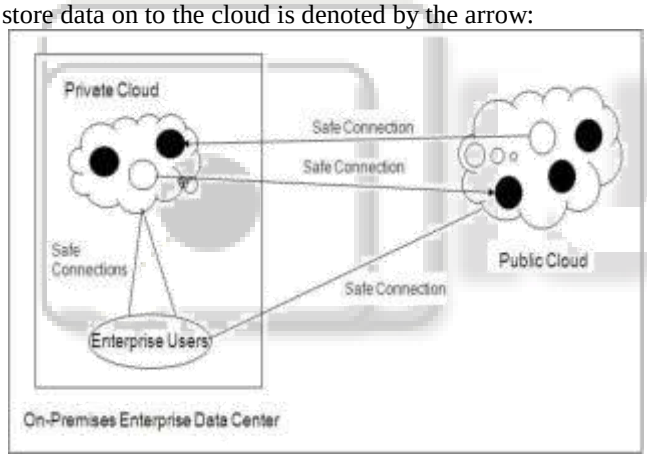

Fig 1: Example of Hybrid cloud Environment.

The ebony circles shows active virtual server images and white circles shows virtual server images which have been migrated by utilizing safe connections. The arrows designate that the direction of migration. Utilizing safe connections Enterprise users are connected to the clouds, which can be secure HTTP browsers or virtual private networks (VPNs) .A hybrid cloud could withal can consist of multiple public or/and private clouds. [3]

1) Integration

One or more private and public clouds integrate to compose a hybrid system and it will be more challenging compared to integrating the on premises systems noted Rex Wang. As different clouds conventionally will have distinct APIs, private, integrating public and legacy systems will often require custom code, verbally expressed UT Dallas' Kantarcioglu.

2) Models:

There are mainly two primary hybrid cloud deployment models.

3) Management

Hybrid cloud computing has a technological key called as management. Systems are expeditiously migrating from single cloud environment to multiple cloud management system. Then later they must manage all types of cloud applications such as platform as an accommodation, infrastructure as an accommodation and software as an accommodation through the whole development and additionally deployment life cycles.

4) Security

In order to secure hybrid clouds, companies use special techniques such as authentication, access control policies and encryption in both private clouds and public clouds.

These will include the coalescence of cloud-predicated security accommodations and managed hosted appliances. Some approaches such as intrusion detection systems and firewalls are always implemented in the hosted environment categorically for utilization of hybrid cloud architecture, verbalized Jonathan Hogue. Since we cannot disclose the sensitive data, companies will require keeping limit for the amount of sensitive data which they outsource or they will have to encrypt the sensitive data afore outsourcing in public clouds, kantarcioglu expounded. Encryption predicated approaches will bulwark sensitive data when it outsourced to public cloud processing this encrypted data is customarily more costly and in volute. [1]

III. IMPLEMENTATION

A. Cloud Accommodation Provider

In this module, we develop Cloud Accommodation Provider module. This is an entity that provides a data storage accommodation in public cloud. The S-CSP provides the data outsourcing accommodation and stores data on behalf of the users. To reduce the storage cost, the S-CSP eliminates the storage of redundant data via de duplication and keeps only unique data. In this paper, we surmise that S-CSP is always online and has abundant storage capacity and computation puissance.

B. Data Users Module

A utilizer is an entity that wants to outsource data storage to the S-CSP and access the data later. In a storage system fortifying de duplication, the utilizer only uploads unique data but does not upload any duplicate data to preserve the upload bandwidth, which may be owned by the same utilizer or different users.

In the sanctioned de duplication system, each utilizer is issued a set of privileges in the setup of the system. Each file is bulwarked with the convergent encryption key and privilege keys to realize the sanctioned de duplication with differential privileges.

C. Private Cloud Module

Compared with the traditional de duplication architecture in cloud computing, this is an incipient entity introduced for facilitating user's secure utilization of cloud accommodation. Categorically, since the computing resources at data utilizer/owner side are restricted and the public cloud is not planarity confided in practice, private cloud is able to provide data utilizer/owner with an execution environment and infrastructure working as an interface between utilizer and the public cloud.

The private keys for the privileges are managed by the private cloud, who answers the file token requests from the users. The interface offered by the private cloud sanctions utilizer to submit files and queries to be securely stored and computed respectively.

D. Secure De Duplication System

We consider several types of privacy we require for fend, that is, i) un forge ability of duplicate-check token: There are two types of adversaries, that is, external adversary and internal adversary. As shown below, the external adversary can be viewed as an internal adversary without any privilege. If a utilizer has privilege p, it requires that the adversary cannot forge and output a valid duplicate token with any other privilege p′ on any file F, where p does not match p′. Furthermore, it additionally requires that if the adversary does not make a request of token with its own privilege from private cloud server, it cannot forge and output a valid duplicate token with p on any F that has been queried.

IV. EXPERIMENTAL WORK

Fig 2: Data user send file token to public cloud.

Fig 3: Data User encrypting data before storing cloud.

Fig 4: Data user upload encrypted file to cloud

Fig 5: User downloading file from cloud.

V. CONCLUSION

The cerebration of sanctioned information de duplication was proposed to ascertain the information security by counting differential benefits of clients in the duplicate copy check. The presentation of a few incipient de duplication developments fortifying sanctioned duplicate copy in hybrid cloud architecture, in that the duplicate check tokens of documents are engendered by the private cloud server having private keys. Security check exhibits that the methods are secure regarding insider and outsider assaults detailed in the proposed security model. As an issue verification of conception, the developed model of the proposed sanctioned duplicate copy check method and tested the model. That showed the sanctioned duplicate copy check method experience minimum overhead comparing convergent encryption and data transfer.

REFERENCES

[1] Bugiel, Sven, et al. "Twin clouds: Secure cloud computing with low latency." Communications and Multimedia Security. Springer Berlin Heidelberg, 2011.
[2] Anderson, Paul, and Le Zhang. "Fast and Secure Laptop Backups with Encrypted De-duplication." LISA. 2010.
[3] Bellare, Mihir, Sriram Keelveedhi, and Thomas Ristenpart. "DupLESS: server-aided encryption for deduplicated storage." Proceedings of the 22nd USENIX conference on Security. USENIX Association, 2013.
[4] Bellare, Mihir, Sriram Keelveedhi, and Thomas Ristenpart. "Message-locked encryption and secure deduplication." Advances in Cryptology–EUROCRYPT 2013. Springer Berlin Heidelberg, 2013. 296-312.
[5] Bellare, Mihir, Chanathip Namprempre, and Gregory Neven. "Security proofs for identity-based identification and signature schemes." Journal of Cryptology 22.1 (2009): 1-61.
[6] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In *USENIX Security Symposium*, 2013.
[7] K. Zhang, X. Zhou, Y. Chen, X.Wang, and Y. Ruan. Sedic: privacyaware data intensive computing on hybrid clouds. In *Proceedings of the 18th ACM conference on*

Computer and communications security, CCS'11, pages 515–526, New York, NY, USA, 2011. ACM.

[8] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, *ACM Conference on Computer and Communications Security*, pages 491–500. ACM, 2011.

