International Journal For Technological Research In Engineering
Volume 1, Issue 10, June-2014 *ISSN (Online): 2347 - 4718*

PREVENTION OF INFORMATION LEAKAGE FROM INDEXING IN CLOUD

Mr Hardik Patel[1] (Student), Mr Indr jeet Rajput[2]
Deptartment of Computer Engineering
Hasmukh Goswami College of Engineering.
Gujarat, India.

Abstract: Cloud computing has to provide virtual, pay as your computing and storage services over the Internet, where the usage cost directly depends on as-needed basis. Content delivery is also one of the most important applications whose goal is to provide fast information retrieval with providing privacy. Also we know that in cloud handle of data management so it becomes onerous to manage and retrieval also incurs delay. While cloud computing is enlarge rapidly and it used by many private and public organizations internationally, data prevention issues in the cloud have not been carefully addressed at current stage. User have phobia of confidential data leakage and loss of privacy in the cloud may become a significant barrier to the wide adoption of cloud services. We explore a newly emerging problem of information leakage caused by indexing in the cloud. We design system that provides privacy to data using client virtual machine id through method call. We develop a portable technique to ensure strong actuation of users' privacy requirements at server side.

Index Terms: Cloud Computing, indexing, prevention techniques (Key words)

I. INTRODUCTION

Cloud computing is emerging concept. Cloud computing means which highly scalable, technology-enabled services can be easily consumed over the Internet on an as per our uses [1]. Cloud computing is a phrase used to describe a variety of computing concepts that involve a large number of computers connected through a real-time communication network such as the Internet [2]. As a metaphor for the Internet, "the cloud" is a familiar cliché, but when combined with "computing", the meaning gets bigger and fuzzier. Some analysts and vendors define cloud computing narrowly as an updated version of utility computing: basically virtual servers available over the Internet. It refers to network-based services, which appear to be provided by real server hardware, and are in fact served up by virtual hardware, simulated by software running on one or more real machines in vast group users, cross platforms and cross enterprise. All services are available if and only if internet is available. There are number of commercial and individual cloud computing services, e.g., from Amazon, Google, Microsoft, Yahoo, and Sales force and a top database vendors, like Oracle, are adding cloud support to their databases. The Cloud Computing includes a number of implementations, based on the services they provide, from application service provisioning, grid and utility computing, Services are like as software as a services(SAAS), Infrastructure as a services(IAAS) ,Platform as a service(PAAS)[2]. Cloud can be of individuals, organizations or enterprises, is processed remotely in unknown machines that users do not own or operate. This technology allows for much more efficient computing by centralizing storage, memory, processing and bandwidth. The efficiency of this approach is privacy and security risks. A significant barrier to of the cloud services is the user's fear of confidential data leakage and loss of privacy in the cloud, which may prove dangerous to many different types of cloud services. To user's privacy concerns, technical mechanisms which strongly implement user's privacy policies at any time anywhere when the user's data is being accessed by the service providers are desired. One of the interesting and very important privacy problems is caused by data indexing. Index may give information about amount of information about the data itself. Index are built after the cloud service provider receives the user's data and decides to build indexes to improve performance of searching, user does not know how data are stored in database. In this paper, we have tried to solve critical privacy problem caused by data indexing of data. Descend from root to leaf in a tree manner for data insertion and retrieval. When we insert a data at that time append client virtual machine Id with system generated index Id when data is marked as important. If data is marked as important client virtual machine id will be passed through method call else virtual machine will be null. When client request for data then check requested client virtual machine's Id with client Id which is append with data. If client id and virtual machine's Id is not matching then return null value. If client id and virtual machine's Id is match then get the data. In our approach does not introduce any further authentication technique to guarantee proper execution. It introduces little overhead. The rest of the paper is in order as follows. Section II introduces a brief background about the cloud. Section III gives the problem statement. Section IV presents our proposed mechanism.

II. BACKGROUND

A. The Cloud

In the Cloud Computing, Most cloud service provider are Amazon [4], Google [3], Microsoft [5], Saleforce.com [6] and Sun. But they are representing only small portion of providers. Other Cloud service provides is Proofpoint [7], Right Scale [8], and Workday [9]. Amazon (AWS) provide a number of infrastructure-related web services, Services provide by

Amazon are the Elastic Computing Cloud (EC2), Simple Storage Service (S3), CloudFront, SimpleDB and Simple Query Service (SQS). Amazon EC2 provides resizable compute capacity in the cloud. It is designed to make web-scale computing easier for developers and system administrators. EC2 provides deployment of applications by providing web services interfaces through which customers can create virtual machines. EC2's simple web service interface allows you to obtain and configure capacity with minimal friction. It provides complete control of your computing resources and run on Aws proven computing environment. EC2 provides developers the tools to build failure resilient applications and isolate themselves from common failure scenarios. S3 (Simple Storage Service) provides a fully redundant data storage infrastructure for storing and retrieving any amount of data, at any time, from anywhere on the Web. S3 is used to store and retrieve unlimited amount of a data. Amazon Glacier is an extremely low-cost storage service that provides secure and durable storage for data archiving and backup. Amazon EBS (Elastic Block Store) provides block level storage volumes for use with Amazon EC2 instances. Amazon EBS volumes are off-instance storage that persists independently from the life of an instance. Simple DB provides core database functions of data indexing and querying. SQS (simple query services) is a provide queue messaging service that supports the sending of messages via web services applications. While Amazon provides an infrastructure, Google App Engine and Microsoft Azure Services provide platforms as a service for building and hosting web applications on the web infrastructure. Amazon provides services Such as messaging; provide securing email systems, application development. For example, client can develop their applications by using the cloud services without their own infrastructure. Sun is offer an open cloud philosophy. Their Open Cloud Platform offer an infrastructure related to servers, storage and databases. Developers will access the Sun cloud services from a web browser to provision resources on their platform of choice. It will stand out for supporting the various operating systems, programming languages and virtual data center capabilities.

B. Privacy Issues in Cloud Computing

Cloud computing raises a range of important privacy issues by a number of recent work [10], [11], [12], [13]. in the cloud computing, user's data and applications reside on the cloud cluster which is maintained by a third party or a service provider. Cloud is not clear about that how their personal information is requested or how it will be used or Passed through other parties. According to privacy concerns of end users and then develop a simple solution. Privacy manager depend upon obfuscation techniques. In this technique user's private and important data is sent to cloud in the encrypted form, and further process is done on the encrypted data or information. When retrieve data at that time de- obfuscation techniques apply by the privacy manager to reveal of correct result. Privacy manager provide limited future for privacy. It does not provide guarantee to protect data. There are many techniques investigated in last few time. [14], [15], [16] and many cryptographic approaches for ensuring data integrity have been proposed. We like to mention that we do not provide the data storage security that only provides solution for privacy issue.

C. Indexing in Cloud

The most common technique for efficient search over distributed data is to build a centralized inverted index. In the technique, a set of documents that contain the term, and based on that term the searcher to obtain a list of matching documents. This scheme is used in web search engines [17]. This technique can support access controlled search by propagating access policies along with data to the indexing host. The index host must apply filtering technic search results appropriately. Filter provides high efficiency to searches in indexing. A centralized index exposes content providers to anyone who has access to the index structure. so hacker can easily access the data using a indexing structure and we have no longer trust for storing data on server. Hackers could lead to a complete and devastating privacy lost. Another alternative architecture is Decentralized indexing. In the decentralized indexing, used to identify a set of documents matching the searcher's query. Hosts are contacted directly by the searcher to retrieve the matching documents. Access control can be support their access policies before providing the documents. Index are hosted by entrusted machine whom the providers have no control. In order to overcome this issue, we have explored the possibility of creating private indexes by relying on predicate-based cryptography [18], [19].

D. Policy Enforcement

Our research on policy shares some equality with approaches aiming at protects objects using strong coupling policies and data. Sticky policies based on cryptographic algorithms to guarantee data protection [20]. Once the access rights are evaluated and access is granted, the user's data is fully available at the authorized party, and there is no way to control its usage. An approach to deploy sticky policies is based on identity based encryption technology, and requires trusted platforms to ensure accountability [21]. Encryption is applied only to text files in the predefined structure, policy is relatively easy to corrupt and a skilled hacker may tamper the file. So that makes the policy illegible. Security is required for guarantee the confidentiality of the data and the policy, our strategy is carried out by executable policies that are safely encoded. Although it's limitation may be overcome by adding auditing or monitoring techniques [22].

III. PROBLEM STATEMENT

A user has subscribed certain cloud service, usually that send his data and retrieve his data as well as associated access control policies to the service provider. When data is received by the service provider, they will have granted access rights on the data and access right are included like read, write and both(read, write). For the purpose of improving search performance, service provider construct index based on user data. This index contains sensitive keyword extracted from user's actual data and store in separate file. This index file is

not generally protected using access control policy. So they are chance to information leakage through indexing. Even If user's confidential data is stored with protection but sensitive information is disclose through unsecure index file. For user's data privacy, we aim to develop technique to satisfy the user requirement. A user may have different levels of privacy concerns over his data file. For example, only a portion of the user's data file contains sensitive information. That data is marked as important so for that data, user obtain the highest level of protection for that particular portion while allow indexing on other parts of the file. Therefore, the proposed techniques should be able to accommodate to a variety of data protection needs. Users have certain degree of ability to control the server's usage of indexing techniques over the data files. The proposed technique is flexible and adapts to possible modification of access control policies without requiring fully recompilations each time of policy update.

IV. A TECHNIQUE FOR PREVENTING INFORMATION LEAKAGE

To prevent information leakage from indexing, we use client id to solve the problem of leakage. Suppose we want to give prevention to data in indexing at that time data is mark as important data.

```
public void processEvent(SimEvent ev , Vm vm)
        {
                this.vm = vm;
                processEvent(ev);
        }
```

For important data this method will be called. This will set virtual machine and call process Event (SimEVent ev) method . If data is not mark as important no need to pass virtual machine object.

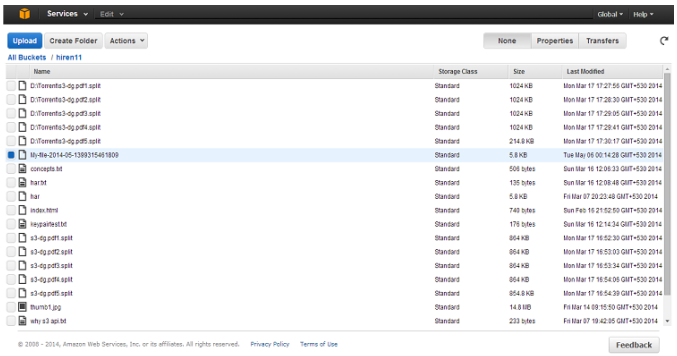
```
if(vm = = null)  { // if data is not mark as important no need
                        to pass virtual machine  object
                            return ;
                }
                secondId = vm.getId() + "_" + id;
                            // second id consist of both
```

if provide privacy to data then client vm id will be passed through method call otherwise virtual machine will be null. if vm is not null. i.e. data is important than append id with system generated indexing id .

When we want to retrieve data from the data server at that time get client Id from second Id.

```
int position = secondId.indexOf("_");
int clientid = Integer.parseInt(secondId.substring(0, position));
        if (vm == null || vm.getId() != clientid)
        {
                // other client is trying to read data
        return false;
        }
```

Check if vm == null i.e. client is not passing vm to take important data or client id and vm id is not matching that return null. i.e. though data is there client is not authenticated to take this so no need to return data else Give data. Server provide data when both Id match. Through this way we can provide prevention of information leakage from indexing in cloud.



In above figure show that my file-2014-05 name folder that contains information about File or data is append with client machine Id. That provides privacy to data in indexing in cloud.

V. CONCLUSION AND FUTURE WORK

In this paper, we explore the new technique data privacy Problems in the cloud caused by indexing. Data privacy issues in the cloud have been carefully addressed at current scenario. Our proposed system will let the user to be assured their data which is stored in cloud had no chance for leakage of information. in propose system virtual machine id match with Client virtual machine ID from second ID which is append with system generated index id of data. Compare both Id if match then give data otherwise not. This system will also provide the privacy in indexing from any unauthorized user tries to access the data. An unauthorized user can not access data. Finally, temporal based data binding approaches will be explored. We add more parameter to provide privacy in future.

REFERENCES

[1] T. Mather, S. Kumaraswamy, and S. Latif. Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance Privacy: An Enterprise Perspective on Risks and Compliance (Theory in Practice). O' Reilly, first edition, 2009.
[2] http://en.wikipedia.org/wiki/Cloud_computing
[3] Google Application Engine code.google.com/appengine/
[4] Amazon Web Services. http://aws.amazon.com/
[5] Microsoft Azure platform.http://www.microsoft.com/windowsazure/?wt.srch=1.
[6] Sales force. http://www.salesforce.com
[7] Proof point. http://www.proofpoint.com
[8] Right Scale Cloud Computing. Delivered. http://www.rightscale.com/.
[9] Workday. http://www.workday.com/
[10] Cavoukian. Privacy in the cloud. Identity in the information Society, 1,2008
[11] P.T.jaeger, J. Lin, and J. M. Grimes. Cloud computing and information policy: Computing in a policy cloud? journal of Information Technology and politics, 5(3), 2009.

[12] B.R Kandukuri, R. P. V., and A. Rakshit. Cloud security issues. In IEEE International Conference on services Computing (SCC), pages 517–520, 2009.
[13] L. M. Kaufman. Data security in the World of Cloud Computing. IEEE Security and Privacy, 7(4):61–64, 2009.
[14] M. Lillibridge, S. Elnikety, A. Birrell, M. Burrows, and M. Isard. A cooperative internet backup scheme. In USENIX Annual Technical Conference, pages 29–41, 2003.
[15] T.J .E. Schwarz and E. L. Miller. Store, forget, and check: Using algebraic signatures to check remotely administered storage. In IEEE International conference on Distributed Systems, page 12, 2006.
[16] Giuseppe Ateniese, Randal Burns, Reza Curtmola, Joseph Herring, Lea Kissner, Zachary Peterson, and Dawn Song. Provable data possession at untrusted stores. In Proc, of ACM conference on Computer and communications security, pages 598-609,2007.
[17] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In Seventh International World-Wide Web Conference (WWW 1998), 1998.
[18] Y.-C. Chang and M. Mitzenmacher. Privacy preserving keyword searches on remote encrypted data. February 2004.
[19] Zhiqiang Yang, Sheng Zhong, and Rebecca N. Wright. Towards privacy-preserving model selection. In PinKDD, pages 138–152, 2007.
[20] W. Tang. On using encryption techniques to enhance sticky policies enforcement. Technical Report (TR-CTIT-08-64), Centre for Telematics and Information Technology, 2008.
[21] M. C. Mont, S. Pearson, and P. Bramhall. Towards accountable management of privacy and identity information. In Proc. of the European Symposium on Research in Computer Security (ESORICS), pages 146–161, 2003.
[22] S. Pearson and A. Charlesworth. Accountability as a way forward for privacy protection in the cloud. Hewlett-Packard Development Company (HPL-2009-178), 2009.
[23] Anna Squicciarini, Smitha Sundareswaran, Dan Lin, Preventing Information Leakage from Indexing in the Cloud,978-0-7695-4130-3/10©2010 IEEE DOI0.1109 /CLOUD.2010 82.

