DATA STORAGE SECURITY SYSTEM BASED ON CLOUD COMPUTING

ABSTRACT:

Cloud computing is a computer model that offers numerous advantages, but it also presents various challenges, with security being a primary concern. Certain companies, driven by the unique characteristics of their regions, have particularly high data security requirements. Any loss or damage to this data can result in severe consequences for the responsible party. To ensure accurate and reliable data transfer, it is crucial to implement robust security measures that not only safeguard the data but also enhance the reputation and safety of the organization. Cloud computing is an increasingly important topic, and effective and standardized management of data is vital for healthy and efficient operations. This document proposes a secure data storage system design technology that utilizes encryption to guarantee information security on cloud platforms. By leveraging the powerful computing capabilities of the cloud, the system significantly improves processing efficiency, system performance, and the functionalities of individual modules. Once implemented successfully, this system can be effectively applied to information management tasks.

Keywords: Cloud computing, Data security, Encryption, Data storage, Security Challenges, Data loss prevention, Reliable data transfer, OTP generation.

Introduction

Cloud computing has transformed various industries, including information technology, social organizations, and financial transactions. It provides a cost-effective solution for the storage and retrieval of vast amounts of data over extended periods. Data storage in cloud networks typically employs two approaches: distributed and centralized. Distributed data storage offers advantages such as fault tolerance and reduced bandwidth consumption when compared to centralized storage.

Cloud computing empowers users with the ability to access computational and storage resources on-demand, regardless of their location or the time. This convenience has resulted in the widespread popularity of cloud storage services provided by prominent companies like Apple's iCloud, Microsoft's Azure, and Amazon's S3. However, along with

the benefits of cloud computing, concerns about security arise, which are of utmost importance to cloud users. When entrusting data to a cloud server, users aim to retain control over data access, ensuring that only authorized individuals can share and retrieve the outsourced data.

Addressing these security concerns necessitates the establishment of a data sharing system that guarantees confidentiality and backward secrecy, limiting access to shared data to authorized users exclusively. Furthermore, implementing forward secrecy requires regular updates to the encrypted form of the shared data. Nevertheless, employing the secret key for frequent updates would prove impractical, as it would entail a burdensome process of downloading, decrypting, re encrypting, and uploading the data.

Hence, an alternative approach is required to update the encrypted version of shared data. This approach should confine the secret key's usage solely to decryption, thereby obviating the need for periodic updates. By discovering an alternative technique for updating the ciphertext of shared data, the data provider can enhance system efficiency and security without compromising the confidentiality and secrecy of the shared data.

LITERATURE SURVEY

TITLE: Heterogeneous Data Storage Management with Deduplication in Cloud Computing

YEAR: 2017

AUTHOR: Zheng Yan; Lifang Zhang; Wenxiu DING; Qinghua Zheng

DESCRIPTION:

Cloud storage is essential in cloud computing, enabling users to overcome resource limitations and expand their storage capacity without device upgrades. However, encrypting data for security introduces challenges in storage efficiency and data sharing. Existing deduplication schemes lack flexibility to meet diverse data sensitivity requirements. This paper proposes a novel scheme that combines deduplication management and access control across multiple Cloud Service Providers (CSPs) to manage heterogeneous data storage. The scheme addresses limitations of traditional deduplication by providing enhanced flexibility based on data sensitivity. Through

security analysis, comparisons, and practical implementation, the scheme demonstrates its security, effectiveness, and efficiency, offering a promising solution for encrypted data storage and management with deduplication.

TITLE: Encrypted Data Management with Deduplication in Cloud Compu ng

YEAR: 2016

AUTHOR: Zheng Yan; Mingjun Wang; Yuxiang Li; Athanasios V. Vasilakos

DESCRIPTION:

In the context of the Internet of Things (IoT), cloud computing plays a crucial role in data storage, processing, and management. To protect user privacy, data is typically stored in encrypted form, which can lead to duplicate data stored with different encryption schemes, resulting in inefficient storage utilization. Existing data deduplication schemes lack security and flexibility in supporting secure data access control, limiting their practicality. This article presents a novel approach that utilizes attribute-based encryption (ABE) to deduplicate encrypted data in the cloud while ensuring secure data access control. The scheme is evaluated comprehensively, demonstrating its efficiency, effectiveness, and scalability for real-world deployment. The results highlight its potential in optimizing storage resource utilization while maintaining data security and access control in cloud-based environments.

TITLE: Lightweight Cloud Storage Auditing with Deduplication Supporting Strong Privacy Protection

YEAR: 2020

AUTHOR: Wenting Shen; Ye Su; Rong Hao

DESCRIPTION:

In cloud storage auditing with deduplication, ensuring data integrity while maintaining a single copy of duplicated files is crucial. However, existing schemes are vulnerable to brute-force dictionary attacks, compromising user privacy. This paper presents a novel approach that addresses this vulnerability in cloud storage auditing. Our scheme focuses on strong privacy protection, even for predictable or small-space files. We introduce a unique method for generating file indexes and a novel strategy for generating file encryption keys. The scheme offers lightweight computation for users, enabling efficient

generation of data authenticators, integrity verification, and file retrieval from the cloud. Through rigorous security analysis and comprehensive performance evaluation, we demonstrate the effectiveness and efficiency of our proposed scheme. It provides a robust solution for cloud storage auditing with deduplication, ensuring user data privacy and maintaining data integrity in the cloud.

TITLE: Investigating the Adoption of Hybrid Encrypted Cloud Data Deduplication With Game Theory

YEAR: 2020

AUTHOR: Xueqin Liang; Zheng Yan; Robert H. Deng; Qinghua Zheng

DESCRIPTION:

This article focuses on the practical deployment of hybrid encrypted cloud data deduplication (H-DEDU) and its economic implications. Existing economic models for cloud storage do not adequately support H-DEDU due to the complex interactions among stakeholders. To address this gap, a formal economic model is established, considering the utilities of data holders, data owners, and Cloud Storage Providers (CSPs). A multistage Stackelberg game is constructed to capture stakeholder interactions, allowing for a comprehensive analysis of the system dynamics. The conditions for a sub-game perfect Nash Equilibrium are analyzed, and a gradient-based algorithm is proposed to assist stakeholders in selecting near-optimal strategies. Extensive experiments demonstrate the feasibility of the proposed algorithm in achieving the Nash Equilibrium. The study also investigates the impact of various parameters on the adoption of H-DEDU. The research provides valuable insights and recommendations for stakeholders regarding optimal strategies for adopting H-DEDU.

TITLE: Deduplication on Encrypted Big Data in Cloud

YEAR: 2016

AUTHOR: Zheng Yan; Wenxiu Ding; Xixun Yu; Haiqi Zhu; Robert H. Deng

DESCRIPTION:

Cloud computing has transformed service delivery by leveraging internet resources, with data storage being a key cloud service. To ensure data privacy, encrypted data storage has become common practice. However, encrypting data poses challenges for

efficient deduplication in cloud environments, which is crucial for large-scale data storage and processing. Existing deduplication schemes are ineffective with encrypted data, lacking security and flexibility for data access control. This paper introduces a novel scheme that integrates ownership challenge and proxy re-encryption for deduplicating encrypted cloud data while supporting access control. Extensive analysis and simulations demonstrate the scheme's efficiency and effectiveness, particularly for big data deduplication in cloud storage. By combining deduplication, access control, and encryption, our scheme optimizes storage resources, enhances data privacy, and enables efficient data management. The evaluation confirms the practical viability of our proposed scheme for encrypted data deduplication in cloud storage scenarios.

TITLE: Game Theoretical Study on a Client-Controlled Deduplication Scheme

YEAR: 2020

AUTHOR: Xueqin Liang; Zheng Yan; Wenxiu Ding; Robert H. Deng

DESCRIPTION:

Data deduplication is vital for efficient storage in cloud services, but existing schemes overlook the complex dynamics and conflicting interests among stakeholders. This research paper presents an incentive mechanism tailored for a client-controlled deduplication scheme. Game theory is used to evaluate the incentives provided to stakeholders and their impact on scheme adoption and participation. Experimental results using real-world data demonstrate the effectiveness of the proposed mechanism in increasing deduplication percentage and discouraging free-riding behaviors. This study contributes to the field by addressing the design of incentive mechanisms for data deduplication, providing valuable insights into stakeholder dynamics and motivations in deduplication contexts.

TITLE: LEVER: Secure Deduplicated Cloud Storage With Encrypted Two-Party Interactions in Cyber--Physical Systems

YEAR: 2020

AUTHOR: Zahra Pooranian; Mohammad Shojafar; Sahil Garg; Rahim Taheri; Rahim Tafazolli

DESCRIPTION:

Cloud-enabled cyber-physical systems (CCPS) leverage the interaction between cyber components and cloud computing, with cloud storage playing a crucial role in optimizing storage utilization through data deduplication. However, existing solutions face challenges related to security and privacy when dealing with independently encrypted data. This article introduces LEVER, a novel protocol that establishes secure communication between CCPS users and cloud storage, reconciling encryption and data deduplication. LEVER is the first brute-force resilient encrypted deduplication protocol, relying solely on cryptographic two-party interactions. Numerical analysis demonstrates its high performance and practical applicability, making LEVER an innovative solution for the encryption-deduplication dilemma in CCPS. It ensures data security while achieving efficient data deduplication, enabling improved performance and usability in real-world CCPS applications.

TITLE: Efficient Computationally Private Information Retrieval from Anonymity or Trapdoor Groups

YEAR: 2010

AUTHOR: Jonathan Trostle & Andy Parrish

DESCRIPTION:

This paper introduces a novel family of computationally efficient Private Information Retrieval (PIR) protocols. Unlike existing protocols, our protocols focus on reducing the computational complexity on the database server while offering improved security and performance properties. By treating the database as a bit sequence, we only require addition operations on the server. Our general protocol relies on the Hidden Modular Group Order (HMGO) assumption or sender anonymity for security. Additionally, we present two specialized protocols based on these assumptions. Through implementation and performance evaluation, we demonstrate the feasibility and competitive performance of our protocols, offering enhanced efficiency and security compared to previous cPIR solutions.

TITLE: Message-Locked Proofs of Retrievability with Secure Deduplication
YEAR: 2016

AUTHOR: Dimitrios Vasilopoulos, Melek Onen, Kaoutar Elkhiyaoui, Refik Molva
DESCRIPTION:

This paper introduces the concept of message-locked Proof of Retrievability (PoR) to reconcile storage correctness with file-based cross-user deduplication in cloud computing systems. The message-locked PoR approach ensures that PoR does not affect duplicate data and depends solely on the value of the data segment. We present two instances of existing PoRs and highlight the modification in the setup phase to support messagelocked PoR. Additionally, we propose a server-aided message-locked key generation technique that enhances security guarantees. By adopting these approaches, we successfully address the challenge of reconciling PoR with deduplication, demonstrating improved security and effectiveness.

TITLE: A Secure Two-Phase Data Deduplication Scheme
YEAR: 2015

AUTHOR: Pierre Meye; Philippe Raipin; Frederic Tronel; Emmanuelle Anceaume

DESCRIPTION:

This paper presents a deduplication scheme that focuses on countering attacks from malicious clients and ensuring data security. By combining intra-user and inter-user deduplication techniques, our approach establishes a secure storage system while maintaining imperceptibility to clients. The scheme achieves significant storage space savings, reduces bandwidth consumption, and optimizes network bandwidth utilization. Evaluation results demonstrate the scheme's effectiveness in addressing security concerns and providing efficient storage utilization.

EXISTING SYSTEM:

Users can establish their own encrypted protected area within the storage environment, which is represented as a file stored on the physical disk. When the encrypted protected area is unlocked, it is automatically mounted as a virtual disk. For unauthorized users or those without legitimate identities, the encrypted security protection remains as an

inaccessible encrypted disk file. Furthermore, the data security system incorporates a file safe deletion function, utilizing random copy technology. When users employ this tool to delete files, the deleted files become irrecoverable, and the original data cannot be accessed even when examining the physical disk. Additionally, an automatic hiding function is implemented within the encrypted protection zone.

DISADVANTAGES OF EXISTING SYSTEM:

- Less collusion resistance
- Reduced performance
- Less flexibility and scalability

PROPOSED SYSTEM

This project presents a novel technology for designing a data storage security system that incorporates encryption to ensure the confidentiality and integrity of information on cloud platforms. By leveraging the robust computing capabilities offered by the cloud, the system maximizes processing efficiency, resulting in improved data processing, system performance, and enhanced functionalities of individual modules. Once the system is fully developed and optimized, it can be effectively utilized for various information management tasks.

ADVANTAGES OF PROPOSED SYSTEM

- Improved performance
- Confidentiality
- Collusion resistance
- Flexibility and scalability

AES ALGORITHM

The more popular and widely adopted symmetric encryption algorithm likely to be encountered nowadays is the Advanced Encryption Standard (AES). It is found at least six time faster than triple DES.

A replacement for DES was needed as its key size was too small. With increasing computing power, it was considered vulnerable against exhaustive key search attack. Triple DES was designed to overcome this drawback but it was found slow.

The features of AES are as follows −

- Symmetric key symmetric block cipher
- 128-bit data, 128/192/256-bit keys
- Stronger and faster than Triple-DES
- Provide full specification and design details
- Software implementable in C and Java

Operation of AES

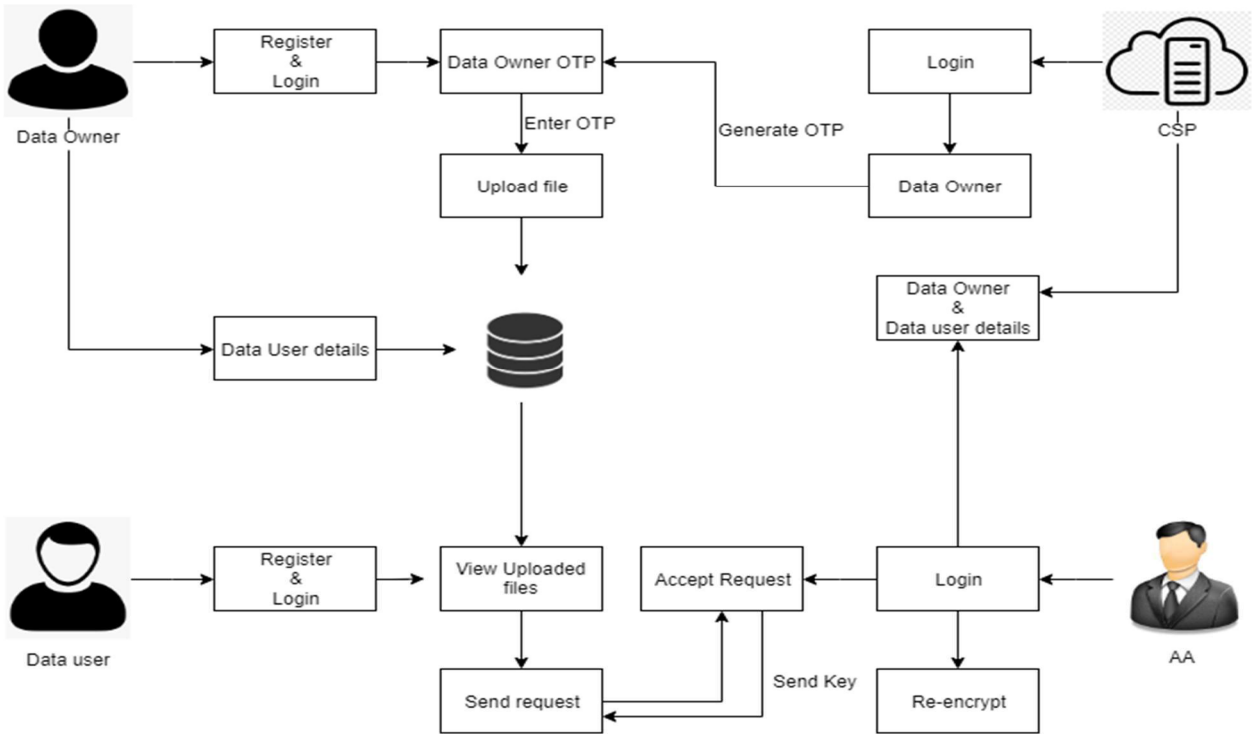
AES is an iterative rather than Feistel cipher. It is based on ‘substitution–permutation network’. It comprises of a series of linked operations, some of which involve replacing inputs by specific outputs (substitutions) and others involve shuffling bits around (permutations).

Interestingly, AES performs all its computations on bytes rather than bits. Hence, AES treats the 128 bits of a plaintext block as 16 bytes. These 16 bytes are arranged in four columns and four rows for processing as a matrix −

Unlike DES, the number of rounds in AES is variable and depends on the length of the key. AES uses 10 rounds for 128-bit keys, 12 rounds for 192-bit keys and 14 rounds for 256-bit keys. Each of these rounds uses a different 128-bit round key, which is calculated from the original AES key.

ARCHITECTURE DIAGRAM

The below Diagrams describe how activities are coordinated to provide a service which can be at different levels of abstraction. Typically, an event needs to be achieved by some operations, particularly where the operation is intended to achieve a number of different things that require coordination.



EXPERIMENTAL RESULTS:

Performance Evaluation:

- Measure the processing efficiency of the data storage security system by recording the time taken for various operations, such as data encryption, decryption, integrity checking, and duplication detection.
- Compare the performance metrics of the system with different data sizes and types to assess scalability.
- Present the results in terms of execution time, throughput, and resource utilization.

Security Analysis:

- Perform a security evaluation to assess the effectiveness of the encryption techniques implemented in the system.
- Evaluate the system's resistance against known attacks, such as brute-force attacks, cryptographic attacks, and data tampering attempts.
- Measure the system's ability to protect data confidentiality and integrity.
- Provide evidence of successful encryption, decryption, and integrity checking for sample data sets.

Duplication Detection:

- Demonstrate the accuracy of the duplication detection mechanism by creating duplicate copies of files and verifying their detection by the system.
- Measure the detection time and accuracy for different file sizes and types.
- Present the results in terms of false positives, false negatives, and detection rate.

Integration with Cloud Service Providers:

- Test the compatibility and integration of the system with popular cloud service providers, such as Amazon S3, Microsoft Azure, or Google Cloud.
- Evaluate the system's performance and security when deployed in a cloud environment.
- Measure data transfer speeds, latency, and resource utilization.

Usability Evaluation:

- Conduct user testing to assess the system's usability and user experience.
- Collect feedback from users regarding the system's ease of use, intuitiveness, and overall satisfaction.
- Identify any areas for improvement or additional features based on user feedback.

CONCLUSION:

In this study, the focus was on the development of a data storage security system that tackles the challenges of integrity checking and duplication detection. By integrating the TDICP protocol and a novel challenge-response mechanism, the system enables independent integrity verification by data holders and accurate duplication detection. Through rigorous security and performance analysis, it was demonstrated that the system is effective and efficient in ensuring data integrity and minimizing duplication in storage environments.

FUTURE SCOPE:

The developed data storage security system presents several avenues for future exploration and improvement. These include enhancing the system's scalability to handle large-scale deployments, exploring advanced encryption techniques for stronger data protection, integrating with various cloud service providers' platforms, enabling dynamic policy management for fine-grained access control, investigating the integration of blockchain technology, incorporating multi-factor authentication mechanisms, and addressing compliance and regulatory considerations. By focusing on these areas, the system can evolve to meet the growing demands and challenges of data storage security, providing enhanced functionality, security, and usability for users in diverse settings.

REFERENCES:

[1] Z. Yan, L. F. Zhang, W. X. Ding, and Q. H. Zheng, "Heterogeneous data storage management with deduplication in cloud computing," IEEE Transactions on Big Data, pp. 1–1, 2017.

[2] Z. Yan, W. X. Ding, and H. Q. Zhu, "A scheme to manage encrypted data storage with deduplication in cloud," in International Conference on Algorithms and Architectures for Parallel Processing, 2015.

[3] Z. Yan, M. J. Wang, Y. X. Li, and A. V. Vasilakos, "Encrypted data management with deduplication in cloud computing," IEEE Cloud Computing, vol. 3, no. 2, pp. 28–35, 2016.

[4] W. Shen, Y. Su, and R. Hao, "Lightweight cloud storage auditing with deduplication supporting strong privacy protection," IEEE Access, vol. 8, pp. 44 359–44 372, 2020.

[5] Q. Zheng and S. Xu, "Secure and efficient proof of storage with deduplication," in CODASPY '12, New York, NY, USA, 2012, p. 1–12.

[6] A. Giuseppe, R. Burns, and C. Reza, "Provable data possession at untrusted stores," in Proceedings of the 14th ACM Conference on Computer and Communications Security, 2007, pp. 598–609.

[7] G. Ateniese, R. Burns, R. Curtmola, J. Herring, O. Khan, Z. Peterson, and D. Song, “Remote data checking using provable data possession,” ACM Transactions on Information and System Security, vol. 14, pp. 1– 34, 2011.

[8] Z. Wen, J. Luo, H. Chen, J. Meng, X. Li, and J. Li, “A verifiable data deduplication scheme in cloud computing,” in INCOS ’14, USA, 2014, p. 85–90.

[9] P. Meye, P. Raïpin, F. Tronel, and E. Anceaume, “A secure two-phase data deduplication scheme,” in HPCC ’14, CSS ’14, ICESS ’14, 2014, pp. 802–809.

[10] D. Vasilopoulos, M. Önen, K. Elkhiyaoui, and R. Molva, “Message locked proofs of retrievability with secure deduplication,” in Proceedings of the 2016 ACM on Cloud Computing Security Workshop, 2016, pp. 73– 83.