Retrieval of Data from Cloud Storage Using Asymmetric Group Key Agreement

J. RADHAKRISHNAN[1], R. PRAVAL RAAJ[2], R. PRAKASH[3]
[1, 2, 3] Computer science engineering department, DR. MGR Educational and Research Institute, Chennai, India

Abstract- Cloud storage allows users in the shared group to upload and access data in the cloud. The storage in cloud has become a growing fashion these days that boosts the secure remote information auditing. Latterly, some analysis thought-about in shared dynamic information that there's a haul of secure and adequate public information integrity auditing. Nevertheless, there are schemes in sensible storage systems that don't seem to be secure against the connivance of cloud storage server and annul cluster user through user revocation which supports the users. Alternately our project plan is based on the verifier-local revocation cluster signature, the connivance attack in existing scheme is the work out associated equipped an adequate public integrity auditing scheme with secure cluster user revocation with a signature or group key, where the group key signature provides designed re-signature. A haul of secure and adequate public information integrity auditing scheme is the work out associated equipped an adequate public integrity auditing scheme with secure cluster user revocation with a signature

I. INTRODUCTION

Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network. The name comes from the common use of a cloud-shaped symbol as an abstraction for the complex infrastructure it contains in system diagrams. Cloud computing entrusts remote services with a user's data, software and computation. Cloud computing consists of hardware and software resources made available on the Internet as managed third-party services. These services typically provide access to advanced software applications and high-end networks of server computers.

The goal of cloud computing is to apply traditional supercomputing, or high-performance computing power, normally used by military and research facilities, to perform tens of trillions of computations per second, in consumer-oriented applications such as financial portfolios, to deliver personalized information, to provide data storage or to power large, immersive computer games.

The cloud computing uses networks of large groups of servers typically running low-cost consumer PC technology with specialized connections to spread data-processing chores across them. This shared IT infrastructure contains large pools of systems that are linked together. Often, virtualization techniques are used to maximize the power of cloud computing.

To reduce the overhead at data owner side many researchers proposed different strategies. These strategies ensure the security and privacy of the cloud data but do not support dynamic data operations such as insertion, deletion, append and update in single and multi-user cloud environment. These strategies are more productive only when the respectability of the data is checked by the public verifier.

Therefore, in cloud storage dynamic data operations are very frequently used to reduce the computational cost. In the present work, the privacy-preserving public integrity auditing scheme with algebraic signature is used in conjunction with dynamic data operations such as insertion, deletion, modification, append and update for effectual computational cost shredding in cloud environment

II. RETRIEVAL OF DATA

Data retrieval is the retrieval of items (objects, Web pages, documents, etc.) which satisfy specific conditions set in a regular expression like query. While IR aims at satisfying a user information need usually expressed in natural language, data retrieval aims at determining which documents contain the exact terms of the user queries Instead of saving and storing files on a computer and accessing them with a type of drive, Cloud storage allows users to store and access and work in files online simultaneously. Explore the definition and concept of Cloud storage, review some examples of Cloud storage, and compare the benefits and disadvantages of Cloud storage.

III. JAVA

The Java programming language is a high-level language that can be characterized by all of the following buzzwords:

- Simple
- Architecture neutral
- Object oriented
- Portable
- Distributed
- High performance
- Interpreted
- Multithreaded
- Robust
- Dynamic
- Secure

With most programming languages, you either compile or interpret a program so that you can run it on your computer. The Java programming language is unusual in that a program is both compiled and interpreted. With the compiler, first you translate a program into an intermediate language called Java byte codes the platform-independent codes interpreted by the interpreter on the Java platform. The interpreter parses and runs each Java byte code instruction on the computer. Compilation happens just once; interpretation occurs each time the program is executed. The following figure 2.4.1 illustrates how this works

IV. REQUIREMENTS SPECIFICATION

Hardware Requirements

Processor:	Pentium Dual Core 2.3 GHz or Higher
Hard Disk:	20 or Higher
Ram:	8 GB (Min)
Monitor:	15” VGA Colour
Other peripherals:	Mouse & Keyboard

Software Requirements

Operating System:	Windows 10
Language:	Java
Database:	MySQL
IDE:	NetBeans 7.4

V. TECHNOLOGY USED

The Java Programming Language

The Java programming language is a high-level language that can be characterized by all of the following buzzwords:

- Simple
- Architecture neutral
- Object oriented
- Portable
- Distributed
- High performance
- Interpreted
- Multithreaded
- Robust
- Dynamic
- Secure

With most programming languages, you either compile or interpret a program so that you can run it on your computer. The Java programming language is unusual in that a program is both compiled and interpreted

The Java Platform

A platform is the hardware or software environment in which program runs. We’ve already mentioned some of the most popular platforms like Windows 2000, Linux, Solaris, and MacOS. Most platforms can be described as a combination of the operating system and hardware. The Java platform differs from most other platforms in that it’s a software-only platform that runs on top of other hardware-based platforms.

The Java platform has two components:

- The Java Virtual Machine (Java VM)
- The Java Application Programming Interface (Java API)

The Java API is a large collection of ready-made software components that provide many useful capabilities, such as graphical user interface (GUI) widgets. The following figure depicts a program that's running on the Java platform. As the figure 2.4.2 shows, the Java API and the virtual machine insulate the program from the hardware

VI. LITERATURE SURVEY

SURVEY PAPER 1

TITLE:	PUBLIC INTEGRITY AUDITING FOR SHARED DYNAMICS CLOUD STORAGE
AUTHOR:	M. RABIN
YEAR:	2020

An Information Dispersal Algorithm (IDA) is developed that breaks a file F of length L = ↿ F↾ into n pieces Fi, l ≤ i ≤ n, each of length ↿Fi↾ = L/m, so that every m pieces suffice for reconstructing F. Dispersal and reconstruction are computationally efficient. The sum of the lengths ↿Fi↾ is (n/m) · L. Since n/m can be chosen to be close to l, the IDA is space efficient. IDA has numerous applications to secure and reliable storage of information in computer networks and even on single disks, to fault-tolerant and efficient transmission of information in networks, and to communications between processors in parallel computers. For the latter problem provably time-efficient and highly fault-tolerant routing on the n-cube is achieved, using just constant size buffers.

MERITS:

- Supports dynamic data operations.

DEMERITS:

- Need an automatic classification algorithm.

Multiple CPS problems.

SURVEY PAPER 2

TITLE:	DYNAMIC PROVABLE DATA POSSESSION
AUTHORS:	C. ERWAY, A. KUPCU
YEAR:	2016

We consider the problem of efficiently proving the integrity of data stored at untrusted servers. In the provable data possession (PDP) model, the client pre-processes the data and then sends it to an untrusted server for storage, while keeping a small amount of meta-data. The client later asks the server to prove that the stored data has not been tampered with or deleted (without downloading the actual data). However, the original PDP scheme applies only to static (or append-only) files. We present a definitional framework and efficient constructions for dynamic provable data possession (DPDP), which extends the PDP model to support provable updates to stored data. We use a new version of authenticated dictionaries based on rank information. The price of dynamic updates is a performance change from O(1) to O(logn) (or O(n$^{\varepsilon}$log n), for a file consisting of n blocks.

MERITS:

- Dynamic data possession is possible.

DEMERITS:

- Does not provide batch auditing.

Works only with constant size integrity proof.

SURVEY PAPER 3

TITLE:	EFFICIENT DISPERSAL OF INFORMATION SECURITY
AUTHORS:	C. Wang, Q. Wang, K. Ren, and W. Lou
YEAR:	2015

Cloud Computing is the long-dreamed vision of computing as a utility, where users can remotely store their data into the cloud so as to enjoy the on-demand high quality applications and services from a shared pool of configurable computing resources. By data outsourcing, users can be relieved from the burden of local data storage and maintenance. However, the fact that users no longer have physical possession of the possibly large size of outsourced data makes the data integrity protection in Cloud Computing a very challenging and potentially formidable task, especially for users with constrained computing resources and capabilities which extends the PDP model to support provable updates to stored data.

MERITS:

- Provides facility to verify cloud data storage with multiple replicas.
- Supports public and dynamic auditing.

DEMERITS:

- Communication cost is greater.
- High computation cost.

VII. PROPOSED SYSTEM

- The deficiency of above schemes motivates us to explore how to design an efficient and reliable scheme, while achieving secure group user revocation. To the end, we propose a construction which not only supports group data encryption and decryption during the data modification processing, but also realizes efficient and secure user revocation.
- Our idea is to apply vector commitment scheme over the database. Then we leverage the Asymmetric Group Key Agreement (AGKA) and group signatures to support ciphertext data base update among group users and efficient group user revocation respectively.
- Specifically, the group user uses the AGKA protocol to encrypt/decrypt the share database, which will guarantee that a user in the group will be able to encrypt/decrypt a message from any other group users. The group signature will prevent the collusion of cloud and revoked group users, where the data owner will take part in the user revocation phase and the cloud could not revoke the data that last modified by the revoked user.

VIII. SYSTEM DESIGN AND IMPLEMENTATION

SYSTEM DESIGN

System design is the process of defining the architecture, components, modules, interfaces and data for the system to satisfy specified requirements. System design could be seen as the application of systems theory to product development. There is some overlap with the disciplines of systems analysis, system architecture and systems engineering

SYSTEM OVERVIEW

System design is the process of planning a new business system or one to replace or component an existing system but before this planning can be done, we must thoroughly understand the old system and determine can best be used to make its operation more effective

SYSTEM IMPLMENTATION

The principle aphorism of ring signatures is to so hide the identity of the underwriter on each square all together to keep private and delicate data un-revealed to open verifier. In any case, the conventional ring marks does not bolster square less verifiability thus the verifier needs to download the whole information from the cloud to check the accuracy of the common information which thus expends more data transmission and additional time. In this manner, it outlines another homomorphic authenticable ring mark (HARS) plot, which is stretched out from great ring mark conspire. HARS produced ring marks are most certainly not just ready to safeguard character protection but on the other hand can bolster piece less verifiability

SYSTEM MODULES

The modules of this project are

1. Registration and Login
2. Group signature
3. Revoked group user
4. Public auditing

IX. TESTING

Testing is consider to be the least creative phase of the whole cycle of system design. In the real sense it is the phase, which helps to bring out the creativity of the other phase makes it shine. Testing is a process of executing a program with intent of undiscovered error. A successful test is one that uncover an undiscovered error.

TYPES OF TESTS

- WHITE BOX TESTING

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

- BLACK BOX TESTING

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box you cannot see into it. The test provides inputs and responds to outputs without considering how the software works

- FUNCTIONAL TEST

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals. Functional testing is centred on the following items:
Valid Input : identified classes of valid input must be accepted.
Invalid Input : identified classes of invalid input must be rejected.
Functions : identified functions must be exercised.
Output : identified classes of application outputs must be exercised.

In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

- INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

X. OBJECTIVES

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.
2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.
3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus, the objective of input design is to create an input layout that is easy to follow.

OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each

output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2. Select methods for presenting information.
3. Create document, report, or other formats that contain information produced by the system.

RESULTS

All the test cases mentioned above passed successfully. No defects encountered.

CONCLUSION

In summary, an efficient public integrity auditing scheme for data storage in cloud computing based on the algebraic signature that also supports dynamic operations like insertion, append, deletion, and update is developed. The scheme incurs a minimum computational cost at data owner side compared to recently reported literature. The results indicate that proposed method has the potential to be extended for an integrity auditing system for large archival files in distributed cloud storage systems and a system with data traceability.

5.2 FUTURE ENHANCEMENT

- To increase the speed and accuracy of data updation.
- To reduce the cost of data retrieval.
- To reduce the resource requirement.