DATA INTEGRITY PROTECTION IN CLOUD COMPUTING

Nadar Chitra Kaviarasan
MCA Final year,
Department of Computer Applications
Francis Xavier Engineering College
kavichitra9833@gmail.com

Mrs. E. Sahaya Chithra
Assistant Professor
Department of Computer Applications
Francis Xavier Engineering College
surenchithra@gmail.com

Abstract - Now a day's use of big data processing is rapidly increasing. Big data infrastructure is being a common solution adopted by large organizations for storing and accessing data. It provides current need for data storage with a flexible and dynamic storage that can grow. Information integrity is the preservation and the guarantee of the accuracy and consistency of, data over its entire life-cycle. Data locality is the main feature for providing fast recovery of data in the storage environment. In the existing work, Meta Data Indexing and Integrity Checking are used for traffic load balancing and recovery of lost data part using remote check in cloud storage. The main drawback of the earlier system, it uses distributed access for checking and recovery of data, which may sometimes leads to time delay. In the proposed system, we use TPA based Integrity Verification and Data Recovery, which may help in reducing the time delay and traffic mismatch errors. The system uses Third Party Auditor, that will be verify the status of the servers in every periodic intervals for the lost connection or data. The user files will be segmented and forwarded to the servers and the index will be saved in the TPA. If the user activated the downloading process, the TPA will easily analyze the segments in the various servers and make them ready for the users in less time delay. The main advantages of the proposed system is more efficient, higher analytical of data records, time consuming. This system provides higher result in time consumption and reduced computation overhead.

Keywords — cloud computing; TPA; Integrity

I. INTRODUCTION

Today sharing and storing the data in cloud is easy. Once a user creates shared data in the cloud, every user in the group is able to access and modify shared data, but cannot share the updated details of the shared data with the remining people of the group. To protect the integrity of data in the cloud, number of mechanisms has been proposed. In this proposal, a signature is attached to each block of data, and the integrity of data relies on the correctness of all the signatures. The most important and common features of these proposal is to allow a Public Auditor to proficiently check data integrity in the cloud without getting the entire data, mentioned to as public auditing. This Public Auditor could be a client who would like to use cloud data for specific purposes (e.g., search, computation, data mining, etc.) or Third Party Auditor(TPA) who is capable to proceed authentication services on data integrity to users. With shared data, once a user changes a block, that user also needs to work out a new signature for the edited block. Due to the modifications from different users, altered blocks are signed by different users.

For safety reasons, when a user leaves the group or misbehaves, this user must be withdrew from the group. As a result, this revoked user should no longer be able to use and edit shared data, and the signatures created by this revoked user are no longer valid to this group. As a result, the content of shared data is not changed during user revocation, the blocks, which were signed earlier by the revoked user, still need to be re-signed by an present user in the group. As a result, the integrity of the entire data can still be confirmed with the public keys of present users only. Many public auditing mechanisms were introduced for proficient integrity proving. During public auditing it fails to preserve the identity privacy on shared data and the results in reviling important confidential information to Public Auditor. In existing system once the user is revoked from the system, the blocks which were signed earlier by this revoked user used to be resigned with the help of straightforward method. In which the Public Auditor requested the existing user to first download the blocks signed earlier by the revoked user, then it verifies the correctness of the blocks, then re-sign these blocks and finally upload the re-signature to the cloud. This method creates enormous amount of communication and computation resources by downloading and confirming the blocks, but the content of the block remains identical. This method is not secured because the confidential data of revoked user is misused by an existing user.

This proposal permits a Public Auditor to efficiently check the data integrity in the cloud without downloading the entire data. This mechanism preserves the confidentiality of the shared data by using the proxy re-signature mechanism. In this mechanism the blocks which were previously assign to revoked user will be re-signed by the existing user. For the security purpose secret key will be provided while login. Cloud Computing is a delivery model in which a pool of resources are available to clients and they can access them via internet. Cloud computing and storage provided to users and enterprises having features to store and process their data in either privately owned, or third-party data centers that may be located far from the user are cost effective. The biggest advantage of cloud computing is cost effectiveness. It works on pay as you use format where in client associated with the cloud service provider will be charged based on services chosen and per hour cost as set by cloud service provider. This provides many services where in the most important one is storage i.e cloud provides facility for clients to keep their data into the cloud storage. Cloud storage is important because it reduces the burden on the client by maintaining the data intact and secured. Storage service given by cloud helps its users to manage their data efficiently and in flexible way without keeping a copy of data in their local system. Outsourcing of data to cloud storage servers is developing as a trend among many firms and users owing to its economic advantages. Users today regularly access files without knowing or needing to know on what machines or in what geographical locations their files reside. Specifically, users can process their data on their PCs, outsource the processed data to cloud servers, and use the data on other devices (for example, mobile phones). The great convenience provided by such services is leading to a growing number of cloud storage providers. Owing to the advantages of cloud storage, there exist difficulties in providing such a service to the user. Difficulty which a client face is data resided in the cloud is whether intact and is it possible for an external advisory to breach the security of the cloud. This security concern is called data integrity. Cloud service provider may hide the data loss or may discard the data which is rarely used by the user. To overcome this difficulty many schemes were proposed. First scheme which was proposed is checking the data after very interval of time by the user for data consistency. The biggest disadvantage of this scheme turned to be the burden on the user to check for data integrity even if it the data is intact which results in computation overhead and in increase of communication costs. To reduce the verification burden another scheme was proposed where in an external and independent authority will periodically verifies data integrity on users' behalf. Public verification techniques allow the users to outsource their data into the cloud and consistency Of the data is checked by a trusted third party called auditor. Objective of the public verification scheme is to avoid external adversary attacks on the data outsourced by the owner.

A. THIRD PARTY AUDITOR (TPA)

For well organization it is very essential that cloud that allows investigation from a single party audit the outsource data to ensure data security and save the user's computation and data storage. It is very important to provide public auditing service for cloud data storage, so that the user trusts an independent third party auditor (TPA). TPA checks the integrity of data on cloud on the behalf of users, and it provides the reasonable way for users to check the validity of data in cloud. Public auditing in addition to user provides the external party to verify the correctness of stored data against external attacks it's hard to find. However these schemes, as in don't involve the privacy protection of the data. It is a main disadvantage which affect the security of the protocols in cloud computing. So clients who rely upon TPA for their security stockpiling need their information to be shielded from outside evaluators. Cloud specialist organization has huge storage room and calculation asset to keep up the clients' information. It likewise has aptitude in building and overseeing disseminated distributed storage servers and capacity to possess and work live distributed computing frameworks. Clients who put their substantial information documents into distributed storage servers can assuage weight of capacity and calculation. In the meantime, it is critical for clients to guarantee that their information are being put away effectively and security check.

Clients ought to be outfitted with certain security implies so they can ensure their information is sheltered. Cloud service provider is always online & assumed to have abundant storage capacity and computation power. The third party auditor is invariably online, too. It makes every data access be in control.

The concept of public auditability used to check the data integrity. Since cloud service providers (CSP) are separate administrative entities, data outsourcing

is truly relinquishing user's ultimate control over the destiny of their data. As a result, the accuracy of the data within the cloud is being place at hazard because of the subsequent reasons. Firstly, although cloud infrastructures is much more powerful and consistent than personal computing devices, still they are facing the broad range of each internal and external threats for information integrity. Secondly, {there is chance that because of numerous motivations for CSP to behave unfaithfully toward the cloud users concerning their outsourced information status. For examples, CSP could rescue storage for economic reasons by discarding information that have not been or are seldom accessed, or maybe hide data loss incidents to maintain a reputation. Due to these reasons data owners would worry that the data could may be lost within the cloud. Thus, enabling public auditability for cloud storage is of important importance in order that users will resort to a thirdparty auditor (TPA) to examine the integrity of outsourced data and be worry free.

II. LITERATURE SURVEY

A distributed repair method where network traffic and the recovery of the data on a failure of the system. Recovery of the data by regenerating codes has been relying much on the storage problems. Here the erasure code reduces such burdens of data recovery and network traffic. By combining this technique with network topology, a novel repair tree to minimize repair traffic is introduced. A repair tree is also generated for the network traffic problem by combining the network topology along with the distributed codes. Evaluation and analysing of the repair tree is done so that they bandwidth of the regenerating code is reduced.[1]

Abouzarjomehri et.al proposed a hybrid cloud environment where privileged access, availability, data location, investigative supports are focused. The three basic principles of confidentiality, availability, integrity are used in all data security techniques. Here the principle of confidentiality deals with the sharing of data with an authenticated user and the data is shared only when the user request for it. The principle of integrity gives the highquality assurance of the data and also the accessibility of it. The principle of availability deals with the obtainability of the data when the authorized user requires it.[3] focuses on the details of protection methods and approaches used to ensure maximum data protection which can be achieved by reducing the risks and threats from the intruders. This study is based on all levels of SaaS, PaaS and IaaS. The data are protected by encryption for example; encrypting the keys for data in transmission can be short lived. Different cryptographic techniques are also being used like block ciphers, stream cipher, hash functions etc…Henry, Chen and Patrick [4] , proposed a practical data integrity protection (DIP) scheme where the integrity of the regenerating codes under a particular network is checked. This also repairs the failed system and works to avoid the traffic problem of the data. While regenerating codes a system may fall in possible failure and that might create a traffic issue for the rest of the network. Thus, this scheme is based on mobile model where the user can verify the integrity of the data against the corruption of the system. This DIP scheme also works as a mathematical model. Gary Grider et al states that all the storage device where the erasure code was implemented was checked on the basis of processing time, power utilization and the energy cost. Generally, the erasure coding and decoding involves multiple intensive operations which lead to system failure. Thus, the proposed approach is cost efficient for all the cloud storage device.

Liu, Huang and the fellow members[6] developed a public auditing scheme for the regenerating-codebased cloud storage. This scheme is an all-or-nothing transform-based encryption and a variant of ElGamal-based proxy re-encryption algorithm are used to provide security of the code on transmission. Here the data owner himself is unable to send or decrypt the re-encrypted data to other users. To solve the regeneration problem of failed authenticators in the absence of data owners, a proxy is introduced, which is privileged to regenerate the authenticators, into the traditional public auditing system model. This typically reduces the burden of the data owner. Extensive security analysis shows that the scheme is provable secure under random oracle model and experimental evaluation

Y. Wang, D. Wei, X. Yin and X. Wang [7] reconsidered the problem of minimizing regeneration time in networks with heterogeneous link capacities. It derives the minimum amount of data to be transmitted through each link to preserve data integrity and prove that, building an optimal regeneration tree is NP-complete and propose a heuristic algorithm for a near-optimal solution. A flexible regeneration scheme is introduced, which allows providers to generate different amount of coded data. Simulation results show that the flexible tree-structured regeneration scheme can reduce the regeneration time significantly.Chen, Patrick, Yang

Tang [8] presented a proxy-based storage system called NCCloud which is a fault tolerant for multiple cloud storage. This NCCloud is laid on top of the network coding called functional minimum-storage regenerating (FMSR) codes. This maintains the data redundancy and fault tolerance of the system. The main advantage of this is that the repair of the network coding takes place on failure with the code in the storage remains untouched. The FMSR provides less cost to repair and also a better performance time.

Wang, Chow and fellow members [9] proposed a secure cloud storage system supporting privacypreserving public auditing. It extends the result to enable the TPA to perform audits for multiple users simultaneously and efficiently. The TPA can efficiently provide authentication through token by enabling public audit ability for cloud storage. Thus, the TPA checks for the integrity of the data and also maintains the queue of the user request. The practical implementation on Amazon E2C proves that this is secured and efficient for all cloud storage medium. Demakis, Martin, Godfrey and Kannan Ramachandran [10] proposed the notion of regenerating codes, which allow a new node to communicate functions of the stored data from the surviving nodes. It shows that regenerating codes can significantly reduce the repair bandwidth. The regenerating od code will relatively reduce the bandwidth and repair it. The repair bandwidth and storage are characterized by using flow arguments which is actually a trade-off. By this optimal trade-off the generation of the network coding and the regeneration of code id achieved.

III. SYSTEM MODEL

A. *PROBLEM DEFINITION*

- Data that is stored in the cloud could suffer from the damage on transmitting to/from cloud data storage.
- Since the data and computation are outsourced to a remote server, the data integrity should be maintained using Third Party Auditor and checked constantly in order to prove that data and computation are intact.
- Data integrity means data should be kept from unauthorized modification. Any modification to the data should be detected.
- Any deviation from normal computation should be detected. Integrity should be checked at the data level and computation level.
- Data integrity could help in getting lost data or notifying if there is data manipulation

B. IMPLEMENTATION

UPLOAD OPERATION

Step 1 : Generate the per-file secrets.
Step 2 : Split the file into four parts according to size.
Step 3 : Encoded each code chunk with BLOWFISH.
Step 4 : Store into the four respective cloud servers.
Step 5 : Update the metadata file and upload.

Download Operation

Step 1 : Check the metadata file.
Step 2 : Decodes the encoded chunk for file F.
Step 3 : Merge and downloads the decoded chunk for file F. Repair Operation.

TPA PROCESS

Step 1 : Periodic update the server status
Step 2 : Update the fail index for download selection.
Step 3 : Data index validation prepare the selection process with higher accuracy.

Recovery Operation

Step 1 : Check the metadata file.
Step 2 : Regenerate the file. In particular, if there is only one failed server, then instead of trying to download k(n-k) chunk from any k server, download one chunk from Backend server.
Step3 : Decodes the encoded chunk for file F.
Step 4 : Merge and downloads the decoded chunk for file F.

C. *MOTIVATION*

- The existing system used distributed metadata access, which leads to communication overhead if the user queue is higher.
- The server and data validation will only be checked after the download process starts. The integrity of metadata processing will be slower than anticipated.
- The existing system uses remote integrity using data encoding and regeneration code in decentralized methodology.

D. *OBJECTIVE*

- To provide faster load balancing recovery of metadata processing

- TPA based communication will result less delay processing.
- Verification of server and data in regular periodic interface using index based packets

E. Modules

1. Meta Data Processing
2. Meta Indexing
3. Third Party Auditor
4. Regeneration Code
5. Integrity Verification

In this module, we apply metadata processing suitable for operating data intensive and computational intensive applications. There is a serious requirement to deal with the data security issues for preserving the data integrity, privacy and trust in the security environment. While security concerns are protecting some organizations from adopting cloud computing at all. In this module, data owners first encode the metadata files by using regenerating code, and then store the coded file across multiple cloud servers. The multiple cloud web servers may locate in the same provider or different service providers. Data owners may perform block-level active functions on the outsourced data.

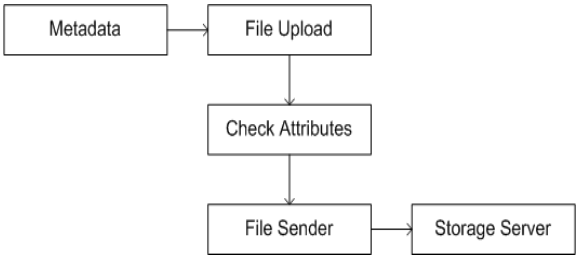

Figure 1: Metadata processing

In this module, meta indexing are proposed using data structure to support dynamic data update operations in which the data owner needs to store block index and block logical location for each block of the outsourced file. The main advantage of this method is that it is able to efficiently support dynamic update operations efficiently due to the node re-balancing problem.

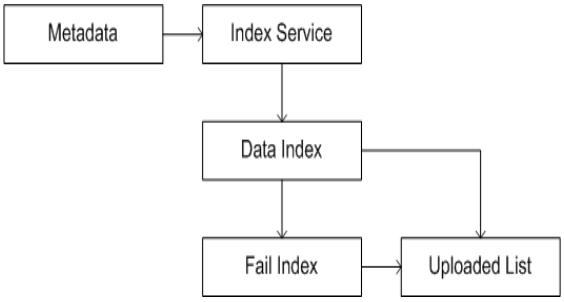

Figure 2: Meta Indexing

In this module, for data integrity confirmation use a third get together auditor, specifically a sole third party auditor. TPA helps an end user verify the metadata. TPA can gain access to control should be applied to determine traditional users and minimize the possibility of unauthorized users. The communication and computation expense should be reduced. Information integrity with high security may be ensured when blocks of information are distributed between multiple auditors for verification.

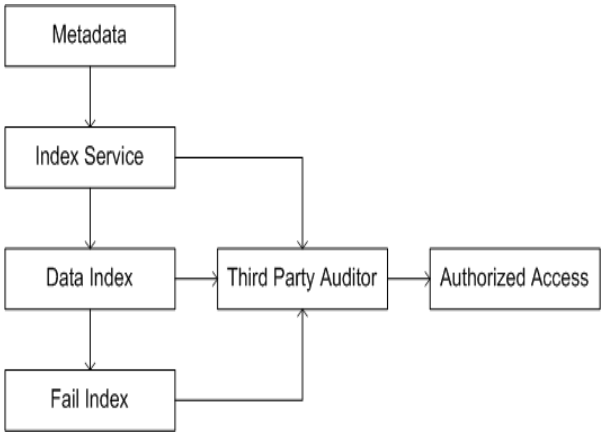

Figure 3: Third party auditor

In this module, the storage that holds data and information on the cloud is obligated on data integrity. Data integrity depends on the assurance pursued by the user that data are unaltered on the provider infrastructure. Data integrity threats involve both malicious third party occurrences and hosting infrastructure weaknesses. Protecting data from loss and leakage involves integrity of many parties involved in providing the resources. Some schemes and mechanism are needed to ensure the data and information kept on the cloud is unaltered or removed. It is suggested to practice auditing techniques such as proof-of-retrievability and proof-of-data possession to enable verification.

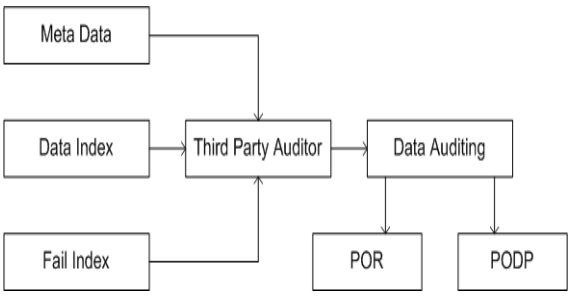

Figure 4: Regeneration codes

In this module, integrity verification provides guarantee that the data will always be available autonomously regardless of hardware failures, corrupted physical disks or downtime. Hardware failures can happen at any time. This includes failures

caused by environmental failures such as a natural disaster, flood or even fire. A hardware design should be built on a basis of having redundancy and minimum single points of failure. At the design phase, the analyst creates a physical hardware map that shows all the connection points for server, storage, network and software.

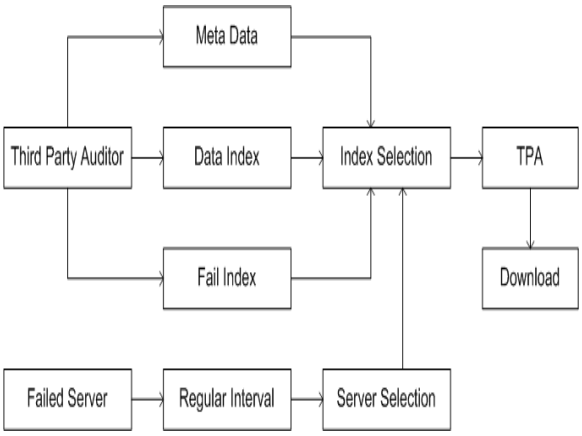

Figure 5: Integrity verification

F. Algorithm

TPA - CHECK FAIL SERVER

The data blocks stored at servers are coded as linear combinations of the original blocks supposing that the curious TPA has recovered m coded blocks by elaborately performing Challenge-Response procedures and solving systems of linear equations. The TPA still requires solving another group of linearly independent equations to derive the native blocks.

ATTRIBUTES

N = No of Servers

C = Connections

Begin

Step 1: for each j = 0 to N

Step 2: check if CF(Sj) > 0

Step 3: for each i = j+1 to N

Step 4: check if (C(Sj)*CF(Si) > C(Si)*CF(Sj)) Then

Step 5: j = i (End of if)

Step 6: return Sj;

Step 7: end if

Step 8: end for

Step 9: end for

End

IV. RESULT

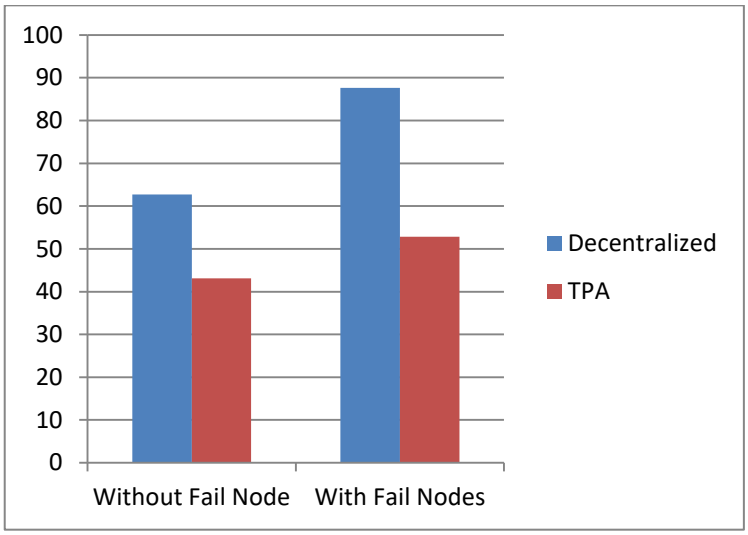

Figure 5: COMMUNICATION PROCESS

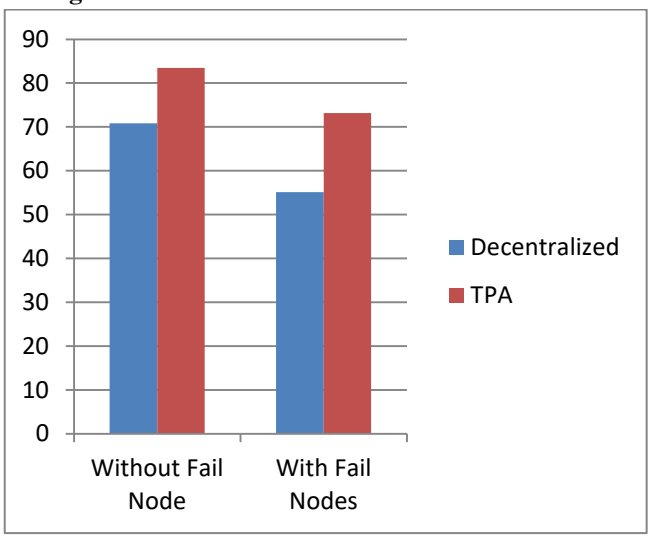

Figure 6: INTEGRITY VERIFICATION

V. FUTURE SCOPE

In the future work, a backup or replication to the TPA can provide higher data retrieval and indexing in very less period. The security can be added to the system will help in protecting the more privacy to the user data and files. An efficient machine learning algorithms like ADABOOST can be implemented in the system which will help the system in time consumption and increase the accuracy of the retrieval.

VI. CONCLUSION

In this paper, a TPA based Integrity Verification and Data Recovery has been proposed, which helps reducing the computation time delay and traffic mismatch errors. The system mainly depends on Third Party Auditor (TPA) which will verify the status of the servers in regular interval for the lost connection or data. The system will gain more

efficient, higher analytical of data records, time consuming. This system provides higher result in time consumption and reduced computation overhead which compared to the previous results.

REFERENCES

[1] H. Abu-Libdeh, L. Princehouse, and H. Weatherspoon. RACS: A Case for Cloud Storage Diversity. In *Proc. of ACM SoCC*, 2010.

[2] R. Ahlswede, N. Cai, S.-Y. R. Li, and R.W. Yeung. Network Information Flow. *IEEE Trans. on Information Theory*, 46(4):1204–1216, Jul 2000.

[3] Amazon Elastic Compute Cloud. http://aws.amazon.com/ec2/.

[4] Amazon Simple Storage Service. http://aws.amazon.com/s3/.

[5] 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 view of cloud computing. *Communications of the ACM*, 53(4):50–58, 2010.

[6] 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. on Information and System Security*, 14:12:1–12:34, May 2011.

[7] G. Ateniese, R. Di Pietro, L. V. Mancini, and G. Tsudik. Scalable and Efficient Provable Data Possession. In *Proc of SecureComm*, 2008.

[8] G. Ateniese, S. Kamara, and J. Katz. Proofs of Storage from Homomorphic Identification Protocols. In *Proc. of ASIACRYPT*, 2009.

[9] A. Bessani, M. Correia, B. Quaresma, F. Andr´e, and P. Sousa. DEPSKY: Dependable and Secure Storage in a Cloud-of-Clouds. In *Proc. of ACM EuroSys*, 2011.

[10] J. Black and P. Rogaway. Ciphers with arbitrary finite domains. In *Topics in Cryptology – CT-RSA 2002*, volume 2271 of *LNCS*, pages 114–130. Springer, 2002.

[11] K. Bowers, A. Juels, and A. Oprea. HAIL: A High-Availability and Integrity Layer for Cloud Storage. In *Proc. of ACM CCS*, 2009.

[12] K. Bowers, A. Juels, and A. Oprea. Proofs of Retrievability: Theory and Implementation. In *Proc. of ACM CCSW*, 2009.

[13] B. Chen, R. Curtmola, G. Ateniese, and R. Burns. Remote Data Checking for Network Coding-Based Distributed Storage Systems. In *Proc. of ACM CCSW*, 2010.

[14] H. C. H. Chen and P. P. C. Lee. Enabling Data Integrity Protection in Regenerating-Coding-Based Cloud Storage. Technical report, CUHK, 2012.

[15] R. Curtmola, O. Khan, and R. Burns. Robust remote data checking. In *Proc. of ACM StorageSS*, 2008.

[16] R. Curtmola, O. Khan, R. Burns, and G. Ateniese. MR-PDP: Multiple-Replica Provable Data Possession. In *Proc. of IEEE ICDCS*, 2008.

[17] A. Dimakis, P. Godfrey, Y. Wu, M. Wainwright, and K. Ramchandran.Network Coding for Distributed Storage Systems. *IEEE Trans. onInformation Theory*, 56(9):4539–4551, 2010.

[18] Y. Dodis, S. Vadhan, and D. Wichs. Proofs of Retrievability via Hardness Amplification. In *Proc. of TCC*, 2009.

[19] C. Erway, A. K¨upc¸ ¨u, C. Papamanthou, and R. Tamassia. Dynamic Provable Data Possession. In *Proc. of ACM CCS*, 2009.

[20] O. Goldreich. *Foundations of cryptography: Basic tools*, volume 1. Cambridge Univ Pr, 2001.

[21] O. Goldreich. *Foundations of cryptography: Basic applications*, volume 2. Cambridge Univ Pr, 2004.

[22] Y. Hu, H. Chen, P. Lee, and Y. Tang. NCCloud: Applying Network Coding for the Storage Repair in a Cloud-of-Clouds. In *Proc. of USENIX FAST*, 2012.