An Approach for Secure Data Storage in Cloud Environment

M. Gobi and R. Sridevi

***Abstract*—Cloud computing provides the facility to access shared resources and common infrastructure, offering services on demand over the network to perform operations that meet changing business needs. The location of physical resources and devices being accessed are typically not known to the end user. It also provides facilities for users to develop, deploy and manage their applications 'on the cloud', which entails virtualization of resources that maintains and manages itself. Confidentiality, Integrity, Authenticity and Privacy are essential concerns for both Cloud Service Providers and consumers as well. This paper proposes a scheme for enhancing the data authenticity, privacy and integrity of cloud data with the aid of an encryption scheme and a hash function which uses message digest.**

***Index Terms*—Authenticity, cloud computing, cloud service provider, confidentiality, encryption, hash function, integrity, message digest, privacy.**

I. Introduction

Cloud computing is defined by the National Institute of Standards and Technology (NIST) as "a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction". It is a prototype for large-scale distributed computing that makes use of existing technologies such as virtualization, service-orientation, and grid computing. It offers a different way to acquire and manage IT resources on a large scale. There is growing interest in cloud computing from consumers and Cloud Service Providers. For example, cloud service spending worldwide rose by over 20% in 2009 when overall IT spending dropped by about 4%. In 2010, most cloud consumers are small enterprises, but large enterprises were exploring the paradigm. Business or major activity users rely in some form, on IT and IT services. These services need to be enabling and appliance-like, and there must be an economy-of-scale for the total-cost-of-ownership to be better than it would be without cyber infrastructure. Technology needs to improve end user productivity and reduce technology-driven overhead. For example, unless IT is the primary business of an organization, less than 20% of its efforts not directly connected to its primary business should have to do with IT overhead; even though 80% of its business might be conducted using electronic means [1].

Manuscript received August 16, 2012; revised September 28, 2012.

The authors are with the Computer Science Department, Government Arts College, Udumalpet, Tamilnadu, India (e-mail: mgobimail@yahoo.com, srinashok@gmail.com).

II. Cloud Computing

Cloud computing is a computing paradigm, where a large pool of systems are connected in private or public networks, to provide dynamically scalable infrastructure for application, data and file storage. With the advent of this technology, the cost of computation, application hosting, content storage and delivery is reduced significantly. Cloud computing is a practical approach to experience direct cost benefits and it has the potential to transform a data centre from a capital-intensive setup to a variable priced environment. The idea of cloud computing is based on a very fundamental principal of reusability of IT capabilities. The difference that cloud computing brings compared to traditional concepts of "grid computing", "distributed computing", "utility computing", or "autonomic computing" is to broaden horizons across organizational boundaries[2].

More and more, small businesses are moving to cloud computing, signing up with Cloud Service Providers that make sophisticated applications more affordable as well as setting up their own accounts with public social media sites like Facebook. The trend is confirmed by Microsoft in its global SMB Cloud Adoption Study 2011, which found that 49% of small businesses expect to sign up for at least one cloud service in the next three years. Although cloud computing can offer small businesses significant cost-saving benefits, the service does come with certain security risks. The top five security concerns in cloud computing are: Secure data transfer, Secure software interfaces, Secure stored data, Authorized user access and Data separation. These risks should be addressed before publishing the cloud data to its servers and applications. Cloud computing offers small businesses too many benefits to dismiss out of hand.

A. Cloud Characteristics

The important five characteristics of cloud computing includes [3]:

1) *Service on demand*: This property involves valid customers using a web site or similar control panel interface to provision computing resources such as additional computers, network bandwidth or user email accounts, without requiring human interaction between customers and the Cloud Service Provider.
2) *Internetworking:* The internetworking enable the customers to access computing resources over networks such as the internet from a broad range of computing devices such as laptops and smart phones.
3) *Virtualization of resources*: This characteristic of virtualization involves the vendors using shared computing resources to provide cloud services to multiple customers. Virtualization and multi-tenancy

mechanisms are typically used to both segregate and protect each customer and their data from other customers, and to make it appear to customer that they are the only user of a shared computer or software application.

4) *Flexible processing*: This property enables the rapid and automatic increase and decrease to the amount of available computer processing, storage and network bandwidth as required by customer demand.
5) *Pay-for-use service*: This pay – for - use service make customers only pay for the computing resources that they actually use, and being able to monitor their usage. This is analogous to household use of utilities such as electricity.

B. Cloud Service Models

Cloud Service Providers offer their services according to three fundamental models as indicated in Fig. 1. They are Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) where IaaS is the most basic and each higher model abstracts from the details of the lower models [4], [5].

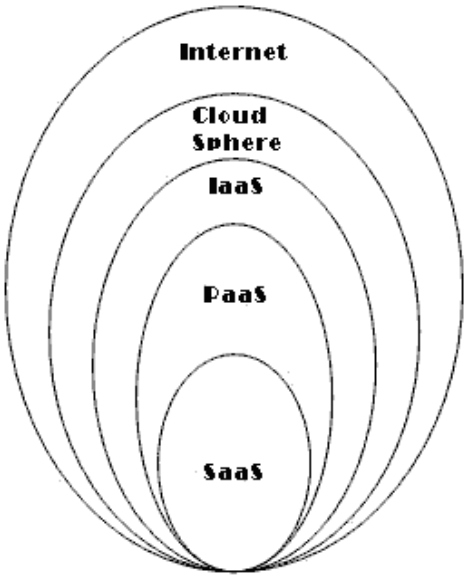

Fig. 1. Cloud service models

Infrastructure as a Service (IaaS): This involves the Cloud Service Provider provides physical computer hardware including CPU processing, memory, data storage and network connectivity. The Cloud Service Provider may share their hardware among multiple customers referred to as "multiple tenants" using virtualization software. IaaS enables customers to run operating systems and software applications of their choice. Typically the Cloud Service Provider controls and maintains the physical computer hardware. Typically the customer controls and maintains the operating systems and software applications. Example for IaaS vendor services includes Amazon Elastic Compute Cloud (EC2), GoGrid and Rackspace Cloud [6].

Platform as a Service (PaaS): The service involves the vendor provides Infrastructure as a Service plus operating systems and server applications such as web servers. PaaS enables customers to use the Cloud Service Provider's cloud infrastructure to deploy web applications and other software developed by the customer using programming languages supported by the Cloud Service Provider. Typically the Cloud Service Provider controls and maintains the physical computer hardware, operating systems and server applications. Typically the customer only controls and maintains the software applications developed by the customer. Example PaaS vendor services include Google App Engine, Force.com, Amazon Web Services Elastic Beanstalk, and the Microsoft Windows Azure platform.

Software as a Service (SaaS): This type of service involves the Cloud Service Provider using their cloud infrastructure and cloud platforms to provide customers with software applications. Example applications include email and an environment for users to collaboratively develop and share files such as documents and spreadsheets. These end user applications are typically accessed by users via a web browser, eliminating the need for the user to install or maintain additional software. Typically the Cloud Service Provider controls and maintains the physical computer hardware, operating systems and software applications. Typically the customer only controls and maintains limited application configuration settings specific to users such as creating email address distribution lists. Example SaaS vendor services include Salesforce.com Customer Relationship Management (CRM), Google Docs and Google Gmail and Microsoft Office 365 (formerly called Business Productivity Online Suite) which consists of Microsoft Office Web Apps, Microsoft Exchange Online, Microsoft SharePoint Online, Microsoft Dynamics CRM Online and Microsoft Lync.

III. Cloud Security

Cloud computing offers potential benefits like cost savings and improved business outcomes for the business concerns. However, there are a variety of cloud data security risks that need to be carefully considered. Risks will vary depending on the sensitivity of the data to be stored or processed, and how the chosen Cloud Service Provider has implemented their specific cloud services [7].

The enterprises are looking toward cloud computing horizons to expand their on-premises infrastructure, but most cannot afford the risk of compromising the security of their applications and data. Security ranked first as the greatest challenge or issue of cloud computing. Security and Privacy affect the entire cloud computing stack, since there is a massive use of third-party services and infrastructures that are used to host important data or to perform operations that are very critical [8].

Corporate companies and individuals are concerned about how security and compliance integrity can be maintained in this new environment. As cloud computing encompasses many technologies including networks, databases, operating systems, virtualization, resource scheduling, transaction management, load balancing, concurrency control and memory management, there are numerous security issues for cloud computing. Therefore, security issues for many of these systems and technologies are applicable to cloud computing. The important keys to security of cloud computing encompasses Security, Confidentiality, Authenticity, Privacy and Integrity of the data stored in

cloud.

Security refers to Confidentiality, Integrity and Availability, which pose major issues for Cloud Service Providers. Confidentiality refers to limiting cloud data access and disclosure to authorized users and preventing access by or disclosure to unauthorized ones. Underpinning the goal of confidentiality are authentication methods like user-IDs and passwords that uniquely identify a data system's users, and supporting control methods that limit each identified user's access to the data system's resources.

Integrity refers to the trustworthiness of cloud data resources. It includes the concept of data integrity, namely, that data have not been changed inappropriately, whether by accident or deliberately. Integrity of cloud data includes only preservation without corruption of whatever was transmitted or entered into the system. Availability refers to the availability of cloud data. Availability, like other aspects of security, may be affected by purely technical issues (e.g., a malfunctioning part of a computer or communications device), natural phenomena (e.g., wind or water), or human causes (accidental or deliberate).

One benefit of cloud computing is that client software security does not need to be enforced as strictly as before. This aspect concerns the view of cloud computing as software as a service, as it becomes more important to ensure security of data transfer rather than a traditional secure application life cycle [9], [10].

IV. Cryptography Based Cloud Computing

The cloud computing model for delivering computing services offers less expensive access to a variety of standardized services from various providers. But after outsourcing a service to the cloud, the owner no longer controls the platform on which the service runs. The user is bound to trust the Cloud Service Provider for Correctness, Privacy, and Integrity of its data and computations. Cryptographic mechanisms can reduce such trust by allowing the user to protect its data and computations, as well as to verify aspects of remote computation [11].

As with other hosted services, data protection is an issue when considering cloud computing. The main data protection risks to the cloud data are loss of data by third-party service providers, unauthorized access to the cloud data, malicious activities targeting Cloud Service Provider and poor internal IT security compromising data protection. Before introducing a cloud computing system, a risk assessment of these hazards and their potential impact on the data should be carried out [12].

High levels of data protection are necessary for cloud applications, and the associated security measures have to be taken to protect the private data on the cloud from these risks.This paper addresses the following areas of risks in cloud computing:

1) Secure *storage* of data in the cloud environment
2) *Protecting* data from unauthorized access

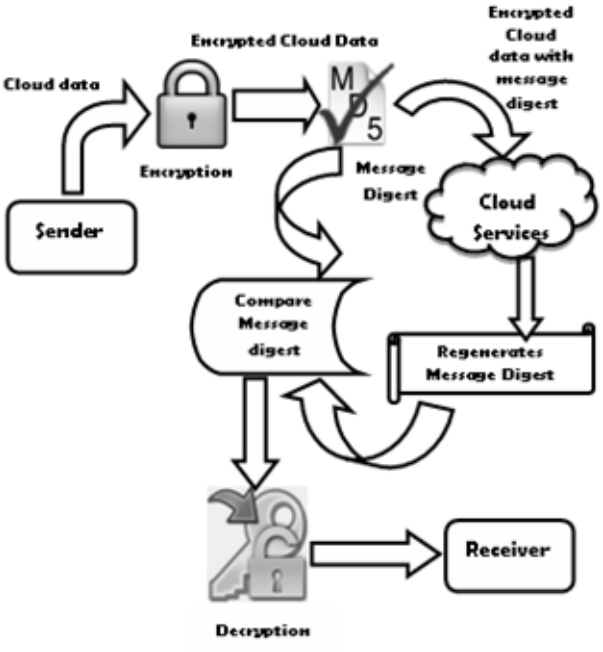

Fig. 2. Cryptography based cloud computing.

Fig. 2 explains the approach which is based on asymmetric cryptography and the message digest. The data stored in the cloud may initially be encrypted using an asymmetric cryptographic algorithm like RSA to produce a cipher text and then a message digest is been generated for that cipher text using MD5 Algorithm. Whenever the client access the data stored in the cloud, the message digest is been regenerated using the same hash function to check the correctness of data in the cloud. If the correctness is achieved, then the cipher text is decrypted to obtain the original plaintext or data previously stored [13].

A. Data Encryption using RSA

The data stored in cloud is encrypted using an asymmetric algorithm (RSA algorithm) before storage to enforce data integrity in cloud environment. The algorithm follows public key cryptography which has two different keys public and private key one for encryption and another one for decryption. The algorithm is briefly discussed below [14].

1) *Key generation:* Each user generates a unique public/private key *pair* by[15], [16]:
   - Selecting two large primes at random - p, q
   - Computing their system modulus
     $N= p.q$
     Note: $\phi(N) = (p-1)(q-1)$
   - Selecting at random the encryption key e where
     $1<e<\phi(N)$, gcd (e, $\phi(N)$) =1
   - Solve following equation to find decryption key *d*
     $e.d=1 \bmod \phi(N)$ and $0 \le d \le N$
   - Publish their public encryption key:
     $KU= \{e, N\}$
   - Keep secret (private) decryption key:
     $KR= \{d, p, q\}$
2) *Usage of Keys:* To *encrypt* a message *M*, the Sender [15]:
   - Obtains public key of recipient $KU=\{e,N\}$
   - Computes: $C=M^e \bmod N$, where $0 \le M < N$

To decrypt the cipher text *C*, the Receiver:

- Uses their private key $KR=\{d,p,q\}$
- Computes: $M=C^d \bmod N$

This enforces the privacy of client data over the cloud and makes the other users not to access the original cloud data since it has been encrypted.

B. Message Digest – MD5

The authenticity of the cloud data may be ensured with the generation of message digest using a hash function of MD5. This algorithm generates 128 bit message digest before publishing the data and after encryption in the cloud. The same algorithm is used to regenerate the message digest to ensure the authenticity of data during the storage. The algorithm is given as [18], [19]:

Step 1. Append padding bits: The input message is "padded" (extended) so that its length (in bits) equals to 448 mod 512. Padding is always performed, even if the length of the message is already 448 mod 512.

Step 2. Append length: A 64-bit representation of the length of the message is appended to the result of step1. If the length of the message is greater than 2^{64}, only the low-order 64 bits will be used.

Step 3. Initialize MD buffer: A four-word buffer (*A*, *B*, *C*, *D*) is used to compute the message digest. Each of *A*, *B*, *C*, *D* is a 32-bit register. These registers are initialized to the following values in hexadecimal, low-order bytes first):

Word A: 01 23 45 67
Word B: 89 ab cd ef
Word C: fe dc ba 98
Word D: 76 54 32 10

Step 4. Process message in 16-word blocks: Four functions will be defined such that each function takes an input of three 32-bit words and produces a 32-bit word output. At the end of this process, a sum of 128 bit message digest is been generated from all the four functions.

F (*X*, *Y*, *Z*) = *XY* or not (*X*) *Z*
G (*X*, *Y*, *Z*) = *XZ* or *Y* not (*Z*)
H (*X*, *Y*, *Z*) = *X* xor *Y* xor *Z*
I (*X*, *Y*, *Z*) = *Y* xor (*X* or not (*Z*))

The client when retrieving the cloud data from the

Cloud Service Provider, a message digest is been regenerated and has been verified with the previously generated message digest. If the 128 bit message digest is matched with the previously generated digest, then the data is decrypted using RSA algorithm to get the original data stored in the cloud.

V. Conclusion

This paper proposes a more effective and distributed two level security scheme to address the data storage security issue in cloud computing. As it rely on the asymmetric cryptography for protecting user data including encryption prior to storage, user authentication procedures prior to storage or retrieval, and building secure channels for data transmission, this method achieves the Reliability, Authenticity and Integrity of the cloud data. This approach of security model is expected to provide more security to user's data in cloud computing during storage and against unauthorized data modification attacks.

References

[1] M. A. Vouk, "Cloud computing – Issues, research and implementations," *Journal of Computing and Information Technology - CIT 16*, vol. 4, pp. 235–246, Sep 2008.

[2] V. S. Rao, N. K. N. Rao, and E. K. Kumari, "Cloud computing: An overview," *Journal of Theoretical and Applied Information Technology*, 2005 – 2009.

[3] J. A. Mukundrao and G. P. Vikram, "Enhancing security in cloud computing," *Information and Knowledge Management*, ISSN 2224-5758 (Paper) , ISSN 2224-896X (Online), vol. 1, no. 1, 2011.

[4] I. Sriram and A. K. Hosseini, "Research agenda in cloud technologies," 2009.

[5] K. D. Kadam, S. K. Gajre, and R. L. Paikrao, "Security issues in cloud computing," *National Conference on Innovative Paradigms in Engineering and Technology (NCIPET-2012), Proceedings published by International Journal of Computer Applications(IJCA).*

[6] P. Arora, R. C. Wadhawan, and E. S. P. Ahuja, "Cloud computing security issues in infrastructure as a service," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, no. 1, January 2012.

[7] T. Andrei, "Cloud computing challenges and related security issues," April 2009. [Online]. Available: http://www.cse.wustl.edu/~jain/cse571-09/index.html

[8] C. Wang, Q. Wang, and W. Lou, "Towards secure and dependable storage services in cloud computing," *17th IEEE International Workshop on Quality of Service (IWQoS'09)*.

[9] A. A. Friedman and D. M. West, "Privacy and security in cloud computing," *Issues in Technology Innovation*, October 2010.

[10] K. Hamlen, M. Kantarcioglu, L. Khan, and B. Thuraisingham, "Security issues for cloud computing," *International Journal of Information Security and Privacy*, vol. 4, no. 2, pp. 39-51, April-June 2010.

[11] W. Jansen and T. Grance, "Guidelines on security and privacy in public cloud computing," *Draft NIST Special Publication*, Jan 2011.

[12] C. Hota, S. Sanka, M. Rajarajan, and S. K. Nair, "Capability-based cryptographic data access control in cloud computing," *Int. J. Advanced Networking and Applications*, vol. 3, no. 3, pp. 1152-1161, 2011.

[13] J. A. Jose, G. C. Sajeev, and C. Suyambulingom, "Implementation of data security in cloud computing," *International Journal of P2P Network Trends and Technology*, vol. 1,no. 1, 2011.

[14] Top 5 Security Risks of Cloud Computing. [Online]. Available: http://blogs.cisco.com/smallbusiness/the-top-5-security-risks-of-cloud-computing/

[15] B. Schneier, "Applied cryptography: Protocols, Algorithms and source code in C," 2008.

[16] E. F. Brickell, "Survey of hardware implementations of RSA," *Advances in Cryptology-CRYPTO '89 Proceedings*, Springer- Verlag, 1990, pp. 368-370.

[17] T. K. Jung, A. K. Lenstra, D. Page, and N. P. Smart, "Using the cloud to determine key strengths," May 2012.

[18] Len Adleman. [Online]. Available: http://imps.mcmaster.ca/courses/SE-4C03-07/wiki/wrighd/rsa_alg.html

[19] MD5 Algorithm. [Online]. Available: http://infohost.nmt.edu/~sfs/Students/HarleyKozushko/Presentations/MD5.pdf

Dr. M Gobi is an Assistant Professor, Department of Computer Science in Government Arts College, Udumalpet, India. He earned his PhD in Computer Science from Bharathiar University, Coimbatore. He teaches courses for UG and PG of Computer Science. His research areas of interest include Cryptography, Information security, and Network security.

R. Sridevi is a Part Time PhD Scholar in the Department of Computer Science, Government Arts College, Udumalpet. She is also working as an Assistant Professor in Computer Science, PSG College of Arts & Science, Coimbatore. She finished her Post Graduate and MPhil under Bharathiar University. Her area of specialization includes Network Security, Cryptography and Data Mining.