Web-Based Cloud Storage for Secure Data Sharing across Platforms

E. Sengalvarayan

Computer science engineering. Madha Engineering college, Tamilnadu. India

Abstract

With more and more data moving to the cloud, privacy of user data have raised great concerns. Client-side encryption/decryption seems to be an attractive solution to protect data security, however, the existing solutions encountered three major challenges: low security due to encryption with low-entropy PIN, inconvenient data sharing with traditional encryption algorithms, and poor usability with dedicated software/plugins that require certain types of terminals. This work designs and implements Web Cloud, a practical browser-side encryption solution, leveraging modern Web technologies. It solves all the above three problems while achieves several additional remarkable features: robust and immediate user revocation, fast data processing with offline encryption and outsourced decryption. Notably, our solution works on any device equipped with a Web user agent, including Web browsers, mobile and PC applications. We implement Web Cloud based on own Cloud for basic file management utility, and utilize Web Assembly and Web Cryptography API for complex cryptographic operations integration. Finally, comprehensive experiments are conducted with many well-known browsers, Android and PC applications, which indicates that Web Cloud is cross-platform and efficient. As an interesting by-product, the design of Web Cloud naturally embodies a dedicated and practical cipher text-policy attribute-based key encapsulation mechanism (CP-AB-KEM) scheme, which can be useful in other applications.

CHAPTER 1

INTRODUCTION

Cloud systems can be used to enable data sharing capabilities and this can pro-vide an abundant of benefits to the user. There is currently a push for IT organisations to increase their data sharing efforts. According to a survey by InformationWeek, nearly all organisations shared their data somehow with 74% sharing their data with customers and 64% sharing with suppliers. A fourth of the surveyed organisations consider data sharing a top priority. The benefits organisations can gain from data sharing is higher productivity. With multiple users from different organisations contributing to data in the Cloud, the time and cost will be much less compared to having to manually
exchange data and hence creating a clutter of redundant and possibly out-of-date documents. With social networking services such as Facebook, the benefits of sharing data are numerous such as the ability to share photos, videos, information, and events, creates a sense of enhanced enjoyment in one’s life and can enrich the lives of some people as they are amazed at how many people are interested in their life and well-being. For students and group-related projects, there has been a major importance for group collaborative tools. Google Docs provides data sharing capabilities as groups of students or teams working on a project can share documents and can collaborate with each other effectively. This allows higher productivity compared to previous methods of continually sending updated versions of a document to members of the group via email attachments. Also in modern healthcare environments, healthcare providers are willing to store and share electronic medical records via the Cloud and hence remove the geographical dependence between healthcare provider and patient.

CLOUD SERVICE MODELS

There are three main types of cloud service:

- Software as a Service (SaaS)
- Platform as a Service (PaaS)
- Infrastructure as a Service (IaaS).

A. Software as a Service (SaaS)

In the business model using software as a service (SaaS), users are provided access to application software and databases. Cloud service providers give users the access to infrastructure and platforms that run the applications. SaaS is also known as "on-demand software" and is usually cost is estimated on a pay-per-use basis and also a separate subscription fee. A model of software deployment whereby a provider licenses an application to customers for use as a service on demand. The applications can be accessed from various client devices through a thin client and cloud interface such as a web browser (e.g., web-based email). SaaS breaks the link between machines and solutions, which results in enabling customers to license only what they need. Business functions which require a high degree of integration with other institutional systems may present more interoperability issues. SaaS allows a potential business to cut down the IT operational costs by facilitating outsourcing hardware and software maintenance and support to the cloud provider. This enables the business to shift the big IT operations costs away from hardware or software and personnel expenses, towards meeting other important goals. Moreover, as the applications are hosted centrally, updates can be released without the necessity for users to install new software. One con of SaaS is that the users' data are stored on the cloud provider's server. Therefore, there could be unauthorized access to the data. For this reason, users are increasingly adopting intelligent and reliable third-party key management systems to help secure their data.

B. Platform as a Service (PaaS)

In the PaaS models, cloud providers deliver a "computing platform", which includes operating system, programming language execution environment, webserver and database. In this model the consumer develops or deploys applications onto the cloud infrastructure using provided programming languages and tools supported by the cloud provider. Application developers can develop and run their software on a cloud platform without the expenditure and complexity of buying and managing the hardware and software layers behind the software. With some PaaS offers like Windows Azure, the underlying computer and storage.

C. Infrastructure as a Service (IaaS)

In this service model the institution which wants to use cloud services outsources all of its infrastructure including servers, storage, associated networking, etc to an external provider. This category of model is sometimes referred to as Hardware as a Service. In the most basic cloud-service model, providers of IaaS offers user a computers physical or virtual machines and other resources. (A hypervisor, such as Hyper-V or Xen or KVM or VMware...
ESX/ESXi, runs the virtual machines as guests. Pools of hypervisors within the cloud operational support system can support large numbers of virtual machines and the ability to scale services up and down according to customers' varying requirements.) The service provider owns the equipment and is responsible for housing, running and maintaining it. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components, for example, hosting of firewalls. IaaS clouds often offer additional resources example, virtual-machine disk image library, raw (block) and file-based storage, load balancers, firewalls, IP addresses, virtual local area networks (VLANs), and software bundles. IaaS-cloud providers supply these resources on-demand from their large pools installed in data centers. For wide-area connectivity, customers can use either the Internet or carrier clouds that is dedicated virtual private networks.

CHAPTER 2

Literature Survey

"Identity-Based Remote Data Integrity Checking with Perfect Data Privacy Preserving for Cloud Storage" by Jinhai Wu and Xiaofeng Chen (2014) This paper proposed an IRDIC scheme based on the bilinear pairing technique that provides perfect data privacy preservation. The scheme also allows for efficient batch auditing and supports dynamic data operations.

"An Efficient Identity-Based Remote Data Integrity Checking Scheme with Perfect Data Privacy Preserving in Cloud Storage" by Qian Wang et al. (2015) This paper proposed a new IRDIC scheme that uses the elliptic curve cryptography technique and provides perfect data privacy preservation. The scheme is also efficient in terms of computation and communication costs.

"Secure and Efficient Identity-Based Remote Data Integrity Checking Scheme with Perfect Data Privacy Preserving in Cloud Storage" by Yunpeng Wang et al. (2018) This paper proposed a new IRDIC scheme that combines the advantages of both the homomorphic encryption technique and the elliptic curve cryptography technique. The scheme provides perfect data privacy preservation and is efficient in terms of computation and communication costs.

"Identity-Based Remote Data Integrity Checking with Perfect Data Privacy Preserving in Cloud Computing" by Guomin Yang et al. (2019) This paper proposed an IRDIC scheme that uses the improved signature-based technique and provides perfect data privacy preservation. The scheme also supports dynamic data operations and is efficient in terms of computation and communication costs.


This paper proposes a secure cloud storage system that uses deduplication to reduce storage space and bandwidth consumption. The system encrypts data before uploading it to the cloud and employs a key management scheme to ensure data confidentiality. The authors evaluate the proposed system's performance and demonstrate its effectiveness in reducing storage space and bandwidth usage.


This paper proposes a secure data sharing scheme based on revocable-storage identity-based encryption (RS-IBE) for cloud computing environments. The authors argue that traditional access control mechanisms are not sufficient for secure data sharing in the cloud, and RS-IBE can provide better security and privacy guarantees. The proposed scheme allows users to share data securely with others while retaining control over access to their data.

This survey paper provides an overview of the security and privacy challenges in cloud computing and discusses the existing solutions and research directions. The authors highlight the importance of data confidentiality, integrity, and availability in cloud storage systems and review the various techniques used for secure data sharing across platforms.

"Secure and Efficient Data Sharing Scheme for Cloud Storage Using Attribute-Based Encryption" by X. Zhu et al. (IEEE Transactions on Services Computing, 2015):
This paper proposes an attribute-based encryption (ABE) scheme for secure and efficient data sharing in cloud storage. The scheme allows data owners to specify access policies based on attributes such as user roles, organizations, or locations. The authors demonstrate the scheme's effectiveness in terms of security, efficiency, and flexibility.

This paper proposes a proxy re-encryption (PRE) scheme for secure and efficient data sharing in cloud computing environments. The scheme allows a data owner to delegate access to their data to another user without revealing the data's content. The authors evaluate the scheme's security and efficiency and demonstrate its effectiveness in facilitating secure data sharing across platforms.

CHAPTER 3
SYSTEM STUDY

FEASIBILITY STUDY

A well-designed study should offer a historical background of the business or project, such as a description of the product or service, accounting statements, details of operations and management, marketing research and policies, financial data, legal requirements, and tax obligations. Generally, such studies precede technical development and project implementation. A feasibility study evaluates the project’s potential for success; therefore, perceived objectivity is an important factor in the credibility of the study for potential investors and lending institutions. There are five types of feasibility study—separate areas that feasibility study examines, described below.

1.1 Technical Feasibility - this assessment focuses on the technical resources available to the organization. It helps organizations determine whether the technical resources meet capacity and whether the technical team is capable of converting the ideas into working systems. Technical feasibility also involves evaluation of the hardware, software, and other technology requirements of the proposed system. As an exaggerated example, an organization wouldn’t want to try to put Star Trek’s transporters in their building—currently; this project is not technically feasible.

1.2. Economic Feasibility - this assessment typically involves a cost/ benefits analysis of the project, helping organizations determine the viability, cost, and benefits associated with a project before financial resources are allocated. It also serves as an independent project assessment and enhances project credibility—helping decision makers determine the positive economic benefits to the organization that the proposed project will provide.

3. Operational Feasibility - this assessment involves undertaking a study to analyze and determine whether—and how well—the organization’s needs can be met by completing the project. Operational feasibility studies also analyze how a project plan satisfies the requirements identified in the requirements analysis phase of system development.
CHAPTER 4

System Analysis

Existing System

Cloud computing is a natural evolution of the widespread adoption of virtualization, service-oriented architecture and utility computing. Details are abstracted from consumers, who no longer have need for expertise in, or control over, the technology infrastructure in the cloud" that supports them. The relative security of cloud computing services is a contentious issue which may be delaying its adoption. Issues barring the adoption of cloud computing are due in large part to the private and public sectors unease surrounding the external management of security based services. Organizations have been formed in order to provide standards for a better future in cloud computing services. One organization in particular, the Cloud Security Alliance is a non-profit organization formed to promote the use of best practices for providing security assurance within cloud computing. Cloud provider vulnerabilities could be platform-level, such as an SQL-injection or cross-site scripting vulnerability in salesforce.com, phishes and other social engineers have a new attack vector. The cloud user must protect the infrastructure used to connect and interact with the cloud, a task complicated by the cloud being outside the firewall in many cases. The enterprise authentication and authorization framework does not naturally extend into the cloud. Potential vulnerabilities in the hypervisor or VM technology used by cloud vendors are a potential problem in multi-tenant architectures. Investigating inappropriate or illegal activity may be difficult in cloud computing because logging and data for multiple customers may be co-located may also be geographically spread across an ever-changing set of hosts and data centers.

Disadvantage

- Organization shares with customers by 64% and 74% with suppliers with respect to the total existing data of organizations. But these type of data transmission always is in possibilities from threat.
- Divisions in scalar multiplication are discontinuous, what we are more concerned about is the computing time of one single division, and the latency is obviously intolerant.

Proposed System

Elliptic curve cryptography (ECC) is an approach to public-key cryptography based on the algebraic structure of elliptic curves over finite fields. Elliptical curve cryptography (ECC) is a public key encryption technique based on elliptic curve theory that can be used to create faster, smaller, and more efficient cryptographic keys. ECC generates keys through the properties of the elliptic curve equation instead of the traditional method of generation as the product of very large prime numbers. The technology can be used in conjunction with most public key encryption methods, such as RSA, and Diffie-Hellman. According to some researchers, ECC can yield a level of security with a 164-bit key that other systems require a 1,024-bit key to achieve. Because ECC helps to establish equivalent security with lower computing power and battery resource usage, it is becoming widely used for mobile applications

Advantage

- Elliptic Curve Cryptography is a one of the secure public key crypto systems where there are two different type of key such as public key and private key.
- The smaller key size also makes possible much more compact implementations for a given level of security, which means faster cryptographic operations, running on smaller chips or more compact software.
System Requirement

Hardware Requirement

- Processor: Intel Core/i3
- Hard Disk: 500GB • RAM: 4GB
- Operating System: 32 bit Windows 7

Software Requirement

- Software: java, Angular, Cloud Storage
- Technology: Cloud

Architecture

Setup Phase: In this phase, the TTP generates the public and private keys for the system. The TTP keeps the private key, and the public key is distributed to the data owner and the cloud server.

Data Preprocessing Phase: In this phase, the data owner preprocesses the data and generates a tag using a hash function. The tag is sent to the cloud server along with the data.

Tag Generation Phase: In this phase, the cloud server receives the data and the tag from the data owner. The cloud server generates a new tag using the received data and sends it back to the data owner.

Tag Verification Phase: In this phase, the data owner verifies the tag generated by the cloud server. If the tag matches the one generated by the data owner, then the integrity of the data is verified. If the tags do not match, then the data has been tampered with.

CHAPTER 5

System Design

Introduction

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

Perfect data privacy against the TPA

The class diagram is the main building block of object-oriented modeling. It is used for general conceptual modeling of the structure of the application, and for detailed modeling translating the models into programming code. Class diagrams can also be used for data modeling. The classes in a class diagram represent both the main elements, interactions in the application, and the classes to be programmed.
Check Proof. Upon receiving m0 from the server, the verifier checks if \( IDS(r||fname) \) is a valid identity-based signature from the data owner on the message \( r||fname \). If not, the proof is invalid. Otherwise, the verifier checks if \( m_0 = H_3(Y \prod_{i \in I} e(H_2(fname||i) vi , r_\rho) \) . If the equality holds, the verifier accepts the proof; Otherwise, the proof is invalid.

V. SECURITY ANALYSIS OF THE NEW PROTOCOL In this section, we show that the proposed scheme achieves the properties of completeness, soundness and perfect data privacy preserving.

A. Completeness If both the data owner and the cloud server are honest, for each valid tag \( \sigma_i \) and a random challenge, the cloud server can always pass the verification. The completeness of the protocol can be elaborated as follows.

CHAPTER 6

System Architecture

Setup Phase: In this phase, the TTP generates the public and private keys for the system. The TTP keeps the private key, and the public key is distributed to the data owner and the cloud server.

Data Preprocessing Phase: In this phase, the data owner preprocesses the data and generates a tag using a hash function. The tag is sent to the cloud server along with the data.
CHAPTER 7

Class Diagram

The class diagram is the main building block of object-oriented modeling. It is used for general conceptual modeling of the structure of the application, and for detailed modeling translating the models into programming code. Class diagrams can also be used for data modeling. The classes in a class diagram represent both the main elements, interactions in the application, and the classes to be programmed.
DataFlow Diagram
B. System Components and its Security

Six algorithms namely Setup, Extract, TagGen, Challenge, ProofGen and Proof Check are involved in an identity-based RDIC system.

• Setup(1k) is a probabilistic algorithm run by the KGC. It takes a security parameter k as input and outputs the system parameters param and the master secret key msk.

• Extract(param, msk, ID) is a probabilistic algorithm run by the KGC. It takes the system parameters param, the master secret key msk and a user’s identity ID ∈ {0, 1}∗ as input, outputs the secret key skID that corresponds to the identity ID.

• TagGen(param, F, skID) is a probabilistic algorithm run by the data owner with identity ID. It takes the system parameters param, the secret key of the user skID and a file F ∈ {0, 1}∗ to store as input, outputs the tags σ = (σ1, · · · , σn) of each file block mi, which will be stored on the cloud together with the file F.

• Challenge(param, F n, ID) is a randomized algorithm run by the TPA. It takes the system parameters param, the data owner’s identity ID, and a unique file name F n as input, outputs a challenge chal for the file named F n on behalf of the user ID.

• ProofGen(param, ID, chal, F, σ) is a probabilistic algorithm run by the cloud server. It takes the system parameters param, the challenge chal, the data owner’s identity ID, the tag σ, the file F and its name F n as input, outputs a data possession proof P of the challenged blocks.
• ProofCheck\((\text{param}, \text{ID}, \text{chal}, P, F_n)\) is a deterministic algorithm run by the TPA. It takes the system parameters \(\text{param}\), the challenge \(\text{chal}\), the data owner’s identity \(\text{ID}\), the file name \(F_n\) and an alleged data possession proof \(P\) as input, outputs 1 or 0 to indicate if the file \(F\) keeps intact. We consider three security properties namely completeness, security against a malicious server (soundness), and privacy against the TPA (perfect data privacy) in identity-based remote data integrity checking protocols. Following the security notions due to Shacham and Waters [7], an identity-based RDIC scheme is called secure against a server if there exists no polynomial-time algorithm that can cheat the TPA with non-negligible probability and there exists a polynomial-time extractor that can recover the file by running the challenges response protocols multiple times. Completeness states that when interacting with a valid cloud server, the algorithm of Proof Check will accept the proof. Soundness says that a cheating prover who can convince the TPA it is storing the data file is actually storing that file. We now formalize the security model of soundness for identity-based remote data integrity checking below, where an adversary who plays the role of the untrusted server and a challenger who represents a data owner are involved. Security against the Server. This security game captures that an adversary cannot successfully generate a valid proof without possessing all the blocks of a user ID corresponding to a given challenge, unless it guesses all the challenged blocks. The game consists of the following phases [37].

• Setup: The challenger runs the Setup algorithm to obtain the system parameters \(\text{param}\) and the master secret key \(\text{msk}\), and forwards \(\text{param}\) to the adversary, while keeps \(\text{msk}\) confidential.

• Queries: The adversary makes a number of queries to the challenger, including extract queries and tag queries

**Work flow Diagram**

A data flow diagram (DFD) is a way of representing a flow of a data of a process or a system (usually an information system) The DFD also provides information about the outputs and inputs of each entity and the process itself. A data flow diagram has no control flow, there are no decision rules and no loops. Specific operations based on the data can be represented by a flowchart.

![Data Flow Diagram](image-url)
CHAPTER 8

Implementation

Data should always be encrypted when stored (using separate symmetric encryption keys) and transmitted. If this is implemented appropriately, even if another tenant can access the data, all that will appear is gibberish. Shared symmetric keys for data encryption should be discouraged, yet tenants should be able to access their own encryption keys and change them when necessary. Cloud providers should not have ready access to tenant encryption keys. Data is not persistent in local system. So a Storage account is created with a cryptographic key. This storage account consists of container, Table, Queue. The container has a feature called blob, which is similar to files in Windows. Security Issues that arise in the Cloud Security issues in cloud fall into two general classes: security issues confronted by cloud providers (associations giving programming, platform, or framework as-a-benefit by means of the cloud) and security issues confronted by their clients (organizations or associations who have applications or store data on the cloud). i) Guaranteeing Data isolation In order to optimize resources, cut costs, and maintain efficiency, Cloud Service Providers store multiple customers’ data on the same server. This leads to a chance that user's private data can be viewed by each other. To avoid such sensitive situations, cloud service providers must ensure proper data isolation and logical storage separation. ii) Guaranteeing Secure Data Transfer In a Cloud environment, the physical location and reach are out of control of the end user, where the resources are hosted. iii) Ensuring Secure Interface In the unsecure internet environment the integrity of information during transfer, storage and retrieval needs to be ensured. iv) Security of Stored Data The issue of controlling the encryption and decryption by either the end user or the Cloud Service provider is still doubtful. v) User Access Control Web data logs are needed to be provided to compliance auditors and security managers for web based transactions (PCI DSS)

Module 1:

Cloud Storage server It contains there are four functional blocks for data storage and accessing data from data centers in public cloud servers such as, data owner, Cloud Service Provider(CSP), authorized users, and Trusted Third Party. The functions of these functional blocks are as follows;

Data owner: The data owner can be any organization for generating outsourcing data to store in data center of public cloud model for the external use on the demand of the authorized users on the basis of pay per usage. Cloud Service Provider: Manage the cloud servers and data centers in the public cloud and provide the storage infrastructure to the data owner for storage of outsourced data in data center on the payment based on the requested storage capacity. It coordinates the trusted third party to verify the authorized users and to retrieve the data from cloud server to make them available for authorized user on demand.

Users: the set of authorized users to access the remote data stored in cloud server through trusted third party and cloud service provider. All the users are the clients of the data owner. Trusted Third Party: an entity who is trusted by all other entities of the system such as CSP, data owner and users. The functions of TTP is to verify whether the requested user in authorized or not, updating the block status table of file and calculate the hash value for file and block status table. We consider three participating entities: the client (C), the user (U) and the cloud service provider (CSP). The client has a collection of data files stored on cloud servers after the preprocessing procedure. The user who shares the stored data with the client may challenge the cloud storage server to provide a proof of possession. KeyGen – given a selected security parameter λ, this algorithm outputs the data owner public and secret keys (pk, sk), where pk is a public elliptic curve point.

• Setup – given a data file D ∈ {0, 1} * and the public key pk, the setup algorithm generates the data file identifier IDF and the corresponding public elements (σ1, σ2).
• GenChal – this algorithm generates a randomized challenge c.

• ChalProof – given the challenge c, and the original version of the file data D, the ChalProof algorithm produces a proof P = (y1, y2). Note that y1 and y2 are two elliptic curve points.

• Verify – given the proof P, the public elements and the private key Verify checks the data possession and outputs a result as either accept or reject

Module 2: Security Model For our technique to be efficient in cloud storage applications, we have to consider realistic threat models. We first point out the case where an untrusted cloud provider has a malicious behaviour. In such cases, the storage server claims that it possesses the data file, even if the file is totally or partially corrupted. To model this situation, our scheme is based on two important requirements proposed. On one hand, there exists no polynomial-time algorithm that can fool the verifier with non-negligible probability. On the other hand, there exists no polynomial-time algorithm that can recover the original data files by carrying out multiple challenge response exchanges.

VII. SNEAKPEEZKS OF IMPLEMENTATION

Fig 3: Login Page
Fig 4: file Upload page

Fig 5: data or file verification screen
CHAPTER 9

Conclusion

In this paper, we investigated a new primitive called identity-based remote data integrity checking for secure cloud storage. We formalized the security model of two important properties of this primitive, namely, soundness and perfect data privacy. We provided a new construction of this primitive and showed that it achieves soundness and perfect data privacy. Both the numerical analysis and the implementation demonstrated that the proposed protocol is efficient and practical. ACKNOWLEDGEMENTS. This work is supported by the National Natural Science Foundation of China (61501333, 61300213, 61272436, 61472083), Fok Ying Tung Education Foundation (141065), Program for New Century Excellent Talents in Fujian University (JA14067).

CHAPTER 10

Reference


