Survey of Data Security Challenges in the Cloud

G.Dakshayani
Assistance Professor, Computer Engineering Department
FCRIT,Vashi

Kavita Shelke
Assistance Professor, Computer Engineering Department
FCRIT,Vashi

Abstract:- **Cloud Computing has become one of the most talked about technologies in recent times and has got lots of attention from media as well as analysts because of the opportunities it is offering. Most of enterprises are striving to reduce their computing cost through the means of virtualization. This demand of reducing the computing cost has led to the innovation of Cloud Computing. Cloud Computing offers better computing through improved utilization and reduced administration and infrastructure costs. Cloud Computing is the sum of Software as a Service (SaaS) and Utility Computing.**

Small and Medium scale Enterprises (SMEs) are key economic growth drivers .Over the years, the information technology (IT) needs of these SMEs have resulted in increased requirements for enterprise IT solutions that are more efficient and have high availability and scalable over time. Cloud computing presents such solutions that can be of immense benefits to SMEs. The main concern of users in the cloud environment is data security and privacy. Security and privacy within and outside the cloud is shared responsibility of service providers and users. These concerns could be mitigated by ensuring that effective security controls are put in place by both parties. This research allows for informed decision to be made about the security risks and benefits of using cloud computing.

Keywords: Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS), Small and Medium scale Enterprises (SMEs)

I. INTRODUCTION

In today's highly competitive business environments, businesses especially SMEs are finding paradigm to operate efficiently with reduced cost and maximum profit called cloud computing[1]. This is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., network, servers, storage, applications and services) that can be rapidly provisioned and released with minimum management effort or service provider interaction.

A new paradigm of computing, cloud computing has emerged to change the old ways of computing. Cloud computing has emerged as one of the enabling technologies that allows the Information Technology world to use computer resource effectively and more efficiently. This is important and attractive for the enterprise, the necessary precautions must be taken to ensure that confidentiality, integrity and available of information and information systems are not compromised in the cloud environment. During the last a few years, data security and integrity in cloud computing has emerged as a significantly important research area that has attracted increasing attention from both industry and academia. The virtual environment of cloud computing allows users to access computing power that exceeds what is contained within their own physical worlds. To enter this virtual environment, cloud users must transfer data throughout the cloud. Typically, cloud users know neither the exact location of their data nor the other sources of the data collectively stored with theirs. Consequently, several data security and integrity concerns have arisen, including key management, access control, searchable encryption techniques, remote integrity checks and proof of ownership in the cloud. n this paper we will be addressing these critical security challenges.

A. Cloud computing architecture:-

nist (National Institute of Standards and Technology) defines the Cloud Computing architecture by describing five essential characteristics, three cloud services models and four cloud deployment models (Cloud Security Alliance, 2009, p14).

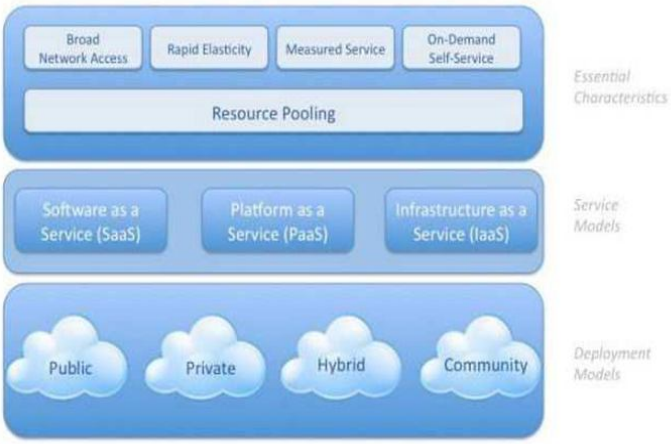

Figure 1 - Visual model of NIST Working Definition of Cloud Computing (Cloud Security Alliance, 2009, p14)

B. Essential characteristics of cloud computing:-

As described above, there are 5 essential characteristics of Cloud Computing which explains there relation and difference from the traditional computing.

1. *On-demand-self-service*:-Consumer can provision or un-provision the services when needed, without the human interaction with the service provider.
2. *Broad Network Access*: - It has capabilities over the network and accessed through standard mechanism.
3. *Resource Pooling*:-The computing resources of the provider are pooled to serve multiple consumers which are using a multi-tenant model, with various physical and virtual resources dynamically assigned, depending on consumer demand.
4. *Rapid Elasticity*:-Services can be rapidly and elastically provisioned
5. *Measured Service* :-Cloud Computing systems automatically control and optimize resource usage by providing a metering capability to the type of services (e.g. storage,processing, bandwidth, or active user accounts) .

C. Cloud type and services providers:-

Cloud services have made it possible for organizations to operate across institutional boundaries. This has led to overcoming the physical barrier present in isolated systems. Cloud services give companies the flexibility to purchase infrastructure, applications, and services, from third-party providers with the goal of freeing up internal resources and recognizing cost savings. Cloud computing it's based in the offer of services, we found 3 different kinds of services:

1. Software as a Service (SaaS) :-

These services are applications over Internet. Normally the user can run these applications using a web-browser. User abstract totally about the hardware and software that is using and simply access to a interface with a web browser and from there he have access to some information and functionalities. It's dedicated to current users; an example to this kind of services may be Google Docs

2. Platform as a Service (PaaS): -

These services are focused on the deployment of applications or services online letting to the developer manage the hardware or software necessary, including also a solution stack. This service includes all the life-cycle of the deployment of application/ service such as design, implementation, testing, deployment, integrity with databases, etc
There are three characteristic points in this service:
a)Services for deployment, testing and maintenance of applications.
b)Multi-user architecture, in other words, scalability.
c)Collaborative tools.
An example of these services is Google App engine.

3. Infrastructure as a Service (IaaS):

These services are focused to offer a computer infrastructure. All the servers, connections, software and other resources are offered by the providers. And the users see it like an entire infrastructure hosted in the same organization.

D. Types of Clouds:-

A. Public :-Public cloud (also known as external cloud), is the traditional way, where services are provided by a third part via Internet, and they are visible to everybody (it doesn't mean that they have to be free). So in the cloud it's the information of lots of users but they can't access of course to the information of the others.

B. Private :-This cloud consists on the hosting of private applications, storage, or computation in the same company emulating a cloud in Internet but only for private use (private networks). The coast of infrastructure and maintenance of it is the same that having it in normal way but the scalability and the sharing of the coasts is better...

C. Hybrid :-It's a combination of public and private cloud. An organization can have a part of their services in its own infrastructure but also in public cloud. Or can use the public just when have peaks of usage. It's a good option when you want to have your data or application in local and don't want to invest too much in infrastructure.

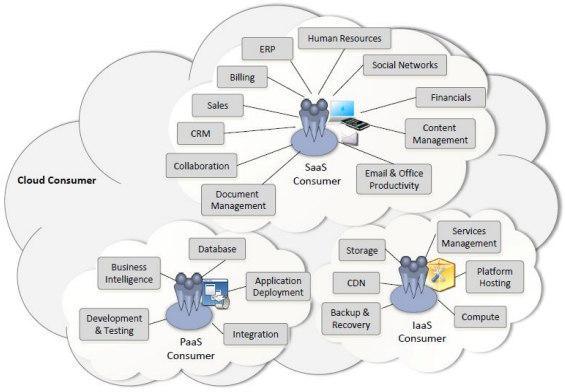

Figure 2: Service available for a cloud consumer

II. DATA SECURITY CHALLENGES IN THE CLOUD

2.1 Key Management:-

Data encryption before outsourcing to the cloud is a common and simple way to protect data privacy. Although the encryption algorithms are public, information encrypted under these algorithms is secure because the key used to encrypt the data remains secret. As a result, key management is a critical element in cloud computing. It is the ability to correctly assign, secure and monitor keys that defines the level of operational security provided by any encryption implementation

2.2 Access Control:-

Unlike the traditional access control in which the data users and Storage servers are in the same trusted domain, access control techniques are very different in cloud computing because the cloud servers are not seen as trustworthy by most cloud users, especially large enterprises and organizations. One possible method to enforce data access control without relying on cloud servers could be to encrypt data individually and disclose the corresponding decryption keys only to the privileged users, but that causes high performance costs. A fine grained access control which is efficient and secure is important in cloud computing.

2.3 Searchable Encryption Techniques:-

As the data is usually encrypted before being outsourced to cloud servers, how to search the encrypted data in the cloud has recently gained attention and led to the development of searchable encryption techniques. This problem is challenging however, because meeting performance, system usability and scalability requirement is extremely difficult.

2.4 Remote Integrity Check:-

Storing data in remote cloud servers has become common practice. As clients store their important data in remote cloud servers without a local copy, it is important to check the remote data integrity (RIC). While it is easy to check data integrity after completely downloading the data, it is a large waste of communication bandwidth. Hence, designing client remote integrity check protocols without downloading the data is an important security issue in the cloud.

2.5 Proof Of Ownership:-

Beyond storage correctness, proof of ownership (POW) is another security issue related to cloud data storage. Client-side de duplication allows an attacker to main access to arbitrary-size files when he has small hash signatures of the files. To overcome such attacks, the technique of POW allows a user to efficiently prove to a cloud server about his ownership, rather than short information about the file such as a hash value.

III. KEY MANAGEMENT IN CLOUD:-

The classical tree-based hierarchy schemes such as RFC2627 [3] and the scheme proposed by Wong et al.[5] have been widely used in group key management. In RFC2627, the hierarchical tree approach is the recommended approach to address the multicast key management problem. Many key management methods of access hierarchies for data outsourcing have been proposed [4, 6]. These methods provide some useful solutions to minimize the number of cryptographic keys which have to be managed and stored. with the intention to provide secure and efficient access to outsourced data, [4] proposed a tree-based cryptographic key management scheme for cloud storage. This tree-based key management structure is similar to a traditional one, where a single root node holds the master key that can be used to derive other node keys. Each node key can be used to derive the keys of its children in the tree hierarchy. In this scheme, a data block stored in the cloud can be deleted or updated by a party who holds either the specific decryption key or a node key corresponding to one of its parents. If there is an outsourcing server authorized to manage a node (not the root node) that has several child nodes, then the outsourced party is granted the node key, which can be used to derive all sub-keys for its child nodes. In other words, once a parent node in the tree is given, all the child nodes will be known. This is a common problem which exists in many tree-based key management schemes. Existing ones such schemes can work perfectly, only if when all legitimate node users are authorized to access all the child nodes under the specific parent node.

IV. CONTOLLING ACCESS IN CLOUD:-

An existing feasible solution to achieve fine-grained access control of outsourced data in cloud computing is to encrypt the data through certain cryptographic primitives and only disclose the private keys to authorized users. Without the appropriate decryption keys, unauthorized users including the cloud providers, cannot decrypt the data. This solution has been widely used and most schemes using it are deployed by introducing a per file groups for efficiency. However, the complexity of these schemes [7] is proportional to the system scale and the number of users. Additionally, this solution lacks scalability and flexibility, especially if the number of authorized users becomes large. Vimercatiet l.[10] proposed an access control scheme for securing data stoppage on un trusted servers based on key derivation methods [9]. In this scheme, the owner created corresponding public tokens to grant access for a user. With his secret key, the user was able to derive decryption keys for desired files. However, the complexity of operations of file creation and user grant/revocation is also linear to the number of users. Atenieseet al.[8] proposed a secure distributed storage scheme based on proxy re-encryption. Specifically, the data owner encrypted blocks of content with a master public key, which could only be decrypted by the master private key. The data owner then generated proxy re-encryption keys by using his master private key and the user's public key. With the proxy re-encryption keys, the semi-trusted proxies could convert the cipher text into another cipher text for a specific user. Thus, they achieved access control. However, the main problem with this scheme is that user access privilege is not protected from the proxy .Recently; ABE has been seen as an ideal technique for achieving flexible, scalable and ne-grained access control mechanisms in the cloud. Wang et al.[11] pro-posed a hierarchical attribute-based encryption scheme to achieve fine-grained access control in cloud storage services by combining hierarchical identity-based encryption and cipher text-policy attribute-based encryption (CP-ABE). In their scheme, they assumed that all attributes in one conjunctive clause were administered by the same domain master. However it is difficult to implement in practice so that the same attribute can be administrated by multiple domain masters

according to specific policies. Another difficulty is that their scheme cannot support compound attributes efficiently and does not support multiple value assignments.

V. SEARCHABLE ENCRYPTION TECHNIQUES:-

The work of Goldreichet al.[12] on oblivious random-access machines (RAMs) could resolve the problem of doing (private) searches on remote encrypted data. They enabled a client to store only a constant amount of data in local storage. Meanwhile, the identities of the remote data files were hidden when the client accessed them. Oblivious RAM is often cited as a powerful tool, which can be used, for example, for search on encrypted data or for preventing cache attack. However, oblivious RAM is also commonly considered to be impractical due to its overhead. Suppose the client stores n data files in remote storage, then each data request is replaced by $O(\log^4 n)$ or $O(\log^3 n)$ requests. Additionally, O(nlogn) external memory is required in order to store then data files .In an effort to reduce the round complexity associated with oblivious RAMs,Songet al.[13] presented a solution for searchable encryption. After that, the question of how to do keyword searches on encrypted data efficiently was raised. In [13], they achieved searchable encryption by constructing a special two-layered encryption for each word. Given a trapdoor, the server could strip the outer layer and ascertain whether the inner layer was in the correct form. The limitations in this construction are as follows. First, it is not compatible with existing encryption schemes and a specific encryption method must be used. Second, while the construction is proven to be a secure encryption scheme, it is not proven to be a secure searchable encryption scheme. Third, the distribution of the underlying plaintext is vulnerable to statistical attacks. Their approach may leak the locations of the keyword in a file. Finally, their searching time is linear in the length of the document collection.

VI. REMOTE INTEGRITY CHECKING

The challenging problem of data integrity verification without explicit knowledge of the full file was first proposed in broad generality by Blumet al.[14], who explored the task of checking the correctness of a memory-management program because they require the data to be transmitted in its entirety to the verifier. The latest verification schemes concentrated on the problems of securing data integrity at remote servers and securing cloud storage applications. These schemes can be classified into`Proof of retrievability'(POR) schemes (e.g., [15,16,17]) and `Provable data possession'(PDP) schemes (e.g., [19,20]). A POR scheme is a challenge-response protocol. In POR schemes, a cloud provider demonstrates the file retrievability (i.e., recoverability without any loss or corruption) to a client. PDP schemes are similar protocols which only detect a large amount of corruption in outsourced data. Several advances have already been proposed [15,19]. In Shahet al.[21], introduced a third-party verifier who could delegate the periodic task of checking data integrity. In their scheme, they reduced the client's storage, communicational and computational cost. This simple solution, however, requires a third-party verifier to keep a lot of hash values of the data blocks. In Opreaet al.[22], allowed a client to detect the modification of data blocks by a remote and untrusted server. Their protocol did not bring additional storage cost to the server and the client, but the entire file had to be retrieved during the verification executions. In addition, the communication complexity is linear in the file size. Deswarteet al.[23] and Filhoet al.[24] proposed techniques to verify data integrity using RSA-based hash functions. This schemes allowed a verifier to perform multiple challenges using the same metadata. The limitation of their algorithms lies in the computational complexity at the server. The computation costs must exponentiate the entire data file.

VII. PROOF OF OWNERSHIP.

Proof-of-ownership (POW) is closely related to two other similar problems: proof of retrievability (POR) and proof of data possession (PDP). POR schemes [15, 16, 18] are challenge-response protocols. In POR schemes, a cloud provider demonstrates the FIle retrievability (i.e., recoverability without any loss or corruption) to the client. PDP schemes [19] are similar protocols which only detect a large amount of corruption in outsourced data.

The main difference between POW and POR/PDP is that the latter usually uses a pre-processing step on the client side while the former does not.POW protocols are proposed for client-side data de-duplication which enables the storage server to store a single copy of repeating data. Client-side data de duplication has become popular and important as it removes data redundancy and data replication, but it brings many data privacy and security issues for the user. Douceuret al.[25] rst studied the problem of de duplication in a multi-tenant system in which de duplication had to be reconciled with confidentiality. Their proposed convergent encryption enabled two users to produce a single cipher text for de duplication. As there are many security problems with convergent encryption,

Storeretal.[26] proposed a security model for secure data de duplication. Recently,Harniket al.[27] formally identified the security problems of client-side de duplication as follows: 1) The first kind of attacks attempted to fool the storage server and abuse the storage system. A malicious user with the hash signature of a file could convince the cloud server that he owns the file. By accepting the hash value as a `proxy' for the entire file, the cloud server allowed anyone who held the hash value to access the entire le. 2) The second kind of attacks targeted the privacy and con dentiality of users of the storage system. A malicious user could check whether another honest user had already outsourced a data le by trying to upload it as well. 3) The third kind of attacks focused on subverting the intended use of a storage system. For example, two malicious users tried to use the cloud storage for a covert channel as they might not have a direct interaction channel. The two users first pre-agreed on two different files. Second, one malicious user outsourced one of the two files. Then the other user could detect which file had been de duplicated and output either 0 (for the first file) or 1 (for the second file). In this way, two malicious users successfully exchange a bit of information without a direct transmitting channel. To overcome such attacks, Haleviet al.[28] introduced the notion of POW for client-side deduplication. In addition, they presented Merkle tree-based schemes to allow a user to efficiently prove his ownership to the server, rather than some short information. However, their scheme cannot be adopted for encrypted file scenario, because encryption of the same file by different users with random keys results in different cipher texts. The server cannot store the same hash root value for the ownership verification.

VIII. CONCLUSION

In this paper we have studied various data security challenges such as key management, access control, searchable encryption techniques, remote integrity check and proof of ownership. we have analyzed that existing system has provision for improvement like *key management technique* can work perfectly, only if when all legitimate node users are authorized to access all the child nodes under the specific parent node. Were as outsourced database cannot remain private and secure. So there is a need for technique in which even if some selected data and key nodes are shared with other parties, database is still private and secure. *Access control* the major problem of defining and assigning keys to users based on different attribute sets is concerned. So there is a need for an encryption system to achieve, flexible and fine-grained access control on outsourced data. To reduce the search cost on encrypted data and there is a need to design a practical searching mechanism in cloud computing, an efficient *keyword search* scheme for cloud computing system. *Remote integrity check* proves that the public verifier cannot learn the target data; the tags themselves can leak some information about the data. So need for achieves public verifiability without disclosing any information. proof of a multiparty ownership solution need to be proposed with the encrypted data in the cloud.

REFERENCES

[1] Mell P., Grance T., 2011 NIST Special Publication 800-145: The NIST Definition of Cloud Computing. Available at: http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

[2] Chung Kei Wong, Mohamed G. Gouda, and Simon S. Lam. Secure group communications using key graphs. In SIGCOMM , pages 68-79,1998.

[3] D. Waller, E. Harder, and R. Agee. Key management for multicast:Issues and architectures. Technical report, RFC2627, 1999

[4] Weichao Wang, Zhiwei Li, Rodney Owens, and Bharat K. Bhargava.ecure and ecient access to outsourced data. In CCSW , pages 55-66,2009.

[5] Chung Kei Wong, Mohamed G. Gouda, and Simon S. Lam. Secure group communications using key graphs. In SIGCOMM , pages 68-79,1998

[6] Ernesto Damiani, Sabrina De Capitani di Vimercati, Sushil Jajodia, Stefano Paraboschi, and Pierangela Samarati. Balancing con dentiality and effciency in untrusted relational dbmss. InACM Conference on Computer and Communications Security, pages 93-102, 2003.

[7] Kevin D. Bowers, Ari Juels, and Alina Oprea. Hail: a high-availability and integrity layer for cloud storage. In ACM Conference on Computer and Communications Security , pages 187-198, 2009.

[8] Giuseppe Ateniese, Kevin Fu, Matthew Green, and Susan Hohenberger.Improved proxy re-encryption schemes with applications to secure distributed storage.IACR Cryptology ePrint Archive , 2005:28, 2005.

[9] Mikhail J. Atallah, Keith B. Frikken, and Marina Blanton. Dynamic and effient key management for access hierarchies. In ACM Conference on Computer and Communications Security , pages 190-202, 2005.

[10] Sabrina De Capitani di Vimercati, Sara Foresti, Sushil Jajodia, Stefano Paraboschi, and Pierangela Samarati. A data outsourcing architecture combining cryptography and access control. In CSAW , pages 63-69,2007.
[11] uojun Wang, Qin Liu, and Jie Wu. Hierarchical attribute-based encryption for ne-grained access control in cloud storage services. In ACM Conference on Computer and Communications Security , pages 735-737, 2010.
[12] Oded Goldreich and Rafail Ostrovsky. Software protection and simulation on oblivious RAMs.J. ACM , 43(3):431-473, 1996.
[13] Dawn Xiaodong Song, David Wagner, and Adrian Perrig. Practical techniques for searches on encrypted data. In IEEE Symposium on Security and Privacy , pages 44-55, 2000.
[14] Manuel Blum, William S. Evans, Peter Gemmell, Sampath Kannan,and Moni Naor. Checking the correctness of memories.Algorithmica,12(2/3):225-244, 1994.
[15] Ari Juels and Burton S. Kaliski Jr. Pors: proofs of retrievability for large les. In ACM Conference on Computer and Communications Security , pages 584-597, 2007.
[16] Kevin D. Bowers, Ari Juels, and Alina Oprea. Hail: a high-availability and integrity layer for cloud storage. In ACM Conference on Computer and Communications Security , pages 187-198, 2009.
[17] Yevgeniy Dodis, Salil P. Vadhan, and Daniel Wichs. Proofs of retrievability via hardness ampli cation. In TCC, pages 109-127, 2009.
[18] Hovav Shacham and Brent Waters. Compact proofs of retrievability.In ASIACRYPT , pages 107, 2008.
[19] Giuseppe Ateniese, Randal C. Burns, Reza Curtmola, Joseph Herring,Osama Khan, Lea Kissner, Zachary N. J. Peterson, and Dawn Song.Remote data checking using provable data possession. ACM Trans.Inf. Syst. Secur., 14:1-34, 2011.
[20] C. Christopher Erway, Alptekin K up↑ c u, Charalampos Papamanthou,and Roberto Tamassia. Dynamic provable data possession. In ACM Conference on Computer and Communications Security , pages 213-222, 2009.
[21] Mehul A. Shah, Mary Baker, Je rey C. Mogul, and Ram Swaminathan.Auditing to keep online storage services honest. In HotOS , 2007.
[22] Alina Oprea and Michael K. Reiter. Space efficient block storage integrity. In NDSS , 2005.
[23] Yves Deswarte, Jean Jacques Quisquater, and Ayda Saidane. Remote integrity checking. In Sushil Jajodia and Leon Strous, editors,Integrity and Internal Control in Information Systems VI , volume 140 of IFIP International Federation for Information Processing , pages 1-11.Springer Boston, 2004.
[24] Decio Luiz Gazzoni Filho and Paulo Sergio Licciardi Messeder Barreto.Demonstrating data possession and uncheatable data transfer.IACR Cryptology ePrint Archive , pages 150-159, 2006.
[25] John R. Douceur, Atul Adya, William J. Bolosky, Dan Simon, and Marvin Theimer. Reclaiming space from duplicate les in a serverless distributed le system. In ICDCS , pages 617-624, 2002.
[26] Mark W. Storer, Kevin M. Greenan, Darrell D. E. Long, and Ethan L.Miller. Secure data deduplication. In Storage SS, pages 1-10, 2008.
[27] Danny Harnik, Benny Pinkas, and Alexandra Shulman-Peleg. Side channels in cloud services: deduplication in cloud storage.IEEE Security & Privacy , 8(6):40-47, 2010.
[28] Shai Halevi, Danny Harnik, Benny Pinkas, and Alexandra Shulman-Peleg. Proofs of ownership in remote storage systems. In ACM Conference on Computer and Communications Security , pages 491-500,2011