International Journal of Scientific Research

Indexed with International ISSN Directory, Paris

Volume 1 | Issue 2 | July 2012

ISSN No. 2277 – 8179

A Multi-Subject Journal

ISSN No. 2277 – 8179

International Journal of Scientific Research

Journal for All Subjects

1. Thoughts, language vision and example in published research paper are entirely of author of research paper. It is not necessary that both editor and editorial board are satisfied by the research paper. The responsibility of the matter of research paper/article is entirely of author.
2. Editing of the **International Journal of Scientific Research** is processed without any remittance. The selection and publication is done after recommendations of atleast two subject expert referees.
3. In any condition if any National/International University denies accepting the research paper published in IJSR then it is not the responsibility of Editor, Publisher and Management.
4. Only the first author is entitle to receive the copies of all co-authors
5. Before re-use of published research paper in any manner, it is compulsory to take written permission from the Editor-IJSR, unless it will be assumed as disobedience of copyright rules.
5. All the legal undertaking related to **International Journal of Scientific Research** is subject to Ahmedabad Jurisdiction.
7. The research journal will be send by normal post. If the journal is not received by the author of research papers then it will not be the responsibility of the Editor and publisher. The amount for registered post should be borne by author of the research paper in case of second copy of the journal.

Editor,
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH
3, SUHANA, Nr. Rubi Apartment, B/H NID, Rajnagar Road,
Paldi – 380007. Ahmedabad-Gujarat. (INDIA)
Contact: +91 98247 02127, +91 88660 03636
Www.theglobaljournals.com | ijsr@theglobaljournals.com

INDEX

Research Paper

Volume : 1 | Issue : 2 | July 2012 • ISSN No 2277 - 8179

Data Security and Protection in Cloud Computing

Computer Science

KEYWORDS : Encryption, Remanence, Phishing, Lineage, Access Control, Sanitization

Shameena Begum
V.Ratna Vasuki
K.V.V.Srinivas

Department of Computer Science, Sir CRR College for Women, Eluru

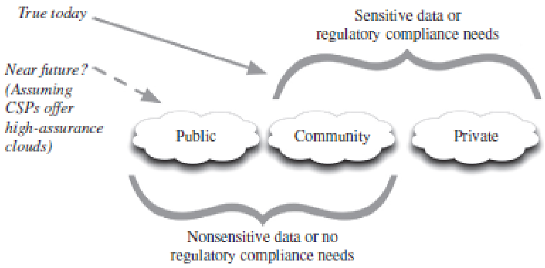
ABSTRACT

Cloud Computing is compelling, it is a rapidly growing trend in IT, and it is forcing significant advances in supporting technologies. The cloud model allows greater scalability and the change from a capital-heavy model of IT spending toward an operating model that is subscription-based brings new opportunities for a broader set of users and tenants to place larger bets with lower risk. Cloud computing allows providers to develop, deploy and run applications that can easily grow in capacity (scalability), work rapidly (performance), and never (or at least rarely) fail (reliability), without any concerns on the properties and the locations of the underlying infrastructures. The penalties of obtaining these properties of Cloud Computing are to store individual private data on the other side of the Internet and get service from Cloud providers and consequently result in security and privacy issues. The primary concerns for cloud security are around cloud infrastructure, software platform and user data. When a cloud is implemented with appropriate security, there cloud security could be equal to or exceed traditional IT implementations. This paper deals with the data security in cloud computing with appropriate methods and approaches.

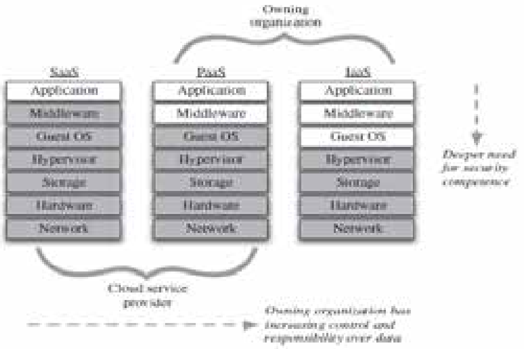
I. Introduction

Cloud computing is a network-based environment that refers to both the applications delivered as services over the Internet and the hardware and software in the datacenters that provide those services. In cloud, costumers must only pay for what they use. The three types of cloud environments are Public, Private, and Hybrid. A Public cloud is standard model by which providers provide several resources, such as applications and storage, available to the public. Private Cloud refers to internal services of a business that is not available for ordinary people. Hybrid cloud, a combination of private and public cloud is an environment that a company provides and controls some resources that are only accessible by certified staff and protected by firewalls from unauthorized outside access and has some others for public use. Cloud offers three major types of services: SaaS, PaaS, and laaS[1].

Like every other proposed technology, there are security concerns around storing and processing sensitive data in a public or hybrid cloud. Data security requirements vary depending on the service models, the deployment models, as well on risk tolerance. Cloud data security is far more than simply data protection. Some kinds of data are simply too sensitive and that the consequence of data exposure is too great for some customers to seriously consider using a public cloud for processing. Cost of providing such security assurance is incompatible with the cost model of a public cloud. If such data security needs prevail, community or private cloud may be more appropriate. Figure 1 shows Meeting security needs: Public, Community and Private Clouds



When data resides in private cloud, appropriate data assurance is not practiced. When data is stored with a CSP, the CSP assumes at least partial responsibility (PaaS) in the role of data custodian. A data owning organization can benefit from their CSP having control and responsibility for customer data in the SaaS model. The data owning organization is progressively responsible beginning with PaaS and expanding with IaaS. But appropriate data assurance can demand significant security competence for the owning organization [11].



The less control the data owning organization has the more concern and the greater the need for assurance that the CSPs security mechanisms and practices are effective for the level of data sensitivity and data value. The owning organization's responsibility for security runs deeper into the stack for the owning organization as they move from SaaS to PaaS and again to IaaS.

Risks to data security in clouds are presented to two states of data: Data that is at rest and data that is in motion.[8] The security triad (confidentiality, integrity, and availability) along with risk tolerance drives the nature of data protection mechanisms, procedures, and processes. The key issue is the exposure that data is subject to in these states.

II. Characteristics of Data Security

A. Encryption

Use of encryption is a key component for cloud security but if the keys are exposed or if encryption endpoints are insecure, it is pointless. Control over these endpoints vary depending on the service and deployment model. Encrypting data-at-rest is possible if you are using an IaaS cloud service (public or private) for simple storage (Amazon's S3). However, encrypting data-at-rest that a PaaS or SaaS cloud-based application is using (e.g., Google Apps, Salesforce.com) as a compensate control is not always feasible.

Data-at-rest used by a cloud-based application is generally not encrypted, because encryption would prevent indexing or searching of that data. Data-in-transit might be encrypted during transfer to and from a cloud provider. An organization's data is definitely not encrypted if it is processed in the cloud (public or private). For any application to process data, that data must be unencrypted. Until June 2009, there was no known method for fully processing encrypted data. In June 2009, IBM announced that one of its researchers had developed a fully homomorphic encryption scheme which allows data to be processed without being decrypted.[5][14][15].

It will have a significant positive impact on cloud computing as soon as it moves into deployment. Although the homomorphic scheme has broken the theoretical barrier to fully homomorphic encryption, it required immense computational effort.

B. Data Lineage

Whether the data an organization has put into the cloud is encrypted or not, it is useful and might be required to know exactly where and when the data was specifically located within the cloud. For example, the data might have been transferred to a cloud provider, such as Amazon Web Services (AWS), on date x1 at time y1 and stored in a bucket on Amazon's S3 in example1.s3.amazonaws.com, then processed on date x2 at time y2 on an instance being used by an organization on Amazon's Elastic Compute Cloud (EC2) in ec2-67-202-51-223.compute-1.amazonaws.com, then restored in another bucket, example2.s3.amazonaws.com, before being brought back into the organization for storage in an internal data warehouse belonging to the marketing operations group on date x3 at time y3.[11][16].

Following the path of data is known as data lineage, and it is important for an auditor's assurance However, providing data lineage to auditors or management is time-consuming, even when the environment is completely under an organization's control. Trying to provide accurate reporting on data lineage for a public cloud service is really not possible.

C. Data Provenance

Even if data lineage can be established in a public cloud, for some customers there is an even more challenging requirement and problem: Proving data provenance.

Provenance means not only that the data has integrity, but also that it is computationally accurate. [16][5][17].

D. Data Remanence

A final aspect of data security is data remanence. "Data remanence is the residual representation of data that has been in some way nominally erased or removed. This residue may be due to data being left intact by a nominal delete operation, or through physical properties of the storage medium. Data remanence may make inadvertent disclosure of sensitive information possible, should the storage media be released into an uncontrolled environment

The risk posed by data remanence in cloud services is that an organization's data can be inadvertently exposed to an unauthorized party—regardless of which cloud service you are using (SaaS, PaaS, or IaaS). When using SaaS or PaaS, the risk is almost certainly unintentional or inadvertent exposure. In spite of the increased importance of data security, the attention that cloud service providers (CSPs) pay to data remanence is strikingly low. [16][18][5]

III. Concerns of Data Security

For data stored in the cloud (i.e., storage-as-a-service), we are referring to IaaS and not data associated with an application running in the cloud on PaaS or SaaS. In today's competitive economy, data is the primary asset enterprises and individuals possess. In cloud computing, level of Security offered by various cloud providers are not same and raises the security issues. For this reason many companies are concerned over the protection of data as their entire data is being controlled by a central authority. Data Security is concerned with three broad categories: Confidentiality, Integrity and Availability.

A. Confidentiality

Confidentiality is a goal to achieve adequate security in Cloud. Keeping all confidential data of users' secret in the Cloud systems is a big obstacle for users to step into it, as many users said "My sensitive corporate data will never be in the Cloud" in the article named "Above the Cloud" [19]. Cloud Vendors adopted the two approaches called Physical isolation and Encryption to achieve Confidentiality.

Virtual Local Area Networks, Firewalls, Packet filters should be deployed to achieve the virtual physical isolation.For Example, Vertica [20] deploys its database on the Amazon EC2 and provides VPN and firewall to secure its database, as shown in Figure 3 [20]. When a Vertica database instance is provisioned by the Amazon EC2, it provides users full root access so users can secure the system as they see it. They chose to create a VPN between their enterprise users and their Vertica for the Cloud instance and set up a firewall to the outside world. Aside from the VPN port and software, they blocked off all external communication.

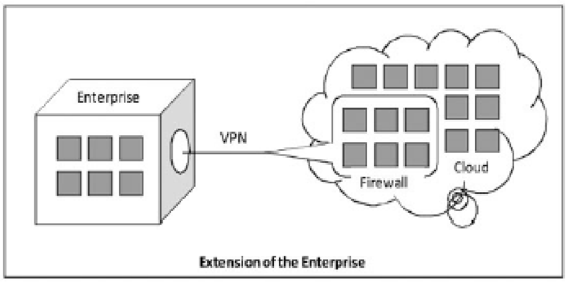


Fig. 3. Vertica Provides VPN and Firewall to Secure Its Database

Encrypted storage is another choice to enhance the confidentiality. For example, encrypting data before placing it in a Cloud may be even more secure than unencrypted data in a local data center. Symmetric encryption involves the use of a single secret key for both the encryption and decryption of data. Only symmetric encryption has the speed and computational efficiency to handle encryption of large volumes of data. It would be highly unusual to use an asymmetric algorithm for this encryption use case. Asymmetric algorithm is not used in data storage encryption.

With symmetric encryption, the longer the key length , the stronger the encryption. Although long key lengths provide more protection, they are also more computationally intensive, key lengths should be a minimum of 112 bits for Triple DES (Data Encryption Standard) and 128-bits for AES (Advanced Encryption Standard)—both NIST-approved algorithms.[16]

Another confidentiality consideration for encryption is key management. It is not recommended that you entrust a cloud provider to manage keys—at least not the same provider that is handling your data. This means additional resources and capabilities are necessary. That being said, proper key management is a complex and difficult task. .

B. Integrity

Encryption alone is sufficient for confidentiality, but integrity also requires the use of message authentication codes (MACs) [21]. The simplest way to use MACs on encrypted data is to use a block symmetric algorithm in cipher block chaining (CBC) mode, and to include a one-way hash function. Integrity of data should be maintained to ensure that neither data was lost nor modified by unauthorized users. Another aspect of data integrity is important, especially with bulk storage using IaaS. Once a customer has several gigabytes (or more) of its data up in the cloud for storage, there are IaaS transfer costs associated with moving data into and back down from the cloud, as well as network utilization (bandwidth) considerations for the customer's own network. It is necessary to validate the integrity of its data while that data remains in the cloud—without having to download and reupload that data.

Cloud Computing system provides volume data in terms of Tera Bytes or Peta Bytes. To scale up the data storage in the Cloud

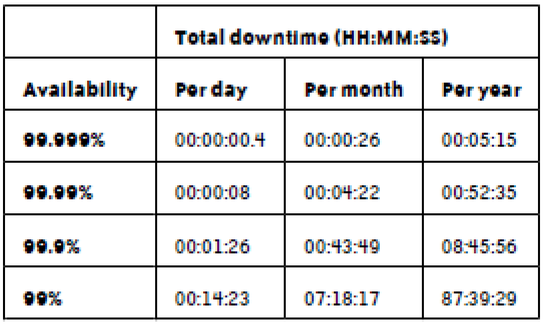
Computing systems, vendors need to increase the population of hard drives. This may consequently result in increased high probability of either node failure or disk failure or data corruption or even data loss. Secondly, disk drives are getting bigger and bigger in terms of their capacity, while not getting much faster in termsof data access. The Zetta system provided by Zetta mainly focus on data integrity for Cloud Computing storage services, which has a similar idea to RAID systems [22][23].

Digital signature is a commonly used technique for data integrity testing. The widely adopted distributed file systems GFS and HDFS usually divide data in large volumes into a set of blocks, each of which has a default size (e.g., 64MB, 128Mb). When a block of the data is physically stored on, a digital signature is attached to it. This digital signature is able to test the integrity of the data, and recover from corruption.

C. Availability

In simple terms, availability means that an organization has its full set of computing resources accessible and usable at all times. Availability can be affected temporarily or permanently, and a loss can be partial or complete. Denial of service attacks, equipment outages, and natural disasters are all threats to availability. Despite employing architectures designed for high service reliability and availability, cloud computing services can and do experience outages and performance slowdowns.[8][16].

There are currently three major threats in this regard. The first threat to availability is network-based attacks. The second threat to availability is the CSP's own availability. No CSPs offer the sought-after "five 9s" (i.e., 99.999%) of uptime. A customer would be lucky to get "three 9s" of uptime. As Table 1 shows, there is a considerable difference between five 9s and three 9s [16]. A number of high-profile cloud provider outages have occurred. For example, Amazon's S3 suffered a 2.5-hour outage in February 2008 and an eight-hour outage in July 2008. In addition to service outages, in some cases data stored in the cloud has actually been lost. In February 2009, cloud provider Coghead suddenly shut down, giving its customers fewer than 90 days (nine weeks) to get their data off its servers—or lose it altogether.



	Total downtime (HH:MM:SS)		
Availability	Per day	Per month	Per year
99.999%	00:00:00.4	00:00:26	00:05:15
99.99%	00:00:08	00:04:22	00:52:35
99.9%	00:01:26	00:43:49	08:45:56
99%	00:14:23	07:18:17	87:39:29

Table 1. Percentage of uptime

Finally, prospective cloud storage customers must be certain to find out just what services their provider is actually offering.

IV. Cloud data protection methods

Protecting data in the cloud is just like protecting data within a traditional data center. Besides Encryption, Authentication and identity, access control, , secure deletion, integrity checking, and data masking are all data protection methods that have applicability in cloud computing.

Authentication and Identity

Authentication of users can be performed by means of cryptography. Stronger authentication requires additional factors; for instance, two factor authentication is based on two authentication factors (such as a pin and a fingerprint)[11].

Access Control Techniques

Access control mechanisms are a key means by which we maintain a complex IT environment that reliably supports separation and integrity of different levels of information belonging to multiple parties. The most common access control models are:

- Discretionary Access Control (DAC): With DAC, access control is determined by the owner of the object who decides who will have access and what privileges they will have. Permission management in DAC can be very difficult to maintain; furthermore, DAC does not scale well beyond small sets of users.

- Role Based Access Control (RBAC) Access policy is determined by the system. RBAC access is based on the role of the subject. A subject can access an object or execute a function only if their set of permissions—or role—allows it [8][10].

- Mandatory Access Control (MAC) Access policy is determined by the system and is implemented by sensitivity labels, which are assigned to each subject and object. A subject's label specifies its level of trust, and an object's label specifies the level of trust that is required to access it. If a subject is to gain access to an object, the subject label must dominate—be at least as high as—the object label[11].

Deletion of Data

When it is time to delete sensitive or valuable data in a cloud, it is important to understand how that data is deleted. The two key aspects of data deletion, as stated in DoD 5220.22-M, National Industrial Security Program Operating Manual 3:

a. Clearing. Clearing is the process of eradicating the data on media before reusing the media in an environment that provides an acceptable level of protection for the data that was on the media before clearing. All internal memory, buffer, or other reusable memory shall be cleared to effectively deny access to previously stored information.

b. Sanitization. Data Sanitization is the removal of sensitive data from a storage device in various situations, such as when a storage device is removed from service or moved elsewhere to be stored. It also applies to backup copies made for recovery and restoration of service, and residual data remaining upon termination of service. In a cloud computing environment, data from one subscriber is physically commingled with the data of other subscribers, which can complicate matters. For example, with the proper skills and equipment, it is possible to recover data from failed drives that are not disposed of properly by service providers.[8].

Data Masking

Data masking is a technique that is aimed at reducing the risk of exposing sensitive information. A common data masking technique involves substitution of actual data values with keys to an external lookup table that holds the actual data values. In operation, such resulting masked data values can be processed with lesser controls than if the original data was still unmasked. But data masking must be performed carefully, or the resulting masked data can still reveal sensitive data. By example, if you mask salary data in an HR database by tokenizing what originally were employee names with name look up keys, the highest salary will probably be the CEOs [16].

V. Conclusions

As a customer of a cloud service, if a CSPs security practices are not in line with the value of our information, then that cloud service does not meet our security needs. If no other CSP will meet our cloud security needs, we probably need to consider building a private cloud or implementing a more expensive and sophisticated custom IT infrastructure. However, today this paradigm is already changing as CSPs and communities of interest are starting to recognize that there is opportunity for higher entrance-cost clouds that cater to communities of specialized security and privacy needs.

Data must be secured while at rest, in transit, and in use, and access to the data controlled. Standards for communications protocols and public key certificates allow data transfers to be protected using cryptography. Security strategies should be deployed in the Cloud environment to achieve the triad (Availability, Confidentiality, Data integrity). Success will be measured in Cloud Computing as the security issues are resolved.

Data that a user chooses to store in the cloud may not require protection if it is not sensitive or if it can easily be recovered. But generally, protecting data is a universal requirement regardless of its value. There is a need to protect data, usually by means such as file permissions, encryption, or more sophisticated container approaches. Identity-based access controls should be followed to support organizational access policies. Procedures are also necessary to limit exposure of such data when we create copies or backups. Also, we need mechanisms to detect when the valuable resource is accessed in ways that warrant concern. Data or information labeling is one information security technique that has been used to great success for classified information such as the hierarchical categories of Unclassified, Confidential, Secret, Top Secret, and Compartmented.

Data security is a significant task, with a lot of complexity, Customers must ensure that CSP is protecting user data as well as its own data. The only viable option for mitigation is to ensure that any sensitive or regulated data is not placed into a public cloud. Given the economic considerations of cloud computing today, as well as the present limits of cryptography, CSPs are not offering robust enough controls around data security. It may be that those economics change and that providers offer their current services, as well as a "regulatory cloud environment".

REFERENCE

[1] Minqi Zhou, Rong Zhang, Wei Xie , " Security and Privacy in Cloud Computing : A Survey", IEEE Sixth Interferance on Semantics, Knowledge and Grids, 2010. | [2] Jiang Wang, Yan Zhao, Shou Jiang, Jiajin Le, "Providing Privacy Preserving in Cloud Computing", IEEE ,pp. 472-475, 2010. | [3] Ranjita Mishra, Sanjit Kumar Dash, "A Privacy Preserving Repository for Securing Data across the Cloud", IEEE, 2011. | [4] Cong Wang, Qian Wang, and Kui Ren, "Towards Secure and Effective Utilization over Encrypted Cloud Data", 31st International Conference on Distributed Computing Systems Workshops, IEEE, 2011. | [5] Shubhashis Sengupta, Vikrant Kaulgud, Vibhu Saujanya Sharma, "Cloud Computing Security - Trends and Research Directions "World Congress on Services, IEEE, 2011. | [6] Farzad Sabahi," Cloud Computing Security Threats and Responses", IEEE 2011. | [7] I-Hsun Chuang, Syuan-Hao Li, Kuan-Chieh Huang, Yau-Hwang Kuo, "An Effective Privacy Protection Scheme for Cloud Computing", IEEE, pp.13-16, ICACT, Feb 2011. | [8] Wayne A. Jansen, "Cloud Hooks: Security and Privacy Issues in Cloud Computing", NISTProceedings of the 44th Hawaii International conference on System Sciences – 2011 | [9] Ziyuan Wang ," Security and privacy issues within the Cloud Computing", International Conference on Computational and Information Sciences IEEE 2011. | [10] Hassan Takabi and James B.D. Joshi ,Gail-Joon Ahn , "Security and | Privacy Challenges in Cloud Computing Environments", copublished | by the ieee computer and reliability societies, IEEE 2010. | [11] Vic (J.R.) Winkler (2011)," Securing the Cloud: Cloud Computer | Security Techniques and Tactics", USA, Elsevier. | [12] www.cloudsecurityalliance.org/guidance/csaguide-dom12.pdf | [13] http://www.schneier.com/blog/archives/2010/06/data_at_rest_vs.html; 2010 [accessed 7.10.10]. Information Sciences IEEE 2011. | [14] A.Greenberg, IBM's Blindfolded Calculator, Forbes Magazine, July 12, 2009. | [15] IBM Homomorphic Encryption research page, http://domino.research.ibm.com/comm/research_projects.nsf/pages/security.homoenc.html | [16] Tim Mather , Subra Kumaraswamy, Shahed Latif, " Cloud Security and Privacy", Orielly, 2009. | [17] Reddy, K.K.M, Macko, P., and Seltzer, M., Provenance for the Cloud. Proceedings of 8th USENIX Conference on File and Storage technologies, 2010. | [18] http://onlinesecurityservices.blogspot.in/2008/03/data-remanence.html accessed on 20th June 2012. | [19] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, et al., "Above the clouds: A Berkeley view of cloud computing," University of California, Berkeley,Tech. Rep, 2009. | [20] Vertica, "Vertica for the Cloud," http://www.vertica.com/cloud, 2008. | [21]. Wassim Itani, Ayman Kayssi, Ali Chehab , " Privacy as a Service: Privacy – Aware Data Storage and Processing in Cloud Computing Architectures", IEEE Eighth International Conferance on Dependable, Autonomic and Secure Computing, 2009. | [22] Zetta, "Zetta: Enterprise cloud storage on demand," | http://www.zetta.net/, 2008. | | [23] P. Chen, E. Lee, G. Gibson, R. Katz, and D. Patterson, "RAID: Highperformance, reliable secondary storage," ACM Computing Surveys | (CSUR), vol. 26, no. 2, pp. 145–185, 1994.

Sara Publishing Academy
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH
Journal for All Subjects

The Editor,
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH
8-A, BANAS, Opp. SLU Girls College, NR. Congress Bhavan,
Paldi – 380006. Ahmedabad-Gujarat. (INDIA)
Contact: +91 98247 02127, +91 88660 03636
Website : www.theglobaljournals.com
Email Id: ijsr@theglobaljournals.com

Printed at Unique Offset, Novatsing Rupam Estate, Opp. Abhay Estate, Tavdipura, Shahibaug, Ahmedabad