ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

Digitally Signed Token Based Data Integrity Checking and Auditing of Cloud Data Using Inter-relation Mechanism

Balamurugan, B., P. Venkata Krishna, Geetha Priya, P., Tamilarasi, T., Silambarasi, J.

School of Information Technology VIT University.

ARTICLE INFO

Article history:
Received 12 November 2014
Received in revised form 26 December 2014
Accepted 29 January 2015
Available online 10 February 2015

Keywords:
Data Storage, Token mechanism, Diffie-Hellman algorithm, Digital Signature, Inter-relation mechanism.

ABSTRACT

Eminent applications are moving towards cloud for data storage. But issues surrounding data storage on cloud are more complicated. When the data is stored in cloud, the user sees a virtual server; it appears as if the data is stored in particular place with specific name. In fact it is imaginary; the cloud manages the storage space dynamically. In this paper we focus on verifying the integrity of the data stored in cloud by means of imposing challenges to the cloud storage provider. And also to audit that the storage provider has not deleted or modified the third party files accidently or intentionally, the user inter-relates the blocks such that identifying the server's misbehaviour is with higher probability. Additionally to overcome the communication and computation overhead, the interactions are done through tokens. All the tokens are digitally signed with the help of Diffie Hellman Key Exchange algorithm which surmounts man in the middle attack.

To Cite This Article: Balamurugan, B., P. Venkata Krishna, Geetha Priya, P., Tamilarasi, T., Silambarasi, J., Digitally Signed Token Based Data Integrity Checking and Auditing of Cloud Data Using Inter-relation Mechanism. ***Aust. J. Basic & Appl. Sci.,*** *9(6): 34-37, 2015*

INTRODUCTION

With increase in computing capabilities and technology there comes a challenge which affects the service quality of the cloud storage provider. A recent study by citric showed that two third of the UK companies where computing in cloud and one third stated that they had concerns over data storage in cloud. Fear of loss of control over user's data is the major challenge that prevents the end user from migrating to the cloud service. If data owners can verify the integrity of data while storing and also after some months or years, the service becomes trustworthy. Data protection Act of 2003 states that there needs a security measure to make sure that the data does not go into others hand. So, there emerges a method to partition the data into blocks and store it in the distributed database. In order to solve the problems in checking the data integrity, several schemes have been proposed. The latest research states that the possession checking protocol must satisfy at least the five requirements.

2. Related work:

Juels and Kaliski (2007) proposed proof of retrievability protocol where sentinels are used to store and access. Sebe *et al* (2008) gave a protocol based on the Diffie Hellman Algorithm. The advantage of using this is unlimited number of verification of data possession can be made. But this protocol requires the cloud storage provider to access the entire data blocks. Tribhwan *et al* (2010) state that their scheme reduces the communication and storage overhead but the drawback is the user does the job of pre-computation of token which makes the computational complexity greater. Chen *et al* (2011) gave PDP scheme was based on homomorphic hashing but the drawback lies in the data dynamics. Guiseppe Ateniese *et al* proposed PDP(Provable Data Possession) the advantage of this model is, only small portions of blocks in the data are accessed. But if the blocks that are deleted are not queried, the servers will never let the client that a particular part of the data is deleted.

3. Proposed work:

Outsourcing the cloud data brings lots of challenges to the security, so it becomes mandatory to protect the users data in all possible means Generally performance is calculated by considering the computational complexity, block access complexity, communication complexity.

We propose an integrity checking and auditing model using digitally signed token with the help of Diffie Hellman Key algorithm as in Fig.1.Even generating the challenges for random set of blocks

Corresponding Author: Balamurugan, B., School of Information Technology VIT University.

will give less probability of owning the data. So to overcome this, to the Diffie Hellman Algorithm we add a mechanism which inter-relates all blocks in a random set such that by querying one block it's associated are queried for checking integrity. Assume that the details about the inter-relation are not known to the storage provider. Notations used to explain the concept are given in Table.1.

Table 1: Notations used in this paper.

S.No	Notations	Description
1.	$\tau_S\ \tau_A\ \tau_C\ \tau_R$	Authorisation ,Acknowledgment , Challenge, Response token
2	DID,DOID/IP	Data ID, Data owner ID/IP
3.	FID,FN,FS	File ID, File Name, File size
4.	Bit D,F,S	Delay,Failed,Success

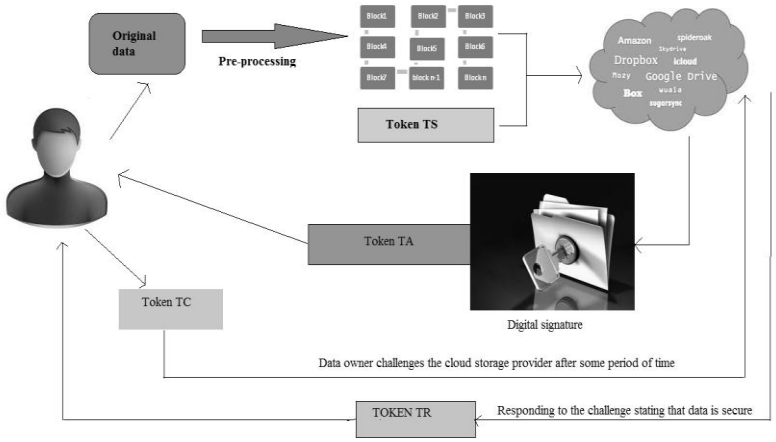

Fig. 1: Overall structure of the process.

Registration phase:

Data owner requests the storage provider for service. The storage provider generates unique token TS for individual user and sends it to the user. TOKEN TS contains the Data owners ID and Data ID and Validity period.

$$DE:(\mathrm{do})_k \xrightarrow{request} (\mathrm{cs})_l$$

$$DE:(\mathrm{do})_k \xleftarrow[(\tau_S)_m]{} (\mathrm{cs})_l\ ;m\in\mathbb{Z}^+$$

TS= (DOID || DID | VP)

Pre-Processing phase:

When the data owner is ready to store large amount of data, the data is pre-processed to make it into blocks and also they are inter-related with each other. Meta data of each block is stored locally to verify at the time of arrival of response.

Let dbl be the block of data from the Large data ,

$$(\mathrm{dbl})_o \in \mathrm{Data}\ ;\text{for each}\ ;1\le \mathrm{dbl}\le 1024\mathrm{kb}\ \ o\le \frac{\mathrm{Data}}{1024k\mathrm{b}}$$

And for every dbl there exists Meta data with which the response is going to be verified in future and each metadata and dbl is indexed with r and o to differentiate from the large chunk.

Let us assume CS be the cloud storage provider such that, $(\mathrm{CS})_i\ ;i\in\mathbb{Z}^+$

Let DO be the Data owner where, $(\mathrm{DO})_j\ ;\ j\in\mathbb{Z}^+$

Let DE be the Data exchange between CS and DO such that $DE:(\mathrm{do})_k\rightarrow(\mathrm{cs})_l$ where

$(\mathrm{do})_k\in(\mathrm{DO})_j\ ;1\le k\le j$

$(\mathrm{cs})_l\in(\mathrm{CS})_i\ ;1\le l\le i$

$(\mathrm{md})_r\in(\mathrm{dbl})_o\ ;1\le r\le o$

$(\mathrm{md})_r\in(\mathrm{do})_k\ ;1\le r\le o$

Data blocks are inter-related in such way that the same data block is not inter related with itself i.e.

$(\mathrm{dbl})_o=(\mathrm{dbl})_p\ \Lambda\ (\mathrm{dbl})q\ ;$

for every $p\neq q$ and $1\le p\le 0, 1\le q\le 0\ ;$

where Λ is interrelation for every $p\neq q$ and $1\le p\le 0, 1\le q\le 0\ ;$

Storing phase:

Data owner sends the blocks of data with Token TS to the storage provider and it checks the validity of the Token TS using its credentials and sends acknowledgement Token TA to the data owner. If the storage provider is not getting the data within validity period mentioned in Token TS server sets the bit Delay or Fail respectively and sends it to the data owner else it sends the Token TA by setting the Bit S.

$$DE:(\mathrm{dbl})_o+(\tau_S)_m \xrightarrow{send\ data} (cs)_l$$

After receiving the Acknowledgement Token TA the data owner deletes the original file and keeps only the meta-data to verify the responses given by the storage provider.TA= (SIP || DOIP || FID || FN || FS ||

ST || D || F || S). Then the Data owner sends the data block along with the Token TS to the CS

$$DE:(\mathrm{do})_k \xleftarrow[(\tau_A)_m]{} (cs)_l ; m \in \mathbb{Z}^+$$

Where m is the count given for the Acknowledgement token as for every block of data it stores it gets the acknowledgement.

Data auditing phase:

If the data owner wants to check the integrity of data after a period of time he has to produce a challenge to certain blocks with token TC.Thus deleting any part of data reveals the user about the server's malfunctioning as it is inter-related.

TC= (SID || DOIP || FID || BID || FN || FS ||ST).

$$DE:(\mathrm{do})_k \xrightarrow{DHchallenge(\tau_C)} (cs)_l$$

The data owner stores the following pre-computed value.

$M=x^m \bmod y \rightarrow$ (1)

x,y large prime numbers known to both owner and storage provider,m-index of meta data.The data owner chooses a random value r and sends the following value A as a challenge to the storage provider with the file to be verified.

$A=x^r \bmod y \rightarrow$(2)

r-randomly generated(secret key).

Response phase:

Server computes **$B=A^m \bmod y \rightarrow$(3)** and sends B to the data owner along with the Token TR by setting the bits in the Token TR bitD is set by computing the value fromaverage receiving time and start time ST from token TC.bitF is set by checking the specified file size in TC and the received TC ,bitS set only if all the received credentials are correct.

$$DE:(\mathrm{do})_k \xleftarrow{Respond(\tau_R)} (cs)_l$$

Verification phase:

Data owner checks the TOKEN TR for serverIP with the one given while registration,bitD,bitF,bitS.The verifier i.e, Data owner computes

$C=M^r \bmod y \rightarrow$ (4)

And verifies the equations (3) & (4) ,If they are equal File integrity is Preserved as [$B=A^m \bmod y = A^m \bmod y = x^r \bmod y = C$] .

5. Validation:

Take a scenario where a storage server has deleted or modified some 20 blocks out of 1000 blocks. Let us assume first 20 blocks in the sample 1 are deleted or modified by the server.

Table 2: Comparison of probability of normal data block and inter-related data block.

SNO	No of deleted or modified blocks	Normal data block $*10^{-4}$	Inter-related data block$*10^{-3}$
1.	20	9.811	9.811
2.	40	9.617	9.617
3.	45	9.569	9.569
4.	60	9.426	9.426
5.	75	9.286	9.286
6.	80	9.240	9.240
7.	100	9.056	9.056
8.	120	9.010	9.010
9.	125	8.833	8.833

X=20, p=1/1000=0.001Using the geometric distribution $g(20;0.001)=(0.001)(1-0.001)^{20-1}=0.0009811$[Normal model of checking integrity].

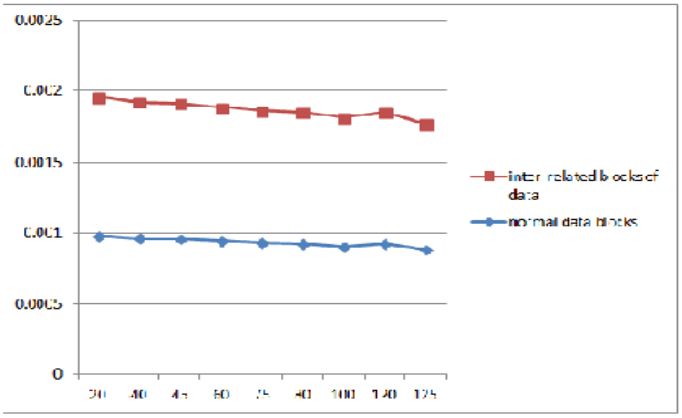

Fig. 2: Probability comparison Normal data block with inter-related data block.

So when blocks are inter-related with each other by 10, one query can test 10 blocks that are present/absent/modified. So 0.0009811*10 =0.00981 is the final probability obtained from the inter-related model. Therefore increase in the probability of owning the data with less number of queries is obtained when compared to normal model can be seen mathematically in Table.2 and graphically in Fig.2.

6.Conclusion:

Recently,outsourcing of the data has brought many challenges to cloud.So in order to protect the data inside the cloud and from the outside attackers,we proposed a model which helps the data owners to verify the integrity of their data with better time and computational complexity.

REFERENCES

Book of Cloud Computing principles and paradigms –Wiley publications, 2009. Rajkumar Buyya, James Broberg (University of Melbourne), Andrzej Goscinski (Deakin University).

Chen Lanxiang, 2011. A homomorphic hashing based provable data possession. Electronic Information Technology.

Data Integrity Proofs in Cloud Storage 2011 Communication System and Networks(COMSNETS) IEEE Sravan Kumar R and Ashutos Saxen Software Engineering and Technology labs Infosys Technologies Ltd Hyderabad, India.

Ensuring Data Storage Security in cloud Computing Through Two-Way Handshake based on Token Management M.R Tribhuwan,V.A.Bhuyar,Shabana Pirzade,2010 International conference on Advances in recent technologies in communication and computing.(ARTCom).

Integrity and Internal Control in Information Systems VI IFIP TC11 / WG11.5 Sixth Working Conference on Integrity and Internal Control in Information Systems (IICIS) 13–14 November 2003, Lausanne, Switzerland.

Juels, A., B.S. Kaliski, 2007. Pors: proofs of retrievability for large files. In: Proc of ACM-CCS ’07.

Legal Issues surrounding Data Storage on the Cloud A Fitzpatrick, M McGrath and R G Lennon Computing Department Letter Kenny Institute of Technology.K. Elissa. Oct 2012 Published in: Tier 2 Federation Grid,Cloud & High Performance Computing.

Partition-Based Cloud Data Storage and Processing Model Proceedings of IEEE CCIS2012(Cloud Computing and Intelligence System)Yawei Zhao, Yong Wang. Oct 30 2012-Nov 1 2012.(Volume 01).

Provable Data Possession at Untrusted Stores by Giuseppe Ateniese, Randal Burns, Reza Curtmola ,Joseph Herring, Lea Kissner , Zachary Peterson, Dawn Song Published in Proceedings of the 14th ACM Conference on Computer and Communication Security.

Sebe, F., J.F. Domingo, A.B. Martinez, Y. Deswarte, J. Quisquater, 2008. Efficient remote data possession checking in critical information infrastructures. IEEE Trans Knowledge Data Eng.

Towards a Theory of Data Entanglement by James Aspnes,Joan Feigenbaum, Aleksandra Yampolskiy , and Sheng Zhong ,Department of Computer Science, Yale University.