International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 08 | Aug 2018 www.irjet.net p-ISSN: 2395-0072

Condense Data Replication on Encrypted File Storage

Saro J

[1]*Saro J Mail id: srijaictk2815@gmail.com& Address: Sendalai, Thanjavur*
[2] *Professor: Dr .M. Manimekalai, M.sc., PGDCA, M.sc(IT), M.phil(CS), Ph. D(CS)..,Dept. of Computer Science, Shrimati Indira Gandhi College, Tamil Nadu, India*

Abstract - *Distributed computing offers another method for benefit arrangement by re-orchestrating different assets over the Internet. The most imperative and prevalent cloud benefit is information stockpiling. Keeping in mind the end goal to protect the security of information holders, information are regularly put away in cloud in an encoded frame. Be that as it may, scrambled information present new difficulties for cloud information deduplication, which ends up significant for huge information stockpiling and preparing in cloud. Customary deduplication plans can't deal with scrambled information. Existing arrangements of scrambled information deduplication experience the ill effects of security shortcoming. They can't adaptably bolster information get to control and denial. Consequently, few of them can be promptly conveyed by and by. This idea propose a plan to deduplicate scrambled information put away in cloud in light of proprietorship test and intermediary re-encryption. It coordinates cloud information deduplication with get to control. I assess its execution in view of broad investigation and PC reenactments. The outcomes demonstrate the predominant proficiency and viability of the plan for potential down to earth sending, particularly for enormous information deduplication in distributed storage.*

Key Words:Information, Cloud, Deduplication, Scrambled, Stockpiling, etc…

1.INTRODUCTION

With the potentially infinite storage space offered by cloud providers, users tend to use as much space as they can and vendors constantly look for techniques aimed to minimize redundant data and maximize space savings. A technique which has been widely adopted is cross-user deduplication. The simple idea behind deduplication is to store duplicate data (either files or blocks) only once. Therefore, if a user wants to upload a file (block) which is already stored, the cloud provider will add the user to the owner list of that file (block). Deduplication has proved to achieve high space and cost savings and many cloud storage providers are currently adopting it. Deduplication can reduce storage needs by up to 90-95 percent for backup applications and up to 68 percent in standard file systems. Along with low ownership costs and flexibility, users require the protection of their data and confidentiality guarantees through encryption. Unfortunately, deduplication and encryption are two conflicting technologies. While the aim of deduplication is to detect identical data segments and store them only once, the result of encryption is to make two identical data segments indistinguishable after being encrypted. This means that if data are encrypted by users in a standard way, the cloud storage provider cannot apply deduplication since two identical data segments will be different after encryption. On the other hand, if data are not encrypted by users, confidentiality cannot be guaranteed and data are not protected against curious cloud storage providers. A technique which has been proposed to meet these two conflicting requirements is convergent encryption whereby the encryption key is usually the result of the hash of the data segment.

Although convergent encryption seems to be a good candidate to achieve confidentiality and deduplication at the same time, it unfortunately suffers from various well-known weaknesses including dictionary attacks: an attacker who is able to guess or predict a file can easily derive the potential encryption key and verify whether the file is already stored at the cloud storage provider or not.

1.1 Big Data

Huge information is a sweeping term for the non-conventional methodologies and innovations expected to assemble, arrange, process, and accumulate bits of knowledge from vast datasets. While the issue of working with information that surpasses the figuring force or capacity of a solitary PC isn't new, the inescapability, scale, and estimation of this sort of processing has enormously extended lately.

The essential prerequisites for working with enormous information are the same as the necessities for working with datasets of any size. Notwithstanding, the enormous scale, the speed of ingesting and handling, and the attributes of the information that must be managed at each phase of the procedure show huge new difficulties when planning arrangements. The objective of most huge information frameworks is to surface experiences and associations from substantial volumes of heterogeneous information that would not be conceivable utilizing customary strategies.

a)Volume

The sheer size of the data handled characterizes huge information frameworks. These datasets can be requests of size bigger than customary datasets, which requests more idea at each phase of the preparing and capacity life cycle.

Regularly, in light of the fact that the work prerequisites surpass the capacities of a solitary PC, this turns into a test of pooling, apportioning, and organizing assets from gatherings of PCs. Bunch administration and calculations fit for breaking errands into littler pieces turn out to be progressively imperative.

b) Velocity

Another manner by which huge information contrasts fundamentally from other information frameworks is the speed that data travels through the framework. Information is much of the time streaming into the framework from different sources and is regularly anticipated that would be handled progressively to pick up experiences and refresh the present comprehension of the framework.

This attention on close moment criticism has pushed numerous huge information professionals from a cluster situated approach and more like a constant gushing framework.

Information is always being included, kneaded, prepared, and broke down to stay aware of the inundation of new data and to surface profitable data early when it is generally significant. These thoughts require strong frameworks with exceptionally accessible segments to prepare for disappointments along the information pipeline.

c) Variety

Huge information issues are regularly one of a kind in view of the extensive variety of both the sources being handled and their relative quality.

Information can be ingested from interior frameworks like application and server logs, from web based life bolsters and other outside APIs, from physical gadget sensors, and from different suppliers. Huge information tries to deal with possibly valuable information paying little respect to what standpoint it's maintaining by merging all data into a solitary framework.

The configurations and kinds of media can change essentially too. Rich media like pictures, video documents, and sound chronicles are ingested close by content records, organized logs, and so forth. While more conventional information preparing frameworks may anticipate that information will enter the pipeline effectively marked, arranged, and composed, enormous information frameworks generally acknowledge and store information closer to its crude state.

In a perfect world, any changes or changes to the crude information will occur in memory at the season of preparing.

Different people and associations have recommended growing the first three Vs, however these recommendations have had a tendency to portray challenges as opposed to characteristics of enormous information. Some basic increments are:

- Veracity: The assortment of sources and the intricacy of the preparing can prompt difficulties in assessing the nature of the information (and therefore, the nature of the subsequent examination)
- Variability: Variation in the information prompts wide variety in quality. Extra assets might be expected to distinguish, process, or channel low quality information to make it more valuable.
- Value: a definitive test of enormous information is conveying esteem. Here and there, the frameworks and procedures set up are sufficiently perplexing that utilizing the information and extricating real esteem can end up troublesome.

So how is information really prepared when managing a major information framework? While ways to deal with execution vary, there are a few shared traits in the methodologies and programming that we can discuss for the most part. While the means introduced underneath won't not be valid in all cases, they are broadly utilized.

The general classes of exercises required with huge information handling are:

- Ingesting information into the framework
- Persisting the information away
- Computing and Analyzing information
- Visualizing the outcomes

2.EXISTING SYSTEM

A procedure which has been proposed to meet clashing necessities is united encryption whereby the encryption key is generally the consequence of the hash of the information fragment. Albeit merged encryption is by all accounts a decent possibility to accomplish classification and deduplication in the meantime, it shockingly experiences different surely understood shortcomings including lexicon assaults: an assailant who can figure or anticipate a document can without much of a stretch determine the potential encryption key and confirm

whether the record is as of now put away at the distributed storage supplier or not.

Touchy development in the quantity of passwords for online applications and encryption keys for outsourced information stockpiling very much surpass the administration furthest reaches of clients.

In this way, outsourcing keys (counting passwords and information encryption keys) to proficient watchword supervisors (fair however inquisitive specialist organizations) is pulling in the consideration of numerous clients.

2.1 Techiniques utilized as a part of Existing System:

TRIPLE DES

Triple Data Encryption Standard (DES) is a kind of mechanized cryptography where square figure calculations are connected three times to every datum square. The key size is expanded in Triple DES to guarantee extra security through encryption abilities. Each square contains 64 bits of information. Three keys are alluded to as package keys with 56 bits for every key. There are three entering choices in information encryption models:

1. All keys being autonomous
2. Key 1 and key 2 being free keys
3. All three keys being indistinguishable

Key choice #3 is known as triple DES. The triple DES key length contains 168 bits yet the key security tumbles to 112 bits.

Before utilizing 3TDES, client initially produce and convey a 3TDES key K, which comprises of three unique DES keys K1, K2 and K3. This implies the real 3TDES key has length 3×56 = 168 bits. The encryption plot is represented as follows –

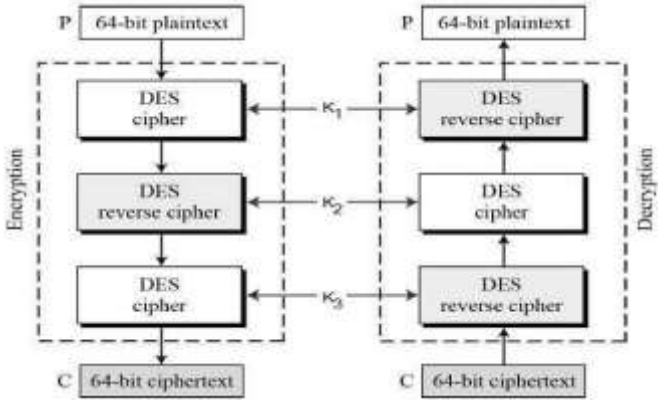

Fig -1: Data Encryption Standard (DES)

The encryption-making an interpretation of process is as per the going with –

- Encrypt the plaintext squares utilizing single DES with key K1.
- Now unscramble the yield of stage 1 utilizing single DES with key K2.
- Finally, scramble the yield of stage 2 utilizing single DES with key K3.
- The yield of stage 3 is the ciphertext.
- Decryption of a ciphertext is a rotate methodology. Client at first unscramble utilizing K3, by then scramble with K2, at last unravel with K1.

Because of this plan of Triple DES as an encrypt– decrypt–scramble process, it is conceivable to utilize a 3TDES (equipment) usage for single DES by setting K1, K2, and K3 to be a similar respect. This gives in reverse closeness DES.

Second assortment of Triple DES (2TDES) is dubious to 3TDES next to that K3is supplanted by K1. At the day's end, client encode plaintext foils with key K1, by then unscramble with key K2, in conclusion scramble with K1 once more. Thusly, 2TDES has a key length of 112 bits.

Triple DES structures are all around more secure than single DES, in any case these are obviously a much slower process than encryption utilizing single DES.

Grand Access Algorithm

Give Access is the specific control of access to a place or other asset. The show of getting to may mean eating up, entering or utilizing. Locks and login affirmations are two fundamentally indistinguishable to systems of access control. Once the key has been made, the arrangement of encryption and deciphering are unassumingly prompt and computationally direct. Every individual's own private and open keys must be numerically related where the private key is utilized to unscramble a correspondence sent utilizing an open key and the an alternate way. Some conspicuous strayed encryption checks depend upon the cryptosystem. The private key must be kept absolutely private by the proprietor, in any case the comprehensive network key can be circled in an open list, for example, with an accreditation ace.

2.2 Survey

[1] The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes

Authors:

1. Joseph Bonneau, Univ. of Cambridge, Cambridge, UK
2. Cormac Herley, Microsoft Res., Redmond, WA, USA
3. Paul C. van Oorschot, Carleton Univ., Ottawa, ON, Canada

To evaluate two decades of proposals to replace text passwords for general-purpose user authentication on the web using a broad set of twenty-five usability, deployability and security benefits that an ideal scheme might provide. The scope of proposals we survey is also extensive, including password management software, federated login protocols, graphical password schemes, cognitive authentication schemes, one-time passwords, hardware tokens, phone-aided schemes and biometrics. In comprehensive approach leads to key insights about the difficulty of replacing passwords. Not only does no known scheme come close to providing all desire benefits: none even retains the full set of benefits that legacy passwords already provide. In particular, there is a wide range from schemes offering minor security benefits beyond legacy passwords, to those offering significant security benefits in return for being more costly to deploy or more difficult to use. Conclude that many academic proposals have failed to gain traction because researchers rarely consider a sufficiently wide range of real-world constraints. Beyond our analysis of current schemes, our framework provides an evaluation methodology and benchmark for future web authentication proposals.

Issues:
It is unable to meet the
1) Confidentiality and privacy of keys;
2) Search privacy on identity attributes tied to keys
3) Owner controllable authorization.

Solutions to Overcome the issues are :

- ✓ Created a Meta data manager to generate private keys for confidentiality
- ✓ Privacy OTA for data sharing for data authorization

[2] Practical techniques for searches on encrypted data.

Authors:

Dawn Xiaodong Song

David Wagner Adrian Perri

It is desirable to store data on data storage servers such as mail servers and file servers in encrypted form to reduce security and privacy risks. But this usually implies that one has to sacrifice functionality for security. For example, if a client wishes to retrieve only documents containing certain words, it was not previously known how to let the data storage server perform the search and answer the query, without loss of data confidentiality. We describe our cryptographic schemes for the problem of searching on encrypted data and provide proofs of security for the resulting crypto systems. The techniques have a number of crucial advantages. They are provably secure: they provide provable secrecy for encryption, in the sense that the untrusted server cannot learn anything about the plaintext when only given the cipher text; they provide query isolation for searches, meaning that the untrusted server cannot learn anything more about the plaintext than the search result; they provide controlled searching, so that the untrusted server cannot search for an arbitrary word without the user's authorization; they also support hidden queries, so that the user may ask the untrusted server to search for a secret word without revealing the word to the server. The algorithms presented are simple, fast (for a document of length n, the encryption and search algorithms only need O(n) stream cipher and block cipher operations), and introduce almost no space and communication overhead, and hence are practical to use today.

Issues:

Minor security benefits above legacy passwords, to those contribution significant security benefits in appearance for being more costly to deploy or more difficult to use.

Solutions to Overcome the issues are :

Passwords are protected by encrypting Database entirely for confidentiality and privacy.

[3] Public Key Encryption with Keyword Search.

Authors:

Dan Boneh∗ Giovanni Di Crescenzo Stanford University Telcordia.

I have study the problem of searching on data that is encrypted using a public key system. Consider user Bob who sends email to user Alice encrypted under Alice's public key. An email gateway wants to test whether the email contains the keyword "urgent" so that it could route the email accordingly. Alice, on the other hand does not wish to give the gateway the ability to decrypt all her messages. Have been define and construct a mechanism that enables Alice to provide a key to the gateway that

enables the gateway to test whether the word "urgent" is a keyword in the email without learning anything else about the email. We refer to this mechanism as *Public Key Encryption with keyword Search*. As another example, consider a mail server that stores various messages publicly encrypted for Alice by others. Using our mechanism Alice can send the mail server a key that will enable the server to identify all messages containing some specific keyword, but learn nothing else. To define the concept of public key encryption with keyword search and give several constructions.

Issues:

The CloudKeyBank provider or the attacker in the middle may derive the private intent of the user from his/her submitted search query (Search privacy).

Solutions to Overcome the issues are :

Our 3 steps of cryptography does not allow attacker in the middle to access the data in anyway.

[4] Anonymous hierarchical identity-based encryption (without random oracles).

Authors:

Xavier Boyen , Brent Waters- 2006.

Have been present an identity-based cryptosystem that features fully anonymous cipher texts and hierarchical key delegation. To give a proof of security in the standard model based on the mild Decision Linear complexity assumption in bilinear groups. The system is efficient and practical, with small cipher texts of size linear in the depth of the hierarchy. Applications include search on encrypted data, fully private communication, etc.

Our results resolve two open problems pertaining to anonymous identity-based encryption, our scheme being the first to offer provable anonymity in the standard model, in addition to being the first to realize fully anonymous HIBE at all levels in the hierarchy.

Issues:

Files storage in cloud lacks in data security and data leakage.

Solutions to Overcome the issues are:

This framework proposes the File encryption system to protect the data even if it is accessed by unauthorized users.

[5] Predicate Encryption Supporting Disjunctions, Polynomial Equations, and Inner Products.

Authors:

Tejal Khandave , Ashwini Madane, Aarti Bhoi, Kalpesh Zala,Prof.Rupali Adhau - 2017

Predicate encryption is a new paradigm for public-key encryption generalizing, among other things, identity-based encryption. In a predicate encryption scheme, secret keys correspond to predicates and cipher texts are associated with attributes; the secret key SK corresponding to a predicate f can be used to decrypt a cipher text associated with attribute I if and only if f(I) = 1. Constructions of such schemes are currently known for certain classes of predicates. We construct such a scheme for predicates corresponding to the evaluation of inner products over ZN (for some large integer N). This, in turn, enables constructions in which predicates correspond to the evaluation of disjunctions, polynomials, CNF/DNF formulae, or threshold predicates (among others). Besides serving as a significant step forward in the theory of predicate encryption, our results lead to a number of applications that are interesting in their own right.

Issues:

Owner controllable authorization over the shared keys is not lack in cloud computing.

Solutions to Overcome the issues are :

Privacy OTA for data sharing for data authorization and also generating the notifications for the data owner when the files are accessed by someone.

3. PROPOSED SYSTEM

In this project, cope with the inherent security exposures of convergent encryption and propose Cloud Dedup, which preserves the combined advantages of deduplication and convergent encryption.

The security of ClouDedup relies on its new architecture whereby in addition to the basic storage provider, a metadata manager and an additional server are defined: the server adds an additional encryption layer to prevent well-known attacks against convergent encryption and thus protect the confidentiality of the data; on the other hand, **the metadata manager is responsible of the key management task** since block-level deduplication requires the memorization of a huge number of keys. Therefore, the underlying deduplication is performed at block-level and we define an efficient key management mechanism to avoid users to store one key per block. Have been proposed Cloud Key Bank, the first unified key management framework that addresses all the three goals above. Under our framework, the key owner can perform

privacy and controllable authorization have been proposed **Base64** algorithm for secrete key generate and **AES algorithm** for encrypt and decrypt our data Our experimental results and security analysis show the efficiency and security goals are well achieved.

4. CONCLUSION

Cloud computing is a growing technology as well as it has emerged as a modern computing Paradigm that providing IT infrastructure and can be used. To meet the continuously growing storage and processing requirements of today's scientific applications. The Open Source Cloud platform is most important which provide an alternative to end user for improved **portability, flexibility, scalability** and better performance using open-stack cloud operating environment. The security comparison and analysis prove that this new proposed cloud system is sufficient to support the identified three security requirements which are not be solve in traditional outsourced scenario.

REFERENCES

[1] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for storage security in cloud computing".

[2] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, "Enabling public audit ability and data dynamics for storage security in cloud computing".

[3] M. A. Shah, R. Swaminathan, and M. Baker, "Privacy-preserving audit and extraction of digital contents".

[4] M. Bellare and G. Neven, "Multi-signatures in the plain public key model and a general forking lemma".

[5] R. C. Merkle, "Protocols for public key cryptosystems".