Volume 5, Issue 11, November 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Data Integrity Check and Efficient Data Storage in Cloud Using Hashfunctions, Blowfish and RSA

A. Sirisha, N. Hiranmayee
Computer Science and Engineering &
JNTUK, India

Abstract— The concept of cloud computing is currently being widely adopted by many business and organizations. Cloud computing is the way of providing computing resources in the form of services over internet.The cloud computing allows storing the user's data and to measure the applications and services provided by cloud server.There is an ample data stored at cloud storage server.Security is one of the major issues which reduces the growth of cloud computing. So Cloud computing entails encyclopedic security solutions.This paper presents secure file exchanging on Cloud using Blowfish, RSA and Hash algorithms which are capable of solving data security, authen cation ,and integrity problems of files on the cloud. This paper also points out how third party auditors can be avoided and proposes a specific solution which involves the customer safeguarding the data integrity by himself in a very simple and efficient way by utilizing existing hash generating algorithm .Data security is improved by cryptography algorithms. In our enhanced system we integrate symmetric, asymmetric and Hash algorithms which provide better results for performance parameters. The data owner has to match the Hash code for the integrity of user data in cloud. Data Owner will get a message when the data integrity is lost.

Keywords— cloud, hash, dataintegrity, SaaS, PaaS, IaaS, cloud computing

I. INTRODUCTION

Cloud computing is a new computing paradigm[7], where various computing resources are provided as a service to the end user. Cloud computing as a technology[1] is becoming more popular, both in the academic world and in the IT industry. One of the paramount advantages of Cloud Computing is on-demand self-service, providing dynamic scalability of the infrastructure and services.Cloud computing is a widely used computing technique which provides three main models SaaS(Software/Storage as a service),PaaS(Platform as a Service)and IaaS(Infrastructure as a Service)[6].These models work on apay-per-use system and provide users the ability to access their database and applications remotely.Very common examples are, Amazon's EC2,Microsoft Azure and Sales force CRM.The typical services provided by cloud providers include resources like data storage or software and hardware services.The data owners can reduce the operational cost of installing and maintaining their own new software or hardware by moving their business applications and services into the clouds.

There are different ways to deliver cloud services(Figure1) as described in the following [2][3].At the lowest level there is the possibility to run virtual machines on cloud.

The infrastructure of a Cloud Service Provider(CSP).This is called *Infrastructure as a Service*(IaaS).One level higher there is the possibility to develop and deploy applications on the infrastructure of a CSP.This is called *Platform as a Service*(PaaS).On the highest level there are standardized applications which are delivered as a service.This is called *Software as a Service*(SaaS).Software/Storage as a service(SaaS)-allows users to access software and applications hosted as a service.These services are deployed on clouds by cloud providers and can be accessed remotely across the internet.These services are available to use on pay per month method or pay per use method.

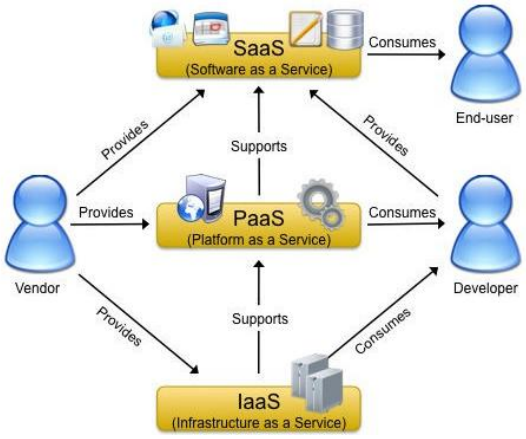

Fig. 1 Service based models of cloud computing[5]

Platform as a service(PaaS)-allows user to use the infrastructure required for an application. Platform as a Service(PaaS)is a cloud delivery model which tries to make the development and deployment of application sales complex and expensive task[2][4].Platform as a Service(PaaS)is mainly designed for the developers to develop their own applications and to deploy it on PaaS environment.

Infrastructure as a service(IaaS)-allows users to use leased infrastructures and use it as if they are using their own hardware and software.Sometimes organizations have to buy infrastructure which they do not use frequently but they still have to have them.Infrastructure as a Service(IaaS) providers give them a solution and save them money by letting them rent their infrastructures.IaaS providers offer users their own separate instance of server, which is also called virtualization. A client can deploy own or vendor supplied virtual machines to run software in the cloud, and pays for the resources (CPU time, memory,storage and network usage) it consumes.

When using cloud services, businesses and organizations will have to trust a third party or technology for safe keeping their data. But the main problem lies in the fact that the data owner cannot always trust the cloud service provider due to the fact that cloud is located outside of the data owner's trusted domain. This factal one creates several potential problems like sensitive data leakage or compromising integrity of data located in the clouds. The data stored in clouds are considered to be very confidential and sensitive by businesses and organizations and must not be disclosed to unauthorized third party. There are several methods that could be used to protect the sensitive data from unauthorized access. Usually, cloud service providers encrypt customer data before it is stored in the cloud[5].But even the best encryption techniques can be compromised. Therefore, there is a need for a process that will check data integrity inorder to make sure that no data modification has occurred. This will raise the confidence of cloud customers that their data is in safe and protected environment.

In order to establish trust between the customer and cloud service provider,a third party auditors should be involved in order to check integrity of customer's data.It is common that third party auditors maintains a challenge/response pair for the data stored in the cloud. From time to time, it checks the data with the challenge/response method for the integrity of data. But in this scenario,it is assumed that the customer should trust third party auditor as well. Therefore, from the customer point of view third party auditors are similar to a cloud provider who has all the access to the customer's private data. Instead of helping the customer to maintain trust with the cloud provider , third party auditor itself could be the weakest link in this security chain and also could be the source of data integrity loss. It would be better for security reasons if this intermediate link (e.g., third party auditor) for integrity check can be avoided and the integrity check can be done by the customers themselves.

The solution provided in this paper, uses a hash algorithm utility which is implemented on the customer's side. The customer pre-computes the hash of the data content ,stores the hash at the local secure hash repository and then sends the file to the cloud server for storage. When the customer wishes to verify the integrity of a data file, the customer takes that data contents and computes the hash and matches it with the pre-computed hash value. If there is any change in the data content ,the two hash values will not match and the customer will know that the data integrity has been compromised. This method is less complicated than the third party auditor scheme and provides a simple way to check data integrity.

To ensure the correctness of users data in the cloud ,we propose an effective mechanism with salient features of data integrity ,confidentiality and authentication. This mechanism uses the concept of Blowfish ,RSA along with Hash function. Hash function improves the file integrity.We compute Hash code at client side and compare this hash code value with hash code value sent as a file to the cloud server.In this scheme encryption is used to provide security to the data while in transmit..Because the encrypted file is stored on the cloud, so user can believe tha this /her data is secure. Only in encrypted form of files transferred over the channel, which reduces the problem of information disclosure.Our enhanced system (Blowfish+RSA+Hashvalue) compares with simple RSA and Blowfish on basis of some performance parameters like:-throughput, encryption time,cipher text and delay time.

This paper mainly concern with the introduction of cloud computing, security issues related to cloud and some basics techniques available for integrity of data in cloud server.First section covers the introduction of cloud computing and security issues related to cloud.Second section of paper covers some related work.Third section consists introduction of Blowfish ,RSA and SHA algorithms.Fourth section covers the proposed methodology.Last section of paper we conclude the better security algorithms for integrity in cloud.

II. RELATED WORK

A. Recent attempts to address cloud security

Many researches have been done so far that have covered the aspects of security in cloud computing and many researches are still on going. Data integrity check by a third party auditing services is one of the newest topic in cloud computing research area.A customer always has reservations in trusting third party cloud service provider. One of the most basic question that they can ask themselves is whether they can integrate a third party auditing service in the existing system to check the integrity of the stored data in a cloud.

B. Public Auditing and data privacy preservation

In order to preserve integrity and privacy of the data stored in the cloud the scheme of the external audit service that checks the integrity of the data stored in a cloud can be employed[8].In this method the public key based authentication is employed and is positioned together with random masking technique to achieve their goal of efficient and privacy preserving auditing. This scheme guarantees that no data is stored locally for third party auditors and it does not create an

extra overhead for the customer.Also,it is claimed that after integration ,there will not be any further weaknesses in the existing security system.This makes the privacy preservation method a very efficient ,secure and of high performance.

C. Digital Content Extraction and Privacy Preserving Audit

This scheme introduces a third party auditor which uses extraction protocol to ensure the integrity of customer data[9].The technique mentioned here does not require from the customer to encrypt the data using some symmetric keys. This is because the keys can be lost over the time from the customer itself and the data is prone to get leaked. One of the big advantages of this solution to privacy preserving of data is that there is no need for customer to generate any secret keys or hash the data or encrypt the data.The customer can just call for the data and retrieve it as and when required.

III. BLOWFISH, RSA, SHA

Blowfish is an encryption algorithm[14] that can be used as a replacement for the DES or IDEA algorithms. It is a symmetric block cipher that uses a variable length key, from 32 bits to 448 bits, making it useful for both domestic and exportable use. Blowfish cryptographic algorithm [15],which was designed by Bruce Schneier in 1993,is a symmetric block cipher that divides a message up into fixed length blocks of 64 bit during encryption and decryption processes. The Blowfish algorithm consists of two parts[18]: a key expansion part and a data encryption part. Key expansion converts a variable length key of at most 448 bits into several subkey arrays,totaling 4168 bytes.The algorithm uses Feistel cipher where the input text is split into two halves. The first half is applied round function using a subkey. The output will be XORed with the second half. Then the two halves will be swapped.In total there are 17 rounds and each round consists of a key dependent permutation and a key and data dependent substitution.

The Feistel network of Blowfish algorithm is one that utilizes a structure that makes encryption and decryption very similar through the use of the following elements [16, 17]:

- P box: Permutation box that performs bit shuffling;
- S box: Substitution box for nonlinear functions;
- XOR: Logic function to achieve linear mixing;

Figure[3] shows a graphical representation of the F function,which has been shown as the most accessed function of the Blowfish algorithm. It requires a 32 bit input data to be decomposed into four 8 bit blocks.

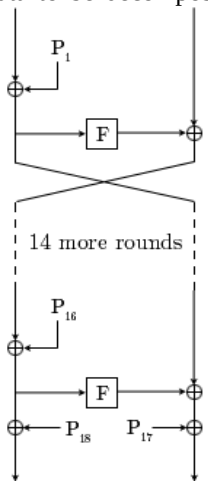

Figure 2: Representation of Blowfish Cryptographic Algorithm

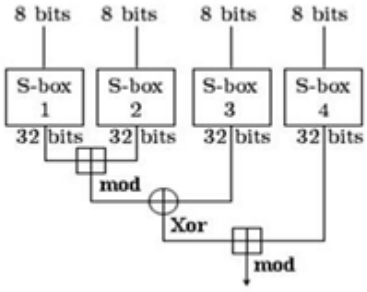

Figure 3: Representation of F function

Each block references an S Box and each entry of the S Box outputs a 32 bit data. First, the output of S Box 1 and S Box 2 are added.Then the result of the addition is XORed with SBox 3. Finally, S Box 4 is then added to the output of the XORed operation and provides a 32 bit output.Blowfish is suitable for applications where the key does not change often, like a communications link or an automatic file encrypter. It is significantly faster than most encryption algorithms when implemented on 32 bit microprocessors with large data caches.

RSA**(**Rivest Shammir Adleman) is public key cryptography algorithm involves two different keys. Public key for encryption and private key for decryption.RSA also provide authentication [18]. RSA Algorithm RSA is a commonly adopted public key cryptography algorithm. RSA can be used for key exchange, digital signatures, or encryption of small blocks of data. RSA uses a variable size encryption block and a variable size key.RSA has been widely used for establishing secure communication channels and for authentication and the identity of service provider over insecure communication medium. file. The RSA algorithm involves three steps:

A. Key generation,
B. Encryption and
C. Decryption.

A. Key generation

RSA involves a public key and a private key.The public key can be known by everyone and is used for encrypting messages. Messages encrypted with the public key can only be decrypted in a reasonable amount of time using the private key.

B. Encryption

Alice transmits her public key (n, e)to Bob and keeps the private key secret. Bob then wishes to send message M to Alice. He then computes the ciphertext c corresponding to : $C \equiv m^e \bmod(n)$

Bob then transmits c to Alice.

C. Decryption

Alice can recover m from c by using her private key exponent d via computing:

$m \equiv c^d \pmod n$

Given m, she can recover the original message M by reversing the padding scheme.

SHA is hash algorithm in which n-bit hash produces a n-bit length finger print from the arbitrary length data[13].SHA- 1,SHA-256 ,SHA-512 produces message digest 160,256 and 512 respectively[18]. The Secure Hash Algorithm is a family of cryptographic hash functions published by the National Institute of Standards and Technology (NIST) as a U.S. Federal Information Processing Standard (FIPS), including:

- *SHA-0*: The original version of the 160-bit hash function published in 1993 under the name "SHA".
- *SHA-1*: A 160-bit hash function which resembles the earlier MD5 algorithm. This was designed by the National Security Agency (NSA) to be part of the Digital Signature Algorithm.
- *SHA-2:* A family of two similar hash functions, with different block sizes, known as *SHA-256* and *SHA-512*. They differ in the word size; SHA-256 uses 32-bit words where SHA-512 uses 64-bit words.
- *SHA-3:* It supports the same hash lengths as SHA-2, and its internal structure differs significantly from the rest of the SHA family.

IV. PROPOSED WORK

The idea of relying on third party auditors for monitoring integrity of the data stored in cloud does not eliminate the "trust problem." In many ways, a third party auditor (TPA) cannot be considered reliable. In order to get the integrity checked by auditors, customers have to reveal the data and the key used to decrypt it must be shared to TPA. However, all these third party auditor schemes cannot assure that the data owner's data is not revealed to the auditors.Using third party auditors makes the whole system even more complicated. Now the dataowner has to deal with the cloud server provider(CSP) as well as the auditor. Communication between dataowner-auditor and auditor-cloud service provider will require more bandwidth and hence creates overhead.

In order to avoid third party auditors , this paper proposes that the integrity check of data stored in cloud can be checked at customer's side i.e. at data owner side. This integrity check can be done by using cryptographic hash functions and algoritms with out involving any third party auditor. In the following is presented and describe our proposed method for data integrity check using hash function.

A. Integrity Check using Hash Function

The trust problem between TPA and Data Owner can be solved, if users can check the integrity of data themselves instead of renting an auditing service to do the same. This can be achieved by hashing the data on user's side and storing the hash values in the local secure hash repository. The below Figure 2 presents the overview of the scheme.

This idea is based on the three properties of a hash function which eliminates the clash between two hash values and makes it possible to check integrity of data using hash. Based on our proposal, the data owner first pre-computes the hash value of file, then sends the file to the cloud server provider(CSP) and the computed hash value is stored in the local secure repository. Whenever the data owner wants to check integrity of the data, they retrieve the file from cloud and computes hash value of the file again and matches it with the pre-computed hash values stored at local hash repository. Since the hash value of a message is considered as its digital fingerprint, any changes in the original message will reflect in the result of its hash value. If the re-computed hash value matches with the pre-computed hash value then the file is intact and if it does not then the file was tampered and its integrity compromised.

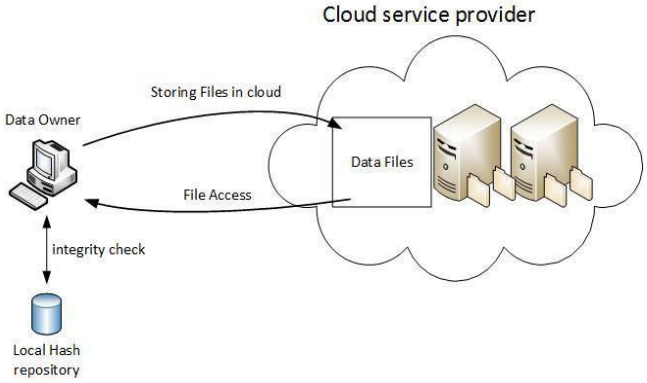

Fig. 2 Data integrity check using hash functions

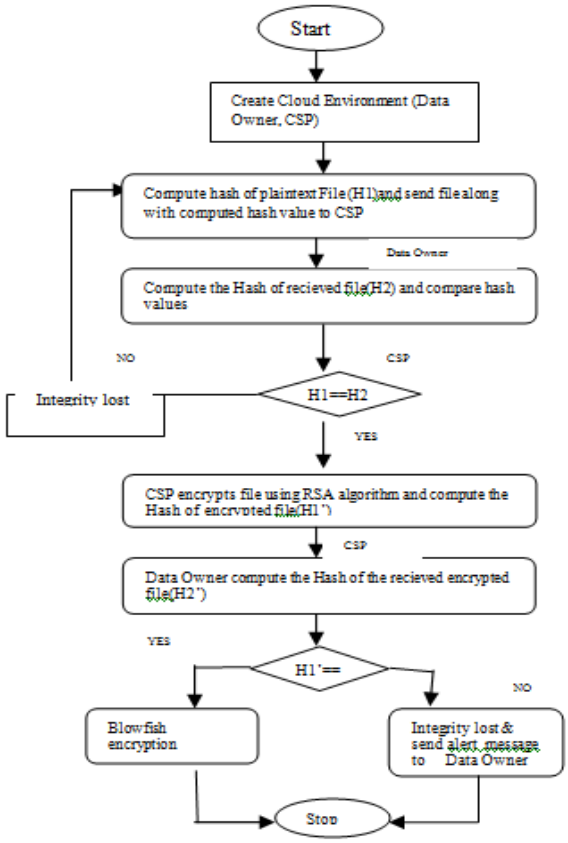

Fig. 1 Example of an image with acceptable resolution

V. CONCLUSIONS

Security of data and trust problem has always been a primary and challenging issue in cloud computing. This paper attempts to point out advantages and security concerns of cloud computing and focuses on avoiding third party auditors for data integrity check. Implementation of proposed utility, which computes hash values of files at the data owner side, can eliminate the need of third party auditors. The resultant hash values from this utility is stored at secure local hash repository. The data file can be retrieved back whenever needed and checked for any arguments among parties involved by re-computing and matching the hash result with the pre-computed hash value.

This idea can be very effective on a small scale where data owner initially want to test the cloud provider and want to establish trust and supplement already existing SLA[11][12]. This idea offers efficiency by minimizing human factor in data integrity checks and replacing them with technological solution which in terms will save money. In aspects of security, a lot more needs to be done to make cloud computing a prominent and reliable platform. The effort made in this paper is very basic and easy to use by dataowner.The proposed system along with integrity also provides authentication and confidentiality to the data stored in cloud computing.

REFERENCES

[1] MichaelArmbrust, ArmandoFox, ReanGriffith, AnthonyD.Joseph, RandyH.Katz, AndrewKonwinski, GunhoLee, DavidA.Patterson, ArielRabkin, IonStoica and MateiZaharia. Above the Clouds :Aberkeley View of Cloud Computing. Technical Report UCB/EECS-2009-28, EECS Department, University of California, Berkeley, Feb 2009.

[2] NIST, "SP800145:The NIST Definition of Cloud Computing. "[Online]. Available:http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf. [Accessed:21-Jan-2013].

[3] L.M.Vaquero, L.Roderomerino, J.Caceres, andM.Lindner, "A break in the Clouds: Towards a Cloud Definition, "Computer Communication Review, vol.39, no.1, pp.50–55, 2009.

[4] CloudDeploymentModels–Private, Community, Public, Hybrid with Examples-By Basant Narayan Singh,October1, 2011.http://www.techno-pulse.com/2011/10/cloud-deployment-private-public-example.html accessed 20 December 2012.

[5] BSI (2008) BSISO/IEC 27005:2008:Information Technology. Security Techniques. Information Security Risk Management. British Standards Institution.

[6] Cloud Computing Security: How Secure is the Cloud? http:// www. www.researchomatic.com/essay/Cloud-Computing-Security-How SecureIsTheCloud92640.aspx accessed: 19/10/2013

[7] CHELLAPA, R. (1997) Intermediaries in Cloud Computing:A new Computing Paradigm. Cluster: Electronic Commerce.

[8] MehulA. Shah, Ram Swaminathan, MaryBaker, "Privacy Preserving Audit and Extraction of Digital Contents, "IACR, 2008, http://eprint.iacr.org/20 08/186.pdf

[9] CATTEDDU, D. & HOGBEN, G. (2009): Cloud Computing: Benefits, risks and recommendations for information security; European Network and Information Security Agency (ENISA).

[10] Ms. Payal P. Kilor and Prof. Vijay B. Gadicha (2014)― DataIntegrity Proofs in Document Management System under Cloud with Multiple Storage‖, International Journal of Engineering& Computer Science, vol. 3

[11] .Cecinio Silva Lacerda, "Service-level agreement (SLA), Searchi T Channel, http://searchitchannel.techtarget.com/definition/service-level- agreement accessed: 24/08/2013

[12] Service Level Agreement in the Data Center, Sun, 2011, http ://www. sun. com/ blueprints/ 0402/ sla. pdf.

[13] Guo, S. Ling, C. Rechberger, and H. Wang, ―Advanced Meetin-the-Middle Preimage Attacks: First Results on Full Tiger, and Improved Results on MD4 and SHA-2,‖ pp. 1–20..

[14] Govinda.K1 Mythili and Geetha Priya(2014),‖ Data Security in Cloud using Blowfish Algorithm‖, International Journal for Scientific Research & Development| Vol. 2, Issue 09.

[15] Gurpreet Kaur and Manish Mahajan (2013), ―Analyzing Data Security for Cloud Computing Using Cryptography Algorithms‖, International Journal Of Engineering Research and Application, Vol.-3,782-786.

[16] Cody, Brian; Madigan, Justin; MacDonald, Spencer; Hsu, Kenneth W.;, "High speed SOC design for blowfish cryptographic algorithm," Very Large Scale Integration, 2007. VLSI SoC 2007. IFIP International Conference on , vol., no., pp.284-287, 15-17 Oct. 2007.

[17] S.Subashini and V.Kavitha(2011), ―A Survey on security issues in service delivery models of cloud computing". Journal of Network andComputer Applications 34,1-11.

[18] Kamak Ebadi,Victor Pena etc.‖High performance implementation and Evaluation of Blowfish Cryptographic Algorithm on Single-Chip Cloud Computer:A Pipelined Approach ‖.

