ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

Secure Access Policies Based Data Deduplication System

Rejinpaul[1], Aravind.S[2], Hemanth.K.B[3]

Assistant Professor, Dept. of CSE, Velammal Institute of Technology, Chennai, Tamilnadu, India[1]

UG Student, Dept. of CSE, Velammal Institute of Technology, Chennai, Tamilnadu, India [2,3]

ABSTRACT: Cloud computing services can be private, public or hybrid. Private cloud services are delivered from a business' data center to internal users. This model offers versatility and convenience, while preserving the management, control and security common to local data centers. Internal users may or may not be billed for services through IT chargeback. In the public cloud model, a third-party provider delivers the cloud service over the internet. Public cloud services are sold on demand, typically by the minute or hour. Customers only pay for the CPU cycles, storage or bandwidth they consume. Leading public cloud providers include Amazon Web Services (AWS), Microsoft Azure, IBM SoftLayer and Google Compute Engine Hybrid cloud is a combination of public cloud services and on-premises private cloud -- with orchestration and automation between the two. Companies can run mission-critical workloads or sensitive applications on the private cloud while using the public cloud for bursting workloads that must scale on demand. The goal of hybrid cloud is to create a unified, automated, scalable environment that takes advantage of all that a public cloud infrastructure can provide while still maintaining control over mission-critical data. Existing solutions of encrypted data de duplication suffer from security weakness. They cannot flexibly support data access control and revocation. Therefore, few of them can be readily deployed in practice. In this paper, we propose a scheme to de duplicate encrypted data stored in cloud based on ownership challenge and proxy re-encryption.

KEYWORDS: Deduplication, Scalable ,Attribute based storage.

I.INTRODUCTION

Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. It makes Use of the commodity hardware Hadoop is Highly Scalable and Fault Tolerant which provides resource management and scheduling for user applications; and Hadoop MapReduce, which provides the programming model used to tackle large distributed data processing -- mapping data and reducing it to a result. Big Data in most companies are processed by Hadoop by submitting the jobs to Master. This type of usage is best-suited to highly scalable public cloud services, The Master distributes the job to its cluster and process map and reduces tasks sequentially. But nowadays the growing data need and the competition between Service Providers leads to the increased submission of jobs to the Master. This Concurrent job submission on Hadoop forces us to do Scheduling on Hadoop Cluster so that the response time will be acceptable for each submission..

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

II.LITERATURE SURVEY

A. Prashant Bhatt1, PG Student, Network Security, GTU PG School, Ahmedabad, Gujarat, IndiaCloud based Storage Drive Forensics. Cloud Storage is recently as emerging topic

in these eras. As the data are increasing, the storage become major issue for the people. There are different kind of Cloud Storage application such as One Drive, Sky Drive, Drop Box and Google Drive. Google Drive is gaining more popularity as it is user friendly than any other Cloud Storage Application. Google Drive is a Cloud Storage Application which allows user to store, share and edit the file in the cloud. In these paper, the authors will perform forensics of Google Drive via di1fferent technique such as using client software, Google Drive access via browser, Memory Analysis, Network Analysis and other techniques. From that the Authors will find, what type of data remnants can be found in user device

B. Amit Sahai,: Fuzzy Identity Based Encryption We introduce a new type of Identity Based

Encryption (IBE) scheme that we call Fuzzy Identity Based Encryption. A Fuzzy IBE scheme allows for a private key for an identity id to decrypt a ciphertext encrypted with another identity id0 if and only if the identities id and id0 are close to each other as measured by some metric (e.g. Hamming distance). A Fuzzy IBE scheme can be applied to enable encryption using biometric measurements as identities. The error-tolerance of a Fuzzy IBE scheme is precisely what allows for the use of biometric identities, which inherently contain some amount of noise during each measurement. In this paper we present a construction of a Fuzzy IBE scheme that uses groups with e±- ciently computable bilinear maps. Additionally, our construction does not use Random Oracles. We prove the security of our scheme under the Selective-ID security model.

C. Martin Abadi Message-Locked Encryption forLock-Dependent Messages Motivated by

the problem of avoiding duplication in storage systems, Bellare, Keelveedhi, and Ristenpart have recently put forward the notion of Message-Locked Encryption (MLE) schemes which sub- sumes convergent encryption and its variants. Such schemes do not rely on permanent secret keys, but rather encrypt messages using keys de-rived from the messages themselves. We strengthen the notions of security proposed by Bellare et al. by con-sidering plaintext distributions that may depend on the public parameters of the schemes. We refer to such inputs as lock-dependent messages. We construct two schemes that satisfy our new notions of security for message-locked encryption with lock-dependent messages. Our main construction deviates from the approach of Bellare et al. by avoiding the use of ciphertext components derived deterministically from the messages. We design a fully randomized scheme that supports an equality-testing algorithm de ned on the ciphertexts. Our second construction has a deterministic ciphertext component that enables more

D. Mihir Bellar, Message-Locked Encryption and Secure Deduplication We 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 de duplication (space-e cient secure outsourced storage), a goal currently targeted by numerous cloud-storage providers. We provide de nitions both for privacy and for a form of integrity that we call tag consistency. Based on this foundation, we make both practical and theoretical contributions. On the practical side, we provide ROM security analyses of a natural family of MLE schemes that includes deployed schemes. On the theoretical side the challenge is standard model solutions, and we make connections with deterministic encryption, hash functions secure on correlated inputs and the sample-then-extract paradigm to deliver schemes under di erent assumptions and for di erent classes of message sources. Our work shows that MLE is a primitive of both practical and theoretical interest.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

E. Farhood Norouzizadeh Dezfoli, Ali Dehghantanha, Ramlan Mahmoud, Nor Fazlida BintiMohd Sani, Farid Daryabar, Digital Forensic Trends and Future.

Nowadays, rapid evolution of computers and mobile phones has caused these devices to be used in criminal activities. Providing appropriate and sufficient security measures is a difficult job due to complexity of devices which makes investigating crimes involving these devices even harder. Digital forensic is the procedure of investigating computer crimes in the cyber world. Many researches have been done in this area to help forensic investigation to resolve existing challenges. This paper attempts to look into trends of applications of digital forensics and security at hand in various aspects and provide some estimations about future research trends in this area.

III.EXISTING SYSTEM:

In the previous deduplication systems have only been maintain single-server setting..A data provider, Bob, intends to upload a file M to the cloud, and share M with users having .certain credentials. In order to do so, Bob encrypts M under an access policy A over a set of attributes, and uploads the corresponding ciphertext to the cloud, such that only users whose sets of attributes satisfying the access policy can decrypt the ciphertext. Later, another data provider, Alice, uploads a ciphertext for the same underlying file M but ascribed to a different access policy A0. Since the file is uploaded in an encrypted form, the cloud is not able to discern that the plaintext corresponding to Alice's ciphertext is the same as that corresponding to Bob's, and will store M twice. Obviously, such duplicated storage wastes storage space and communication bandwidth. So . they requires that the deduplication storage systems provide reliability comparable to other high-available systems. However, as lots of deduplication systems and Storage systems are intended by users and applications for higher reliability, especially in archival storage systems where data are critical and should be preserved over long time periods. Cost increases to the storage of content as well as for the keys storage. Increase bandwidth with upload time.

IV.PROPOSED SYSTEM

An attribute-based storage system which employs ciphertext-policy attribute-based encryption (CP-ABE) and supports secure deduplication. To enable the deduplication and distributed storage of the data across HDFS.And then using two way cloud in our storage system is built under a hybrid cloud architecture, where a private cloud manipulates the computation and a public cloud manages the storage. The private cloud is provided with a trapdoor key associated with the corresponding ciphertext, with which it can transfer the ciphertext over one access policy into ciphertexts of the same plaintext under any other access policies without being aware of the underlying plaintext. After receiving a storage request, the private cloud first checks the validity of the uploaded item through the attached proof. If the proof is valid, the private cloud runs a tag matching algorithm to see whether the same data underlying the ciphertext has been stored. If so, whenever it is necessary, it regenerates the ciphertext into a ciphertext of the same plaintext over an access policy which is the union set of both access policies. like public cloud and private cloud. We have shown the concept of deduplication effectively and security is achieved by means of Proof of Ownership of the file. That is attribute-based storage system ciphertext-policy attribute-based encryption (CP-ABE) and supports secure deduplication

Algorithms used:

- MD5 algorithm.
- Base 64.
- RSA algorithm.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

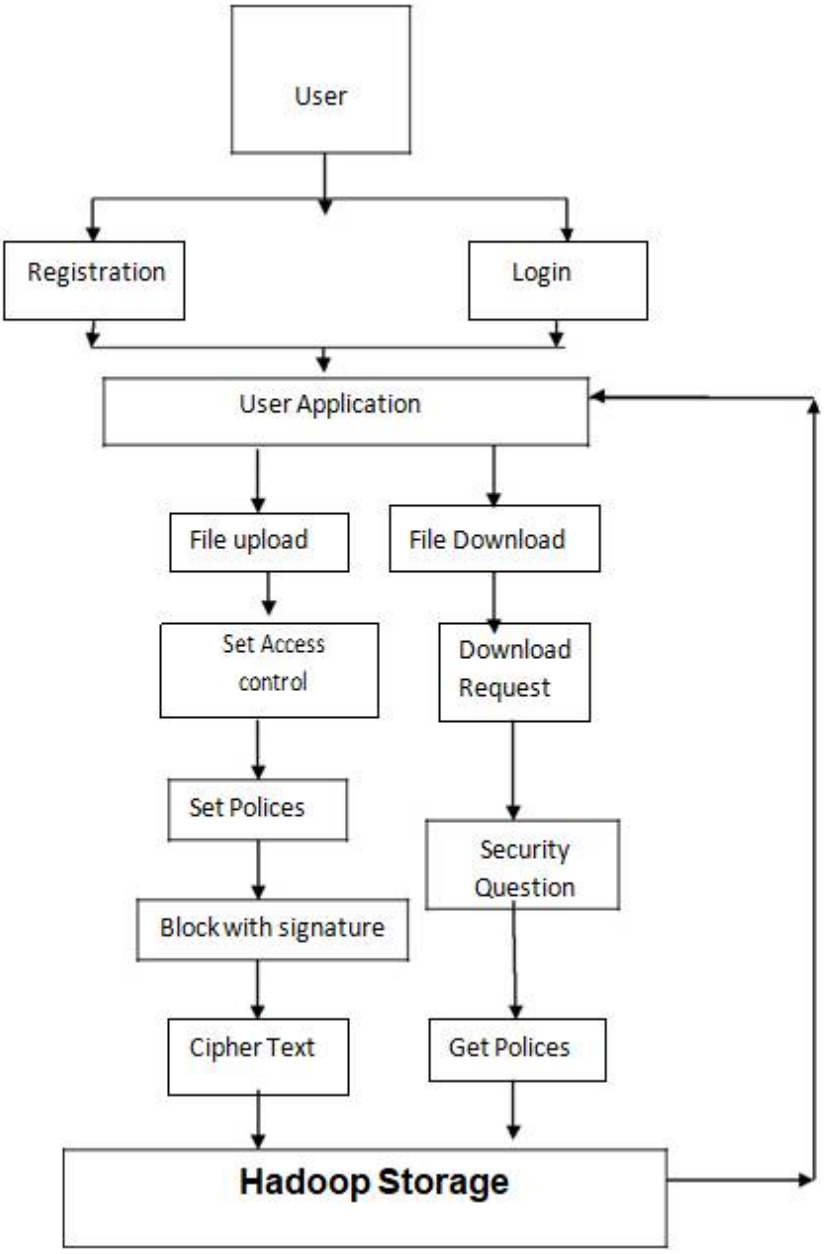
International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

SYSTEM DESIGN



MODULES

Cloud user Registration:
In this module cloud user first register the user details (Name, password, email,access policy,mobile no, dob). And then login the user credential details like username, password. Once user name and password is valid open the user profile screen will be display.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

File upload with access policies:
In this module User will chooses the file and uploads to Storage where the HDFS storage system .In the system will generate a signature in particular file and then split into multiple block. Each block will be generate signature with key . In the signature by using MD5 message-digest algorithm is cryptographic hash function producing a 128-bit hash value typically expressed in text format as 32 digit hex value so that files of same are de-duplicated. After that generate convergent keys for each blocks splitting to store CSV file .like filename, file path, blocks, username, password and block keys .Encrypt the blocks by RSA algorithm is asymmetric cryptography algorithm. Asymmetric actually means that it works on two different keys i.e. Public Key and Private Key. As the name describes that the Public Key is given to everyone and Private key is kept private. Here the plain text is encryption to cipertext and stored in slave system. Blocks are stored in Distributed HDFS Storage Providers. After upload the file to set the access policies with set security question.

Detection Deduplication method:
File-level data deduplication compares a file to be backed up or archived with copies that are already stored. This is done by checking its attributes against an index. If the file is unique, it is stored and the index is updated; if not, only a pointer to the existing file is stored. The result is that only one instance of the file is saved, and subsequent copies are replaced with a reference that points to the original file. Another one signature match checking looks within a file and saves unique iterations of each block. All the blocks are broken into chunks with the same fixed length. Each chunk of data is processed using a hash algorithm such as MD5 or SHA-1.

Download User File:
The final model user request for downloading their own document which they have uploaded in HDFS storage. In this download request will analysis the user attribute once it will matched then ask the security questions for particular file. After complete the process needs proper ownership verification. After verification the original content is decrypted by requesting the Distributed HDFS storage where HDFS storage request key management slave for keys to decrypt and finally the original content is received by the user.

V.RESULTS

The above methods are executed in cloud and the final results are obtained

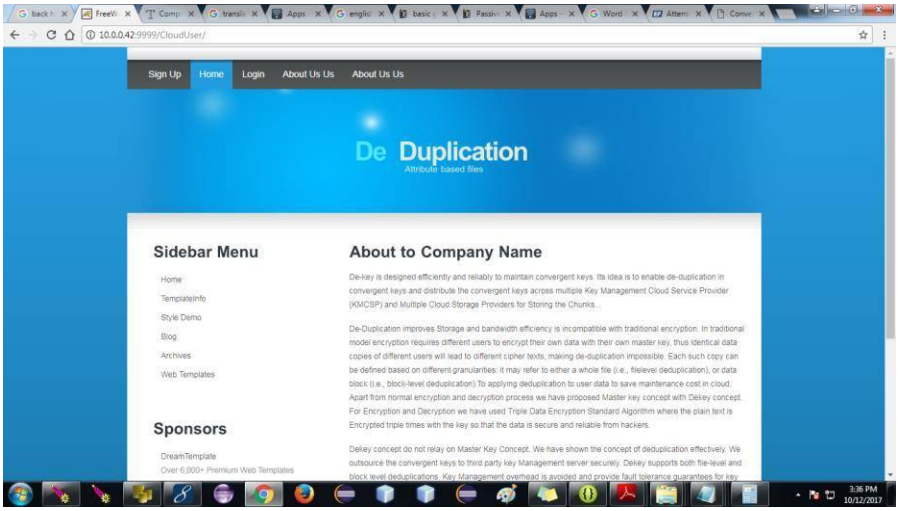


Fig.1.Home page

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

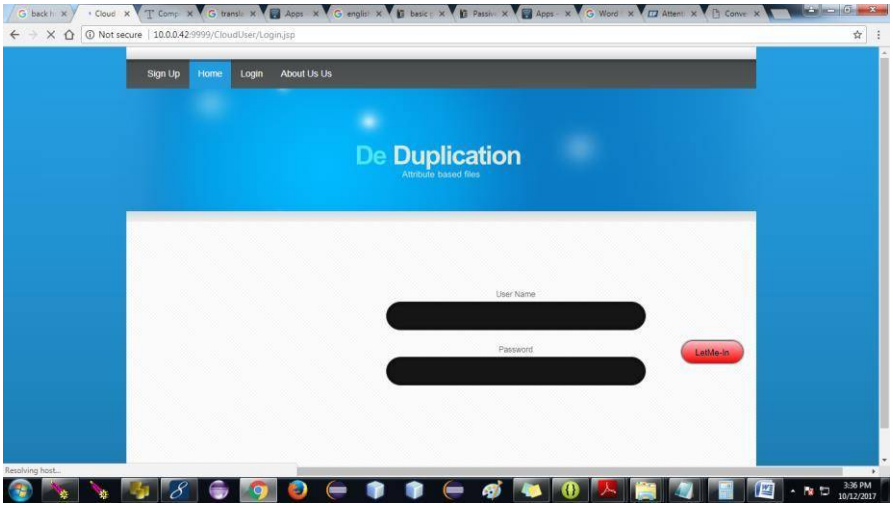


Fig.2.Login page

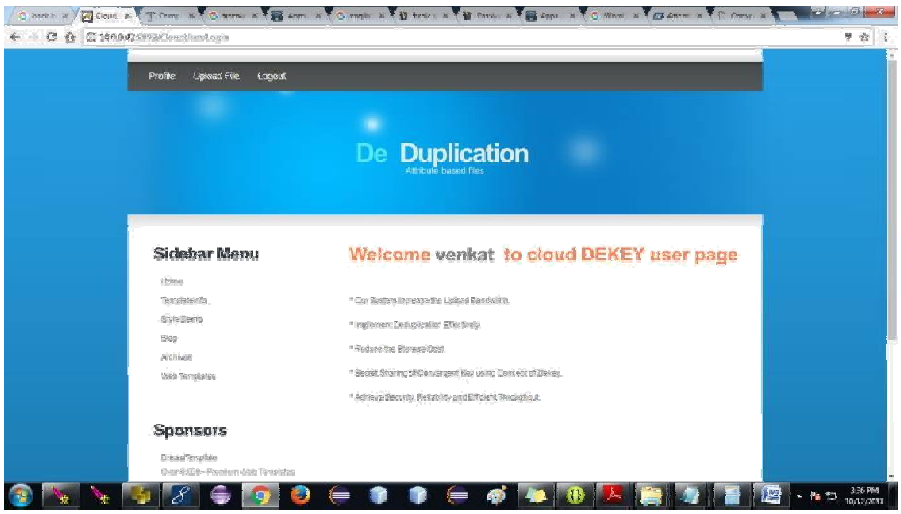


Fig.3.User home page

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

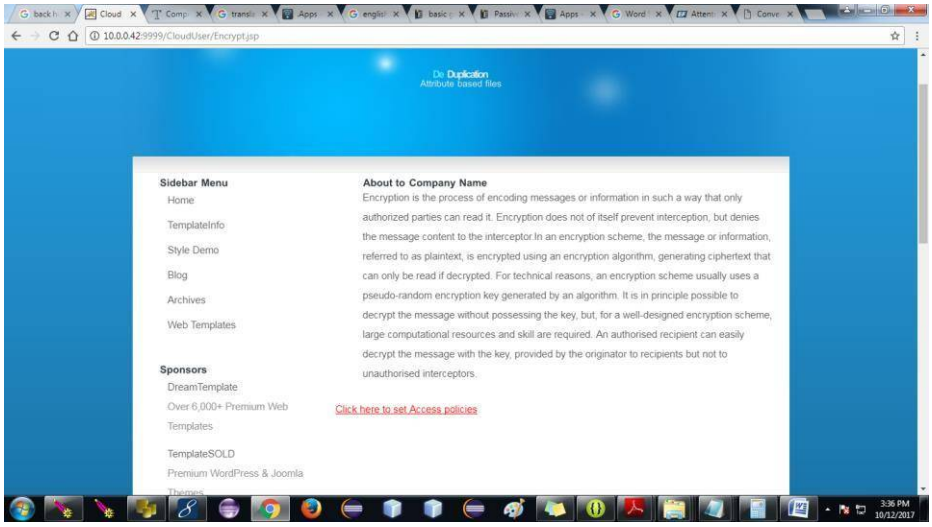


Fig.4.Upload file

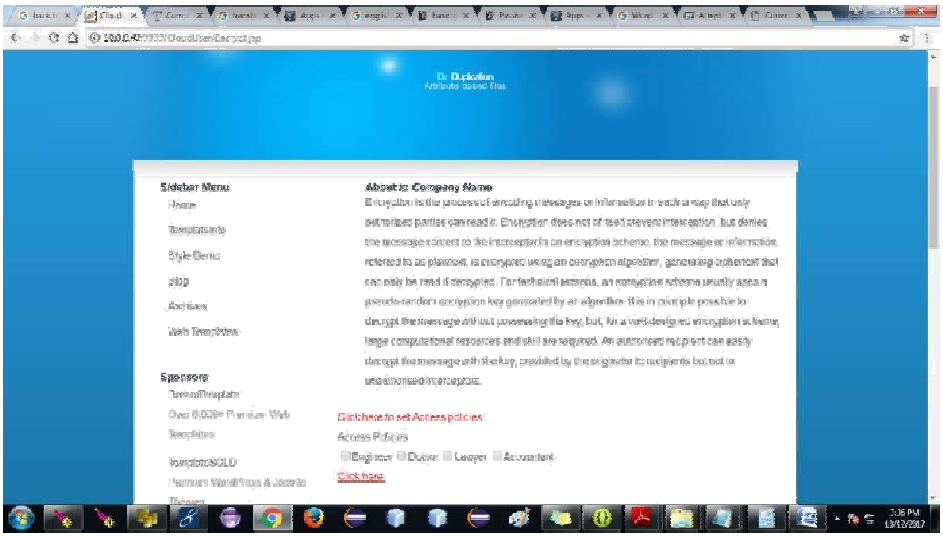


Fig.5.Specifing the access

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

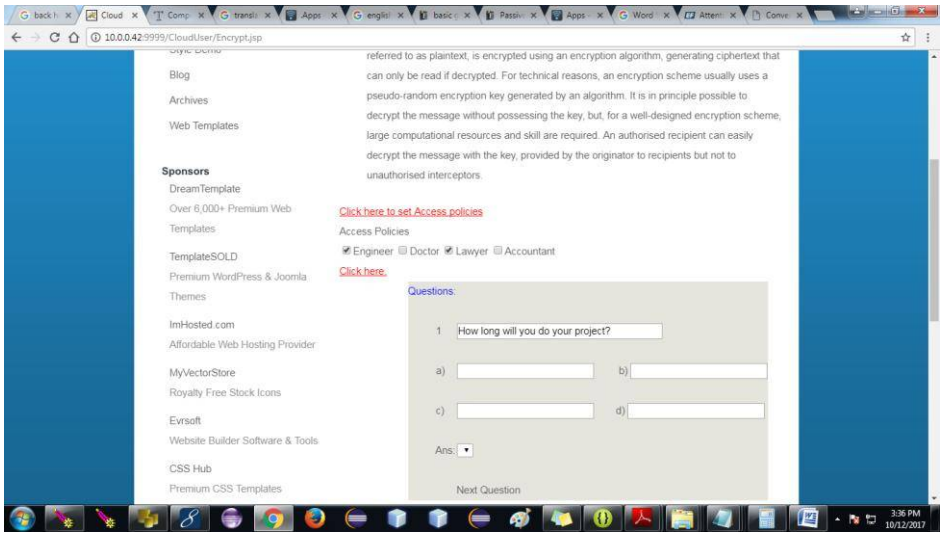


Fig.6.Selecting question for accessing download files

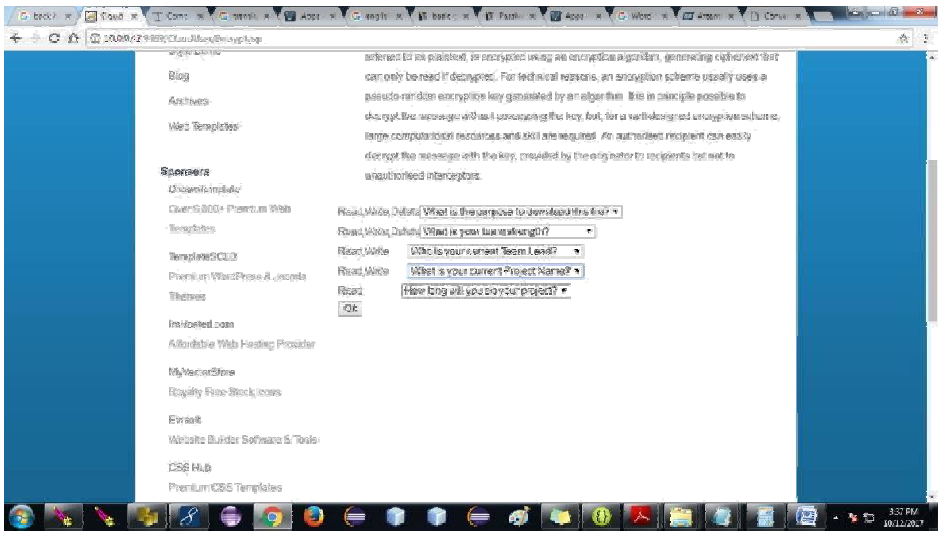


Fig.7.Setting question according to the access policy

VI.CONCLUSION

Attribute-based encryption (ABE) has been widely used in cloud computing where data providers outsource their encrypted data to the cloud and can share the data with users possessing specified credentials. The purpose of the new distributed de-duplication systems with file-level and fine-grained block-level data deduplication, higher reliability in which the data chunks are distributed across HDFS storage, reliable key management in secure de-duplication and the security of tag consistency and integrity were achieved

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

REFERENCES

[1] Amazon, "Case Studies," https://aws.amazon.com/solutions/casestudies/# backup.
[2] J. Gantz and D. Reinsel, "The digital universe in 2020: Big data, bigger digi tal shadows, and biggest growth in the far east," http://www.emc.com/collateral/analyst-reports/idcthe- digital-universe-in-2020.pdf, Dec 2012.
[3] M. O. Rabin, "Fingerprinting by random polynomials," Center for Research in Computing Technology, Harvard University, Tech. Rep. Tech. Report TR-CSE-03-01, 1981.
[4] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer, "Reclaiming space from duplicate files in a serverless distributed file system." in ICDCS, 2002, pp. 617–624.
[5] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in USENIX Security Symposium, 2013.
[6] ——, "Message-locked encryption and secure deduplication," in EUROCRYPT, 2013, pp. 296–312.
[7] G. R. Blakley and C. Meadows, "Security of ramp schemes," in Advances in Cryptology: Proceedings of CRYPTO '84, ser. Lecture Notes in Computer Science, G. R. Blakley and D. Chaum, Eds. Springer-Verlag Berlin/Heidelberg, 1985, vol. 196, pp. 242–268.
[8] A. D. Santis and B. Masucci, "Multiple ramp schemes," IEEE Transactions on Information Theory, vol. 45, no. 5, pp. 1720–1728, Jul. 1999.
[9] M. O. Rabin, "Efficient dispersal of information for security, load balancing, and fault tolerance," Journal of the ACM, vol. 36, no. 2, pp. 335–348, Apr. 1989.
[10] A. Shamir, "How to share a secret," Commun. ACM, vol. 22, no. 11, pp. 612–613, 1979.