WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Implementation of Fine-Grained Multi-Cloud Data Protection and Secure Data Sharing Across Cloud Platforms

S. Monika

PG Scholar, Dept. of MCA, Jayam College of Engineering, Dharmapuri, Anna University, Tamilnadu State, India

A B S T R A C T

In recent times, numerous businesses have increasingly turned to cloud storage as a cost-effective option for outsourcing essential data. However, ensuring the integrity of data remains a major concern due to inherent distrust of cloud servers. To address these challenges, various security schemes have emerged, aiming to facilitate data outsourcing while ensuring efficient integrity checks suitable for devices with limited capabilities. This research focuses on an ocean information service provider operating within a cloud computing framework. Our approach introduces a novel technique for remote data integrity verification: Identity-based Distributed Provable Data Possession (ID-DPDP) across multiple cloud environments. We present formal system and security models and propose an ID-DPDP protocol that utilizes bilinear pairings.

Keywords: Integrity of data, Multiple Cloud Services, Integrity checking model, Data security.

1. Introduction

Cloud computing entails delegating computing tasks to external providers, thereby exposing potential security risks concerning data confidentiality, integrity, and availability. Ensuring that client data remains intact is crucial, especially since clients typically do not retain local copies. Remote data integrity checking is essential to address this concern, especially when data is distributed across multiple cloud servers. Effective integrity checking protocols are essential to accommodate devices with limited computational capabilities. This study focuses on developing a distributed model for remote data integrity checking in multi-cloud environments.

This project specifically investigates an ocean information service corporation operating within the cloud computing framework. The corporation provides various services such as ocean measurement, environmental monitoring, hydrological and marine biological data, and GIS information. Alongside these services, the corporation manages both private and public data, including promotional materials. To enhance security and cost-effectiveness, the corporation employs multiple cloud service providers, each selected based on their reputation and security offerings tailored to different data types and sensitivity levels.

Among the contributions made in this work are:

- This project focuses on an ocean information service corporation operating within a cloud computing environment.
- It addresses security risks related to data confidentiality, integrity, and availability in cloud computing.
- Many verifiers in cloud computing have limited computational capacity. Identity-based public key cryptography simplifies certificate management processes.

2. Objective of the Study

The proposed model generates probabilistic proofs of possession by randomly sampling sets of blocks from the server, thereby significantly lowering input/output (I/O) costs. The client maintains a fixed amount of metadata to validate the proof. Using a challenge/response protocol, the system sends a small, consistent data size, minimizing network traffic. For verification, the Provable Data Possession (PDP) scheme can generate proofs without validating certificates, only ensuring the system's public key is authorized by the service provider. The verifier does not need to retain entire file copies for inspection, ensuring security against malicious provers. The protocol supports unlimited verification runs.

3. System Study

3.1. Existing System:

There is a significant concern about privacy with the current method of sharing data, where there is a risk of identity privacy leaking to public verifiers. Traditional methods require retrieving the entire dataset from the cloud to verify integrity by checking accuracy against markers.

To effectively address the requirement for a reliable third-party auditor, two primary conditions must be fulfilled:

1. The auditor should be capable of efficiently auditing cloud data storage without requiring a local copy of the data, thereby avoiding additional online burdens on the cloud client.
2. The auditing process should not introduce new vulnerabilities to user data privacy.

Cloud computing involves outsourcing computing tasks to third parties, which introduces security risks related to data confidentiality, integrity, and availability. When clients store data on multi-cloud servers, distributed storage, and integrity checks become crucial. The integrity-checking protocol must also be efficient to accommodate devices with limited capacities. Therefore, this study focuses on exploring a distributed model for remote data-integrity checking and presenting a specific protocol tailored for multi-cloud storage environments.

Drawbacks of the Existing System:

- Verifying data becomes complex when using multiple servers.
- Requires significant storage capacity.
- Insufficient measures for preventing data loss.

3.2. Proposed System:

The proposed framework introduces secure evaluation techniques for shared information in cloud environments. Digital signatures are utilized to construct homomorphic authenticators, enabling an external verifier to check data integrity without accessing the entire dataset or identifying individual signers on each block.

To enhance the versatility of verification methods, the toolset is extended to support batch auditing. Future improvements are considered, particularly regarding scenarios where group administrators may need to reveal the identity of endorsers based on metadata verification. This paper focuses on Identity-based Public Key cryptography and distributed provable data possession in multi-cloud storage.

The proposed solution, ID-DPDP (Identity-based Distributed Provable Data Possession), eliminates the need for certificate management, enhancing protocol efficiency. The system and security models are formally defined, and the ID-DPDP protocol is designed using bilinear pairings. In the random oracle model, the ID-DPDP protocol is proven secure. Moreover, the protocol offers flexibility and high efficiency, supporting private, delegated, and public verification modes based on client authorization.

3.2.1. Advantages of Proposed System:

- Offers expanded storage capacity.
- Ensures the security of public data.
- Implements private key generation.

4. System Architecture

This system is structured around a four-step approach facilitated by four key modules: Registration, User Login, Combiner, and Verifier.

The Registration module gathers user details such as username, gender, date of birth, email address, and mobile number. It enables users to set a password of their choice. Upon completion, the module automatically generates a client key essential for subsequent logins. All registration data is securely stored in a database. It's important to note that losing the client key results in the inability to access the system.

Once registered, users utilize the User Login module where they input their username, password, and the client key established during registration. The system verifies these credentials for validity before allowing file uploads. Uploaded files are segmented into blocks for processing.

The Combiner module necessitates users to input their username and password to initiate file combination, prompted by a message from the Verifier module.

The Verifier module finalizes the process by prompting users to log in with their username and password. Users select files for verification, after which the Verifier confirms file integrity. Once verified, the Verifier signals the Combiner to merge the files, enabling users to retrieve them. Forgetting the client key prevents downloading files.

5. System Testing and Implementation

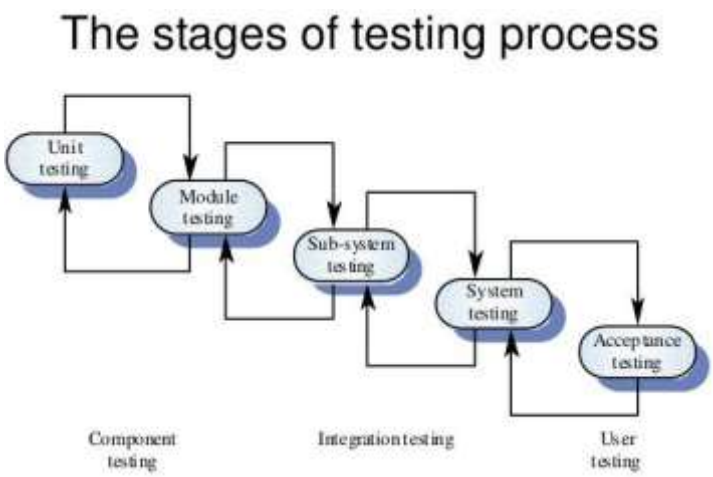
5.1. System Testing

System testing involves the deliberate process of testing software to uncover and resolve errors. This principle is fundamental for web applications as well. Web-based systems and applications operate within a networked environment, interacting with various operating systems, browsers, hardware platforms, and communication protocols. Detecting errors in this context presents a significant challenge.

The distributed nature of client/server environments, performance issues related to transaction processing, the diversity of hardware platforms, complexities in network communication, the necessity to handle multiple clients from centralized databases, and the stringent requirements on servers all contribute to the complexity of testing client/server architectures.

Testing Issues:-

- Client GUI considerations
- Target environment and platform diversity considerations
- Distributed database considerations
- Distributed processing considerations.



Testing Methodologies:

System testing is a critical phase in software development aimed at ensuring the system operates accurately and efficiently before it goes live. It verifies that all components of the software work together seamlessly.

Key activities in system testing include creating a comprehensive test plan that encompasses program testing, integration testing, validation testing, and user acceptance testing. Successfully implementing a newly designed system is crucial for its adoption and success.

Testing is a pivotal stage in the software development lifecycle, validating the code against the functional specifications to ensure the system meets its objectives. The primary goal of testing is to identify and rectify errors.

To achieve this objective, a series of test phases including unit testing, integration testing, validation testing, and system testing are meticulously planned and executed. These tests are essential for ensuring the system functions as expected and meets quality standards before deployment.

The test steps are,

- Unit Testing
- Integration Testing
- Validation Testing
- Output Testing

Quality Assurance

This system “Identity” is a user-friendly, GUI-based client/server application that will automate the different activities involved in the day-to-day activities of the Automobile business organization. An application for completely automating the different activities of the various departments or sections and maintaining the status of different processes.

The quality assurance contains facilities to input the data, by interacting with user-friendly input screens in this system, and requests for the reports are also provided. The system will help in reducing the time and effort in preparation marking lists, Performance, invoices, purchase orders, inwards, outward, and employee management, and generating required reports. It provides an error-free generation of all the reports.

5.2. System Implementation

The implementation phase marks the final stage of a project, where the theoretical design transitions into a functional system. During this critical phase, the primary workload and significant impact on existing practices shift to the user department. If implementation is not meticulously planned and controlled, it can lead to chaos, making it a crucial stage in ensuring the success of the new system and instilling user confidence in its functionality and effectiveness.

Key activities in the implementation phase include careful planning, assessing the current system and its limitations regarding implementation, designing strategies to facilitate procedural changes, and evaluating these methods for their effectiveness. Successful implementation of a website or any system is essential for meeting the company's and customers' needs. The effort invested in developing the system yields success only when the implementation is executed effectively.

5.2.1. Module description

1. Cloud Computing Module: This module involves the creation of a local cloud infrastructure offering affordable storage services. Users can securely upload their data to this cloud. Although the cloud storage is designed to be secure, users may still distrust it due to Cloud Service Providers (CSPs) often operating outside users' trusted domains. The cloud server is assumed to be honest but curious, meaning it won't maliciously alter or delete data but might attempt to access the content and identities of cloud users.

2. Registration Module: This module focuses on user registration to allocate file blocks using tag pairs, allowing registered users (clients) to collaboratively create and store valid files on a cloud server. A client can be an individual or a corporation needing extensive data storage and computation capabilities across multiple clouds.

3. Client Key Segregation Module: This module assigns a unique key accessible only to the client. The Private Key Generator (PKG) is responsible for generating the private key corresponding to the client's identity and securely transmitting it during client registration.

4. Combiner Module: In this module, the combiner acts as an intermediary that does not store or upload user data directly to a cloud server. Instead, it manages storage requests by distributing block-tag pairs to designated cloud servers. When challenged by a verifier, the combiner splits and distributes the challenge among cloud servers. After receiving responses from these servers, it aggregates them and forwards the combined response to the verifier.

5. Verifier Module: The verifier initiates challenges to the combiner, which then distributes these challenges to relevant cloud servers based on storage metadata. Cloud servers respond to these challenges, and the combiner aggregates their responses before sending the aggregate back to the verifier for validation.

6. Conclusions

In the realm of multi-cloud storage, we have formalized the ID-DPDP system model and established security specifications. Our innovation introduces the initial ID-DPDP protocol, validated as secure under the assumption of the Computational Diffie-Hellman (CDH) problem's difficulty. Our protocol eliminates the necessity for certificate management while ensuring notable flexibility and efficiency.

The proposed ID-DPDP protocol supports private verification, delegated verification, and public verification based on client authorization. We meticulously designed, developed, and rigorously tested the system to meet all predefined objectives. The coding was methodically crafted to conform to best practices, facilitating comprehension and ease of maintenance for developers.

During testing, paramount importance was placed on user-friendliness and simplicity, with feedback gathered from external users to ensure broad usability. Our solution is implemented using ASP.NET, a widely embraced technology for robust application development. This project focuses primarily on customer and dealer registration across diverse regions of India, utilizing ASP.NET for the frontend and SQL Server for the backend.

7. Future Enhancement:

This system offers high flexibility, allowing for easy maintenance and adaptation to evolving environments and requirements. Robust security measures are in place to prevent system failures caused by changes. The project is designed to support multi-user environments, enabling simultaneous use by multiple users. Emphasis has been placed on ensuring the system's security and protection against potential risks.

References

1. Appalaraju S., Jasani B., Kota B., Xie Y., Manmatha R. "Doc-Former: End-to-End Transformer for Document Understanding". 2021.
2. T. I. Denk, C. Reisswig. "BERTgrid: Contextualized Embedding for 2D Document Representation and Understanding". 2019.
3. J. Devlin, M.-W. Chang, K. Lee, K. Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT), 2019.
4. M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, C. Pal. "The Importance of Skip Connections in Biomedical Image Segmentation". Deep Learning and Data Labeling for Medical Applications, Springer, 2016.
5. W. Wikimedia Downloads.
6. He K., Gkioxari G., Dollar P., Girshick R. "Mask R-CNN". 2017.
7. Z. Huang, W. Xu, K. Yu. "Bidirectional LSTM-CRF Models for Sequence Tagging". 2015.
8. M. Kerroumi, O. Sayem, A. Shabou. "VisualWordGrid: Information Extraction From Scanned Documents Using A Multimodal Approach". 2020.
9. D. P. Kingma, J. Ba. "Adam: A Method for Stochastic Optimization". 2014.
10. Lin W., Gao Q., Sun L., Zhong Z., Hu K., Ren Q., Huo Q. "ViBERTgrid: A Jointly Trained Multi-Modal 2D Document Representation for Key Information Extraction from Documents". 2021.
11. Liu S., Qi L., Qin H., Shi J., Jia J. "Path Aggregation Network for Instance Segmentation". 2018.
12. X. Liu, F. Gao, Q. Zhang, H. Zhao. "Graph Convolution for Multimodal Information Extraction from Visually Rich Documents". NAACL-HLT, 2019.
13. R. B. Palm, O. Winther, F. Laws. "CloudScan - A Configuration-Free Invoice Analysis System Using Recurrent Neural Networks". Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), 2017.
14. R. Prabhavalkar et al. "Minimum Word Error Rate Training for Attention-Based Sequence-to-Sequence Models". Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018.
15. Redmon J., Farhadi A. "YOLO9000: Better, Faster, Stronger". 2016.
16. S. Ren, K. He, R. B. Girshick, J. Sun. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015.
17. Simonyan K., Zisserman A. "Very Deep Convolutional Networks for Large-Scale Image Recognition". 2015.
18. E. F. Tjong Kim Sang, J. Veenstra. "Representing Text Chunks". Ninth Conference of the European Chapter of the Association for Computational Linguistics (EACL), 1999.
19. Tompson J., Goroshin R., Jain A., LeCun Y., Bregler C. "Efficient Object Localization Using Convolutional Networks". 2014.
20. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., Polosukhin I. "Attention Is All You Need". 2017.
21. Y. Xu, M. Li, L. Cui, S. Huang, F. Wei, M. Zhou. "LayoutLM: Pre-training of Text and Layout for Document Image Understanding". Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD), 2020.
22. Xu Y., Xu Y., Lv T., Cui L., Wei F., Wang G., Lu Y., Florencio D., Zhang C., Che W., Zhang M., Zhou L. "LayoutLMv2: Multimodal Pre-training for Visually-Rich Document Understanding". 2020.
23. Yu F., Koltun V. "Multi-Scale Context Aggregation by Dilated Convolutions". International Conference on Learning Representations (ICLR), 2016.
24. P. Zhang et al. "TRIE: End-to-End Text Reading and Information Extraction for Document Understanding". Proceedings of the 28th ACM International Conference on Multimedia (MM), 2020.