STRE: Privacy-Preserving Keyword Search over Multiple Clouds

Jingwei Li\textsuperscript{1}, Dan Lin\textsuperscript{2}, Anna Squicciarini\textsuperscript{3} and Chunfu Jia\textsuperscript{1}

\textsuperscript{1}College of Computer and Control Engineering, Nankai University
\textsuperscript{2}Computer Science, Missouri University of Science and Technology
\textsuperscript{3}Information Sciences & Technology, Pennsylvania State University

September 24, 2014
Presentation Outline

Introduction

System Model

Our Approach: STRE

Evaluation

Conclusion
Presentation Outline

Introduction

System Model

Our Approach: STRE

Evaluation

Conclusion
Privacy and Reliability Challenge in Cloud

- Cloud computing is growing exponentially, whereby there are now hundreds of cloud service providers (CSPs) of various sizes.
- This multi-cloud environment (also known as Cloud-of-Clouds and InterCloud) offers plenty of new opportunities and avenues to cloud consumers.
- While cloud consumers may enjoy cheaper data storage and powerful computation capabilities offered by multiple clouds, consumers also face more complicated reliability issues and privacy preservation problems of their outsourced data.
More Specific

- Cloud consumers are typically suggested to adopt searchable encryption techniques to encrypt their outsourced data in a way that the encrypted data can be directly searched by the CSPs without decryption [privacy is preserved].
- Despite many efforts devoted to improving efficiency and security of the searchable encryption, there is little consideration on ensuring the reliability of the searchable encrypted data [lack of reliability].
- Existing reliability guarantees solely rely on each CSP’s own backup solution, which however could be a single-point of failure [reliability is necessary].
Motivation and Overview

- **Motivation:** Obtain reliability on searchable encrypted data.
- **Overview:**
  - User distributes files to multiple clouds for storage;
  - Even a subset of clouds are unavailable, the stored files are retrievable, including, 1) The keyword-based search service is still available; 2) The retrieved chunks (from available clouds) can be used for reconstructing the files.
Our Contribution

We propose a privacy-preserving STorage and REtrieval (STRE) mechanism that enables the cloud users to distribute and search their encrypted data in multiple CSPs. STRE is full-featured:

- It not only ensures security and privacy but also provides reliability guarantees for the outsourced searchable encrypted data.
- It also offers the benefit of flexible storage strategy and partially hidden search pattern.
Presentation Outline

Introduction

System Model

Our Approach: STRE

Evaluation

Conclusion
System Model: Multiple Clouds Collaboration

- **User**: stores files and later presents keyword-based query.
- **CSPs**: work together to provide storage service.
Protocols Framework

Storage Protocol
1. A master secret key is generated and given to user;
2. User uploads a collection of files to multiple clouds.

Search Protocol
1. User generates and sends trapdoor for query keyword;
2. Clouds perform collaborative search and return results;
3. User recovers and decrypts returned results.
Presentation Outline

Introduction

System Model

Our Approach: STRE

Evaluation

Conclusion
Three Challenges

We identify three challenges for designing the protocol:

- How to add redundancy, such that the reliability is preserved without violating privacy;
- How to construct the “trapdoor function” for multiple CSPs to search on encrypted data;
- How to design the “collaborative and interaction protocol”, such that the underlying file chunks could be respectively located by these CSPs.
Encrypt clear file $f$ into $c_f$;
Divide $c_f$ into natural chunks $c_1$ and $c_2$;
Make linear combination on $c_1$ and $c_2$ to get $c_3$ and $c_4$.

Even 1st and 3rd CSPs are unavailable, the $c_f$ could still be reconstructed from $c_2$ and $c_4$. 
Technical for Challenge II

- Build and distribute an index [Curtumola et al. 2006] to multiple clouds;
- For searching files related to keyword $w$, make $(n, t)$-secret sharing on $(\pi(\text{msk}_2, w), \text{P}(\text{msk}_1, w))$, $t$ is threshold and $n$ is the number of CSPs.
Technical for Challenge II

More specifically, build two matrices:

$$\begin{bmatrix}
\pi(msk_2, w) & P(msk_1, w) \\
\text{filling random values}
\end{bmatrix}$$

$$\begin{bmatrix}
0 & 0 \\
\quad & \text{equal } S
\end{bmatrix}$$

- Make \((n, t)\)-secret sharing [Bai et al. 2007] on \(S\) and distribute the share \(v_i\) to corresponding CSP [Search pattern is partially hidden under the collusion of at most \(t - 1\) CSPs];
- Publish \(S'\), and the random filled values could be used for checking the correctness of reconstruction in future.
Technical for Challenge III

We use \((n, n)\)-secret sharing:

1. \(i\)th cloud maintains \(\overline{B}_i = [0, v_i, 0]\) and makes \((n, n)\)-secret sharing \(\{\overline{B}_{ij}\}_{j=1}^n\) on \(\overline{B}_i\).

2. Distribute \(\{\overline{B}_{ij}\}_{j\neq i}\), and gather and sum \(\overline{B}'_i = \sum_{j=1}^n B_{ji}\) sent from other clouds.

3. Distribute \(\overline{B}'_i\), and gather and sum \(\overline{B} = \sum_{i=1}^n \overline{B}_i\) sent from other clouds.

\(\overline{B}\) is the matrix including all the shares, and used for reconstructing \((\pi(msk_2, w), P(msk_1, w))\) for search.
Presentation Outline

Introduction

System Model

Our Approach: STRE

Evaluation

Conclusion
Evaluation Setup

- Simulate multiple clouds using several 32-bit Linux T1 Micro servers in Amazon EC2 platform;
- Use a subset of Enron email dataset as corpus;
- Use AES-256 for file encryption;
- Implement in C using PolarSSL library;
- Comparison: STRE vs SSE-1 [Curtumola et al. 2006].
Evaluation - File Storage

- Index construction time is negligible;
- File encryption time of STRE is slightly longer than SSE-1 due to the additional step of encoding;
- Transmission time of STRE is longer than SSE-1 by the reason that the file chunks need to be sent to multiple clouds.
- For file storage, STRE enables to save \((t - 1)/t\) proportion of space (compared with SSE-1) for each individual CSP.
Compared with SSE-1, STRE mechanism requires another two rounds of interaction with the CSPs as well as encrypted file reconstruction.

STRE achieves slightly better efficiency than SSE-1 in file transmission due to concurrency.

In summary, STRE mechanism achieves both security and reliability without introducing significant overhead compared to SSE-1 [execution time is still within seconds].
Presentation Outline

Introduction

System Model

Our Approach: STRE

Evaluation

Conclusion
Conclusion

We propose the STRE mechanism, to promote reliability of outsourced searchable encrypted data. In STRE, user’s searchable encrypted data is strategically distributed to and stored at multiple CSPs, so as to achieve high crash tolerance. Besides reliability, the STRE mechanism also affords efficient and flexible storage properties and partially hidden search pattern.
Ending

😊 Thanks for your attention! Q & A

Jingwei Li
Nankai University
Email: lijw1987@gmail.com