Scientific Journal of Impact Factor (SJIF): 4.72

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

International Journal of Advance Engineering and Research Development

Volume 4, Issue 12, December -2017

Cloud Data Integrity Checking Using Third Party Auditor: A Survey

[1]Prof. S. R. Nalamwar, [2]Shweta Jagtap , [3]Sneha Mahadik, [4]Gauri Ranade, [5]Safiya Shaikh

[1]*Professor, AISSMS COE , Pune, Maharashtra, India*
[2]*BE, Student, AISSMS COE , Pune,Maharashtra,India*
[3]*BE, Student, AISSMSCOE , Pune,Maharashtra,India*
[4]*BE, Student, AISSMS COE, Pune, Maharashtra, India*
[5]*BE, Student, AISSMS COE, Pune, Maharashtra, India*

Abstract —*Remote data integrity checking (RDIC) allows checking the data integrityfor the data stored on cloud. There are number of RDIC protocols have been proposed in the literature, but they suffer from issue of a complex key management. In thisproject, we propose a new construction of identity-based (ID-based) RDIC protocolby making use of security key primitives to reduce the system complexity and the costfor establishing and managing the public key authentication framework in PKI basedRDIC schemes. In our project there are three main concepts that are cloud server,user and third party auditor (TPA). TPA gives the proof that our data integrity is maintain or not, in addition we are also providing the security mechanisms .In caseif the data is gets modified by the cloud server or unauthorized person then user getstheir original data from the dummy server. Third Party Auditor (TPA) is responsiblefor checking the integrity of the cloud data on behalf of the cloud users in case if clouduser does not have time to monitor their resources and integrity of data, and returnsthe auditing report to the cloud user.*

Keywords-*Cloud storage, data integrity, privacy preserving, identity-based cryptography.*

I. INTRODUCTION

Cloud computing, that has received sizable attention from analysis communities in domain likewise as trade, may be a distributed computation model over an outsized pool of shared-virtualized computing resources, like storage, process power, applications and services. Cloud users can use resources as they require in cloud computing atmosphere. This type of latest computation model represents a replacement vision of providing computing services as public utilities like water and electricity. Cloud computing brings variety of advantages for cloud users. For example, Users will scale back cost on hardware, software package and services as a result of they pay just for what they use; Users will get pleasure from low management overhead and immediate access to a good variety of applications; and Users will access their information where they need a network, instead of having to remain near their computers. However, there's an enormous type of barriers before cloud computing are often wide deployed. A recent survey by Oracle referred the information supply from international data corporation enterprise panel, showing that security represents 87 of cloud users' fears. The key security considerations of cloud users is that the integrity of their outsourced files, since they do not physically possess their knowledge and so can lose the management over their knowledge. Moreover, the cloud server isn't totally trustworthy and it's not obligatory for the cloud server to report information loss incidents. Indeed, to establish cloud computing irresponsibleness, the cloud security alliance (CSA) revealed an analysis of cloud vulnerability incidents. The investigation disclosed that the incident of information Loss and escape accounted for twenty fifth of all incidents, stratified second solely to"Insecure Interfaces and APIs". Amazon's immense EC2 cloud services crash for good destroyed some knowledge of cloud users. The information loss was apparently tiny relative to the whole data hold on, however any UN agency that runs an internet site will forthwith perceive however terrific a possibility of any knowledge loss is. Generally it's poor to observe knowledge corruption once accessing the knowledge it'd be too late to recover the corrupted data. As a result, it's necessary for cloud users to check if their outsourced information iskept properly.

II. LITERATURE SURVEY

According to literature survey after studying various IEEE paper, collected some related papers and documents.Some of the pointsdescribe here are:

1. Reclaiming Space from Duplicate Files in a Serverless Distributed FileSystem
Authors : John R. Douceur, A. Adya, W.J. Bolosky, P. Simon, M. Theimer
The Farsite distributed file system provides availability by replicating each file ontomultiple desktop computers. Since this replication consumes significant storage space,it is important to reclaim used space where possible. Measurement of over 500 desktop file systems shows that nearly half of all consumed space is occupied by duplicatefiles. They present a

mechanism to reclaim space from this incidental duplication tomake it available for controlled file replication. Their mechanism includes 1) convergent encryption, which enables duplicate files to coalesced into the space of a single_le, even if the files are encrypted with different users keys, and 2) SALAD, a Self-Arranging, Lossy, Associative Database for aggregating file content and location in-formation in a decentralized, scalable, fault-tolerant manner. Large-scale simulationexperiments show that the duplicate-file coalescing system is scalable, highly effective, and fault-tolerant.

2. DupLESS: Server-Aided Encryption for Deduplicated Storage
Authors : MihirBellare, SriramKeelveedhi, Thomas Ristenpart
Cloud storage service providers such as Dropbox, Mozy, and others perform deduplication to save space by only storing one copy of each file uploaded. Message-lockedencryption (the most prominent manifestation of which is convergent encryption) resolves this tension. However it is inherently subject to brute-force attacks that canrecover files falling into a known set. They propose an architecture that providessecure deduplicated storage resisting brute-force attacks, and realize it in a systemcalled DupLESS. In DupLESS, clients encrypt under message-based keys obtainedfrom a key-server via an oblivious PRF protocol. It enables clients to store encrypteddata with an existing service, have the service perform deduplication on their behalf,and yet achieves strong confidentiality guarantees. They show that encryption for deduplicated storage can achieve performance and space savings close to that of using the storage service with plaintext data.

3. Message-Locked Encryption and Secure Deduplication.
Authors : MihirBellare, SriramKeelveedhi, Thomas Ristenpart
They formalize a new cryptographic primitive, Message-Locked Encryption (MLE),where the key under which encryption and decryption are performed is itself derived from the message. MLE provides a way to achieve secure deduplication (space-efficient secure outsourced storage), a goal currently targeted by numerous cloud-storageproviders. They provide definitions both for privacy and for a form of integrity thatwe call tag consistency. Based on this foundation, they make both practical andtheoretical contributions. On the practical side, they provide ROM security analysesof a natural family of MLE schemes that includes deployed schemes. On the theoretical side the challenge is standard model solutions, and they make connectionswith deterministic encryption, hash functions secure on correlated inputs and thesample-then-extract paradigm to deliver schemes under different assumptions and fordifferent classes of message sources. Their work shows that MLE is a primitive of bothpractical and theoretical interest.

4. CD Store:Toward Reliable,Secure,and Cost-Efficient Cloud Storage viaConvergent Dispersal
Authors : Mingqiang Li, Chuan Qin, Patrick P. C. Lee
They present CD Store,which disperses usersbackup data across multiple clouds andprovides a unified multi-cloud storage solution with reliability,security,and cost- efficiencyguarantees. CD Store builds on an augmented secret sharing scheme called convergent dispersal,which supports deduplication by using deterministic content- derivedhashes as inputs to secret sharing. They present the design of CD Store, and in particular, describe how it combines convergent dispersal with two-stage deduplicationonachieve both bandwidth than the storage saving sand be robust against side-channelattacks. They evaluate the performance of our CD Store prototype using real-worldworkloads on LAN and commercial cloud test beds.Their cost analysis also demonstrates that CD Store achieves a monetary cost saving of 70 over a baseline cloudstorage solution using state-of-the-art secret sharing.

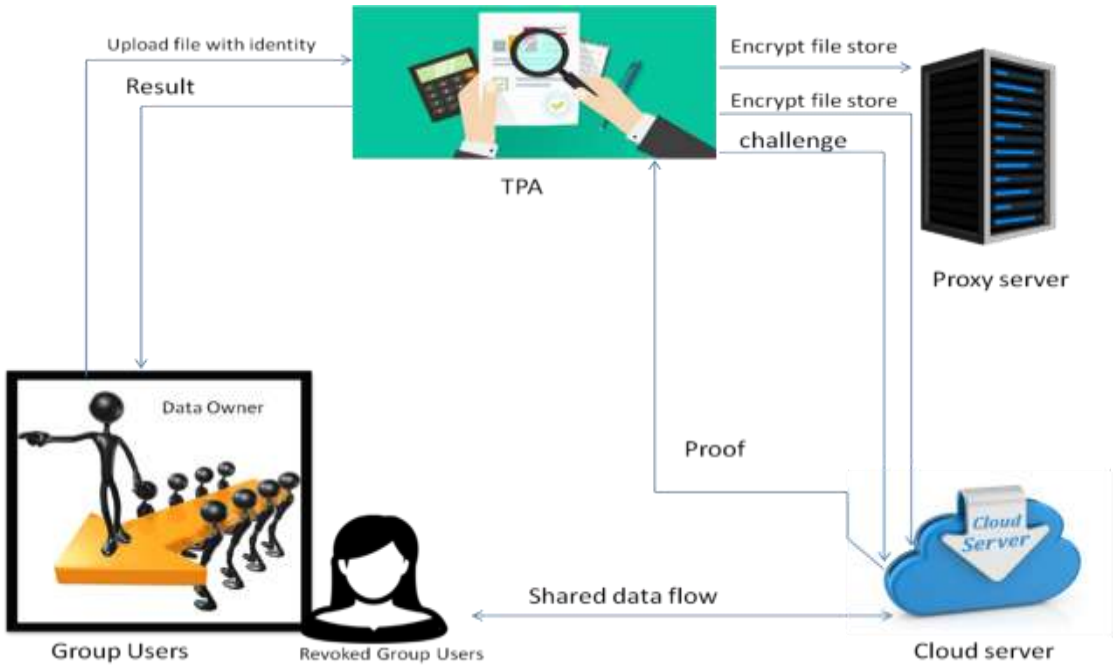
5. Is Naive Bayes a Good Classifier for Document Classification?
Authors : S.L. Ting, W.H. Ip, Albert H.C. Tsang
In this paper the author has highlighted the performance of implementing Naïve Bayes classifier against several other classifiers such as decision tree, neural network,and support vector machines in terms of accuracy and computational efficiency. Intheir study, Naive Bayes classifier has been discussed as the classifier, which satisfiesthe literature result. Through their implementation of different feature selection inWEKA tool, they have demonstrated that preprocessing and feature selection are important two steps for improving the mining quality. To test whether Naive Bayes isthe best classifier among other classifiers, they have applied three different classifiersfor testing. In their experiment, a dataset of 4000 documents are used for evaluation.1200 documents are extracted randomly to build the training dataset for the classifier. The other 2800 documents are used as the testing dataset to test the classifier.They have summarized that Naive Bayes classifier gave 96.9 percent of accuracy whileclassifying.

III. PROPOSED SYSTEM

We propose a replacement construction of identity-based (ID-based) RDIC protocol by creating use of key homomorphism cryptology primitive to cut back the system quality and also the value for establishing and managing the general public key authentication framework in PKI based mostly RDIC schemes. We have a tendency to formalize ID-based RDIC and its security model together with security against a malicious cloud server and zero data privacy against a third party auditor. The planned ID-based RDIC protocol leaks no information of the kept knowledge to the auditor throughout the RDIC method. The new construction is evidenced secure against the malicious server within the generic cluster model and achieves zero data privacy against the auditor. In depth security analysis results demonstrate that the

planned protocol is demonstrably secure and sensible within the real-world applications. We Extend this work with time span based third party auditor system and recovery of file once knowledge integrity checking fault occur.

IV. SYSTEM DESIGN



V. ADVANTAGES

- Remote data integrity checking for secure cloud storage.
- It achieves soundness and perfect data privacy.
- Systemproposes a protocol that is provably secure and practical in the real-world applications.

VI. CONCLUSION

Identity-based remote data integrity checking protocol successfully provides secure cloud storage. The security model provides two important properties ofthis primitive namely, soundness and perfect data privacy. In addition to the previouswork, we added time span based third party audition system and data backup. Thenumerical analysis demonstrated that the proposed protocol (Remote Data IntegrityChecking Protocol) is efficient and practical.

REFRENCES

[1]H. Wang, Identity-based distributed provable data possession in multicloud storage, IEEE Trans. on Service Computing, 8(2), 328–340, 2015.

[2] G. Ateniese, R. C. Burns, R. Curtmola, J. Herring, O. Khan, L. Kissner, Z. N. J. Peterson, and D. Song, Remote data checking using provable data possession. ACM Trans. Inf. Syst. Secur., 14, 1–34, 2011.

[3] A. F. Barsoum, M. A. Hasan, Provable multicopy dynamic data possession in cloud computing systems, IEEE Trans. on Information Forensics and Security, 10(3): 485–497, 2015.

[4] D. Boneh, M. Franklin, Identity-based encryption from the weil pairing, Proc. of CRYPTO 2001, LNCS 2139, 213–229, 2001.

[5] G. Ateniese, R. C. Burns, R. Curtmola, J. Herring, O. Khan, L. Kissner,Z. N. J. Peterson, and D. Song, Remote data checking using provabledata possession. ACM Trans. Inf. Syst. Secur., 14, 1–34, 2011.

[6] A.Juels, and B. S. K. Jr. Pors, proofs of retrievability for large files.Proc.of CCS 2007, 584-597, 2007.

[7] H. Shacham, and B. Waters, Compact proofs of retrievability. Proc. OfCryptology-ASIACRYPT 2008, LNCS 5350, pp. 90-107, 2008.

[8] G. Ateniese, S. Kamara, J. Katz, Proofs of storage from homomorphicidentification protocols. Proc. of ASIACRYPT 2009, 319-333, 2009.

[9] A. F. Barsoum, M. A. Hasan, Provable multicity dynamic data possessionin cloud computing systems, IEEE Trans. on Information Forensicsand Security, 10(3): 485–497, 2015.

[10] J. Yu, K. Ren, C.Wang, V. Varadharajan, Enabling cloud storage auditingwith key-exposure resistance, IEEE Trans. on Information Forensics andSecurity, 10(6): 1167–1179, 2015.

[11] J. Liu, K. Huang, H. Rong, H. M. Wang, Privacy-preserving publicauditing for regenerating-code-based cloud storage, IEEE Trans. OnInformation Forensics and Security, 10(7): 1513–1528, 2015.

[12] Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou, Enabling publicverifiability and data dynamics for storage security in cloud computing.Proc. of ESORICS2009, LNCS 5789, 355–370, 2009.

