Data Deduplication And Data Sharing With Delegated Access Control

Ms. Triveni Bhalerao[1] and **Prof. N. P. Kulkarni**[2]

[1]*Department of Information Technology SKNCOE, Pune*
[1]*Savitribai Phule Pune University, India.*

[2]*Department of Information Technology SKNCOE, Pune.*
[2]*Savitribai Phule Pune University, India.*

***Abstract*—Nowadays Data sharing, maintenance, its security are major challenges in global world. User in the data sharing system upload their file with the encryption using private key. This property is especially important to any large scale data sharing system, as any user leak the key information then it will become difficult for the data owner to maintain security of the information. In this system provide a concrete and efficient instantiate of scheme, prove its security and provide an implementation to show its practicality. There are lots of challenges for data owner to share their data on servers or cloud. There are different solutions to solve these problems. These techniques are very much critical to handle key shared by the data owner. This system will introduce how to reduce burden of data owner, authenticate those who have the access to the data on cloud. SHA algorithm is used by the TTP to generate the key and that key will get share to user as well as the owner. The TTP module receives encrypted file F using AES Algorithm from the data owner and computes hash value using MD-5 algorithm. It stores key in its database which will be used during the dynamic operations and to determine the cheating party in the system (CSP or Owner). Trusted Third Party sends file to Cloud Service Provider to store data on cloud.**

***Index Terms*—AES, Data Security, Deduplication, Cloud, Encryption, Key, Recovery**

I. Introduction

Cloud computing is nothing but Internet computing wherein cloud data is accessible from anywhere over the Internet. Due to this Security is one of the major concern as weak username/password security, hacking which may cause big loss if data is critical or sensitive. Hence high level security should be one of the promise. One very common approach to secure data is Cryptography. Though AES, we ensure no one can access data by observing and monitoring users' request and getting responses.

Cloud computing is presumably the most cost effective technique to utilize, keep up and overhaul. Conventional desktop programming costs organizations a great deal regarding account. Including the authorizing charges for numerous clients can end up being extremely costly for the foundation concerned. The cloud, then again, is accessible at considerably less expensive rates and thus, can fundamentally bring down the organization's IT costs. Plus, there are numerous one-time-payment, pay-as-you-go and other versatile alternatives accessible, which makes it extremely reasonable for the organization being referred to.

Since every one of your information is put away in the cloud, backing it up and reestablishing the same is generally substantially easier than putting away the same on a physical gadget. Moreover, most cloud specialist organizations are typically sufficiently skilled to deal with recuperation of data.

Consequently, this makes the whole procedure of reinforcement and recuperation substantially easier than other conventional techniques for information stockpiling.

The other significant issue while in the cloud is that of security issues. Before embracing this technology, you ought to realize that you will surrender all your organization's touchy data to an outsider cloud specialist co-op. This could conceivably put your organization to incredible hazard. Subsequently, you have to ensure that you pick the most dependable specialist organization, who will keep your data thoroughly secure. Putting away data in the cloud could make your organization powerless against outside hack assaults and dangers. As you are very much aware, nothing on the Internet is totally secure and consequently, there is dependably the sneaking probability of stealth of touchy information. Cloud computing gives upgraded and disentangled IT administration and support abilities through focal organization of assets, merchant oversaw infrastructure and SLA sponsored assentions. IT infrastructure updates and support are dispensed with, as all assets are kept up by the specialist organization. You appreciate a basic online UI for getting to programming, applications and administrations – without the requirement for establishment - and a SLA guarantees the convenient and ensured conveyance, administration and support of your IT administrations. Lastly and above all, cloud computing gives you the upside of speedy arrangement. When you decide on this technique for working, your whole framework can be completely utilitarian in a matter of a couple of minutes. Obviously, the measure of time taken here will rely on upon the correct sort of technology that you requirement for your

business.

ABAC is an intelligent access control show that is discernable on the grounds that it controls access to objects by assessing rules against the characteristics of the substances (subject and question) activities and the earth significant to a demand. Properties might be considered qualities of anything that might be characterized and to which an esteem might be alloted. In its most essential frame, ABAC depends upon the assessment of qualities of the subject, properties of the question, condition conditions, and a formal relationship or access control govern characterizing the permissible operations for subject-protest trait and condition blends.

• Coarse-grained - bigger segments than fine-grained, vast subcomponents. Essentially wraps at least one fine-grained services together into a more coarse grained operation. • Fine-grained - little segments of which the bigger ones are created, bring down level administration SaaS is the most fundamental level of administration with PaaS and SaaS next two above levels of administration. Moving upwards each of the administration acquires abilities and security worries of the model underneath. IaaS gives the foundation, PaaS gives stage advancement condition and SaaS gives working condition.

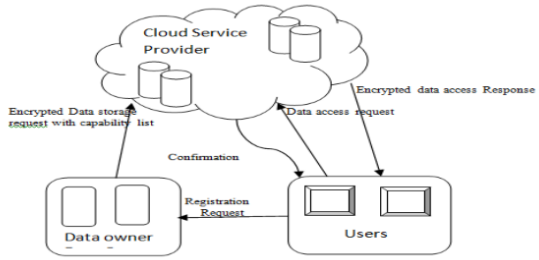

Fig. 1. Existing System Ref [3]

II. Review of Literature

Boyang Wang, Baochun Li, Hui Li, in "Public Auditing for Shared Data with Efficient User Revocation in the Cloud", Make sense of the collusion attack in the leaving plan and give an efficient public integrity auditing plan with secure group client revocation basedon vector responsibility furthermore, verifier-neighborhood revocation group signature [1]. Customer must rehash the costly pre-preparing stage if the noxious server tries to cheat and take in a touch of info.Cheng-Kang Chu, Sherman S.M. Chow, Wen-GueyTzeng, Jianying Zhou, what's more, Robert H Deng, in "Key-Aggregate Cryptosystem for Scalable Data Sharing in Cloud Storage, where Constant-measure ciphertexts and more adaptable than various leveled key task which can just spare spaces if every single key-holder share a comparative arrangement of benefits [2].Allows efficient and adaptable key delegation.In cloud storage, the quantity of cipher texts generally develops quickly significant disadvantage is predefined bound of the quantity of most extreme ciphertext classes.Seung-Hyun Search engine optimization, Mohamed Nabeel, Xiaoyu Ding, Elisa Bertino, in "An Efficient Certificateless Encryption for Secure Data Sharing in Public Clouds", takes care of the key escrow issue in character based encryption and declaration revocation issue in public key cryptography and enhance the effectiveness of encryption at the data owner [3]. Here, downside is at whatever point the group dynamic changes, the rekeying operation requires to redesign the private keys given to existing individuals keeping in mind the end goal to give in reverse/forward secrecy.Mohamed Nabeel and Elisa Bertino, in "Privacy Preserving Delegated AccessControl in Public Clouds" expressed it depends on find grained encryption of data and designating the greater part of the access control authorization to the cloud. An efficient group key administration plot that underpins expressive ACPs.KaipingXue and Peilin Hong, in "A Dynamic Secure Group Sharing Framework in Public Cloud Computing", where Dynamic secure group sharing structure in public cloud computing environment and the sharing documents are secured put away in cloud servers and all the session key are secured in the advanced Envelopes however Node Assignment id troublesome and complex capacity. Tao Jiang, Xiaofeng Chen, and Jianfeng Ma, in "Public Integrity Auditing for Shared Dynamic Cloud Data with Group User Revocation", this Figure out the collusion attack in the leaving plan and give an efficient public integrity auditing plan with secure group client revocation basedon vector responsibility and verifier-neighborhood revocation group signature, however Client must rehash the costly pre-handling stage if the malignant server tries to cheat and take in a touch of information.. Jiawei Yuan and Shucheng Yu, in "Public Integrity Auditing for Dynamic Data Sharing With Multiuser Modification", Allowing multiple cloud clients to modify data with integrity affirmation, public auditing, high blunder recognition likelihood, efficient client revocation and in addition viable computational /correspondence auditing performance.Wei Zhang, Understudy Yaping Lin, Sheng Xiao, Jie Wu, , and Siwang Zhou, in "Privacy Preserving Ranked Multi-Keyword Search for Multiple Data Owners in Cloud Computing", here efficient data client validation protocol, which not just forestalls attackers from listening stealthily mystery keys and putting on a show to be unlawful data clients performing searches, additionally empowers data client validation and revocation. Systematically develop a novel secure search protocol, which not just empowers the cloud server to perform secure positioned keyword search without knowing the genuine data of both keywords and trapdoors, additionally permits data owners to encode keywords with self-picked keys furthermore, permits confirmed data clients to question without knowing these keys. This may mischief and cause velnerability. Xinyi Huang, Joseph K. Liu, Shaohua Tang, Yang Xiang, Kaitai Liang, Li Xu, and Jianying Zhou, in "Cost-Effective Authentic also, Anonymous Data Sharing with Forward Security", where Key exposureNetwork Connections exists and cost is more.

III. Proposed System

Data owner is responsible for preprocessing in setup phase. This fine grained access control increases overhead on Data

owner. In order to reduce burden of data owner knows as Coarse grained access control, data owner could outsource expensive operations to some cloud computing server. Furthermore, considering cloud computing server and data owner as a one system. Hence delegated access control to the cloud computing server reduces overhead of data owner. In real applications, Owner-delegated auditor are another server provides free or paid auditing services to many cloud users. In this attack system expect the attacker has entry to the abnormality score appointed to a picked payload. Besides, it is sensible to expect that some ordinary payloads are known as well. (Consider, for instance, the instance of an IDS examining HTTP asks for sent to a publicly available web server, where countless payloads will be known by the attacker.) Let p be one such ordinary payload. A clear system to distinguish what components of p have a place with the key D comprises of nourishing KIDS with the main byte of p, at that point with the initial two bytes of p, et cetera. At the point when the nextto the last byte happens to be a delimiter, KIDS will identify a move where the left word is probably going to have been seen amid preparing, while the correct word is frequently obscure (since it is truncated). Now, the abnormality score will endure a slight decrement. By advantageously rehashing the method, all the delimiters exhibit in p can be recovered. Notwithstanding the specialized points of interest, the fundamental disadvantage of the naıve technique talked about above is that the attacker will just have the capacity to recover those key components exhibit in the typical payloads accessible, which may well be only a portion of every one of them. In addition, the unpredictability of such an attack is straight in the quantity of payloads and their lengths. System next portray an alternate approach that gets all the key components all the more productively and without specifically depending on typical payloads.Few security issues must be addressed in cloud computing : Data confidentiality as Data confidentiality is lost when the user data has been accessed by other user.Data integrity where retrieved data should be same as the data stored on cloud and no other user should be able to modify or access that data.Data authentication is only authorized user can access data. Data loss occurs when hacker attacking the data to the closest server as data is the most valuable assets of any user. Data location may be an issue if the user does not want his data to be stored somewhere in the world as the uploaded data can be stored anywhere. Cloud service provider (CSP): CSP is an entity responsible for storage of the data. CSP and TTP which helps in encryption of data that are going to be stored on cloud storage. Thrusted third party is responsible for encrypting data and store it on cloud and decrypt when register user access that data. For Data confidentiality – Symmetric and asymmetric key are used for cloud storage.Algorithms like AES, DES, RSA, 3DES, RC4, blowfish are used for encryption and decryption purpose. ASE is a simple symmetric technique. In RSA algorithm is faster algorithm which uses asymmetric key technique. For Data Integrity – To identity data uniquely, numeric value of fixed length which identifies it know as Hash value. To identify which data is modified, comparison between sent data and received data is done. For Authentication – Two factor authentication is the technique where first factor is password which is simple way of authentication along with which second factor what user has is used such as OTP (One Time Password).

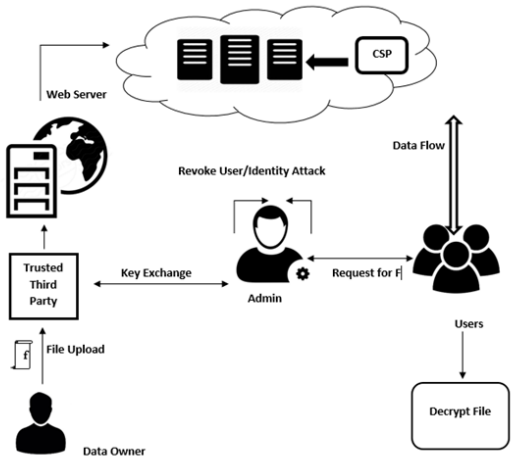

Fig. 2. System Architecture Ref (IPGCON PUNE 2017)

In a few regards, this data is less finegrained than the abnormality score, so it is sensible to anticipate that attacks working under this supposition will be marginally more intricate. The focal thought behind our attack is entirely straightforward. System will give KIDS a normal payload connected with a precisely developed tail. Such a tail contains an extensive number of inconspicuous words isolated by the applicant delimiter. On the off chance that the delimiter does not have a place to the key, the whole tail will be handled as only one word and the irregularity score will be generally like that of the first payload. If so, then the payload will be set apart as normal with high likelihood. On the other hand, if the delimiter belongs to the key, the tail will be divided into countless concealed words and moves. This will contrarily affect the peculiarity score, constantly bringing about an odd payload.

IV. Algorithm

A. AES

Input: Plain Text
Step1:
Byte state[4,Nb]
State = in
AddRoundKey(state, w[0, Nb-1])
Step2:
for round=1 to Nr-1
SubBytes(state)
ShiftRows(state)
MixColumns(state)
AddRoundKey(state, w[round*Nb, round+1)*Nb-1])
end for
Step3:
SubBytes(state)

ShiftRows(state)
AddRoundKey(state, w[Nr*Nb, (Nr+1)*Nb-1)
Output:Cipher Text

B. Key-Recovery on Black-Box KIDS

In this payload will be normal with properly structured tail.The tail contain the large number of unseen words separated with delimeter.In Black Box recovery algorithm attacker tries to recover the w1(word 1) and w2(word 2).For this the attacker tries different combination till the length of payload.If w1 recover then the w2 can be easily recoverd.
Algorithm
Input:
Set of payload Q=qi
word w2 s.t. n(w2)=0
Parameter l>1
for each qi belongs Q do
Di $<--$a
for d=0 to 255 do
p$<---$(qi||d||w2||d||...||d|| w2)
if anom(p)=true then
Di$<--$ Di U d
end if
end for
end for
return D=Di

C. Key-Recovery on Gray-Box KIDS

In this attack assume the attacker has access to the anomaly score assigned to a chosen payload. Furthermore, it is reasonable to assume that some normal payloads are known too.
Algorithm
Input:
w1,w2 such that n(w1)>0,n(w2)=0
D1$<--$a
D2$<--$a
for d=0 to 255 do
p$<---$(w1 || d || w2)
if s(p)=S(w1 || d || w 2) then
D1$<--$ D1 U d
else D2$<--$D2 U d
end if
end for
return D2
return D1

V. Mathematical Model

Input Set S=I,E,D
I= {I0(u0,p0,f0),I1(u1,p1,f1),.....,I0(un,pn,fn)}
Where,u= Username
p=Password
f= File to be stored
Intermediate Output Set
E= E1, E2,E3
Where,E1=Authorized User au if $(au \in (u \cup p))$
E2=File f Successfully uploaded and key k is generated $(E2 \in E1)$
E3=File f shared $(E3 = if(E1 \in u) and (E2 \in E1))$
E=$((I \cap E1) \cup (f \cap E2) \cup E3)$
Final Output Set
D= D1, D2
Where,
D1=Block unauthorized user or for authorise user allow the access to download it.
D1=$(u \notin (E1 \cap u))$
D2=Generation of New Key
D2=new$(k \in u \in f)$

VI. Experimental Setup

The system has actualized as independent system utilizing jdk 1.7. The system utilizes MySql 5.5 as the database engine and Amazon EC2 Cloud. All investigations are run utilizing machine with Intel(R) Core (1M) i3-4005U CPU @ 1.70GHz , and running on Windows 7. Representative data-set is utilized for this analysis.

VII. Efficiency

The run-time cost of data sharing is depend upon the size of the file which data owner upload to the cloud. It also depend on how accurately file get shared to authorised user.System also detect the key recovery attack by unauthorised users.Efficiency of proposed system is calculated by
Efficiency=Execution Time/Cycle Time

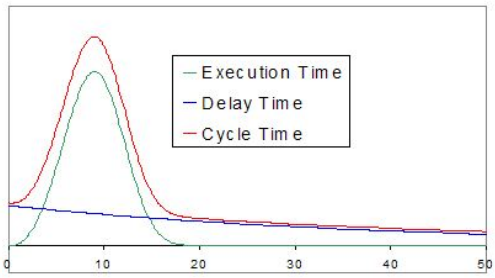

Fig. 3. Efficiency

VIII. Result And Discussion

TABLE I
Result Table

Particular	Precision	Recall	Fitness
File1	0.70	0.30	0.32
File2	0.64	0.36	0.35
File3	0.85	0.15	0.46

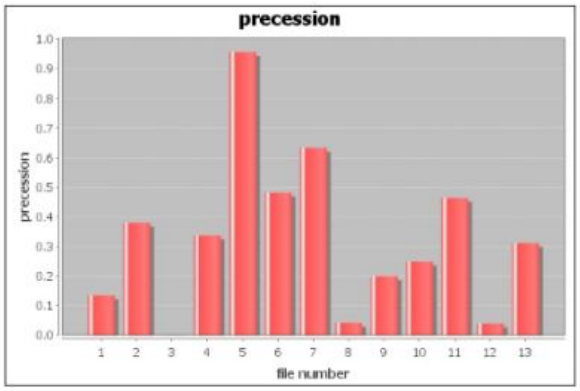

Fig. 4. Precision

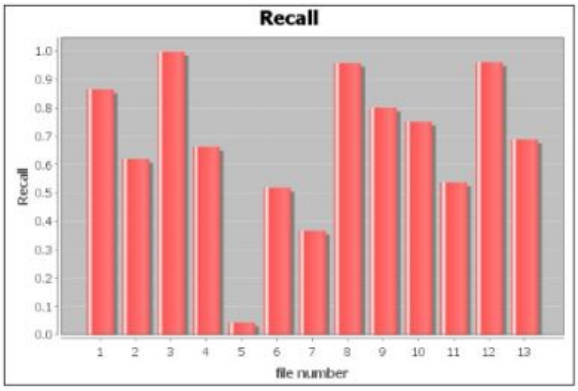

Fig. 5. Recall

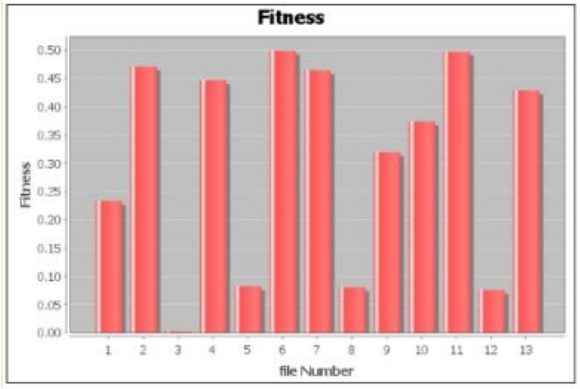

Fig. 6. F Measure

divided by total number of files uploaded by the data owners. precision=no/tcount;
Here no=Number of times file is downloaded correctly by user
tcount=Total number of file uploaded by the data owner

Recall is calculated as total number of files which are not retrieved or incorrectly accessed by the user divided by the total number of files uploaded by the data owner.

Recall=(tcount-no)/tcount

Fmeasure=2*precision* recall

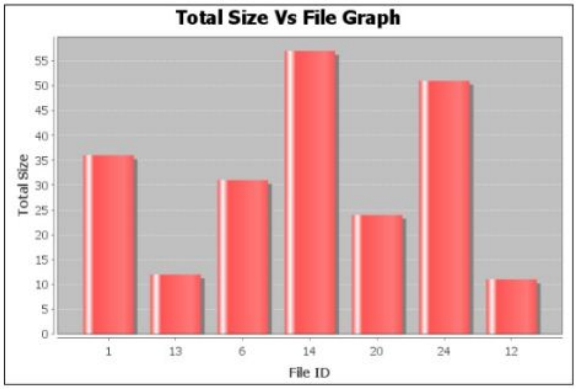

Fig. 7. File Size Without Deduplication

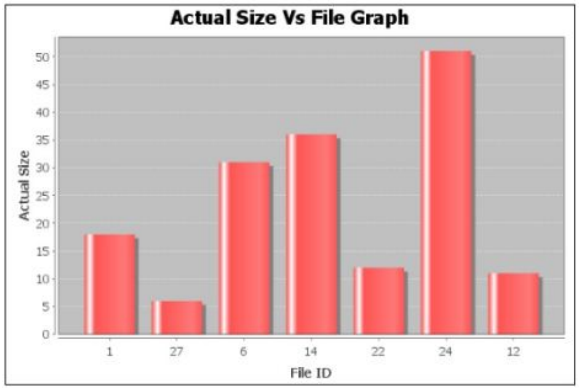

Fig. 8. File Size With Deduplication

This system can access the data files of any type and maximum size allowed is 5 MB. In this system user can upload the data that data may be text file,word file,image etc. This data is encrypted and stored on cloud.User can share the uploaded files to another registered user.The analysis is generated based on total number of files shared by the data owners and their access by users.To measure the accuracy of files downloaded by the user system have taken three analysis parameters.Precision,Recall,Fmeasure.Precision is calculated as total number of (accurate)times file is access

Data deduplication is implemented in two ways, one which is very common and simple by checking filename and other one is to check the contents of file before uploading to the cloud. While uploading any file to cloud hash key is generated and stored so that this is key is used for further verification for duplicate files. MD-5 is used to get hash key of file. Optimizing redundancies with preserving data integrity is crucial task. As cloud service provider charges cost as per the space utilization, hence optimizing space which saves or reduce cost associated with Duplicated data.

IX. Conclusion

Considering useful needs in data sharing, another thought called forward secure ID-based ring signature alongside assigned get to control is postured. It permits an ID-based ring signature plan to have forward security. It is the principal writing to have fine grained encryption on cloud and coarse grained at Data owner alongside the element for ring signature in ID-based setting. This plan gives genuine obscurity and can be demonstrated forward secure indefensible in the arbitrary ,key update,signing and varifying algorithm to share media content in a controllable way. Staying away from duplication to fight cost for cloud storage. It permits an ID-based ring signature plan to have forward security. This plan will be exceptionally helpful in numerous other functional applications, particularly to those require client privacy and verification, for example, specially appointed network, ecommerce exercises and savvy lattice. Current plan depends on the arbitrary presumption to demonstrate its security. SHA-1 and MD5 algorithm is utilized for data encryption. At that point improve security on data sharing and transfer the data on cloud. This system likewise give data depulication which diminishes the cloud storage cost

Acknowledgment

I dedicate all my works to my esteemed guide, Prof. N. P. Kulkarni , whose interest and guidance helped me to complete the work successfully. This experience will always steer me to do my work perfectly and professionally. I also extend my gratitude to (H.O.D.Information Technology) who has provided facilities to explore the subject with more enthusiasm. I express my immense pleasure and thankfulness to all the teachers and staff of the Department of Computer Engineering, for their co-operation and support. Last but not the least, I thank all others, and especially my friends who in one way or another helped me in the successful completion of this system.

References

[1] Boyang Wang, Baochun Li and Hui Li, “Panda: Public Auditing for Shared Datawith Efficient User Revocation in the Cloud”, in *IEEE TRANSACTIONS ON SERVICES COMPUTING*, VOL. 8, NO. 1, pp. 92-106, 2014.

[2] Cheng-Kang Chu, Sherman S.M. Chow, Wen-GueyTzeng, Jianying Zhou, and Robert H. Deng, “Key-Aggregate Cryptosystem for ScalableData Sharing in Cloud Storage", in *IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS*, VOL. 25, NO. 2, pp. 468-477, 2014.

[3] Seung-Hyun Seo,Mohamed Nabeel, Xiaoyu Ding, and Elisa Bertino, “An Efficient Certificateless Encryption forSecure Data Sharing in Public Clouds”, in *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, VOL. 26, NO. 9, pp. 2107-2119, 2014.

[4] Mohamed Nabeel and Elisa Bertino, “Privacy Preserving Delegated AccessControl in Public Clouds”, in *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING"*, VOL. 26, NO. 9, pp.2268-2280, 2014.

[5] Kaitai Liang, Man Ho Au, Joseph K. Liu, Willy Susilo, Duncan S. Wong, Guomin Yang, Tran Viet Xuan Phuong, and Qi Xie, “A DFA-Based Functional Proxy Re-EncryptionScheme for Secure Public Cloud Data Sharing”, in *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY*, VOL. 9, NO. 10, pp. 1667-1680, 2014.

[6] Jiawei Yuan and Shucheng Yu, “Public Integrity Auditing for Dynamic Data Sharing With Multiuser Modification", in *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY*, VOL. 10, NO. 8, pp. 1717-1726, 2014.

[7] Tao Jiang and Jiafeng Ma, “‘Public Integrity Auditing for Dynamic Data Sharing With Multiuser Group Modification", in *IEEE TRANSACTIONS ON TECHNOLOGY*, VOL. 6, NO.7, pp.456-492, 2014.

[8] Jiawei Yuan and Shucheng Yu, Member, IEEE proposed a system on “Public Integrity Auditing for Dynamic Data Sharing With Multiuser Modification", in *IEEE TRANSACTIONS ON SERVICES COMPUTING*, VOL. 4, NO. 2, pp. 1168-1207, 2014.

[9] C. Wang, S. S. M. Chow, Q. Wang, K. Ren, and W. Lou, “Privacy preserving public auditing for secure cloud storage”, in *IEEE TRANS. COMPUT.*, vol. 62, no. 2, pp. 362–375, Feb. 2013.

[10] Xinyi Huang, Joseph K. Liu, Shaohua Tang, Yang Xiang, Kaitai Liang, Li Xu, and Jianying Zhou, “Cost-Effective Authentic and Anonymous Data Sharing with Forward Security", in *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY*, VOL. 9, NO. 10, pp. 1667-1680, 2014.