International Journal of MC Square Scientific Research Vol.6, No.1 Nov 2014

SECURE & EFFICIENT AUDIT SERVICE OUTSOURCING FOR DATA INTEGRITY IN CLOUDS

Gyan Prakash,
Software Engineer, Vee Eee Technologies Solutions Pvt. Ltd.
Email: prkashgyan90@yahoo.com

Bhaskar Vyas
Product Design Specialist, Cognizant Technology Solutions, Kolkata, India

Venkata Reddy Kethu
Associate Director-Projects, Cognizant Technology Solutions, Hyderabad, India

Abstract - Cloud-based outsourced storage relieves the client's load for storage management and maintenance by providing a comparably low-cost, scalable, location-independent platform. Though, the information that clients no longer have physical control of data specifies that they are facing a potentially formidable risk for missing or corrupted data. To avoid the security risks, inspection services are serious to ensure the integrity and availability of outsourced data and to achieve digital forensics and reliability on cloud computing. Provable data possession (PDP), which is a cryptographic method for validating the reliability of data without retrieving it at an untrusted server, can be used to realize audit services. In this project, profiting from the interactive zero-knowledge proof system, the construction of an interactive PDP protocol to prevent the fraudulence of prover (soundness property) and the leakage of verified data (zero knowledge property).To prove that our construction holds these properties based on the computation Diffie–Hellman assumption and the rewindable black-box knowledge extractor. An efficient mechanism on probabilistic queries and periodic verification is proposed to reduce the audit costs per verification and implement abnormal detection timely. Also, we present an efficient method for choosing an optimal parameter value to reduce computational overheads of cloud audit services.

Keywords - security risks, audit services, Provable data possession (PDP), zero knowledge property, Diffie–Hellman assumption.

1. Introduction

In recent existence, the up-and-coming cloud-computing prototype is rapidly gaining momentum as an alternative to usual information technology. Cloud compute provide a scalability atmosphere for on the rise amounts of data and processes that work on various applications and services using on-demand self-services. One essential aspect of this model changing is that data are being centralized and outsourced into clouds. This kind of outsourced storage services in clouds have become a new profit growth point by providing a comparably short-price, scalable, locality-independent stage for managing consumers' data.

The cloud storage service (CSS) relieves the lumber of storage organization and maintenance. Though, if such an imperative service is predisposed to attacks or failures, it would bring permanent losses to users since their data or archives are stored into an unsure storage pool exterior the enterprise. These security risks come from the following reasons: the cloud infrastructures are much more powerful and reliable than personal computing devices.

However, they are still susceptible to security threats both from outside and inside the cloud for the profit of their tenure, there exist various motivations for cloud service providers (CSP) to behave unfaithfully toward the cloud users Moreover, the dispute rarely suffers from the lack of trust on CSP. Accordingly, their actions may not be known by the cloud user, even if this quarrel may affect outcome from the users' improper operations (KO Et Al., 2011). Therefore, it is necessary for cloud service providers to offer an efficient assessment service to check the integrity and availability of the stored data.

Traditional cryptographic technologies for data integrity and accessibility, based on chop function and signature scheme can't work on the outsourced data devoid of a local duplicate of data. Also, it is not a realistic solution for data validation by downloading them due to the expensive transaction, especially for large size files. Furthermore, the solutions to check the reality of the data in a cloud environment can be formidable and expensive for the cloud users (Armrest et al., 2010). Therefore, it is crucial to realize public assessment knack for CSS, so that data owners may resort to a third party assessment (TPA), who has expertise and capabilities that a common user does not cover, for occasional assessment the farm out data. This assessment service is significantly important for digital forensics and data assurance in clouds.

To implement public assessment ability, the notions of proof of irretrievability (POR) (Jules, 2007) and provable data possession (PDP) (Attendees et al., 2007) have been proposed by some investigators. Their loom was based on a probabilistic proof technique for a storage provider to prove that clients' data remain intact without downloading the stored data, which is called "verification without downloading." For ease of use, some POR/PDP scheme toil on a publicly certifiable mode, so that someone can use the certification protocol to prove the availability of the stored data. Therefore, this provides us an efficient loom to accommodate the requirements from public assessment ability. POR/PDP schemes evolved around an untrusted storage offer a publicly accessible remote interface to check the tremendous amount of data.

Although PDP/POR schemes evolved around untrusted storage, offer a publicly accessible remote interface to check and manage the tremendous amount of data, most of the obtainable schemes cannot give a strict security proof against the untrusted CSP's dishonesty and fake, as well as information leakage of verified data in the verification process. These drawbacks greatly affect the crash of cloud assessment services. Thus, new frameworks or models are desirable to enable the security of public verification protocol in cloud assessment services. Another major concern addressed by this paper is how to improve the performance of assessment services.

The assessment performance concerns not only the costs of computation, communication, storage for assessment activities but also the scheduling of assessment actions. No distrust inappropriate arrangement, more or less repeated, causes poor assessment routine, but an efficient arrangement can help provide a better quality of and a more cost-

effective service. Hence, it is serious to examine a well-organized schedule for cloud assessment services.

In practical applications, above conclusions will play a key role in obtaining a more efficient assessment schedule. Further, our optimization algorithm also supports an adaptive parameter selection for different sizes of files (or clusters), which could ensure that the extra storage is optimal for the verification process.

Finally, implement a prototype of an assessment system to evaluate our proposed approach. Our experimental results not only validate the effectiveness of above-mentioned approaches and algorithms but also show our system has a lower computation cost, as well as a shorter extra storage for verification. List the features of our PDP scheme. Also contain an evaluation of related techniques, such as PDP, DPDP, and CPOR. Although the computation and communication overheads of O (t) and O (1) in PDP/SPDP schemes are lower than those of O (t + s) and O(s) in our scheme, our scheme has less difficulty due to the introduction of a scrap arrangement, in which an outsourced file is split into n blocks, and each block is also split into s sector. This means that the number of blocks in PDP/SPDP schemes is times more than that in our scheme and the number of sampling blocks t in our scheme is merely 1/s times more than that in PDP/SPDP schemes. Moreover, the chance of detection in our scheme is much greater than that in PDP/SPDP schemes because of $1-(1-b)^{ts} \geq 1-(1-b)^{t}$. Also, our scheme, similar to PDP and CPOR schemes, provides the ownership proof of outsourced data as a result that it is constructed on the public-key verification knowledge, but SPDP and DPDP schemes cannot offer such a feature because they are only based on the Hash function.

2. Related Works

Arithmetic Operators for Pairing-Based Cryptography

Since their introduction in constructive cryptographic applications, pairings over (hyper) elliptic curves are at the heart of an ever increasing number of protocols. Software implementations being rather slow, the study of hardware architectures became an active research area. Our architecture is based on a unified arithmetic operator which performs addition, multiplication and cubing over F397 [1]. This design methodology allows us to design a compact coprocessor (1888 slices on a Vertex-II Pro 4 FPGA) which compares favorably with other solutions described in the open literature. Then describe ways to extend our approach to any characteristic and any extension field.

Finite Field Arithmetic

Arithmetic in a finite field is different from standard integer arithmetic. There are a limited number of elements in the finite field; all operations performed in the finite field result in an element within that field. While each finite field is itself not infinite, there are infinitely many different finite fields; their number of elements (which is also called cardinality) is necessary of the form p^n where p is a prime number and n is a positive integer, and two finite fields of the same size are isomorphic. The prime p is called the characteristic of the field, and the positive integer n is called the dimension of the field over its prime field [1], [2], [3], [4].

Finite fields are used in a variety of applications, including in classical coding theory in linear block codes such as BCH codes and Reed–Solomon error correction and cryptography algorithms such as the Rijndael encryption algorithm.

Addition, Multiplication, and Frobenius Map over Fpm

The architecture of the operators generated by our program is directly inspired from the unified operator given in Figure 1 and can be adapted to any prime characteristic p and any irreducible polynomial f(x) of degree m. Comparisons against FPGA-based accelerators over F397. The parameter D refers to the number of coefficients processed at each clock cycle by a multiplier [8].

Table 2.1 Addition Frobenius Map over Fpm

	Grabher and Page [13]	Kerins *et al.* [17]	Beuchat *et al.* [6, 7]
Algorithm	Duursma-Lee	Duursma-Lee	η_T pairing
FPGA	Virtex-II Pro 4	Virtex-II Pro 125	Cyclone II EP2C35
Multiplier(s)	1 ($D = 4$)	18 ($D = 4$)	9 ($D = 3$)
Area	4481 slices	55616 slices	$\sim$ 18000 LEs
Clock cycles	59946	12866	4849
Clock frequency	150 MHz	15 MHz	149 MHz
Calculation time	432.3 μs	850 μs	33 μs

Addition over Fp[x]/ (f(x)) is performed in the same way as in the operator over F397 presented: the digits of the two operands are all added in parallel, thus requiring m additions over Fp. In the current version of the generator, those additions over Fp are implemented as simple look-up tables addressed by the bits of the two operands, particularly suited for small values of p (typically p = 2 to 7). For higher characteristics, it will be necessary to resort to more complex methods for modular addition.

Table 2.2 Multiplication Frobenius Map over Fpm

	Ronan *et al.* [25]		Proposed architecture
Algorithm	η_T pairing	η_T pairing	η_T pairing
FPGA	Virtex-II Pro 100	Virtex-II Pro 100	Virtex-II Pro 4
Multiplier(s)	3 ($D = 8$)	2 ($D = 8$)	1 ($D = 3$)
Area	10000 slices	7491 slices	1888 slices
Clock cycles	15113	17190	32618
Clock frequency	70.4 MHz	70.4 MHz	147 MHz
Calculation time	178 μs	203 μs	222 μs

Also as in the original operator, multiplication over $F_p[x]/(f(x))$-relies on a parallel-serial algorithm, with D digits of the multiplier being processed each iteration. The generation of the partial products, which consists in multiplying all the digits of the multiplicand with each digit of the multiplier, requires m multiplications over Fp in parallel for each of the D partial products. Here also, the multiplications over Fp are directly tabulated, as this is the best solution for small characteristics. Once the D partial products are computed, the D – 1 most significant ones along with the accumulator are then multiplied by xk (where k ranges from 1 to D) and reduced modulo f(x). After the modular reductions, the D partial products and the accumulator have added thanks to a binary tree of adders

over Fpm [5], [6], [7]. Consequently, to optimize the critical path of this multi-operand adder, one should choose a parameter D of the form 2n − 1 (typically D = 3, 7, 15 or 31).

$$a(x)^p \bmod f(x) = \sum_{i=0}^{m-1} a_i^p x^{ip} \bmod f(x)$$

$$a(x)^p \bmod f(x) = \sum_{i=0}^{m-1} a_i x^{ip} \bmod f(x)$$

Fig 2.1 Formula for Addition Frobenius Map over Fpm

This general expression of the Frobenius map can then be seen as a sum of elements of Fpm. The coefficients of those polynomials can be directly matched to the coefficients of the operand, possibly multiplied by a constant. It is possible to reuse the partial product generation hardware of the multiplication to compute those polynomials, only some extra wiring being required for the permutation of the coefficients. The sum of all the polynomials can then be computed by the final multi-operand adder. To decrease the number of partial products necessary to compute the Frobenius map, a simple decomposition technique can be applied to share the maximum amount of hardware between these partial products. In case this is still not enough, a second technique can further pack the partial products, at the expense of some additions over Fp.

Efficiency and Security

The proposed PDP scheme relies only on efficient symmetric key operations in both setup (performed once) and verification phases. However, our scheme is more efficient than POR as it requires no bulk encryption of outsourced data and no data expansion due to additional sentinel blocks. Our scheme provides probabilistic assurance of the unhampered data being stored on the server with the exact probability fixed at setup. Furthermore, our scheme is provably secure in the random oracle model (ROM).

3. Proposed System

In this paper, utilize the public Provable data possession (PDP), which is a cryptographic technique for verifying the integrity of data without retrieving it at an untrusted server; can be used to realize audit services. It with random mask technique to achieve a privacy-preserving public auditing system for cloud data storage security while keeping all above requirements in mind. To support efficient Handling of multiple auditing tasks, further explore the technique of bilinear aggregate signature to extend our main result into a multiuser location, where TPA can achieve multiple audit tasks simultaneously. Extensive security and performance analysis shows the proposed schemes are provably secure and highly efficient. And also show how to an extent our main scheme to support batch auditing for TPA upon delegations from multi-users.

Audit Service System

In this module, we provide an efficient and secure cryptographic interactive audit scheme for public audit capacity. We provide an efficient and protected cryptographic interactive retains the consistency property and zero-knowledge property of evidence systems. These two properties make certain that our plan can not only prevent the deception

and forgery of cloud storage provider but also avoid the outflow of outsourced data in the process of verification.

Data Storage Service System

In this module, we considered FOUR entities to store the data in a secure manner:

- **Data Owner (DO)**

A large amount of data to be stored in the cloud.

- **Cloud Service Provider (CSP)**

Who provide data storage service and have sufficient storage spaces and computation resources.

- **Third Party Auditor (TPA)**

Who have capabilities to manage or monitor – outsourced data under the delegation of data owner?

- **Granted Applications (GA)**

Who has the right to access and maneuver stored data? These applications can be whichever inside clouds or faint clouds according to the specific requirements.

Audit Outsourcing Service System

In this module the client (data owner) uses the secret key to pre-process the file, which consists of a collected works of block, generate a set of civic Certification Information that is stored in TPA, transmit the box file and some certification tags to Cloud service supplier CSP, and may remove its local facsimile. At a later time, using a protocol of proof of irretrievability, TPA (as an audit agent of clients) issues a challenge to audit (or check) the integrity and availability of the outsourced data regarding the public certification in sequence. It is essential to give an alarm for irregular events.

Secure and Performance Analysis

In this module, we considered to secure the data and give the performance to the following:

Audit-Without-Downloading

To allow TPA (or other clients with the help of TPA) to verify the correctness of cloud data on demand without retrieving a copy of whole data or introducing additional on-line burden to the cloud users.

Verification-Correctness

To make sure there exists no devious CSP that can pass the audit from TPA without indeed storing users' data intact.

Privacy-Preserving

To make sure that there exist methods for TPA to originate consumer's data from then information collected during the auditing process.

High-Performance

To allow TPA to perform auditing with minimum outlay in storage, communiqué, and computation, and to carry statistical inspect sampling and optimized audit schedule with a long enough period.

4. Algorithm

Our proposed solution and many of the related work are based on the following cryptographic primitives:

One-way hash function

A one-way hash function denoted as $h(.)$, is a hash function that works in one direction: it is easy to compute a fixed-length digest $h(m)$ from a variable-length pre-image m; however, it is hard to find a pre-image that hashes to a given hash value. Examples include MD5 and SHA. Will use the terms hash, a hash value and digest interchangeably.

Digital signature

A digital signature algorithm is a cryptographic tool for authenticating the integrity and origin of a signed message. In the algorithm, the signer uses a private key to generate digital signatures on messages, while a corresponding public key is used by anyone to verify the signatures. RSA and DSA are two commonly-used signature algorithms.

Signature aggregation

Different messages into one signature. Signing a message m involves computing the message hash $h(m)$ and then the signature on the hash value. To aggregate t signatures, one simply multiplies the individual signatures, so the aggregated signature has the same size as each signature. Verification of an aggregated signature involves computing the product of all message hashes and then matching with the aggregated signature.

Signature chain

In a signature, chain scheme is proposed that enables clients to verify the completeness of answers to range queries. A very nice property of the scheme is that only result values are returned, thus ensuring that there is no violation of access control. The scheme is based on two concepts: (a) the signature of a record is derived from its digest as well as its left and right neighbors'. In this way, an attempt to drop any value from the answer of a range query will be detected since it would no longer be possible to derive the correct signature for the record that depends on the dropped value. (b) For the boundaries of the answer, a collaborative scheme that involves both the publisher and the client is proposed – the publisher performs partial computation based on but not revealing the two records bounding the answer and the query range, while the client completes the computation based on the two end points of the query range.

5. Result and Performance Analysis

Signature Chain in Multi-Dimensional Space

The goal of our work is to devise a solution for checking the correctness of query answers on multi-dimensional datasets. The design objectives include:

Completeness: The user can verify that all the data points that satisfy a window query are included in the answer.

Authenticity: The user can check that all the values in a query answer originated from the data owner. They have not been tampered with, nor have spurious data points been introduced.

Precision: Proving the correctness of a query answer entails minimal disclosure of data points that lie beyond the query window. Define precision as the ratio of the number of data Points within the query window, to the number of data points returned to the user.

Security: It is computationally infeasible for the publisher to cheat by generating a valid proof for an incorrect query answer.

Efficiency: The procedure for the publisher to generate the proof for a query answer has polynomial complexity. Likewise, the procedure for the user to check the proof has polynomial complexity.

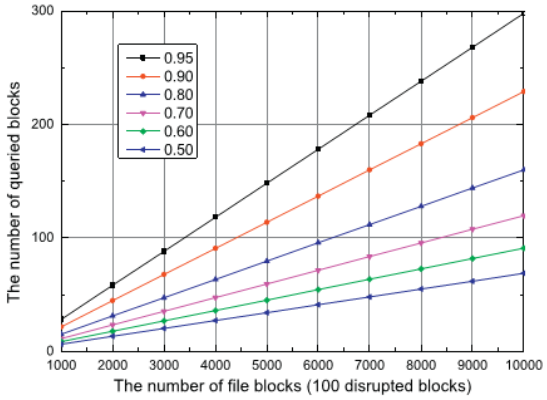

Fig.5.1 Numbers of queried blocks under different detection probabilities and the different number of file blocks.

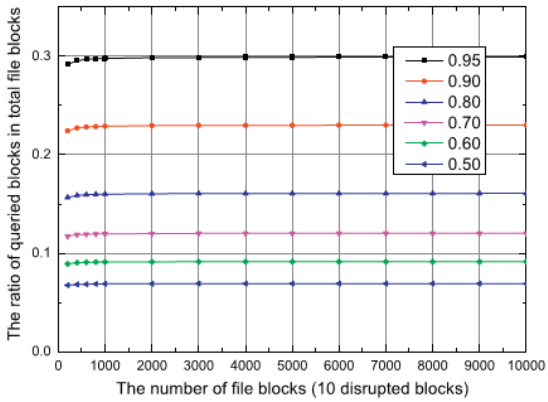

Fig.5.2 Ratio of queried blocks in the total file blocks under different detection probabilities and a different number of disrupted blocks (for ten disrupted blocks).

$$f = \frac{\log(1-P_r)}{w.n.T.\log(1-\rho_b)}$$

Without loss of generality, assume that the data in the multi-dimensional space are split into partitions – this can be done using a spatial data structure. To ensure that the answer for a window query is complete, two issues must be addressed.

First, need to prove that the answer covers all the partitions that overlap the query window. Refer to these partitions as candidate partitions. Second, need to show that all qualifying values within each candidate partition are returned.

The first issue is dependent on the partitioning strategy adopted

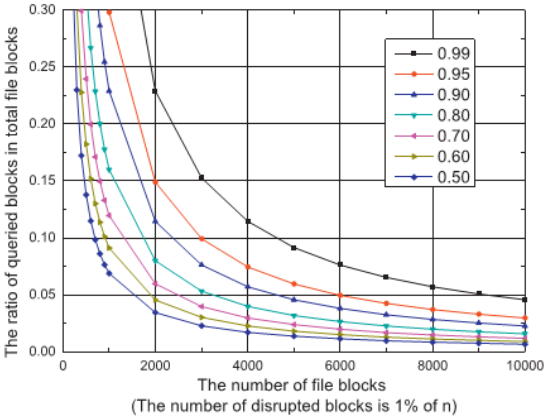

Fig.5.3 Ratio of queried blocks in total file blocks under different detection probabilities and 1% disrupted blocks.

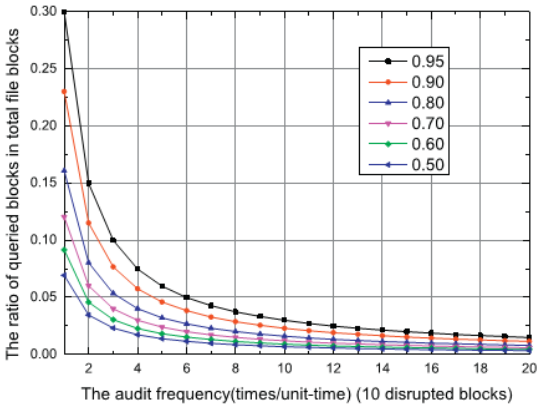

Fig.5.4 Ratio of queried blocks in the total file blocks under different audit frequency for 10 disrupted blocks and 10,000 file blocks.

6. Conclusion

In this dissertation, the construction of an efficient audit service for data integrity in clouds is discussed. In this audit service, the third party inspector, known as a manager of data proprietor, can issue a periodic certification to observe the change of outsourced data by providing an optimized schedule. To realize the audit model, it is necessary to maintain the security of the third party auditor and deploy a lightweight daemon to execute the verification protocol. This technology can be easily adopted in a cloud computing environment to replace the traditional hash-based solution.

More importantly, this dissertation quantified a new audit approach based on probabilistic queries and periodic certifications, as well as an optimization scheme of bound of cloud inspect services. This approach enormously reduces the work stack on the storage servers; while still attain the detection of servers' mischief with a high prospect. The experimental results clearly showed that our approach could minimize computation and communication overheads.

REFERENCES

1. A.D. Rabkin, A., Stoica, I., Zaharia, M., 2010. A view of cloud computing. Common. ACM 53 (4), 50–58.
2. Alam, I., C. Perry. "A Customer-Oriented New Service Development Process." Journal of Services Marketing 16, no. 6 (2002): 515-534.
3. Armbrust, M., A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica and M. Zaharia. "Above the Clouds: A Berkeley View of Cloud Computing." Technical Report, 2009.
4. Ateniese, G., Burns, R.C., Curtmola, R., Herring, J., Kissner, L., Peterson, Z.N.J., Song, D.X., 2007. Provable data possession at untrusted stores. In: Proceedings of the 2007 ACM Conference on Computer and Communications Security, CCS 2007, pp. 598–609.

5. Ateniese, G., Pietro, R.D., Mancini, L.V., Tsudik, G., 2008. Scalable and efficient provable data possession. In: Proceedings of the 4th International Conference on Security and Privacy in Communication Networks, SecureComm, and pp. 1–10.
6. Barreto, P.S.L.M., Galbraith, S.D., O'Eigeartaigh, C., Scott, M., 2007. Efficient pairing computation on supersingular abelian varieties. Des. Codes Cryptogr. 42 (3), 239–271.
7. Beuchat, J.-L., Brisebarre, N., Detrey, J., Okamoto, E., 2007. Arithmetic operators for pairing-based cryptography. In: Cryptographic Hardware and Embedded Systems – CHES 2007, 9th International Workshop, and pp. 239–255.
8. Boneh, D., Boyen, X., Shacham, H., 2004. Short group signatures. In: In Proceedings of CRYPTO 04, LNCS Series. Springer-Verlag, pp. 41–55.Boneh, D., Franklin, M., 2001. Identity-based encryption from the Weil pairing. In: Advances in Cryptology (CRYPTO'2001). Vol. 2139 of LNCS, pp. 213–229.
9. Bowers, K.D., Juels, A., Oprea, A., 2009. Hail: a high-availability and integrity layer for cloud storage. In: 28.ACM Conference on Computer and Communications Security, pp. 187–198.
10. Breidert, C. Estimation of Willingness-to-Pay – Theory, Measurement, Application. Wiesbaden: DUV, 2006.
11. Briscoe, G., and A. Marinos. "Digital Ecosystems in the Clouds: Towards Community Cloud Computing." Digital Ecosystems and Technologies Conference. IEEE Press, 2009.
12. Buyya, R., CS. Yeo, S. Venugopal, J. Broberg, and I. Brandic. "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility." Future Generation Computer Systems 25, no. 6 (2009): 599-616.
13. Cramer, R., Damgård, I., MacKenzie, P.D., 2000. Efficient zero-knowledge proofs of knowledge without intractability assumptions. In: Public Key Cryptography, pp. 354 373.