International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
1st International Conference on Advent Trends in Engineering, Science and Technology
"ICATEST 2015", 08 March 2015

An Approach for Data Integrity in the Cloud Computing: A Survey

Tejas J. Pateliya[1], Manoj S. Chaudhari[2]

[1] *Research Scholar, Department of Computer Science, Priyadarshini Bhagwati College of Engineering, Nagpur, Maharashtra, ptejas.patelia@gmail.com*
[2] *Head of Department, Department of Computer Science, Priyadarshini Bhagwati College of Engineering, Nagpur, Maharashtra, India*

Abstract: Cloud based computing services is among the fastest growing sections of the IT industry .The Cloud computing is a resources and Internet based service providing technology which is used to store, share, retrieve and develop different kinds of applications. It also makes variety of hardware resources available to the users. Due to features such as attractive cost savings for buyers, accessibility & reliability options for users and scalable sales for cloud service providers it is readily being accepted worldwide. This great boom in cloud computing bring a lot of security issues with it. This study brings forth various security related aspects of cloud. Our study also allows researchers, vendors and users to know the current advancements in this filed.

Keywords – Cloud Computing, Cloud Computing Security, Security Issues, Survey Of Cloud Computing

A.
B.

1. INTRODUCTION

C.

"Cloud computing" is nothing but an Internet based service providing facility. Cloud storage moves the user's data to large data centers, which are remotely located, on which user does not have any control. However, this unique feature of the cloud poses many new security challenges which need to be clearly understood and resolved. The benefits of cloud computing are significant—economies of scale, potential cost savings, fast deployment and easy scalability. Cloud computing is a technology which provide you a service through which you can use all the computer hardware and software from your

D.

desktop, or somewhere inside your company's network but they are not actually installed on your computer, it is provided for you as a service by another company and accessed over the internet. [1][2]

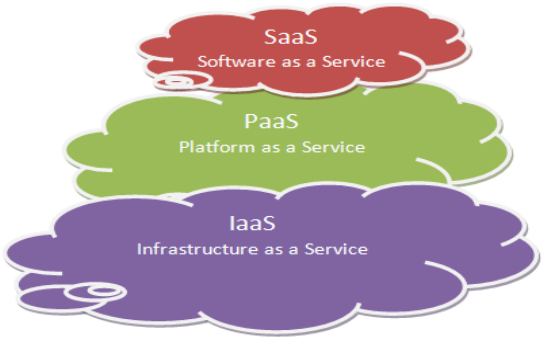
1.1 Cloud Computing Service Model

In the cloud computing, the available service models are (figure 1):

A. Infrastructure as a Service (IaaS): This model allows user to rent processing, storage, network and other resources. The user can deploy and run text the guest OS and application.[3] The user does not manage of control the underlying cloud infrastructure but has control over only OS, storage and deployed application and possibly selected networking components. Examples include Amazon Elastic Computer Cloud (EC2), Microsoft Windows Azure.

B. Platform as a Service (PaaS): Provides the consumer with the capability to deploy onto the cloud infrastructure, consumer created or acquired applications, produced using programming languages and tools supported by the provider [3]. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations. Examples include Google's Apps Engine, Microsoft- Windows Live. Open shift, etc.

C. Software as a Service (SaaS): Provide the user with the capability to use the provider's applications running on a cloud infrastructure. The applications are accessible from various consumer devices, through a thin client interface, such as a web browser. [3] The consumer does not manage or control the underlying cloud infrastructure various resources. Examples include Salesforce.com, VoIP from Skype and Vonage, Google's Gmail and Apps, instant messaging from Yahoo and AOL.



A. Fig. 1: Service Model of Cloud computing

1.2 Deployment Models Of Cloud Computing

Fig. 2 shows the types of cloud deployment models. These are describing in the following section.

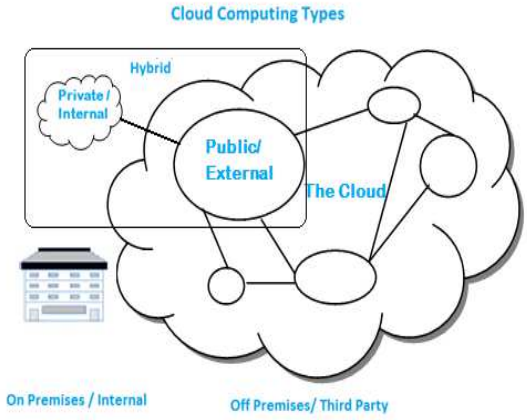


Fig. 2: Deployment model of cloud computing

A. Public Cloud

In a public cloud the computing infrastructure is used by the organization or end user through cloud service providers or vendors. Public clouds are typically offered through virtualization and distributed among various physical machines. [4][5] [6]

B. Private Cloud

In a private cloud the computing infrastructure is dedicated to the particular organizations and not shared with other organization. Private clouds are more secure than public clouds. [5][6]

C. Hybrid Cloud

This is a combination of the other two types of cloud. In hybrid cloud organizations may host critical application on private clouds and applications which are having less security concerns hosted on public clouds. It is also known as cloud bursting. [5][6]

D. Community Cloud

It involves sharing of computing infrastructure in between organizations of the same community. For example all Government organizations within the state of Maharashtra may share computing infrastructure on the cloud to manage data related to the citizens residing in Maharashtra. [5][6]

2. LITERATURE REVIEW

Ryan K L et al [7] gave a new security proposal that focused on the file centric logging for access control and file transfer. In this paper they have identified and tried to create a system which provides a mechanism that creates logs for each file access and transfer over the network. It also records file centric access and transfer information from within the kernel spaces of both virtual machines (VM) and physical machines (PM) in the cloud thus giving full transparency to the data kept in the cloud space. This tool allows an end user to track whether his/her file was touched by an unauthorized user or not. Flog, file centric log has complete detail of the file opened, its time of access and permission. The proposed work seems to be solving the problem to some extent.
The strength of this work is that it can be applied to public as well as private cloud. The concept is quite difficult to implement is one of its weak point.

A Juels & Burton S. Kaliski proposed a scheme called Proof of retrievability for large files using "sentinels" [8]. In this scheme special blocks (called sentinels) are hidden among other blocks in the data file F. To make the sentinels indistinguishable from the data blocks, the file is encrypted and stored in the cloud space. The use of encryption here renders the sentinels indistinguishable from other file blocks. This scheme is best suited for storing encrypted files.
The strength of this system is its simplicity and ease of implementation.
Due to the size of file, encryption becomes very heavy especially for thin clients. Hence, this scheme proves disadvantages to small users with limited computational power (PDAs, mobile phones etc.). There will also be storage overhead at the server, partly due to the newly inserted sentinels and partly

due to the error correcting codes that are inserted. The clients are also supposed to store all the sentinels with it for the stage of verification. Storing these sentinels will also incur storage overhead to the clients.

Sravan Kumar and Ashutosh Saxena [9] describes that data integrity of a file stored in the cloud can be identified using a simple method. In their work, a file is divided into different blocks and some arbitrary bits are chosen from each block to find Meta data. This Meta data is then encrypted using some secret function only known to the verifier. The Meta data is appended to the file before uploading it to the cloud. At the time of verification the CSP is asked to send some specified bits from the Meta data, hence it provides a proof to the verifier whether the file is integrated or not. The concept is easy to implement, it also does not encrypt entire file which has less overhead on the thin clients.
The major problem with this scheme being its capability to only process and check integrity of static data stored in the cloud.

Subramanian Anbazhagan and Dr. K. Somasundaram [10] discusses different cloud related security issues such User Identification & Authentication, Authorization, Confidentiality, Integrity, Non-repudiation, Availability. They have proposed a system which deals with above security issues. The system is based on the symmetric cipher model of encryption. A New Symmetric Key Algorithm has been designed to ensure that the data can be accessed and stored securely.
The strength of this scheme is it is very simple to implement. The weakness is that the proposed model incurs a lot of overhead while processing heavy data.

M. Raykova et al [11] discusses a double layered Access Control List (ACL) architecture. One layer would specify what files a cloud provider should or should not use, another layer (secondary) will specify the access for a user depending on its right to access the file. According to the author, this scheme can be implemented on a public as well as private cloud. As the user data is stored on the public cloud and not on private cloud, ACL must take into consideration the large no. of users accessing the cloud systems and such an access control much be finely provided to the user. Ye et al [12] have also described ACL and said that it could also be implemented in the cloud environment. Wang et al [13] explains ACL as one of the measures for safe guarding data integrity. The strength of the system is it provides different access control mechanism for both different clouds and various users. The proposed schemes are not completely implemented and in development phase.

Schwarz and Miller [14] propose a scheme that allows a client to verify the storage of data across multiple sites. The data possession guarantee is achieved using a special construct, called an "algebraic signature": A function that fingerprints a block and has the property that the signature of the parity block equals the parity of the signatures of the data blocks. The parameters of the scheme limit its applicability: The file access and Computation complexity at the server and the communication complexity are all linear in the number of file blocks (n) per challenge. Additionally, the security of the scheme is not proven and remains in question.

Sebe et al. [15] give a protocol for remote file integrity checking, based on the Diffie-Hellman problem in ZN. The client has to store N bits per block, where N is the size of an RSA modulus, so the total storage on the client is O (n) (which does not conform to our notion of an outsourced storage relationship). Indeed, the authors state that this solution only makes sense if the size of a block is much larger than N. Moreover, the protocol requires the server to access the entire file. Similar techniques were proposed by Yamamoto et al. [16], in the context of checking data integrity through batch verification of homomorphic hash functions.

K. Govinda, Dr. E. Sathiyamoorthy [17]: identity based secure data transfer in cloud using GDS (Group Digital Signature) is introduced. In this schema the group manager communicate with the cloud provider using the secret key generated using the Diffie Hillman key exchange algorithms. Now the group manager receives the member (user in the group) public key. For member who sends the data to the cloud server it can sign the message with the assigned (d, n) private key. Now the message is received by the group manager authenticate the member and then collect the required detail and attach the secret group id and sign and send to the cloud provider. Cloud provider authenticates the message and allows the encrypted message to store in private cloud.

Edna el at [18] discusses various advantages and disadvantages of cloud computing. They also focus on the Computing Architecture. The paper also elaborates different security aspects of cloud such as Security in the cloud, file system security, Trust.
The authors also propose a "High Level Model for File Sharing" which calculates the trust factor of a node based on node storage space, link and processing capacity. When a node wants to share files with other users, it will select trusted nodes to store this file through the following metrics: processing capacity (the average workload processed by the node, for example, if the node's processing capacity is 100%

utilized, it will take longer to attend any demands), storage capacity and link (better communication links and storage resources imply greater trust values, since they increase the node's capacity of transmitting and receiving information). The trust value is established based on queries sent to nodes in the cloud, considering the metrics previously described (figure 3). Each nodes maintains two trust tables: a) Direct trust Table b) Recommended List (consisting of reliable node)

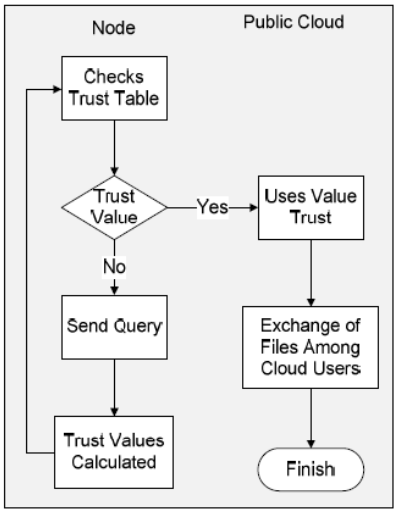


Figure 3. Proposed Trust Model

In this model, a trust rank is established, allowing a node A to determine whether it is possible to trust a node B to perform storage operations in a public cloud. In order to determine the trust value of B, node A first has to obtain basic information on this node. Figure 4 depicts the query exchange process used for gathering the necessary trust information from a node B by a node A.

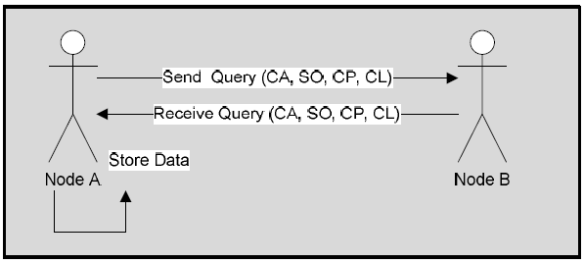


Figure 4. Scenario of Information Request

Miranda and Siani [19]: This paper states that most important obstacle to wide acceptance of ―cloud computing services security and privacy issues in cloud computing, users have serious concerns about confidential data seepage. Privacy is not observed while critical data is being processed in the public accessible cloud. Some practical scenarios has been discussed in this paper, based on these scenarios it is recommended strongly that use of sensitive information must be minimized when data is processed on clouds and privacy to end users must be assured. To address this issue, a client- based privacy manager tool has been proposed in this paper. The proposed reduces security issues, and provides added privacy features. The tool has been tested accordingly in different cloud computing environments. The entire structure of their developed privacy manager tool is shown in Figure 5 [19].

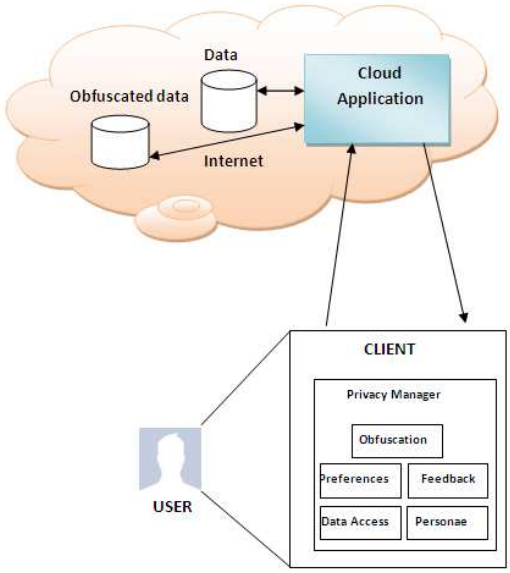


Figure 5. Privacy Manager Tool

Flavi and Roberto [20] stated that clouds are being targeted increasingly day by day. In this paper integrity protection problem in the clouds, sketches a novel Architecture and Transparent Cloud Protection System (TCPS) for improved security of cloud services has been discussed. They claim that they have identified the integrity safety problem in clouds .To address the integrity issues, they have proposed a system, and the system is named as Transparent Cloud Protection System (TCPS) for increased security of cloud resources. According to them their proposed system, TCPS can be used to observe the guests integration and keeping the transparency and virtualization.

The strength of their work is their proposed tool which provides improved security, transparency and intrusion detection mechanism. The limitation of their work is that they did not validate their work nor they have deployed in professional cloud computing scenario.

3. COMPARTIVE STUDY

Lit. Ref	**Context of Research**	**Problem Discussed**	**Technique Used**	**Model/ Tool /Proposed**
7	Cloud computing security	File Access and Transfers within Cloud Computing Environments	FLOGGER (File-centric Logger)	Yes
8	Data-centric cloud security	Proof of Retrievability	Sentinels and Encryption Algorithms	No
9	Data Integrity	Checking Data Integrity of a file	Meta Data and Encryption Algorithm	Yes
10	Complete Data Security	Data Security and Availability	A New Symmetric Key Algorithm	Yes
11, 12, 13	Privacy and Access Control	Cloud & User Access	Access Control List	No
14	Data Storage and Integrity	Remotely Stored Data	Algebraic Signature	Yes
15, 16	Remote file integrity checking	Data integrity of the remotely hosted data	RSA Algorithm	No
17	Data Confidentiality and Authentication	User data Confidentiality and User Authentication	Group Digital Signature	Yes
18	Data Storage and finding a trusted node for file storage	Trust is evaluated based on processing, link and capacity of a node	High Level Model for File Sharing	Yes
19	Privacy Manager for Cloud Computing	Security, Privacy and user concerns.	Privacy Manger tool developed to address security issues at user level.	Yes
20	Transparent Cloud Security	Cloud Security vulnerabilities and Security Attacks	The Transparent Cloud Protection System (TCPS)	Yes

4. CONCLUSION

In this study different security and privacy related research papers were studied in depth. This paper also studies different aspect of cloud securities such as data integrity, storage, transparency, file access and confidentiality. A lot of research has been done on security issue on cloud, but still much concrete work

has to be established so that accessing, retrieving, and using cloud services becomes easy.

REFERENCES

[1]. J. Ruiter and M. Warnier, Privacy regulation for cloud computing, compliance and implementation in theory and practice, article.

[2]. P. Metri and G. Sarote, Privacy Issues and Challenges in Cloud Computing, International journal of Advanced Engineering Sciences and Technologies, Vol. No. 5, Issue No. 1,001-006.

[3]. Dimitrios Zissis, Dimitrios Lekkas "Addressing cloud computing security issues" 2010 Future Generation Computer Systems (Science Direct)

[4]. J. Ruiter and M. Warnier, Privacy regulation for cloud computing, compliance and implementation in theory and practice, article.

[5]. P. Metri and G. Sarote, Privacy Issues and Challenges in Cloud Computing, International journal of Advanced Engineering Sciences and Technologies, Vol. No. 5, Issue No. 1,001-006.

[6]. R. Pandya, K. Sutaria, "An analysis of privacy techniques for data integrity in the cloud environment", International Journal of Computer and Electronics Engineering,(Dec 2012) ISSN: 0975-4202

[7]. Ryan K L Ko, Peter Jagadpramana, Bu sung Lee "Flogger: A File Centric Logger for Monitoring file access and transfers within cloud computing environment" 2011 International Joint conference of IEEE TrustCom 9780-0-7695-4600-1/11

[8]. A. Juels and B. S. Kaliski, Jr., "Pors: proofs of retrievability for large files," in CCS '07: Proceedings of the 14th ACM conference on Computer and communications security. New York, NY, USA: ACM, 2007, pp. 584–597.

[9]. Sravan Kumar and Ashutosh Saxena "Data Integrity Proofs in cloud Storage" 978-1-4244-8953-4/11/$26.00 c 2011 IEEE.

[10]. Subramanian Anbazhagan, Dr. K. Somasundaram "Cloud Computing Security Through Symmetric Cipher Model" International Journal of Computer Science & Information Technology (IJCSIT) Vol 6, No 3, June 2014

[11]. M. Raykova, H. Zhao & "Privacy enhanced access control for the outsourced data sharing" Financial cryptography security, LNCS Vol. 7379, Springer 2012

[12]. S. Yu, C. Wang, K Ren, and W. Lou, "Achieving secure, scalable and fine-grained data access control in cloud computing" in INFOCOM 2010 Proceedings, IEEE, 2010

[13]. G. Wang, Q Li and J. Wu "Hierarchical attribute-based encryption for fine grained access control in cloud storage services" in Proceedings of 17th ACM conference on computer and communication security (CSS'10), 2010

[14]. T. S. J. Schwarz and E. L. Miller. Store, forget, and check: Using algebraic signatures to check remotely administered storage. In Proceedings of ICDCS '06. IEEE Computer Society, 2006.

[15]. F. Sebe, A. Martinez-Balleste, Y. Deswarte, J. Domingo-Ferrer, and J.-J. Quisquater. Time-bounded remote file integrity checking. Technical Report 04429, LAAS, July 2004.

[16]. G. Yamamoto, S. Oda, and K. Aoki. Fast integrity for large data. In Proc. of SPEED '07, 2007.

[17]. K. Govinda, Dr. E. Sathiyamoorthy "Identity Anonymization and Secure Data Storage using Group Signature in Private Cloud" Science Direct 2012

[18]. Edna Dias Canedo and Robson de Oliberia, Rafel Timoteo de Sousa Junior "Review of Trust-based File Sharing in Cloud Computing" in IARIA, 2011. ISBN: 978-1-61208-147-2

[19]. Miranda & Siani, ―A Client-Based Privacy Manager for Cloud Computing, COMSWARE'09, 2009, Dublin, Ireland

[20]. Flavio Lombardi& Roberto Di Pietro, ―Transparent Security for Cloud‖, SAC'10 March 22-26, 2010, Sierre, Switzerland.