SECURE CLOUD DATA DEDUPLICATION WITH EFFICIENT RE-ENCRYPTION

FATHIMA LIBA P[1], JAHANA THASNIM CK[1] ,FATHIMA JUBIN PK [1],HANEENA T [1] , Mr. SHIHABUL HAQ. M [2]

[1]BSC.Computer Science, KAHM Unity Women's College, Manjeri, Malappuram, Kerala, India

[2] Assistant Professor , Department Of Computer Science ,KAHM Unity Women's College Manjeri, Malappuram, Kerala, India

Corresponding author: shihabulhaq2010@hotmail.com

ABSTRACT

Data deduplication technique has been widely adopted by commercial cloud storage providers, which is both important and necessary in coping with the explosive growth of data. To further protect the security of users sensitive data in the outsourced storage mode, many secure data deduplication scheme have been designed and applied in various schemes. Among these schemes, secure and efficient re-encryption for encrypted data deduplication is the best technique.

INTRODUCTION

After the emergence of the cloud architecture, many companies migrate their data from conventional storage i.e., on bare metal to the cloud storage. Since then huge amount of data was stored on cloud servers, which later resulted in redundancy of huge amount of data. Hence in this cloud world, many data de-duplication techniques has been widely used. Not only the redundancy but also made data more secure and privacy of the existing data were also increased. Some techniques got limitations and some have their own advantages based on the requirements. Some of the attributes like data privacy, tag regularity and interruption to brute force attacks. To make data reduplication technique more efficient based on the requirements. This paper will discuss schemes that brace user-defined access control, by allowing the service provider to get information of the information owners. Thus our scheme eliminates redundancy of the data without breaching the privacy and security of clients that depends on service providers. Our latest deduplication scheme after performing various algorithms resulted in conclusion and producing more efficient data confidentiality and tag consistency. This paper has discussion on various techniques and their drawbacks for the effectiveness of the reduplication.

ISBN Number : 978-93-340-1632-1

METHODOLOGY

AGILE methodology is a practice that promoted continuous iteration of development and testing throughout the software development lifecycle of the project. Both development and testing activities are concurrent unlike the Waterfall model. Agile software development emphasizes in four values.

SYSTEM ANALYSIS

System analysis is a general term that refers to an orderly, structured process for identify and solving a problem. The system analysis process is calling the life cycle methodology, since it relates to four significant phases in the life cycle of all business information system: study, design, development and operation. The definition of system analysis includes not only the process but also the process of putting together to form a new system. A system analyst is an individual who performs system analysis during any, or all, of the life cycle phases of a businessinformation system. The system analyst not analyses business information system problems, butalso synthesizes new to solve those problem or to meet other information needs.

The various techniques used in the study of the present system are:

- Observation
- Interviews
- Site visits
- Discussion

EXISTING SYSTEM

A number of deduplication systems have been proposed based on various reduplication strategiessuch as client-side or server-side reduplications, file-level or block-level reduplications. Bellaire et al formalized this primitive as message-locked encryption, and explored its application in space efficient secure outsourced storage. There are also several implementations of convergent implementations of different convergent encryption variants for secure deduplication. Li addressed the key-management issue in block-level deduplication by distributing these keys across multiple servers after encrypting the files. Showed how to protect data confidentiality by transforming the predictable message into a unpredictable message. Data reliability is actually a very critical issue in a deduplication storage system because there is only one copy for each file stored in the server shared by all the owners. Most of the previous deduplication systems have only been considered in a single-server setting.

The traditional deduplication methods cannot be directly extended and applied in distributed and multi-server systems.

PROPOSED SYSTEM

Our proposed constructions support both file-level and block-level deduplications. Security analysis demonstrates that the proposed deduplication systems are secure in terms of the definitions specified in the proposed security model. In more details, confidentiality, reliability and integrity can be achieved in our proposed system. Two kinds of collusion attacks are considered in our solutions. These are the collusion attack on the data and the collusion attack against servers. In particular, the data remains secure even if the adversary controls a limited number of storage servers.

SYSTEM REQUIREMENT SPECIFICATION

Requirement analysis is a software engineering task that bridges the gap between system level software designs. We have done the requirement analysis in order to understand the problem faced in our objectives. The emphasis in requirement analysis is on identifying from the system, not how the system will achieve this goal.

SYSTEM SPECIFICATION

Hardware Specification

The selection of hardware is very important in the existence and proper working of any of the software. When selecting hardware, the size and capacity requirements are also important. The hardware must suit all application developments.

Processor : i3 or above.

System Bus : 32Bit or 64Bit

RAM : 4 GB or Above

HDD : 500 GB or Above

Monitor : 14” LCD or Above

Key Board : 108

Keys Mouse : Any Type of mouse

Software specification

One of the most difficult tasks is selecting software, once the system requirement is find out then we have to determine whether a particular software package fits for those system requirements. This system summarizes the application requirement.

Operating System : Windows 10 Any 32 bit or 64 bit platform

Front End : Python

Back End : My SQL Sever

IDE :Python 3.6 or Above

:PyCharm

SYSTEM DESIGN

System design is the second phase of the system lifecycle. The detailed design of the system selected in the study phase is accomplished in the design phase. The principal activity performed during this phase includes allocation of function between computer programs equipment and manual tasks and data base design and test requirement definition. In the course of design phase, the performance specification is expanded into the design specification. The user oriented baseline prepared in study phase becomes a base line document, oriented to the needs of the programmers and other professional who will actually develop the system. A design phase report is prepared after the completion of design phase activities and the review is held with the user organization in order to determine whether or not the computer based business information system project is ready to the development phase.

INPUT DESIGN

Input design is a part of the overall system design, which requires very careful attention. Often the collection of input data is the most expensive part of the system, in terms of both the equipment used and people involved. If the data going into the system is incorrect, then the processing and out put will magnify the errors. Thus the clear objectives of input design are:

To produce a cost-effective method of input.

To achieve the highest possible level of accuracy.

To ensure that the input is acceptable to and understood by the user.

OUTPUT DESIGN

The output design is done so that the result of processing could be committed to the user and to

provide a hard copy of these results and evaluations for later consultations. Effective output design will improve the clarity and performance of outputs. Output design phase of the system is concerned with the convergence of information' s to the end user friendly manner. The output design should be efficient, intelligible so that system relationship with the end user is improved and there by enhancing the process of decision making. Outputs from the computer systems are

required primarily to communicate the results of the processing to the users. They are also used to provide a permanent copy of these results of processing to the users. They are also used to provide a permanent copy of these results for late consultation. There are various types of output

required by most systems, the main ones are:

External outputs: whose destinations outside the organization and which require special attention because they project the image of the organization, Internal outputs: whose destination is within the organization and which require careful design because they are the user's main interface with the computer. Operational outputs: whose use is purely within the computer department. Turn around outputs, to which the data will beaded before they are returned to the computer for further processing.

FRONT END

- Python
- HTML
- PyCharm

BACK END

- My SQL

SYSTEM IMPLEMENTATION

System implementation is the final stage of software development life cycle. For the successful implementation and cooperation of new systems users must be selected, educated and trained. Unless the users are not trained, the system will become complex it will become feel as a burden for them. A software implementation method is a systematically structured approach to

effectively integrate software based service or component into the workflow of an organizational structure or an individual end-user. A software implementation method is a blueprint to get users and/or organizations running with a specific software product. The method is a set of rules and views to cope up with the most common issues that occur when implementing a software product: business alignment from the organizational view and acceptance from human view. It is stated that the implementation of software consumes up to 1/3 of the budget of a Software purchase. The complexity of implementing product differs on several issues. Examples are: the number of end users that will use the product, the effects that the implementation has on changes of tasks and responsibilities for the end user, the culture and the integrity of the organization where the software is going to be used and the budget get available for acquiring the software. The implementation stage of the system being by preparing a plan for implementation of the system. According to this plan, activities are to be carried out, discussions are to be made regarding the equipment to be required, resources and additional facilities required implementing the system. The most critical stage in achieving a successful system is by giving users confidence that the system will work based on their requirements and be effective. This method also offers the greatest securities since the old system can take over if the errors are found or inability to handle certain transactions while using the new system.

The implementation involves the following formalities:

- Carefull planning
- Investigation of the system and constraints
- Design the methods to achieve the changes
- Training the staffs in the changed phase
- Evaluation of the changeover method

IMPLEMENTATION

Implementation of the system refers to the final installing of the package in its real environment , to the satisfaction of the indeed users and the operation of the system. It is the process of converting a new or revised system design to operation. It is the key stage in achieving successful new system. The process of putting the developed system in actual use is called system implementation. This includes all those activities that take place to convert from the old system to new system. It must therefore be carefully planned and controlled. Proper guidance should be imparted to the users so that he is comfortable in using the application.

CONCLUSION

We proposed the distributed de duplication systems to improve the reliability of data while achieving the confidentiality of the users outsourced data without an encryption mechanism. Four constructions were proposed to support file level de duplication. The security of tag consistency and integrity were achieved.

REFERENCE

- X. Chen, J. Li, J. Weng, J. Ma and W. Lou, "Verifiable computation over large database with incremental updates", *IEEE Trans. Comput.*, vol. 65, no. 10, pp. 3184-3195, Oct. 2016.
- H. Yuan, X. Chen, T. Jiang, X. Zhang, Z. Yan and Y. Xiang, "DedupDUM: Secure and scalable data deduplication with dynamic user management", *Inf. Sci.*, vol. 456, pp. 159-173, 2018.
- T. Jiang, X. Chen, J. Li, D. S. Wong, J. Ma and J. K. Liu, "TIMER: Secure and reliable cloud storage against data re-outsourcing", *Proc. 10th Int. Conf. Inf. Security Practice Experience*, pp. 346-358, 2014.
- J. Wang, X. Chen, J. Li, K. Kluczniak and M. Kutylowski, "TrDup: Enhancing secure data deduplication with user traceability in cloud computing", *Int. J. Web Grid Services*, vol. 13, no. 3, pp. 270-289, 2017
- J. Hur, D. Koo, Y. Shin and K. Kang, "Secure data deduplication with dynamic ownership management in cloud storage", *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 11, pp. 3113-3125, Nov. 2016.
- M. Bellare, S. Keelveedhi and T. Ristenpart, "DupLESS: Server-aided encryption for deduplicated storage", *Proc. 22nd USENIX Conf. Security*, pp. 179-194, 2013.
- Websites:
  - www.wikipedia.com
  - www.learnvisualstudio.net
  - www.google.com
  - www.codeproject.com