

Assignment1

Data Leakage in Cloud Storage

Lecturer

Dr. Mohammed Anbar

Student

Ahmed Mohammed Ahmed Alnahari

251003035

Table of contents

Introduction and A real-life examples of the dangers of data leakage

Summary research papers

Critical Analysis and Connections

Synthesis Matrix for Literature Review

Conclusion

References

Introduction

Cloud storage has revolutionized data management, providing scalability, flexibility, and cost-effectiveness. However, concerns over data leakage persist, posing security risks to users and organizations.

A real-life example of the dangers of data leakage is:

1. Yahoo (2013)

- **Details:** In one of the largest data breaches in history, Yahoo suffered a cyberattack that exposed **all 3 billion user accounts**.
- **Leaked Data:** Names, email addresses, dates of birth, phone numbers, and hashed passwords.
- **Consequences:** Yahoo faced significant reputational damage and financial losses, eventually selling its core business to Verizon at a reduced valuation.

2. Sony PlayStation Network (2011)

- **Details:** A security breach in Sony's PlayStation Network exposed the personal details of **77 million users**.
- **Leaked Data:** Names, addresses, email addresses, usernames, passwords, and potentially credit card details.
- **Consequences:** Sony had to shut down its network for over a month, incurring losses of around **\$171 million**.

3. British Airways (2018)

- **Details:** A cyberattack on British Airways compromised around **500,000 customer records**.

- **Leaked Data:** Payment card details, names, addresses, and travel booking information.
- **Consequences:** The airline was fined **£20 million (\$26 million)** by the UK Information Commissioner's Office (ICO) for failing to protect customer data.

4. Box (2019)

- **Details:** Security researchers found that **sensitive customer data stored on Box's cloud storage service** was accessible due to misconfigured access permissions.
- **Leaked Data:** Passport images, financial documents, personal identification numbers, and confidential business data.
- **Consequences:** Potential exploitation of exposed credentials, raising concerns about misconfigurations in cloud storage.

5. Facebook (2018)

- **Details:** A data breach affected **over 50 million Facebook users** due to a vulnerability in the "View As" feature.
- **Leaked Data:** Personal information, including access tokens that could be used to hijack accounts.
- **Consequences:** Widespread backlash and scrutiny over Facebook's data security practices, leading to regulatory investigations.

These incidents highlight the critical need for robust cybersecurity measures, including encryption, secure authentication, and cloud security best practices, to prevent data leaks and breaches.

we will critically analyzes six research papers addressing data leakage in cloud storage, focusing on objectives, methodologies, results, strengths, and limitations.

Summary research paper's:

Paper 1:

"Why Does Your Data Leak? Uncovering the Data Leakage in Cloud from Mobile Apps" (Zuo et al., 2019)

Objective : Investigates the causes of data leakage in cloud services used by mobile applications.

Methodology : Conducts an empirical study on 1.6 million Android apps using automated scanning tools.

Results : Identified over 15,000 exposed cloud servers due to poor authentication and misconfiguration.

Strengths : Large-scale analysis, automated detection tool (LeakScope).

Weaknesses : Limited to Android apps; lacks real-time protection mechanisms.

Paper 2:

"StoreSim: Optimizing Information Leakage in Multi-cloud Storage Services" (Zhuang et al., 2015)

Objective: Proposes a system to minimize data leakage in multi-cloud environments.

Methodology: Uses MinHash and Bloom filters to classify and distribute data securely.

Results: Reduced data leakage risk by 60% compared to traditional distribution.

Strengths: Effective in reducing exposure across cloud providers.

Weaknesses: Computational overhead in managing data distribution.

Paper 3:

"Secure Data Exchange and Data Leakage Detection in an Untrusted Cloud" (Ulybyshev et al., 2018)

Objective: Enhances secure data exchange while detecting leakage in untrusted cloud environments.

Methodology: Introduces "Active Bundle" technology, embedding encryption and watermarking.

Results: Successfully identified insider threats and unauthorized access attempts.

Strengths: Dual-layer security mechanism.

Weaknesses: Increased computational complexity in handling encrypted bundles.

Paper 4:

"Survey of Techniques on Data Leakage Protection and Methods to Address the Insider Threat" (Montano et al., 2022)

Objective: Reviews of 42 research studies on data leakage prevention and insider threats.

Methodology: Systematic literature review of techniques including encryption, machine learning, and blockchain.

Results: Found encryption and DRM policies most effective, but integration challenges remain.

Strengths: Comprehensive review identifies research gaps.

Weaknesses: Lacks experimental validation.

Paper 5:

"DLPFS: The Data Leakage Prevention FileSystem" (Braghin et al., 2022)

Objective : Develops a file system (DLPFS) to prevent unintentional data leakage.

Methodology: Implements access control layers within the filesystem on cloud platforms.

Results: Prevented unauthorized file sharing while maintaining usability.

Strengths: Low impact on system performance.

Weaknesses: Primarily focuses on accidental leakage, not targeted attacks.

Paper 6:

"Using Sparse Matrices to Prevent Information Leakage in Cloud Computing" (Khan et al., 2018)

Objective: Introduces sparse matrices for privacy-preserving computation in the cloud.

Methodology: Uses obfuscation techniques to mask sensitive data in matrix operations.

Results: Prevented cloud providers from inferring actual data values

Strengths: Lightweight, computationally efficient.

Weaknesses: Limited to specific mathematical operations.

Critical Analysis and Connections

- **Common Findings:**

- Misconfiguration and weak authentication are major causes of cloud data leakage.
- Encryption and access control remain the primary solutions.
- Insider threats require additional mechanisms like watermarking and behavioral analysis.

- **Contradictions:**

- Some studies prioritize efficiency (e.g., sparse matrices), while others emphasize comprehensive security (e.g., Active Bundles).
- The effectiveness of machine learning-based solutions is debated, as they require extensive datasets and training.

- **Common Findings:**

- Misconfiguration and weak authentication are major causes of cloud data leakage.
- Encryption and access control remain the primary solutions.
- Insider threats require additional mechanisms like watermarking and behavioral analysis.

- **Key Connections Between Papers :**

- Papers 1 and 4: Both highlight misconfiguration as a leading cause of data leaks, reinforcing the need for improved security settings.
- Papers 3 and 5: Both introduce security layers (Active Bundles and DLPFS) to enhance file security and prevent unauthorized access.
- Papers 2 and 6: Focus on efficiency in cloud security, with StoreSim reducing leakage through optimized data distribution and sparse matrices enabling privacy-preserving computation.

- **Contradictions and Divergences:**

- Some studies prioritize efficiency (e.g., sparse matrices), while others emphasize comprehensive security (e.g., Active Bundles).
- The effectiveness of machine learning-based solutions is debated, as they require extensive datasets and training.
- Paper 6 provides a lightweight method for protecting cloud computations, whereas Paper 5 focuses on structured access control within file systems, highlighting different levels of security focus.

- **Common Findings:**

- Misconfiguration and weak authentication are major causes of cloud data leakage.
- Encryption and access control remain the primary solutions.
- Insider threats require additional mechanisms like watermarking and behavioral analysis.

Synthesis Matrix for Literature Review

Paper	Year	Authors	Methodology	Key Findings	Limitations
Why Does Your Data Leak?	2019	Zuo et al.	Empirical analysis of Android apps	Identified 15,000+ misconfigured cloud servers	Limited to Android apps
StoreSim	2015	Zhuang et al.	MinHash & Bloom filters	Reduced leakage risk by 60%	Computational overhead
Secure Data Exchange	2018	Ulybyshev et al.	Encryption & watermarking	Detected unauthorized access	High processing requirements
Survey on Data Leakage Protection	2022	Montano et al.	Literature review	Found encryption most effective	Lacks experimental validation
DLPFS	2022	Braghin et al.	File system access control	Prevented accidental leakage	Not designed for targeted attacks
Sparse Matrices for Privacy	2018	Khan et al.	Obfuscation techniques	Protected computation in cloud	Limited application scope

Conclusion

Data leakage in cloud storage is a critical issue requiring a multi-layered security approach. The reviewed studies highlight various solutions, from encryption to access controls and behavioral monitoring. Future research should focus on integrating machine learning, real-time monitoring, and optimizing security without compromising performance.

References

- Zuo, C., Lin, Z., & Zhang, Y. (2019). Why Does Your Data Leak? Uncovering the Data Leakage in Cloud from Mobile Apps. IEEE Security & Privacy Conference. <https://yinqian.org/papers/sp19.pdf>
- Zhuang, H., Rahman, R., Hui, P., & Aberer, K. (2015). StoreSim: Optimizing Information Leakage in Multi-cloud Storage Services. IEEE CloudCom. <https://web.archive.org/web/20170816235949/https://infoscience.epfl.ch/record/212799/files/storesim.pdf>
- Ulybyshev, D., Bhargava, B., & Oqab-Alsalem, A. (2018). Secure Data Exchange and Data Leakage Detection in an Untrusted Cloud. ICACCT (Springer CCIS). https://www.cs.purdue.edu/homes/bb/Papers2018/LeakageDetection/Ulybyshev2018_Chapter_SecureDataExchangeAndDataLeakage-%20FromRealSpringerJournal.pdf
- Montano, I. H., Aranda, J. J. G., Diaz, J. R., Cardín, S. M., de la Torre Díez, I., & Rodrigues, J. J. P. C. (2022). Survey of Techniques on Data Leakage Protection and Methods to Address the Insider Threat. Cluster Computing (Springer).

https://www.researchgate.net/publication/362010417_Survey_of_Techniques_on_Data_Leakage_Protection_and_Methods_to_address_the_Insider_threat/link/62d8314e45865722d7791f97/download

- Braghin, S., Simioni, M., & Sinn, M. (2022). DLPFS: The Data Leakage Prevention FileSystem. Cloud Security & Privacy Workshop.
<https://arxiv.org/pdf/2108.13785>
- Khan, K., Shaheen, M., & Wang, Y. (2018). Using Sparse Matrices to Prevent Information Leakage in Cloud Computing. IEEE FiCloud Conference.

<https://qspace.qu.edu.qa/handle/10576/13165#:~:text=Cloud%20computing%20represents%20the%20promise,protocols%20ensure%20no%20information%20leakage>

- Other references :

<https://me.kaspersky.com/blog/top-five-data-breaches-in-history/12036/>

<https://xygeni.io/ar/sscs-glossary/what-is-data-leakage/>