A Review on Auditing in Cloud

MS Tanuja Sali

M.E. (Computer Networks)
Department of Computer Engineering
G. H. Raisoni College of Engineering & Management, Wagholi
Pune, Maharashtra State, India

Abstract: ***Cloud Computing is the usage of puddle of assets for distant users through internet that is easily available and expanded. The users accessing storage services of cloud has no direct control over their data; hence data security has become one of the major concerns in cloud. Current research work has already allowed the verification of data integrity without changing of the actual data file. This process was done by a trusted third party data auditing by an auditor. However, this scheme has from many common drawbacks. First, there is no proper authorization/authentication process between the auditor and cloud service provider and it may open the data contents to the auditor because it requires the server to send the local data blocks to the auditor. Consistency of the audits can be ensured by doing multiple auditing tasks simultaneously. To have real benefit, an audit program must be designed in such a way that it can be sustainable, both in terms of management commitment and to the ongoing development of talented and capable auditors. A management process addresses the content and scope of the audits, standards and regulations that apply to the facilities, use of audit protocols, the frequency of audits, and training of auditors is integral to an effective and sustainable auditing program. To securely introduce a reliable TPA the following two fundamental conditions have to be fulfilled 1) TPA must be able to audit the cloud data storage without asking the actual copy of data, and hence induce no extra load to the cloud user; 2) Also the third party auditing process should bring in no new problems towards user data privacy. In this paper, I explain importance of audit and its features. Different types of audit practices provide the opportunity to emphasize different aspects of the management program.***

Keywords: Cloud auditing, Public Auditing, cloud data security, provable data possession, authorized auditing

1. Introduction

Cloud computing is typically defined as a part of computing that relies on sharing resources instead of having separate local servers or individual device. Cloud provides all the necessary resources and services required for an application .This services are differentiated as Infra-structure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS) [1], [2], [3].

Many IT companies offer these services to users (from individual to big firms) all over the world; some of the examples are Amazon AWS, Microsoft Azure.

This development, progress and proliferation of cloud computing is rapid, but then also data security/privacy is one of the major concerns in the cloud computing as the users will lose their direct control over their data. Here we can find the method called data auditing where the verification is carried out by a trusted third person. It is also called ‘auditing-as-a-service’. There are several ways to explain an audit. Webster’s defines an audit as a methodical examination and review. It also defines the act of auditing as an examination with the intention to verify.

In an earlier scheme, the cloud storage server (CSS) was unable to give a valid integrity proof for a given data block to the verifier unless all the data processing is completed. The previous prevailing data auditing schemes already had various properties, also some potential risks and inefficiency such as security risks in unauthorized auditing requests and inefficiency in processing small updates still exist.

A cloud data storage service has three part: the user of cloud (U), who handles huge amount of data files to be stored in the cloud; the cloud server (CS), managed by the cloud service provider (CSP) who gives storage space and all the required resources the third party auditor, who has more expertise and capabilities than cloud users do not have. Users rely on the CS for storing and maintaining the data in cloud. Also the users can communicate with the CS for accessing and updating the data stored for various applications. To reduce the resources needed and the online workload, users ask help to TPA for ensure the safe integrity of their outsourced data, also keeping data private from TPA

In this paper I have done a survey of papers on different types of auditing methods for dynamic data updates. As we have seen auditing protocol design should achieve the security and good performance guarantees by including features as

1) Public auditability: To permit TPA to check the validity of the cloud data without passing a copy of the whole data or increasing surplus burden to the users.

2) Storage correctness: Here it is ensures that there is no fraud server that will forward the TPA’s audit without keeping users’ data intact.

3) Privacy-preserving: It ensures that the TPA cannot get users data content from the all the data gathered during the process of auditing.

4) Batch auditing: It enables TPA to do perform multiple auditing from many different users simultaneously also conducting it securely.

5) Lightweight: Here TPA is allowed to do auditing with least communication and computation burden

2. Why TPA is needed?

A third party auditor can;

- Give auditing results which are unbiased
- Good for both data owners and service providers
  Data owners – here owners are ensured about data integrity
  Service providers – it maintains good reputation
- Able to do a good job efficiently
- Professional expertise
- Computing capabilities

3rd Party Auditing System Architecture

Figure shows data owners (owner), the cloud server (server), and the third-party auditor (auditor). The owner generates the data and save their data in the cloud. The cloud server places the owners' data and it gives data access to all other users (data consumers). The auditor which is a trusted third-party is expert and capable enough to give auditing service for both the owners and servers. The auditor can be a trusted one handled by the government, which can provide unbiased auditing result.

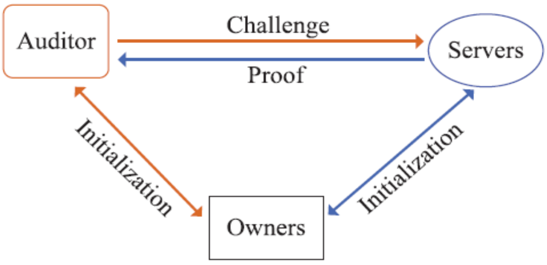


Figure 1: 3rd Party Auditing

Initialization: In these step owner of data sends encrypted data & verified tags to server, and also it forwards index of data to auditor

Challenge: Auditor sends Challenge to cloud server frequently

Proof: Server answers with the Proof

Verification: Auditor will verify accuracy of the Proof

3. Basic concepts

Audit Process

These process includes three elements: 1) the pre-audit steps, which is done in planning time and getting ready for audit 2) the actual audit activities by the team, from collection and start of the audit to final reporting of the results of the audit; and 3) the post audit steps, which involves arranging the audit in an suitable report format and then generating and running a right actions to improve given task.

Public Auditing

Users of cloud cannot physically have direct control over their data. The data integrity verification has to be done without changing the actual data file. When this verification is done by a third party, this verification process is also called public data auditing, and this third party is called an Third Party Auditor (TPA). TPA can be anyone challenging the confidentiality of data stored in Cloud.

PDP Protocol

It is not preserving privacy or cannot support the changing data operations. It is not applicable to cloud storage systems.

Auditor Scheme

This scheme opens the contents of data to the auditor as needs the server to send the linear combinations of data blocks to the auditor. It may incur a heavy storage overhead on the server due to many data tags.

IPDP Scheme

It does not support the batch auditing for multiple owners for creating the tags used by each owner are different. An additional trusted organizer is needed to send a response to the auditor during the multi-cloud batch auditing as it applies the mask technique for data privacy. Additional trusted organiser is not practically seen in cloud storage system. It incurs heavy computation cost of the auditor, which makes the auditor a performance bottleneck.

Source of Data

The major source for getting data is websites, software, fact, figures .Large data regarding company, consumers, producers, retailers, legal documents, data warehouse can be obtained and analysis of data is done. Some of the sources are given below.

A World Wide Web

In web there are huge amount of data is present this data is the big source for researchers and users both in the world of web blogs, comment box, form feeding etc. techniques is used for data extracting and to transfer data emails, face book, twitter is used.

B Sites

In the current era many recognized groups are doing the work of analyzing data and maintaining the sites. Industries are hiring people to performing the respective work. Finally on the bases of obtained data such as price, quantity, ranking the product result occurs.

C Web based Interface

Interface is a medium between user and the web this consist text messaging, digital audio/video, e-mail, links etc. Interface plays an essential role between user and web because without interface it is not possible to connect or communicate.

Cloud Computing has achieved highest precedence currently in market due to central storing of data, efficient security and pool of resources, but the major worry is due to lack of control on precious and sensitive data of cloud user, and also for stored data.

Cloud users store the data on central server and this central server splits big datasets into smaller datasets and stores them in different physical storage servers for reliability, maintaining privacy or for easy transactions.

Some schemes like Provable Data Possession (PDP) and Proofs of Retrievability (POR) supports public data integrity checks by TPA but these schemes cannot assure if TPA is authenticated or not. To implement TPA authorization exchange of signatures is` done between three parties in cloud.

Roles of the Participating Parties

Public data verification is generally seen in PDP and POR schemes. There are three parties involved: client, CSS and TPA. Relations among these three parties are shown in Figure 2.

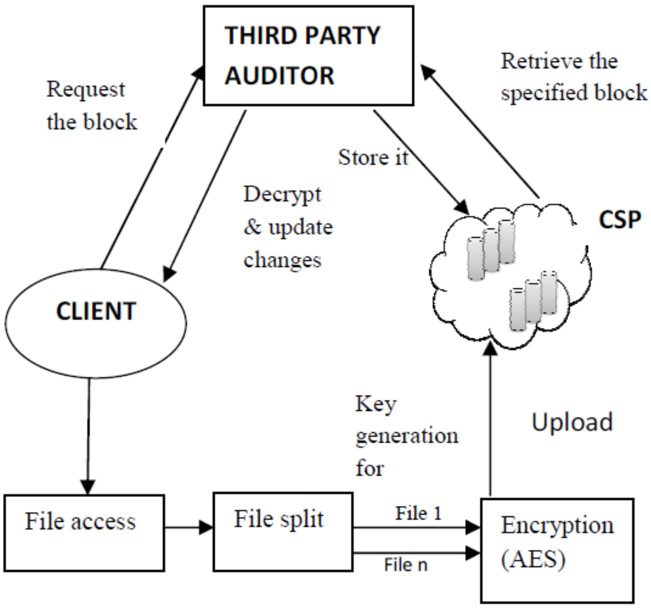
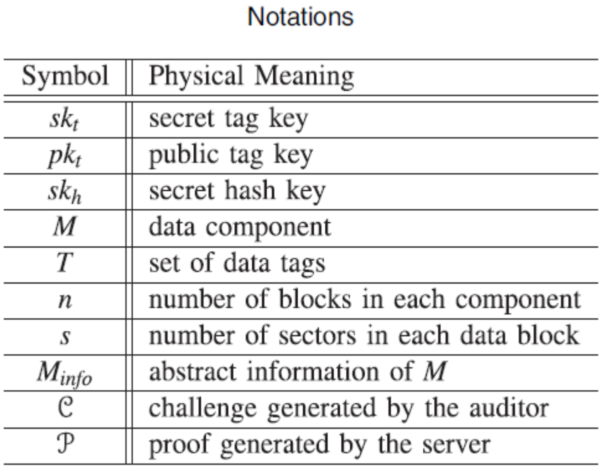


Figure 2: Relation between the participating parties

Both CSS and TPA are only semi-trusted to the client. When CSS is challenged multiple times, an adversary can sometimes get valuable information about user data, or collect the statistical information about current status of cloud service. Though PDP support public verification it cannot satisfy the requirement as a perfect auditing task. Second, investigation is needed to improve the efficiency of verification for frequent small updates. To enhance auditing facilities following steps are taken into consideration.

Auditing Protocol for Storage:

Table 1:


Notations

Symbol	Physical Meaning
sk_t	secret tag key
pk_t	public tag key
sk_h	secret hash key
M	data component
T	set of data tags
n	number of blocks in each component
s	number of sectors in each data block
M_{info}	abstract information of M
$\mathcal{C}$	challenge generated by the auditor
$\mathcal{P}$	proof generated by the server

There are five different algorithms in this auditing protocol.

Let g_1 and g_2 be the symbols of G_1 and G_2, respectively. Let H: {0, 1} $\rightarrow G_1$ be a keyed secure hash function that maps the message information to a point in G1.

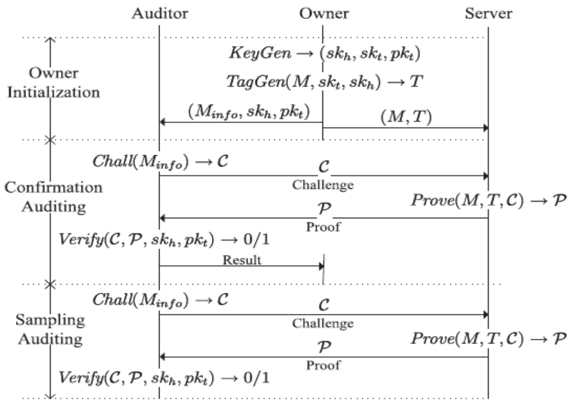
1. KeyGen (p) → (skh; skt; pkt). - This algorithm takes as input only the implicit parameters of security and it outputs a secret hash key skh and a pair of secret-public tag key (skt; pkt).

2. TagGen (M;skt; skh) → T – The tag generating algorithm takes input as 3 parameters that are M, the secret tag key skt, and the secret hash key skh .It calculates a data tag t_i based on skh and skt for each data M. Then it gives outputs a set of data tags/labels. T

3. Chall (Minfo) → C - This algorithm takes as input the required information of the data M_{info} (e.g., file identity, total number of blocks, version number, time stamp, etc.). It gives output as challenge C

4. Prove (M; T; C) → P - The proof algorithm takes three inputs as the file M, the tags T, and the challenge from the auditor C. It outputs a proof P.

5. Verify (C;P; skh; pkt; $_{Minfo}$) → 0/1 - The verification algorithm takes as inputs P from the server, the secret hash key skh, the public tag key pkt, and the abstract information of the data Minfo. It outputs the auditing result as 0 or 1.


Auditor
Owner
Server
Owner Initialization
$KeyGen \rightarrow (sk_h, sk_t, pk_t)$
$TagGen(M, sk_t, sk_h) \rightarrow T$
(M_{info}, sk_h, pk_t)
(M, T)
Confirmation Auditing
$Chall(M_{info}) \rightarrow \mathcal{C}$
$\mathcal{C}$
Challenge
$\mathcal{P}$
Proof
$Prove(M, T, \mathcal{C}) \rightarrow \mathcal{P}$
$Verify(\mathcal{C}, \mathcal{P}, sk_h, pk_t) \rightarrow 0/1$
Result
Sampling Auditing
$Chall(M_{info}) \rightarrow \mathcal{C}$
$\mathcal{C}$
Challenge
$\mathcal{P}$
Proof
$Prove(M, T, \mathcal{C}) \rightarrow \mathcal{P}$
$Verify(\mathcal{C}, \mathcal{P}, sk_h, pk_t) \rightarrow 0/1$

Three steps for auditing protocol which is as follows;

Step 1: Owner Initialization

This is a key generating algorithm. Owner runs KeyGen to create the secret hash key skh, the pair of secret-public tag key (skt; pkt). Then, it executes the tag generation algorithm TagGen to compute the data tag.

Step 2: Confirmation Auditing

This phase involves two-way: Challenge and Proof. During this phase, the owner needs the auditor to recheck whether the data is correctly stored on the server.

The auditor performs the confirmation auditing phase as;

1. Auditor executes the challenge algorithm Chall to generate the challenge C for all the data blocks and sends the challenge C request to the server.
2. When the challenge C from the auditor goes to the server, the server runs the prove algorithm Prove to generate the Proof P → (TP; DP) and sends it back to the auditor.
3. The auditor on receiving the proof P from the server runs the verification algorithm Verify to check the accuracy of P and extract the auditing result.

This auditing result is send by auditor to the owner.

If the result is true, the owner is convinced that its data is correctly stored on the server.

Step 3: Sampling Auditing

This is also a two way communication process between the auditor and the sender. This sampling auditing is periodically done by the auditor by challenging a sample set of data blocks. The frequency of carrying out auditing operation depends on the service contract between the auditor and the owner of data. This sampling auditing also depends on how much faith the data owner has over the server.

To make proposed protocol - A Dynamic Auditing

There are four steps in dynamic auditing protocol which are owner initialization, confirmation auditing, sampling auditing, and dynamic auditing.

The first three phases are almost similar to our privacy preserving auditing protocol which is described in above section. Here the only difference is the last step dynamic auditing. In the tag generation algorithm TagGen and the ITable generation is done in during the owner initialization phase as shown in the Figure 2.

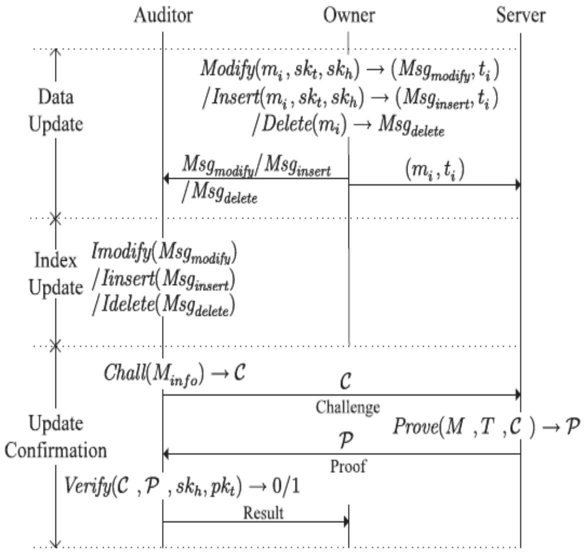

Auditor
Owner
Server
Data Update
$Modify(m_i, sk_t, sk_h) \rightarrow (Msg_{modify}, t_i)$
$/Insert(m_i, sk_t, sk_h) \rightarrow (Msg_{insert}, t_i)$
$/Delete(m_i) \rightarrow Msg_{delete}$
$Msg_{modify}/Msg_{insert}$
$/Msg_{delete}$
(m_i, t_i)
Index Update
$Imodify(Msg_{modifu})$
$/Iinsert(Msg_{insert})$
$/Idelete(Msg_{delete})$
$Chall(M_{info}) \rightarrow \mathcal{C}$
$\mathcal{C}$
Challenge
Update Confirmation
$Prove(M, T, \mathcal{C}) \rightarrow \mathcal{P}$
$\mathcal{P}$
Proof
$Verify(\mathcal{C}, \mathcal{P}, sk_h, pk_t) \rightarrow 0/1$
Result

Figure 2: Framework of auditing for dynamic operations [1]

The dynamic auditing phase again contains three steps: data update, index update, and update confirmation.

Step1: Data Update

3 types of update operations are done on data

The data update step carries out operations that can be used by the owner: modification, inserting, and deleting.

For each update operation, there is;

- **Modify (mi; skt, skh)**.
- **Insert (mi, skt; skh)**
- **Delete (Mi) → Msgdelete**

Step2: Index update

On receiving the three types of update messages, the auditor calls three corresponding algorithms to update the ITable as IModify. Each algorithm is designed as follows:

- IModify (Msgmodify): This is an index modifying algorithm which takes the update message Msgmodify as input. It replaces the old version number Vi by the new one Vi and modifies the tag Ti by the new time stamp tag Ti.

- IInsert (Msginsert): This is an index insertion algorithm which takes input as the update message Msginsert. Then a new record is inserted as (I, Bi; Vi; Ti) in ith position in the ITable. It then moves the original ith record and other records after the ith position in the previous ITable backward in order, with the index number increased by 1.

- IDelete (Msgdelete): The index deletion algorithm takes as input the update message Msgdelete. It deletes the ith record (i; Bi; Vi; Ti) in the ITable and all the records after the ith position in the original ITable moved forward in order, with the index number decreased by 1.

Step3: Update confirmation

Auditor after updating the ITable, it conducts a confirmation auditing for the updated data and sends the result to the owner.

Batch auditing for multiowner and multicloud

Let the involved set of owners and cloud servers in group auditing are denoted as Ochal and Schal respectively.

This batch auditing also consists of three steps: batch challenge, batch proof, and batch verification.

- Batch Challenge - During this step, the auditor executes the batch challenge algorithm BChall to generate a batch challenge C for a set of challenged owners Ochal and a set of clouds Schal.
- Batch proof - Upon receiving the challenge, each server gives a proof P using the group prove algorithm BProve
- The batch prove algorithm takes as inputs the data the received challenge Cl, and the challenge stamp. It generates the tag proof as TP

Phase 3: Batch verification - The auditor executes the following group verifying algorithm on receiving proof from challenged servers BVerify to assure the correctness and accuracy of the proofs. It takes as inputs the challenge C, the proofs, the set of secret hash keys, the public tag keys, and the abstract information of the challenged data blocks For each owner, it calculates the set of identifier hash values for all the chosen data blocks from each challenged server and use these hash values to compute a challenge hash.

Mathematical Model

Input:

U=Set of all available data on the cloud
A=Set of secure data in terms of any security
B=secure cloud data in terms of Auditing and any security algorithm
C=set of cloud secure data in terms of technique or algorithm like cryptography

Output:

D=secure data in terms of public privacy preserving auditing protocol.

Representation of proposed model in mathematical form

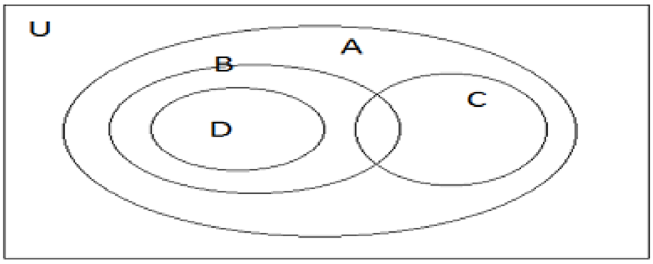


Figure 3: Mathematical Representation

4. Conclusion

The users accessing storage services of cloud has no direct control over their data; hence data security has become one of the major concerns in cloud. In this paper, I have putted light on cloud auditing and its features and also explained schemes that can fully support authorized auditing. Different types of audit practices provide the opportunity to emphasize different aspects of the management program.

References

[1] R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, ‘‘Cloud Computing and Emerging IT Platforms: Vision, Hype, Reality for Delivering Computing as the 5th Utility,’’ Future Gen. Comput. Syst., vol. 25, no. 6, pp. 599-616, June 2009.

[2] M.Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.Katz, A.Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, andM. Zaharia, ‘‘AViewof Cloud Computing,’’ Commun. ACM, vol. 53, no. 4, pp. 50-58, Apr. 2010.

[3] Customer Presentations on Amazon Summit Australia, Sydney, 2012, accessed on: March 25, 2013. [Online]. Available: http://aws.amazon.com/apac/awssummit-au/.

[4] J. Yao, S. Chen, S.Nepal,D. Levy, and J. Zic, ‘‘TrustStore: Making Amazon S3 Trustworthy With Services Composition,’’ in Proc. 10th IEEE/ACM Int’l Symposium on Cluster, Cloud and Grid Computing (CCGRID), 2010, pp. 600-605.

[5] D. Zissis and D. Lekkas, ‘‘Addressing Cloud Computing Security Issues,’’ Future Gen. Comput. Syst., vol. 28, no. 3, pp. 583-592, Mar. 2011.

[6] Q. Wang, C.Wang, K. Ren,W. Lou, and J. Li, ‘‘Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing,’’ IEEE Trans. Parallel Distrib. Syst., vol. 22, no. 5 pp. 847-859, May 2011.

[7] C. Wang, Q. Wang, K. Ren, and W. Lou, ‘‘Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing,’’ in Proc. 30st IEEE Conf. on Comput. and Commun. (INFOCOM), 2010, pp. 1-9.

[8] G. Ateniese, R.D. Pietro, L.V. Mancini, and G. Tsudik, ‘‘Scalable and Efficient Provable Data Possession,’’ in Proc. 4th Int’l Conf. Security and Privacy in Commun. Netw. (SecureComm), 2008, pp. 1-10.

[9] G. Ateniese, R. Burns, R. Curtmola, J. Herring, O. Khan, L. Kissner, Z. Peterson, and D. Song, ‘‘Remote Data Checking Using Provable Data Possession,’’ ACM Trans. Inf. Syst. Security, vol. 14, no. 1, May 2011, Article 12.

[10] G.Ateniese, R.B. Johns,R. Curtmola, J.Herring, L. Kissner,Z. Peterson, and D. Song, ‘‘Provable Data Possession at Untrusted Stores,’’ in Proc. 14th ACM Conf. on Comput. and Commun. Security (CCS), 2007, pp. 598-609.

[11] R. Curtmola, O. Khan, R.C. Burns, and G. Ateniese, ‘‘MR-PDP: Multiple-Replica Provable Data Possession,’’ in Proc. 28th IEEE Conf. on Distrib. Comput. Syst. (ICDCS), 2008

[12] C. Erway, A. Ku¨ pc¸u¨, C. Papamanthou, and R. Tamassia, ‘‘Dynamic Provable Data Possession,’’ in Proc. 16th ACM Conf. on Comput. and Commun. Security (CCS), 2009, pp. 213-222.

[13] Y. Zhu, H. Hu, G.-J. Ahn, and M. Yu, ‘‘Cooperative Provable Data Possession for Integrity Verification in Multi-Cloud Storage,’’ IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 12, pp. 2231-2244, Dec. 2012.

[14] A. Juels and B.S. Kaliski Jr., ‘‘PORs: Proofs of Retrievability for Large Files,’’ in Proc. 14th ACM Conf. on Comput. and Commun. Security (CCS), 2007, pp. 584-597.

[15] H. Shacham and B. Waters, ‘‘Compact Proofs of Retrievability,’’ in Proc. 14th Int’l Conf. on Theory and Appl. of Cryptol. and Inf. Security (ASIACRYPT), 2008, pp. 90-107