DEDUPLICATING DATA AND REMOVING REDUNDANCY IN CLOUD

[1]Arun Singh Kaurav, [2]T.Santhosh Kumar, [3]V.Yadigiri
[1]Assistant Professor, [2]Assistant Professor, [3]Assistant Professor
[1]Computer Science and Engineering
Guru Nanak Institutions Technical Campus, Hyderabad, India

ABSTRACT

Rendering economical storage and security for all knowledge is incredibly necessary for cloud computing. Securing and privacy conserving of information is of high priority once it involves cloud storage. Thus to produce economical storage for cloud knowledge house owners and render high security for knowledge this paper proposes Cloud Computing Secure Framework (CCSF). Therefore CCSF consists of 4 segments: 1) Identity Management 2) Intrusion detection and bar system 3) knowledge deduplication 4) Secure Cloud Storage. Intrusion detection and bar square measure performed manually by network operators within the existing system. In our projected design the intrusion detection and bar is performed mechanically by shaping rules for the main attacks and alert the system mechanically. The main attacks/events includes vulnerabilities, cross web site scripting (XSS), SQL injection, cookie poisoning, wrapping. Knowledge deduplication technique permits the cloud users to manage their cloud space for storing effectively by avoiding storage of continual data's and save information measure. The information square measure finally holdon incloudserver particularly CloudMe. Toconfirm knowledge confidentiality theinformationsquaremeasure holdon in AN encrypted sort mistreatment Advanced cryptography customary (AES) rule

INTRODUCTION

Cloud computing is one in every of the rising technology, that helped many organizations to save lots of cash and time adding convenience to the tip users. Therefore the scope of cloud storage is Brobdingnagian as a result of the organizations will nearly store their data's while not bothering the whole mechanism. Cloud Computing provides key advantage to the tip users like price savings, ready to access the information no matter location, performance and security. In our projected system we have a tendency to invoke a effective user authentication mistreatment fingerprint feature extraction, image based mostly authentication throughout file upload/download, eliminating repetition of information in cloudserverand enforced throughmultiplecloudstorage.

Maximum authentication system has demerits, for which graphical passwords square measure desirable authentication system where users click on pictures to explain themselves. Our projected system states image based mostly effective authentication. Once the Admin uploads the come in the cloud, the admin can split the image into four elements. The admin can hold two elements andthereforethe userofthat individual cluster will read theopposite two elements. Thephotographs squaremeasure spilt haphazardlymistreatment pseudo random generator technique. Once the user tries to transfer a file, the user will send the requisition to the individual admin aspect in conjunction with beside at the side of together with} the user side accessible two elements.

The admin can verify each the elements and if the authentication is passed, the file is going to be sent to the user in AN encrypted approach. Knowledge deduplication is one in every of the techniques that accustomed solve the repetition of information. The deduplication techniques square measure typically utilized in the cloud server for reducing the house of the server. to forestall the unauthorized use {of knowledge of knowledge of information} accessing and make duplicate data on cloud the cryptography technique to cipher the information before hold on on cloud server. Cloud Storage sometimes contains business-critical knowledge and processes; thence high security is that

the solely answer to retain sturdy trust relationship between the cloud users and cloud service suppliers. Therefore to beat the protection threats, this paper proposes multiple cloud storage. Therefore the common kinds of knowledge storage like files and databases of a selected user are split and hold on within the varied cloud storages (e.g. Cloud A and Cloud B).

EXPERIMENTAL LITERATUREREVIEW

Data deduplication in cloud computing may be a paradigm shift within the net technology. Knowledge deduplication will save space for storing and scale back the quantity of information measure of information transfer. Secure and constant price public cloud storage auditing with deduplication system within the cloud storage is employed to cut back the storage size of the tags for integrity check. Fingerprint verification supported trivia features: a review the fingerprint feature extraction and matching is performed mistreatment trivia Map rule (MM). Trivia is that the relevancy bifurcation and termination values of the ridges within the fingerprint. The distribution on the fingerprint
provides anovel signature forevery andeach individual.StorageFileexchangeandfileretrieving Functions of information deduplication it compares objects (usually files or blocks) and removes objects (copies) that exist already within the knowledge set. The deduplication method removes blocks that don't seem to be distinctive. 1. Divide the computer file into blocks or "chunks." 2. Calculate a hash price for every block of information. 3. Use these values to work out if another block of a similar knowledge has already been hold on. 4. Replace the duplicate knowledge with relevancy thethingalready withinthe information.

The actual method of information deduplication will be enforced in a very variety of various ways that. We are able to eliminate duplicate knowledge by merely comparison 2 files and creating the choice to delete one that's older or now not required.

EXSISTING SYSTEM

For verification of integrity the Integrity auditing interactive protocol is to be initialized by an entity except server cloud. During this protocol, the cloud server plays the role of prover, whereas the auditor or shopper works because the booster. This protocol includes 2 phases: section one (cloud client/auditor → cloud server): verifier(i.e., shopper or auditor) generates a group of challenges and sends them to the prove (i.e., cloud server) section two (cloud server → cloud client/auditor)based on the hold on files and file tags, prover (i.e., cloud server)tries to prove that it precisely owns the target file by causing the proof back to booster (i.e., cloud shopper or auditor).At the tip of this protocol, booster outputs true if the integrityverificationis passed.

PROPOSED TECHNIQUE

Deduplication may be a technique wherever the server stores solely one copy of every file, despite what percentage shoppers asked to store that file, specified the disc space of cloud servers yet as network information measure square measure saved. However, trivial shopper facet deduplication results in the run of facet channel data. For example, a server telling a shopper that it needn't send the file reveals that another shopper has the precise same file that can be sensitive data in some case. In order to limit the run of facet channel data, Halevi et al. introduced the proof of possession protocol that lets a shopper expeditiously encourage a server that that the shopper precisely holds this file.

MODULES

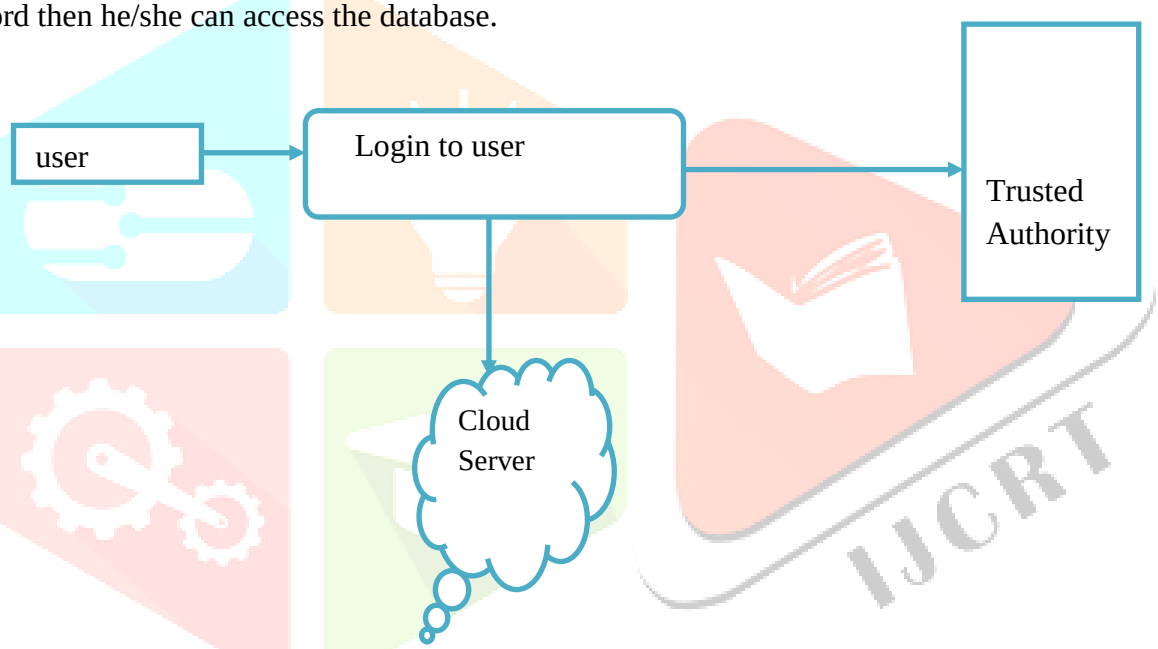
User Interface Design

- Key Generation
- Secure Deduplication
- Secured Auditing
- Proof of ownership

User Interface Design

Explanation

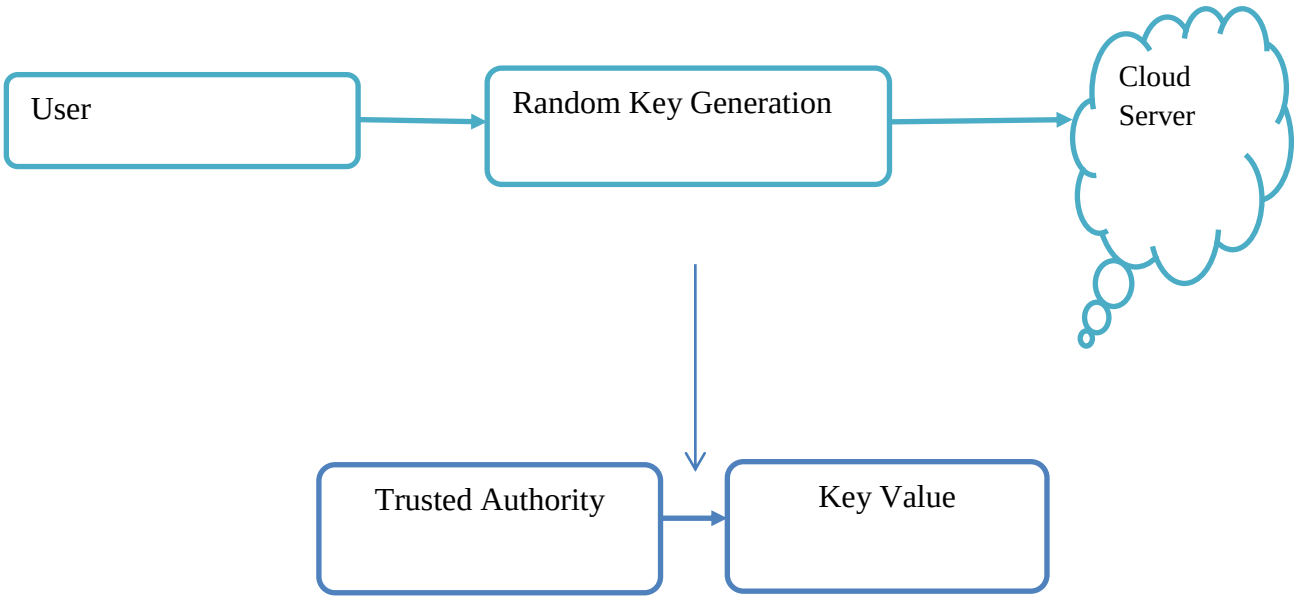
In this module we design the windows for the project. These windows are used to send a message from one peer to another. We use the Swing package available in Java to design the User Interface. Swing is a widget toolkit for Java. In this module mainly we are focusing the login design page with the Partial knowledge information. Application Users need to view the application they need to login through the User Interface GUI is the media to connect User and Media Database and login screen where user can input his/her user name, password and password will check in database, if that will be a valid username and password then he/she can access the database.



Key Generation and Encryption

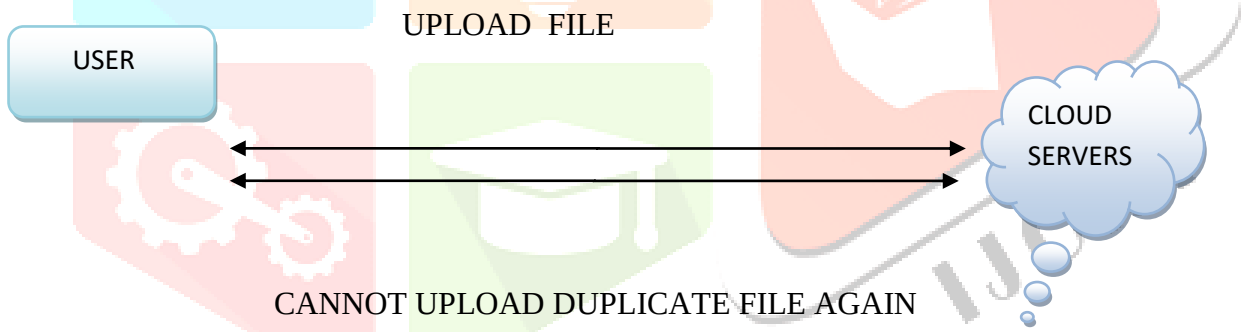
Explanation

Security parameter λ is taken as input in Algorithm and public key PK and the master key MK are the outputs. Master key is secret at PKG.The private key generation algorithm run by PKG, takes as input the master key MK and user's identity ID ∈ {0, 1}∗. Private key SKID respective to the identity ID will be returned...Sender runs the encryption algorithm, which takes the receiver's identity ID and a message M to be encrypted as input. Cipher text CT will be the output. Receiver runs the decryption algorithm, takes as input the cipher text CT and private key SKIDs. Message M or an error will be returned.



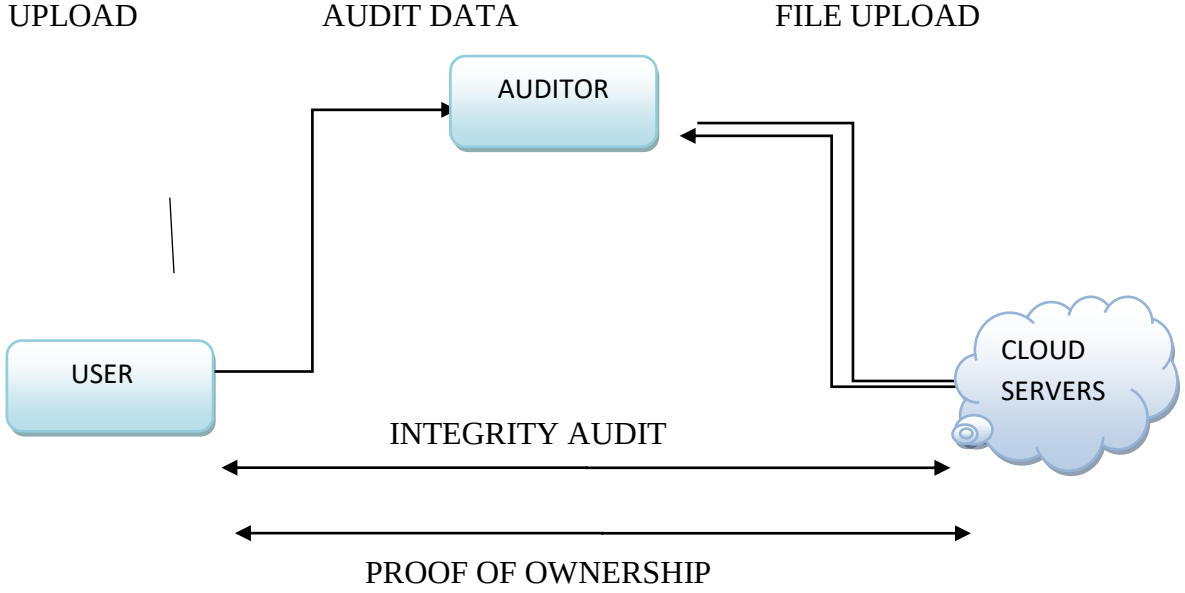
Secure De duplication

De duplication is the mechanism where server stores only single copy of each file, regardless of how many clients are asked to store that file, so that the disk space of servers cloud as well as network bandwidth are saved. However, trivial client side De duplication leads to the leakage of side channel information. For example, a server telling a client that it need not send the file reveals that some other client has the exact same file, which could be sensitive information in some case.



Secured Auditing & Proof of ownership
Explanation

Even the storage of cloud system is being adopted; it still fails to have some requirements like the capability of auditing the cloud files integrity by the clients cloud and finding duplicated files by server cloud. We illustrate both problems below. The first problem is integrity auditing. The cloud server is able to relieve clients from the heavy burden of storage management and maintenance. The most difference of cloud storage from traditional in-house storage is that the data is transferred via Internet and stored in an uncertain domain, not under control of the clients at all, which inevitably raises clients great concerns on the integrity of their data. This interactive protocol is started at the server cloud for verification of that client exactly has a claimed file. This protocol is triggered with file uploading protocol for the prevention of the leakage of side channel information. In the opposite of integrity auditing protocol, in PoW the server cloud works as verifier, whereas the client has the role of prover.



CONCLUSION

Thus this paper compresses {the knowledge the info the information} by removing the duplicate copies of identical data and it's extensively utilized in cloud storage to save lots of information measure and minimize the space for storing. To secure the confidentiality of sensitive knowledge throughout deduplication.The focussed cryptography technique isemployed to cipher th e information before outsourcing. For higher knowledge protection, this paper talks regarding thedifficulty ofinformation deduplicationauthorization.

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 read of cloud computing," Communication of the ACM,vol.53,no.4,pp.50–58,2010.

[2] J. Yuan and S. Yu, "Secure and constant price 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 possession in remote storage systems," in Proceedings of the eighteenth ACM Conference on pc and CommunicationsSecurity.CM,2011,pp.491–500.

[4] S. Keelveedhi, M. Bellare, and T. Ristenpart, "Dupless: Serveraided cryptography for deduplicated storage," in Proceedings of the twenty second USENIX Conference on Security, ser. SEC'13. Washington, D.C.: USENIX Association, 2013, pp. 179–194. [Online]. Available: