Scientific Journal of Impact Factor(SJIF): 3.134

e-ISSN(O): 2348-4470
p-ISSN(P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 2,Issue 11, November -2015

Deduplicating Data and Secure Integrity Verification in Cloud

NIhal Dastgir Mulla [1], Akshay Shetty [2], Varma Shubhangi Alok Kumar [3] Megha Tanaji Bugade[4] Prof. Leena Raut [5]

[1-5] *Department of Computer Engineering, Siddant College of Engineering,*

Abstract — *As the cloud computing innovation creates amid the most recent decade, outsourcing information to cloud administration for capacity turns into an alluring pattern, which benefits in saving endeavours on substantial information upkeep and administration, In any case, following the outsourced cloud stockpiling is not completely reliable, it raises security worries on the most proficient method to acknowledge information deduplication in cloud while accomplishing uprightness examining. In this work, we ponder the issue of honesty examining and secure deduplication on cloud information. Specifically, going for accomplishing both information uprightness and deduplication in cloud, we propose two protected frameworks, to be specific SecCloud and SecCloud+. SecCloud presents an examining substance with an upkeep of a Map Reduce cloud, are assists customers with producing information labels before transferring and in addition review the honesty of information having been put away in cloud. Contrasted and past work, the calculation by client in SecCloud is enormously lessened amid the file transferring and reviewing stages. SecCloud+ is composed propelled by the way that clients constantly need to scramble their information before transferring, and empowers honesty evaluating and secure deduplication on encoded inform*

Keywords- Integrity Auditing, SECCLOUD,SECCLOUD+, Secure Deduplication

I. INTRODUCTION

Cloud storage is a model of arranged venture stockpiling where data is put away in virtualized pools of capacity which are for the most part facilitated by third parties. The primary issue is integrity auditing. The cloud server can alleviate customers from the substantial weight of capacity administration and upkeep. The most contrast of cloud capacity from conventional in-house stockpiling is that the data is exchanged through Internet and put away in a questionable area, not under control of the customers by any stretch of the imagination, which definitely raises customers incredible worries on the uprightness of their information. These concerns begin from the way that the distributed storage is helpless to security dangers from both outside and inside of the cloud [1], and the uncontrolled cloud servers might inactively conceal some information misfortune occurrences from the customers to keep up their notoriety. The second issue is secure deduplication. The fast appropriation of cloud administrations is joined by expanding volumes of information put away at remote cloud servers. Among these remote put away documents, the greater part of them are copied: agreeing to a late overview by EMC [2], 75% of late advanced data is copied duplicates. This raises an innovation in particular deduplication, in which the cloud servers might want to deduplicate by keeping just a solitary duplicate for every record (or piece) and make a connection to the document (or piece) for each customer who claims on the other hand requests that store the same record (or piece). Sadly, this activity of deduplication would prompt various dangers conceivably influencing the stockpiling framework [3][2], for instance, a server telling a customer that it (i.e., the customer) does not require to send the document uncovers that some other customer has the definite same document, which could be delicate some of the time.

SecCloud presents an evaluating substance with an upkeep of a MapReduce cloud, which offers customers some assistance with generating information labels before transferring and in addition review the honesty of data having been put away in cloud. This outline alters the issue of past work that the computational burden at client or reviewer is as well colossal for label era. For culmination of fine-grained, the usefulness of inspecting planned in SecCoud is bolstered on both piece level and segment level. Furthermore, SecCoud additionally empowers secure deduplication. clients dependably need to encrypt their data before transferring, for reasons extending from individual protection to corporate strategy, we bring a key server into SecCloud as with [4] and propose the SecCloud+ outline. Other than supporting trustworthiness inspecting and secure deduplication, SecCloud+ empowers the surety of document privacy. In particular, on account of the property of deterministic encryption in joined encryption, we propose a system for straightforwardly reviewing honesty on encrypted data. The test of deduplication on encrypt is the anticipation of word reference assault [4]. Likewise with [4], we make a change on focalized encryption such that the joined key of document is produced and controlled by a secret "seed", such that any enemy couldn't specifically get the merged key from the substance of document and the word reference assault is forestalled.

Despite the fact that cloud stockpiling framework has been generally embraced, it neglects to oblige some critical emerging needs, for example, the capacities of auditing integrity of cloud files by cloud customers and detecting copied files by cloud servers. We show both issues underneath. The first issue is integrity auditing. The cloud server has the capacity alleviate customers from the substantial weight of capacity administration and maintenance. The most distinction of cloud stockpiling from customary in-house stockpiling is that the data is exchanged by means of Internet and put away in an uncertain domain, not under control of the customers by any stretch of the imagination, which inevitably raises customers extraordinary worries on the integrity of their data. These worries originate from the way that the cloud stockpiling is defenseless to security dangers from both outside and inside of the cloud, and the uncontrolled cloud servers might inactively conceal some data misfortune incidents from the customers to maintain their notoriety. In addition genuine is that for saving cash and space, the cloud servers may even effectively and purposely dispose of once in a while got to data files belonging to an ordinary customer. Considering the substantial size of the outsourced data files and the customers' constrained asset abilities, the first issue is summed up as in what manner can the customer efficiently perform periodical integrity verifications even without the neighborhood duplicate of data files.

II. LITERATURE REVIEW

1) Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing

AUTHORS: Qian Wang

Cloud Computing has been envisioned as the next-generation architecture of IT Enterprise. It 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. This unique paradigm brings about many new security challenges, which have not been well understood. This work studies the problem of ensuring the integrity of data storage in Cloud Computing. In particular, we consider the task of allowing a third party auditor (TPA), on behalf of the cloud client, to verify the integrity of the dynamic data stored in the cloud. The introduction of TPA eliminates the involvement of client through the auditing of whether his data stored in the cloud is indeed intact, which can be important in achieving economies of scale for Cloud Computing. The support for data dynamics via the most general forms of data operation, such as block modification, insertion and deletion, is also a significant step to- ward practicality, since services in Cloud Computing are not limited to archive or backup data only. While prior work on ensure remote data integrity often lack the supports of either public verifiability or dynamic data operation.

2) Proofs of Ownership in Remote Storage Systems

AUTHORS: ShaiHalevi

Cloud storage systems are becoming increasingly popular. A promising technology that keeps their cost down is deduplication, whichstoresonlyasinglecopyofrepeatingdata.Client-sidededuplication attempts to identify deduplication opportunities already attheclientandsavethebandwidthofuploadingcopiesofexisting files to the server. In this work we identify attacks that exploit client-side deduplication, allowing an attacker to gain access to arbitrary-size files of other users based on a very small hash signature of these files. More specifically, an attacker who knows the hash signature of a filecanconvincethestorageservicethatitownsthatfile,hencethe server lets the attacker download the entire file.

3.DupLESS: Server-Aided Encryption for Deduplicated Storage

AUTHORS: MihirBellare

Cloud storage service providers such as Dropbox, Mozy, and others perform deduplication to save space by only storing one copy of each file uploaded. Should clients conventionally encrypt their files, however, savings are lost. Message-locked encryption (the most prominent manifestation of which is convergent encryption) resolves this tension. However it is inherently subject to brute-force attacks that can recover files falling into a known set. We propose an architecture that provides secure deduplicated storage resisting brute-force attacks, and realize it in a system called DupLESS. In DupLESS, clients encrypt under message-based keys obtained from a key-server via an oblivious PRF protocol. It enables clients to store encrypted data with an existing service, have the service perform deduplication on their behalf, and yet achieves strong confidentiality guarantees. We show that encryption for deduplicated storage can achieve performance and space savings close to that of using the storage service with plaintext data.

4.Provable Data Possession at Untrusted Stores

AUTHORS: Giuseppe Ateniese

We introduce a model for provable data possession (PDP) that allows a client that has stored data at an untrusted server to verify that the server possesses the original data without retrieving it. The model generates probabilistic proofs of possession by sampling random sets of blocks from the server, which drastically reduces I/O costs. The client maintains a constant amount of metadata to verify the proof. The challenge/response protocol transmits a small, constant amount of data, which minimizes network communication. Thus, the PDP model for remote data checking supports large data sets in widely-distributed storage systems.

5.**Remote Data Checking Using Provable Data Possession**

AUTHORS: GIUSEPPE ATENIESE

We introduce a model for provable data possession (PDP) that can be used for remote data checking: A client that has stored data at an untrusted server can verify that the server possesses the original data without retrieving it. The model generates probabilistic proofs of possession by sampling random sets of blocks from the server, which drastically reduces I/O costs. The client maintains a constant amount of metadata to verify the proof. The challenge/response protocol transmits a small, constant amount of data, which minimizes network communication. Thus, the PDP model for remote data checking is lightweight and supports large data sets in distributed storage systems. The model is also robust in that it incorporates mechanisms for mitigating arbitrary amounts of data corruption.

III. SURVEY OF PROPOSED SYSTEM

We determine that our proposed SecCloud framework has accomplished both integrity auditing and file deduplication. Be that as it may, it can't keep the cloud servers from knowing the substance of files having been put away. In other words, the functionalities of integrity auditing and secure deduplication are just forced on plain files. In this area, we propose SecCloud+, which takes into account integrity auditing and deduplication on scrambled files. Framework Model Compared with SecCloud, our proposed SecCloud+ involves an extra trusted element, to be specific key server, which is in charge of assigning customers with mystery key (according to the file content) for encrypting files. This construction modeling is in line with the late work. However, our work is distinguished with the past work by allowing for integrity auditing on encoded data. SecCloud+ takes after the same three protocols (i.e., the file uploading protocol, the integrity auditing protocol and the proof of proprietorship protocol) as with SecCloud. The main distinction is the file uploading protocol in SecCloud+ involves an extra stage for correspondence between cloud customer and key server. That is, the customer needs to speak with the key server to get the merged key for encrypting the uploading file before the phase in SeeCloud.

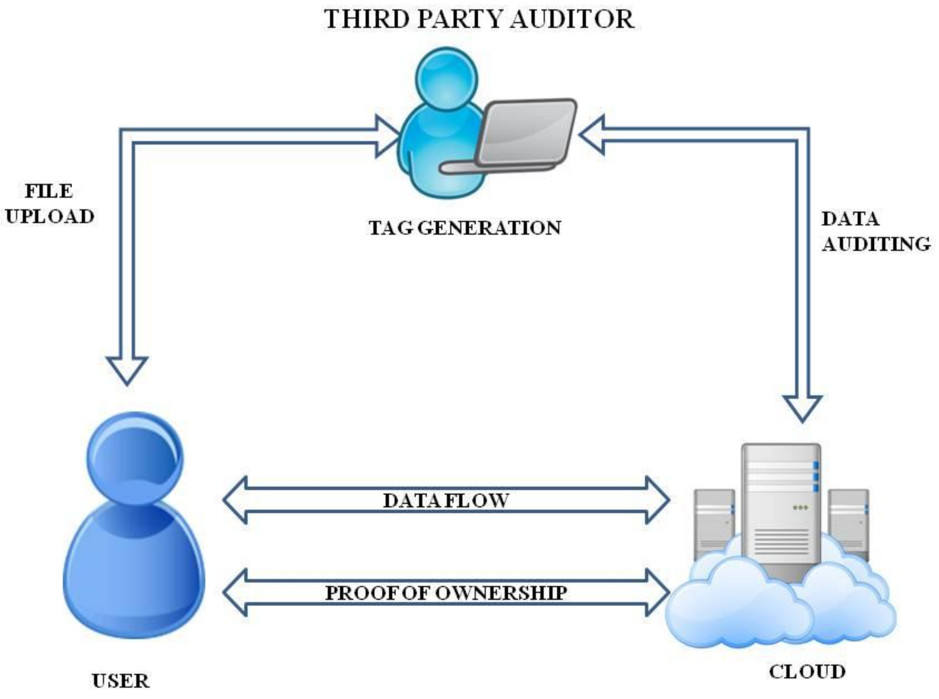
IV Mathematical Model

Let S is the Whole System Consist of
S={U,F,TPA,CSP,DB}
Where,

U= no of users
U={u1,u2,u3,…..un}
F= no of files
F={f1,f2,f3,…..fn}
TPA= Third Party Auditor
TPA={C,PF,V,POW}
Where,
C=challenge
PF =proof by CSP
V= verification by TPA
POW= proof of ownership
CSP= Cloud Service provider
CSP={PF,F}
where
PF=proof
F=files

OUTPUT: Secure auditing & unique Storage of File in Cloud

V SYSTEM ARCHITECTURE



.

VI CONCLUSION AND FUTURE WORK

Aiming at achieving both data integrity and deduplication in cloud, we propose SecCloud and SecCloud+. SecCloud introduces an auditing substance with maintenance of a MapReduce cloud, which assists customers with generating data labels before uploading and additionally reviews the integrity of data having been put away in cloud. Furthermore, SecCoud empowers secure deduplication through introducing a Proof of Ownership protocol and preventing the leakage of side channel information in data deduplication. Contrasted and past work, the calculation by client in SecCloud is incredibly diminished during the file uploading and auditing stages. SecCloud+ is a propelled development persuaded by the way that clients constantly need to encode their data before uploading, and takes into account integrity auditing and secure deduplication straightforwardly on scrambled data.

ACKNOWLEDGMENT

We might want to thank the analysts and also distributers for making their assets accessible. We additionally appreciative to commentator for their significant recommendations furthermore thank the school powers for giving the obliged base and backing.

VII REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “A view of cloud computing,” *Communication of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.

[2] J. Yuan and S. Yu, “Secure and constant cost public cloud storage auditing with deduplication,” in *IEEE Conference on Communications and Network Security (CNS)*, 2013, pp. 145–153.

[3] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, “Proofs of ownership in remote storage systems,” in *Proceedings of the 18th ACM Conference on Computer and Communications Security*. ACM, 2011, pp. 491–500.

[4] S. Keelveedhi, M. Bellare, and T. Ristenpart, “Dupless: Serveraided encryption for deduplicated storage,” in *Proceedings of the 22Nd USENIX Conference on Security*, ser. SEC’13. Washington, D.C.: USENIX Association, 2013, pp. 179–194. [Online]. Available: https://www.usenix.org/conference/usenixsecurity13/technicalsessions/presentation/bellare

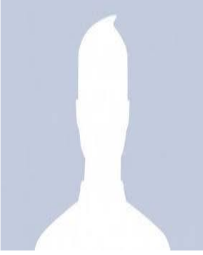
AUTHORS

NIihal Dastgir Mulla, pursuing the B.E degree in Computer Engineering at Siddhant College of Engineering ,Talwade, Pune

Akshay Shetty, pursuing the B.E degree in Computer Engineering at Siddhant College of Engineering ,Talwade, Pune

Varma Shubhangi Alok Kumar, pursuing the B.E degree in Computer Engineering at Siddhant College of Engineering ,Talwade, Pune

Megha Tanaji Bugade, pursuing the B.E degree in Computer Engineering at Siddhant College of Engineering ,Talwade, Pune



Prof. Leena raut, Assistant Professor in Computer Engineering at Siddhant College of Engineering ,Talwade, Pune