IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Identity-based Remote Data Integrity Checking with Perfect Data Privacy Preserving for Cloud Storage

Yuvaraj S
Computer science engineering. Madha Engineering college, Tamil nadu. India

Abstract

Remote data integrity checking (RDIC) enables a data storage server, such as a cloud server, to prove to a verifier that it is actually storing a data owner's data honestly. To date, a number of RDIC protocols have been proposed in the literature, but almost all the constructions suffer from the issue of a complex key management, that is, they rely on the expensive public key infrastructure (PKI), which might hinder the deployment of RDIC in practice. In this paper, we propose a new construction of identity -based (ID-based) RDIC protocol by making use of key-homomorphic cryptographic primitive to reduce the system complexity and the cost for establishing and managing the public key authentication framework in PKI based RDIC schemes. We formalize ID-based RDIC and its security model including security against a malicious cloud server and zero knowledge privacy against a third party verifier. We then provide a concrete construction of ID-based RDIC scheme which leaks no information of the stored files to the verifier during the RDIC process. The new construction is proven secure against the malicious server in the generic group model and achieves zero knowledge privacy against a verifier. Extensive security analysis and implementation results demonstrate that the proposed new protocol is provably secure and practical in the real-world applications.

Keywords: Cloud storage, data integrity, privacy preserving, identity-based cryptography.

Introduction

Cloud computing , which has received considerable attention from research communities in academia as well as industry, is a distributed computation model over a large pool of shared-virtualized computing resources, such as storage, processing power, applications and services. Cloud users are provisioned and de-provisioned recourses as they want in cloud computing environment. This kind of new computing represents a vision of providing computing services as public utilities like water and electricity. Cloud computing brings a number of advantages for cloud users. As examples, this include the following issues: Users can avoid capital expenditure on hardware, software and services because they pay only for what they use; Users can enjoy low management overhead and immediate access to a wide range of applications; and Users can access their data wherever they are, rather than having to stay close to their computers.

CLOUD SERVICE MODELS

There are three main types of cloud service:

- Software as a Service (SaaS).
- Platform as a Service (PaaS) C.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 (eg 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 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.

An Efficient and Privacy-Preserving Identity-Based Remote Data Integrity Checking Protocol for Cloud Storage" by Zhang et al. (2020):

- This paper proposes an efficient and privacy-preserving IRDIC protocol that uses an identity-based signature and a bilinear pairing to ensure data integrity and privacy. The protocol is efficient and secure and can be used for dynamic data.

"Privacy-Preserving Identity-Based Remote Data Integrity Checking Protocol for Cloud Storage" by Zhang et al. (2021):

- This paper proposes a privacy-preserving IRDIC protocol that uses an identity-based signature and a homomorphic hash function to ensure data integrity and privacy. The protocol is efficient and secure and can be used for dynamic data.

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.

1.1Technical 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 AnalysisExisting 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 forexpertise 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, phishesand 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. Solution is to get a contractual commitment to support specific forms of investigation. High-speed constant-time division module in extended binary finite field. Although division is the most computationally intensive arithmetic, it is also an essential component for scalar multiplication. Extended Euclidean Algorithm (EEA) to give results in a constant number of iterations, there have been several works that developed modified versions.

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 beused 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 systemsrequire 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.
- 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 integrityof the data is verified. If the tags do not match, then the data has been tampered with.
- 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.
- We fill the gap that there is no a secure and novel IDbased RDIC scheme to date. Specifically, we propose a concrete ID-based RDIC protocol,
- which is a novel construction that is different from the previous ones, by making use of the idea of a new primitive called asymmetric group key agreement [32], [33].
- To be more specific, our challenge-response protocol is a two party key agreement between the TPA and the cloud server, the challenged blocks must be used when generating a shared key by the cloud server, which a response to a challenge from the TPA
- Usually, data owners themselves can check the integrity of their cloud data by running a two-party RDIC protocol.
- However, the auditing result from either the data owner or the cloud server might be regarded as biased in a twoparty scenario. The RDIC protocols with public verifiability enable anyone to audit the integrity of the outsourced data. To make the description of the publicly verifiable RDIC protocols

- clearly, we assume there exits a third party auditor (TPA) who has expertise and capabilities to do the verification work. With this in mind, the ID-based RDIC architecture is illustrated in Fig 1. Four different entities namely the KGC, the cloud user, the cloud server and the TPA are involved in the system. The KGC generates secret keys for all the users according to their identities. The cloud user has large amount of files to be stored on cloud without keeping a local copy, and the cloud

- server has significant storage space and computation resources and provides data storage services for cloud users

Usually, data owners themselves can check the integrity of their cloud data by running a two-party RDIC protocol. However, the auditing result from either the data owner or the cloud server might be regarded as biased in a twoparty scenario. The RDIC protocols with public erifiability enable anyone to audit the integrity of the outsourced data. To make the description of the publicly verifiable RDIC protocols clearly, we assume there exits a third party auditor (TPA) who has expertise and capabilities to do the verification work. With this in mind, the ID-based RDIC architecture is illustrated in Fig 1.

Four different entities namely the KGC, the cloud user, the cloud server and the TPA are involved in the system. The KGC generates secret keys for all the users according to their identities. The cloud user has large amount of files to be stored on cloud without keeping a local copy, and the cloud server has significant storage space and computation resources and provides data storage services for cloud users. TPA has expertise and capabilities that cloud users do not have and is trusted to check the integrity of the cloud data on behalf of the cloud user upon request. Each entity has their own obligations and benefits respectively. The cloud server could be self-interested, and for his own benefits, such as to maintain a good reputation, the cloud server might even decide to hide data corruption incidents to cloud users. However, we assume that the cloud server has no incentives to reveal the hosted data to TPA because of regulations and financial incentives. The TPA's job is to perform the data integrity checking on behalf the cloud user, but the TPA is also curious in the sense that he is willing to learn some information of the users' data during the data integrity checking procedure

CHAPTER 5

System DesignIntroduction

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.

Architecture Diagram

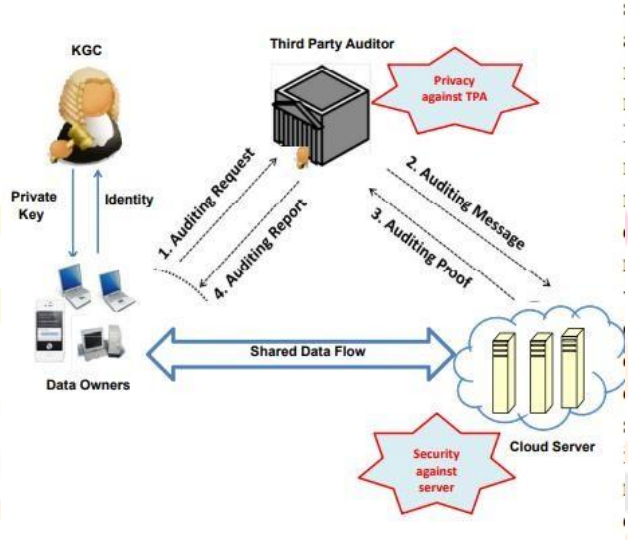


Fig. 1. The system model of identity-based RDIC

B. System Components and its Security Six algorithms namely Setup, Extract, TagGen, Challenge, ProofGen and ProofCheck 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 $\in \{0, 1\} *$ 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 $F \in \{0, 1\} *$ to store as input, outputsthe tags $\sigma = (\sigma 1, \cdots, \sigma n)$ 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(param, ID, chal, P, F n) is a deterministic algorithm run by the TPA. It takes the system parameters param, the challenge chal, the data owner's identity ID, the file name F n and an allegeddata 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 challengesresponse protocols multiple times. Completeness states that when interacting with a validcloud server, the algorithm of ProofCheck 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 param and the master secret key msk, and forwards param to the adversary, while keeps msk confidential.

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

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 classdiagram represent both the main elements, interactions in the application, and the classes to be programmed.

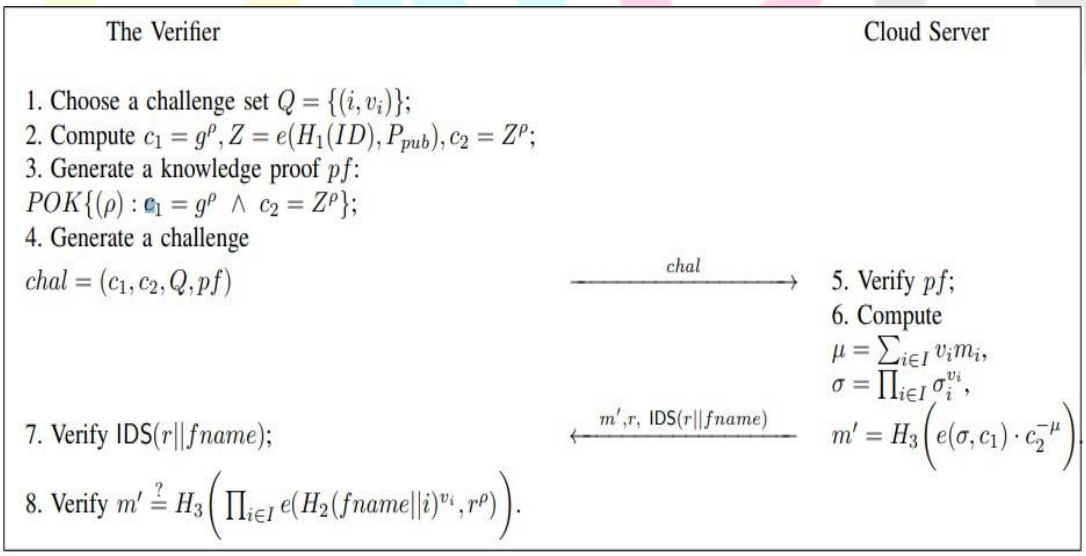


Fig. 2. Identity-based remote data integrity checking protocol

CheckProof. 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 m0 = H3 CY i∈I e(H2(fname||i) vi , rρ) ? . 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. Completeness guarantees the correctness of the protocol while soundness shows that the protocol is secure against an untrusted server. Perfect data privacy states that the protocol leaks no information of the stored files to the verifier. A. Completeness If both the data owner and the cloud server are honest, for each valid tag σi and a random challenge, the cloud server can always pass the verification. The completeness of the protocol can be elaborated as follows.

Implementation Results

The implementation was conducted with pbc-0.5.13 [42] with pbc wrapper-0.8.0 [43] on Intel i7-4700MQCPU @ 2.40GHz. The memory is always sufficient since the scheme only requires a polynomial space. In our implementation, we made use of parameter a.param, one of the standard parameter settings of pbc library. Parameter a.param provides a symmetric pairing with the fastest speed among all default parameters. The implementation time overheads of the protocol are displayed in the following two parts.

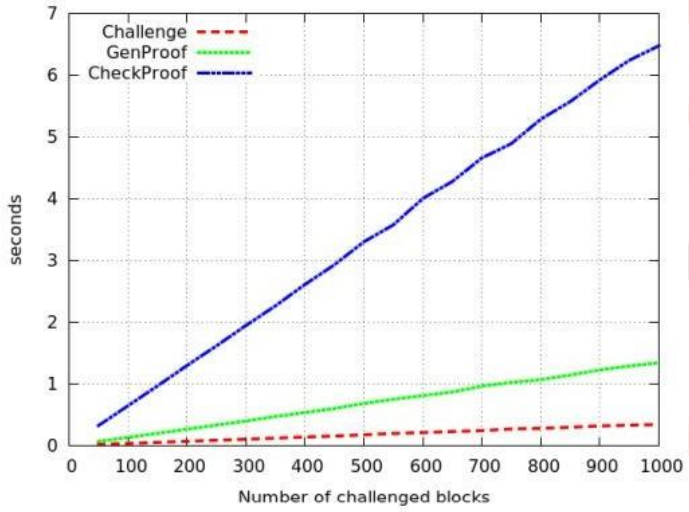


Fig. 3. Increasing number of challenges for fixed size of file

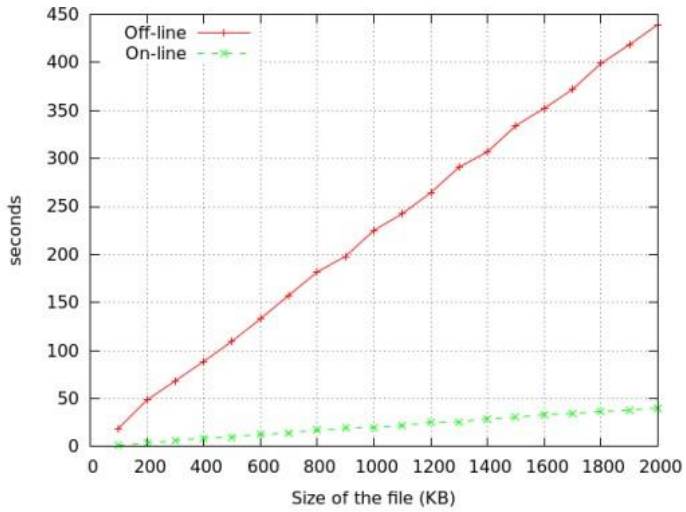
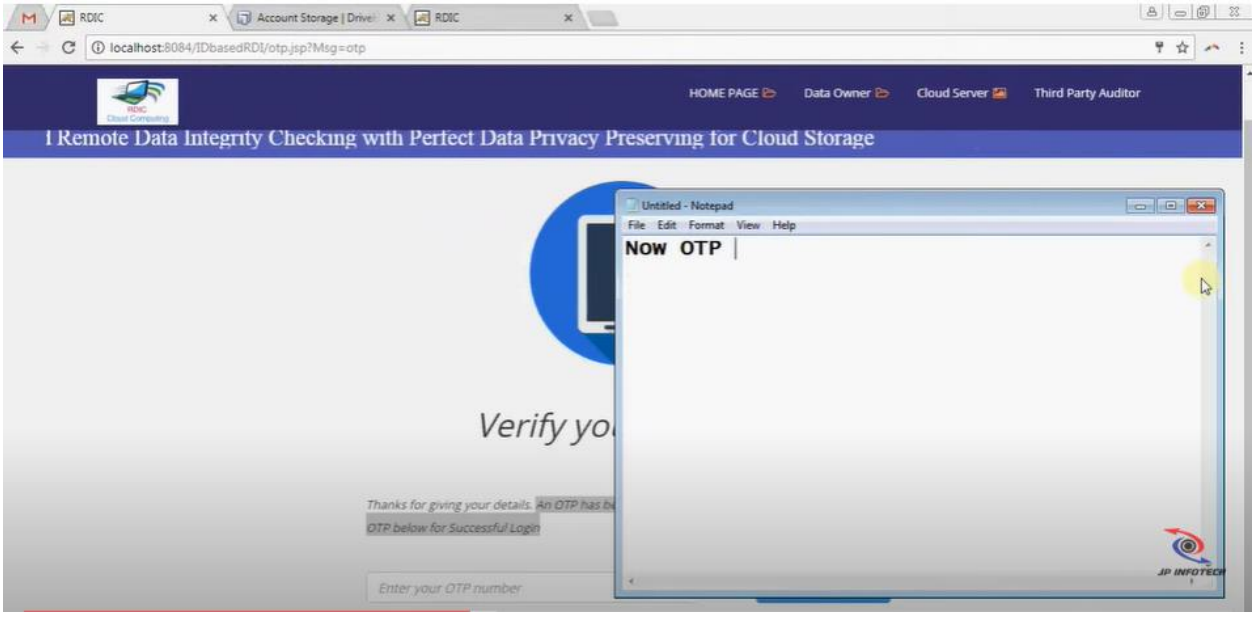


Fig. 4. Tag generation time for increased size of files

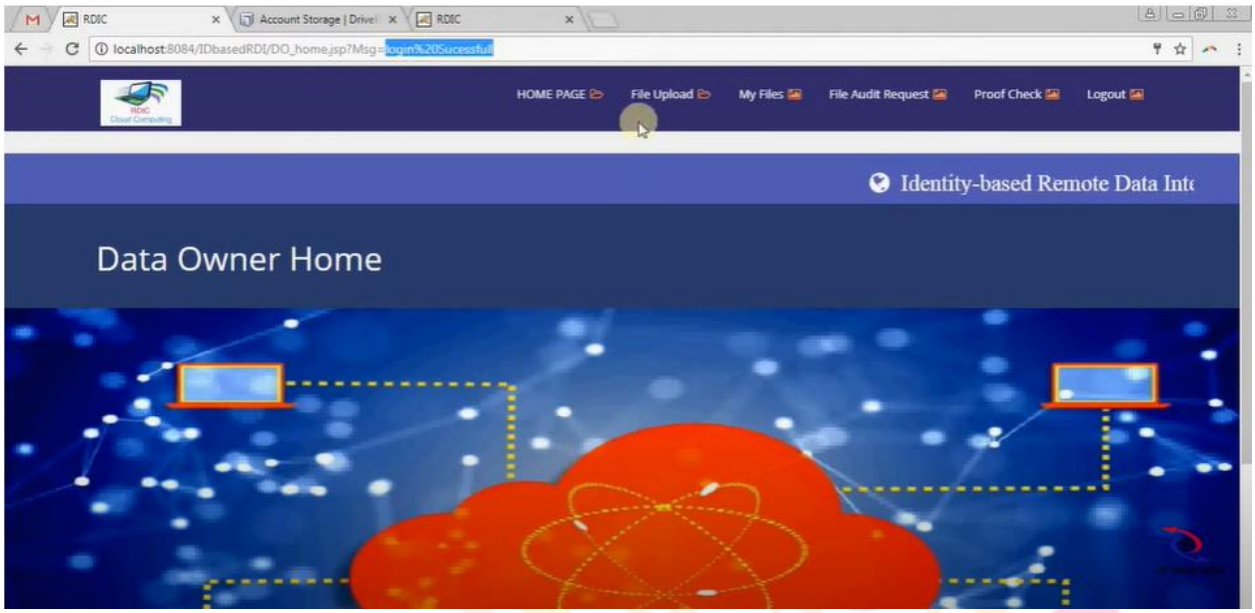
compared to challenging the outsourced data, which will be done repeatedly. Since the cloud users can do the off-line work completely parallelizable in advance, we pay only attention to the on-line cost. We can see that the efficiency of TagGen of our protocol is comparable to that of the existing well-known schemes, say [5]. To generate tags for a 2 MB file, it costs almost 42 seconds. As such, one shall be able to anticipate the time cost of generating tags for any size of files.

Input Screen

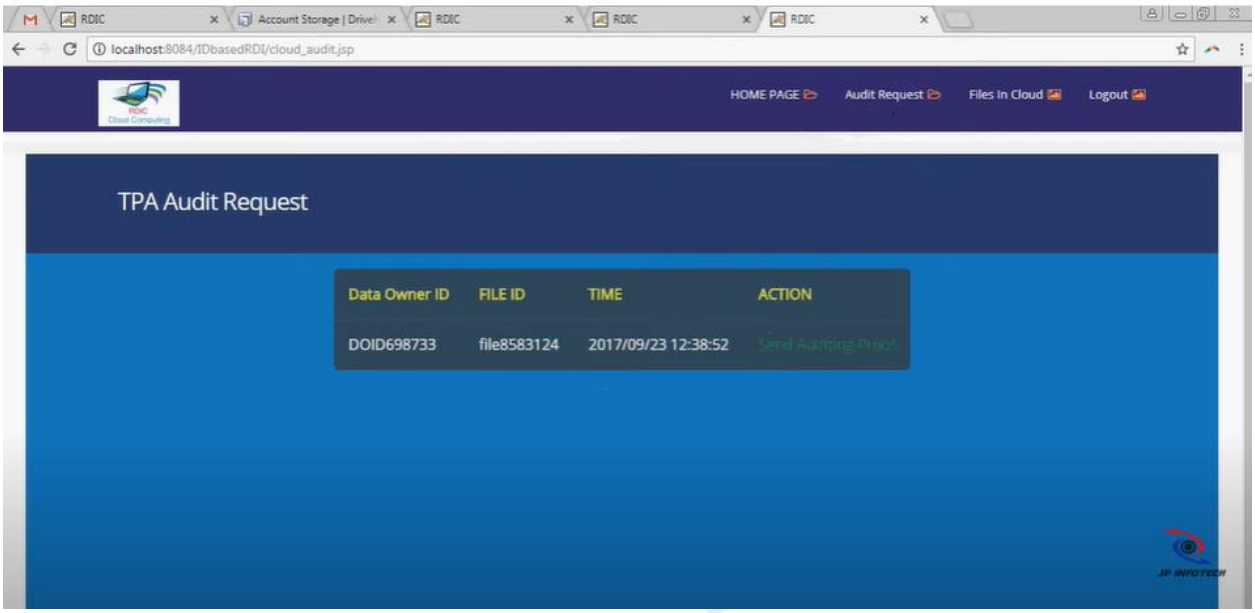
Screen 1:



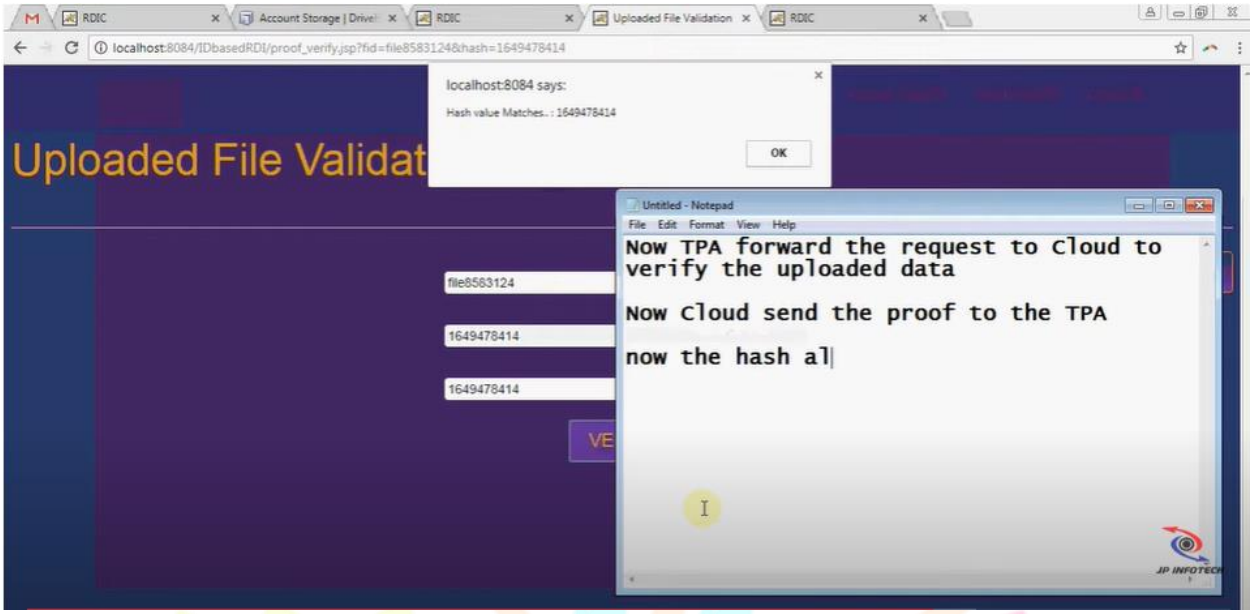
Screen 2:



Screen 3:



Screen 4:



CHAPTER 5

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 of this primitive and showed that it achieves soundness and perfect data privacy. Both the numerical analysis and theimplementation 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).

Reference

[1] P. Mell, T. Grance, Draft NIST working definition of cloud computing, Reference on June. 3rd, 2009. http://csrc.nist.gov/groups/SNC/cloudcomputing/index.html.

[2] Cloud Security Alliance. Top threats to cloud computing. http://www.cloudsecurityalliance.org,2010.

[3] M. Blum, W. Evans, P.Gemmell, S. Kannan, M. Naor, Checking the correctness of memories. Proc. of the 32nd Anual Symposium on Foundations fo Vomputers, SFCS 1991, pp. 90–99, 1991.

[4] G. Ateniese, R. C. Burns, R. Curtmola, J. Herring, L. Kissner, Z. N. J. Peterson, D. X. Song, Provable data possession at untrusted stores. ACM Conference on Computer and Communications Security, 598- 609, 2007.

[5] G. Ateniese, R. C. Burns, R. Curtmola, J. Herring, O. Khan, L. Kissner, Z. N. J. Peterson, and D. Song, Remote data checking using provable data possession. ACM Trans. Inf. Syst. Secur., 14, 1–34, 2011.

[6] A.Juels, and B. S. K. Jr. Pors, proofs of retrievability for large files.Proc. of CCS 2007, 584-597, 2007.

[7] H. Shacham, and B. Waters, Compact proofs of retrievability. Proc. of Cryptology-ASIACRYPT 2008, LNCS 5350, pp. 90-107, 2008.

[8] G. Ateniese, S. Kamara, J. Katz, Proofs of storage from homomorphic identification protocols. Proc. of ASIACRYPT 2009, 319-333, 2009.

[9] A. F. Barsoum, M. A. Hasan, Provable multicopy dynamic data possession in cloud computing systems, IEEE Trans. on Information Forensics and Security, 10(3): 485–497, 2015.

[10] J. Yu, K. Ren, C. Wang, V. Varadharajan, Enabling cloud storage auditing with key-exposure resistance, IEEE Trans. on Information Forensics and Security, 10(6): 1167–1179, 2015.

[11] J. Liu, K. Huang, H. Rong, H. M. Wang, Privacy-preserving public auditing for regenerating-code-based cloud storage, IEEE Trans. on Information Forensics and Security, 10(7): 1513–1528, 2015.

[12] Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou, Enabling public verifiability and data dynamics for storage security in cloud computing. Proc. of ESORICS2009, LNCS 5789, 355–370, 2009.