Data Swapping In Cloud Computing

Sachida Nanda Barik

School of computer engineering, KIIT UNIVERSITY, BHUBANESWAR, ODISHA
sachikjr@gmail.com

ABSTRACT: In the modern era of fast computing data storage and data retrieval is occurred due to cloud computing. Cloud computing is a pool of computing resources available on demand with the use of internet. it is a term used for the consumption of utility computing resources. it is used to maximize the effectiveness of the shared resources which is currently available on cloud. it relaxes the companies from investing money on heavy infrastructure rather they can used the same resources on shared basis..in the wide use of cloud computing threat comes to storage of data in the cloud server as it not known to the cloud user where there data will be stored they usually reluctant to store their data on cloud. to gain their trust cloud providers are facilitating various user confidentiality methods to stored their data..one technology is being used is data swapping. it is a strategy for protecting the micro data confidentially which is yet to be released. the purpose of data swapping is to create uncertainty in the minds of the intruder. it is a method to adjust the data values by swapping fraction of records between the records so that the intruder won't get the actual data rather gets scrambled data.

1. INTRODUCTION

Today we are in the era of fast computing. we are widely using the concept of distributed computing over the internet which is called cloud computing. Cloud computing is a technology which uses the internet or intranet and central remote servers to maintain the data and applications. now-a-days it is widely used to make simpler to daily life problems. cloud computing collects all related information and computing resources and manages them automatically according to the need by using various softwares.through vast data repository around the globe cloud computing gives smart and intelligent services to its users .by using the services of cloud computing the user can buy computing resources rather to set up vast infrastructure for their organisation without bothering about the setup complexity. the applications are upgraded very easily through internet. The user no needs to do it manually using the upgraded version software. Cloud computing is new utility which many enterprises wants to incorporates in order to improve their way of working. It implies sharing of computing resources to handle applications to make faster execution of their tedious works. by using cloud computing, organizations can use services of the provider and the data is being stored at any physical location outside their own control. It provides development of software environment, allocation and reallocation of existing resources when needed, vast storage and networking facility virtually. Cloud computing is privileged by virtualization technology . normally the term cloud computing, was influenced by the symbol of cloud which is generally represented in the Internet using workflow steps and systematic diagrams. A specific movement over the clouds is being occurred over the years with the end users since past, minutely keeping a increasing number of personal data which includes bookmarks, various photographs, unlimited music files and many more, on remote servers made available through a network.[1]

2. CLOUD COMPUTING OVERVIEW

The cloud computing model involves three kinds of functional units or components as stated below:

2.1. Cloud service provider: this is a company, which maintains many Cloud Storage Server(CSS),having large storage space to store the clients data and with high computation and greater accuracy power.its prime focus is to make profit and take care of sensitive information

2.2. **Client/owner:** It is an enterprises or entity or can be individual customer, whom have large data files to be stored in the cloud and relies on the cloud for data maintenance and computation. the client depends on cloud for maintaining their data.

2.3 User: user is an individual who is registered with the service provider and uses the data of the owners data stored on the cloud. The user may be an owner itself also. there are basically two models are available for cloud computing. these are service models and deployment models. under service models there are three types of models are available viz. Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Platform as a Service (PaaS). Infrastructure as a Service (IaaS) provides all necessary infrastructures on behalf of the customer. it offers on demand resources which are highly scalable which works on the principle of pay per use basis generally the amount of cloud resources being used by the customer. it enables the user to deploy and run various software including operating system software and application softwares.and the cloud user has mastery over various operating systems ,vast storage and already installed application and probably partial control over selected networking gadgets. Platform as a Service (PaaS). it provides a ready to use platform to the registered user to develop and manage customized web application having less complexity. here in cloud the user does not bother about underlying cloud infrastructure but it has the control over installed application and over the environment in which it is being installed. in Software as a Service (SaaS) the software's licence is subscribed for a certain period which is being accessed by the user. it gives the opportunity to use the service providers customized applications running on cloud. the cloud applications being accessed via many client gadgets through an user interface like accessing email on the web. it includes many application comes under SAAS are DBMS software ,MIS software,ERP software and Accounting software. in cloud computing architecture basically there are four deployment models are present .these are private cloud model. public cloud model ,community cloud model and hybrid cloud model. according to the definition of Private cloud is being built for a single organisation can gives better control and effectively secure the data. Generally it is of two kinds .first is on site which is hosted within the providers data centre and will be best for the application which needs complete security and another is off site which is hosted outside the data centre of the service provider.Public clouds are maintained and operated by other than the service provider but supported by the service provider. It offers low cost and pay per use model. all the customer is share the pool of infra-

structure within defined configuration. Hybrid cloud is the combination of both the private and public cloud .in this some resources are managed by the service provider and some are managed by the third party. it combines services and data from different cloud model to give a well managed and unified cloud environment. in community cloud the infrastructure is shared among many organisation having common computing goals which forms a specific users community. it is a multi tenant infrastructure [2].

3. CLOUD COMPUTING FEATURES

the benefits of cloud computing are cost effectiveness, unlimited storage, easy access of information and quick deployment.

3.1 Cost effective:-

now a days it is the most cost efficient method to use, maintain and upgrade the software which is being used by the companies. it reduced the cost of licence software for the company because in cloud, everything is available at much cheaper rates and hence, can significantly lower the company's IT expenses.

3.2 Unlimited storage

Cloud computing gives almost unlimited storage capacity given by various clod server companies like amazon web services.

3.3 Easy access of information

Once you stored the information on cloud u can easily access the information with any devices with internet connection.

3.4 Quick deployment

The software needed by the client can be easily downloaded from the web and it can easily installed in the client system within a matter of few minutes.

4. DRAWBACKS OF CLOUD COMPUTING

Besides advantages there were short comings also in cloud computing descibed as follows

4.1 Fluctuation in downloading speed

As the Service Providers of cloud handles a huge traffic of customers every time as a result they get overwhelmed and may face difficulties occurred due to technical glitches and unable to serve request. Business process is being halted and the client unable to access information in that time.

4.2 Prone to security

Although the service providers for Cloud service providers implemented the best and sophisticated security standards and various industry certifications for accommodating sensitive data and files on its external or agent service providers always prone to risks as intruders are always in search of loop holes.

4.3 Less data privacy

The client will always in search of a trust worthy service provider. Once the client upload their data on cloud then they lose direct access to their information and even don't know in which location and server their data is stored. the data can be modified by the service provider itself.

5. PRIVACY PRESERVATION IN CLOUD[7]

While accessing the information the privacy must be maintained as not every information is meant for public users. there are many techniques to preserve data privacy like anonymity method, public auditing schemes, method of multi-keyword ranked search and data combination privacy preservation in cloud to safe guard the interest of the client the service provider adopts many methodologies

6. DATA SWAPPING

After using all the methods by different author one more prevalent method for data privacy is to swap the data or data swapping. it is a disclosure limitation technique. as it is very old methodology used in data mining concept but it can be used in cloud computing to preserve the data values which is being stored in cloud from the hackers. the term Data swapping first introduced by Tore Dalenius and Steven Reiss in the year 1978 as one practical method for protecting disclosure among various data sets micro data (individual record) of databases. again this concept is being modified by the same author in the year 1982 . it preserves confidentiality by partially modifying the fraction of records among many rerecords in a particular database by switching record attributes with another record pair. it is one types of masking technology in which the data values are not known to the intruder as values always changes.it reduces the chances of deletion of data to safeguard the sensitive data from public.[8]The purpose of data swapping is to retain the amount of information, but randomly perturb data values to maintain confidentiality. Data swapping begins by selecting target records at random, then proceeds by finding a swapping partner for each target record with similar characteristics, and, finally, swaps data values between the target records and their swapping partners.[9] the data swapping procedure has many properties. some of them are as follows:-

1. it hides the relationship between the end user and the corresponding record.
2. this method can be applied on sensitive or important variable and no need to disturb the variable not related to this swapping.
3. it has designed to give data preservation where needed mostly..
4. implementation is very simple.

7. SWAPPING METHODS [3]

There are two data swapping methods being widely used like Targeted swapping strategy and random swapping strategy

7.1 Targeted swapping strategy

In this method first check the disclosure level having same control variable between the source and the target variable using a specified swapping rate.

7.2 Random swapping strategy

Here all the elements gets equal chance of being swapped by pairing of one element with other element having same control variable using the iterative loop procedure.

8. STEPS INVOLVED IN DATA SWAPPING [4]

Generally it has three steps

Step-I
first to decide whether to implement data swapping or not .if to use data swapping then it is to be used alone or combined with other available statistical disclosure control strategies.

Step-II
once the data swapping method is being employed disclosure risk function and data utility measure must be selected to calculate the risk associated with the data going to be swapped.[5]

Step-III
For using of data swapping release rate must be well specified by using the following attributes

(1). **swap rate:-**the piece or fraction of records in the database where swapping is being occurred.
(2). **swap attributes:-**the values of the attributes which is randomly exchanged
(3). **constraints:**-it is used to distinguished between swapped and un swapped attributes.

9. DATA SWAPPING ADVANTAGES [6]

Tthe data swapping approach has underlying advantages.

1. Data swapping hides the correct facts about every respective respondent.
2. It performs on all important variables (i.e., the variables whose attributes or values keeping together provides to the linking of records with its respondent),this method swapping abolishes all kinds of contact between their record and its respondent.
3. This process is very easy and need nothing ,just required couple of things a file containg micro data and a routine to generate arbitrary numbers at random number to implement. The programming is quite simple.
4. This swapping process could be applied on a selected sets of one (or more) variables, not disconcerting the response which meant for the insensitive and unidentified fields.
5. the swapping of regular variables delivers safety when most required. occasional and unique reactions are often used to determine respondents. These values are expected to be changed continuously . mostly occurred responses are less expected to be changing for the intruder and have a few chances to be altered during swap.
6. The process is not restricted or limited only to continuous or regular variables but to categorical variables (race, sex, occupation) can be swapped. Care must be taken while swapping occurs in categorical variables ; otherwise the usefulness of the files will be degraded as it loses true and correct information which in result will create a large number of strange combinations.

SUMMARY

Cloud computing is an emerging technology is being used by most of the organisations to save their cost of implementation and building infrasture deployment .the client can download the required software or the information rather to develop all the software or to implement in infrasture as is it is time savvy. as there are many advantages in cloud but there were also major drawbacks are also prevailing in cloud .the main drawback is to secure the information and privacy of data and to make the cloud a trustworthy for clients. to preserve the data over cloud various preservation methodology is being used by many authors .the concept about this proposal is to secure the information over cloud using data swapping methods. but to control the disclosure data swapping methods can be used .in cloud for protection of the data using dataswapping is that it will swap or interchange randomly the fraction of information with another randomly so that the intruder will not be sure that what they got that was the correct piece of information.

REFERENCES

[1] Jensen, Meiko, et al. "On technical security issues in cloud computing." Cloud Computing, 2009. CLOUD'09. IEEE International Conference on. IEEE, 2009.

[2] Mell, Peter, and Tim Grance. "The NIST definition of cloud computing." National Institute of Standards and Technology 53.6 (2009): 50.

[3] Shlomo, Natalie, Caroline Tudor, and Paul Groom. "Data swapping for protecting census tables." Privacy in statistical databases. Springer Berlin Heidelberg, 2010.

[4] Gomatam, Shanti, Alan F. Karr, and Ashish P. Sanil. "Data swapping as a decision problem." JOURNAL OF OFFICIAL STATISTICS-STOCKHOLM- 21.4 (2005): 635.

[5] Skinner, Chris J., and M. J. Elliot. "A measure of disclosure risk for microdata." Journal of the Royal Statistical Society: series B (statistical methodology) 64.4 (2002): 855-867.

[6] Moore Jr, Richard A. "BUREAU OF THE CENSUS STATISTICAL RESEARCH DIVISION."

[7] Wang, Jian, et al. "Providing privacy preserving in cloud computing." Test and Measurement, 2009. ICTM'09. International Conference on. Vol. 2. IEEE, 2009.

[8] Swapping, Categorical Data. "NISS." (2003).

[9] Krenzke, Tom, et al. "Data Coarsening and Data Swapping Algorithms."

[10] Reiss, Steven P. "Data-swapping--A technique for disclosure control." J. Statistical Planning and Inference 6.1 (1982): 73-85.

[11] Fienberg, Stephen E., and Julie McIntyre. "Data swapping: Variations on a theme by dalenius and reiss." Privacy in statistical databases. Springer Berlin Heidelberg, 2004.

[12] Datta, Souptik. On random additive perturbation for privacy preserving data mining. Diss. University of Maryland, 2004.

[13] Carlson, Michael, and Mickael Salabasis. "A data-swapping technique using ranks—a method for disclosure control." Research in Official Statistics 6.2 (2002): 35-64.