Efficiency, Utility, Security, and Privacy Trade-Offs

Dr. Chen Zhang

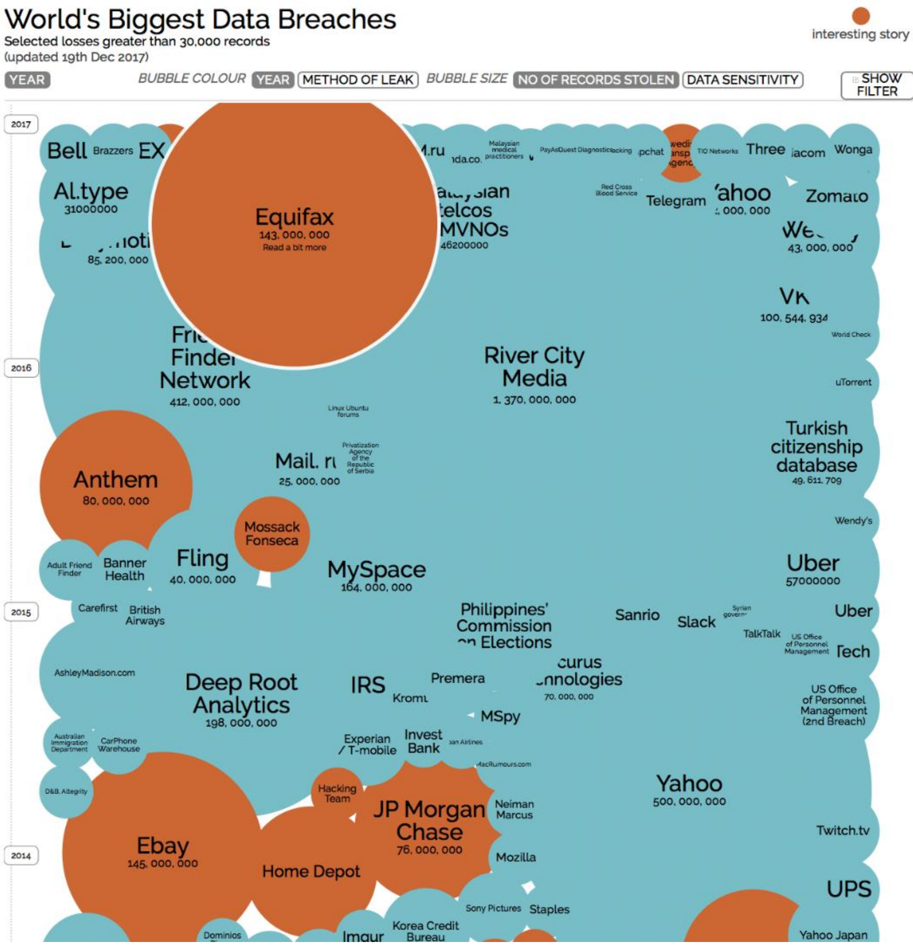
Department of Computer Science
The Hang Seng University of Hong Kong

Data Analytics Over Encrypted Data

Why Encrypted Search?

- – Sensitive data demands encrypted storage.

- – Encrypted search reduces risks of data breaches



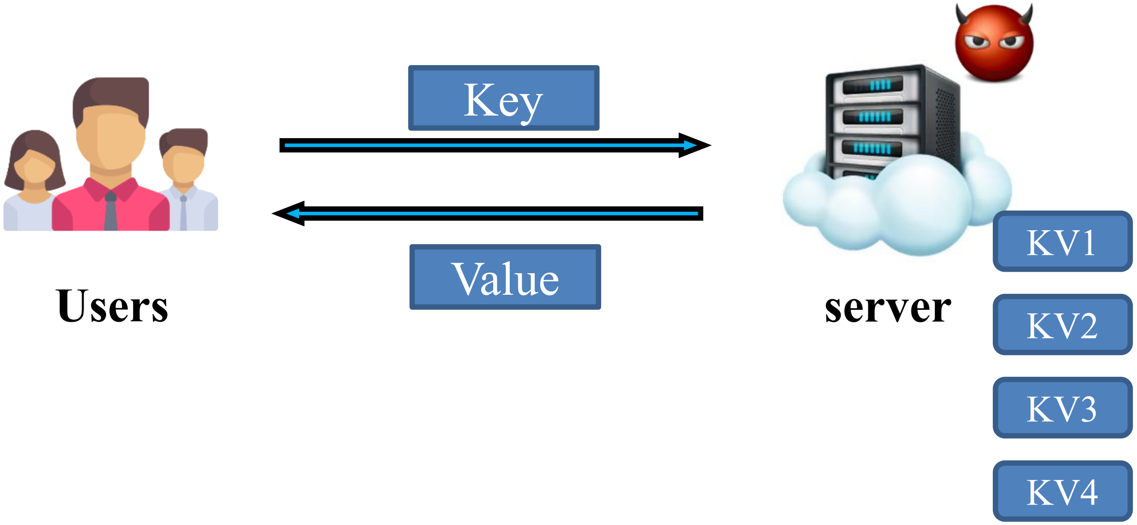
World's Biggest Data Breaches

Selected losses greater than 30,000 records

(updated 19th Dec 2017)

http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/

Cloud Storage



Key

Value

Users

server

KV1

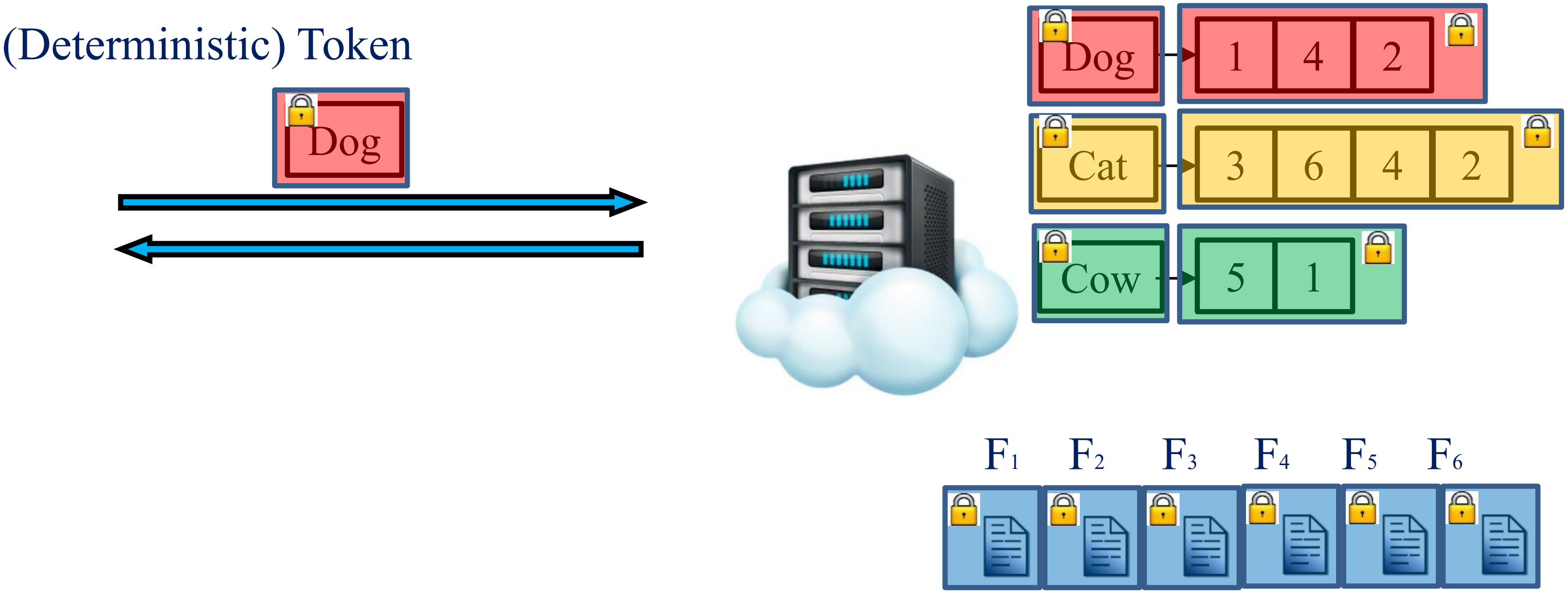
KV2

KV3

KV4

Symmetric Searchable Encryption (SSE) in a Nutshell

SSE: enable untrusted servers to directly search over encrypted data without server-side decryption.



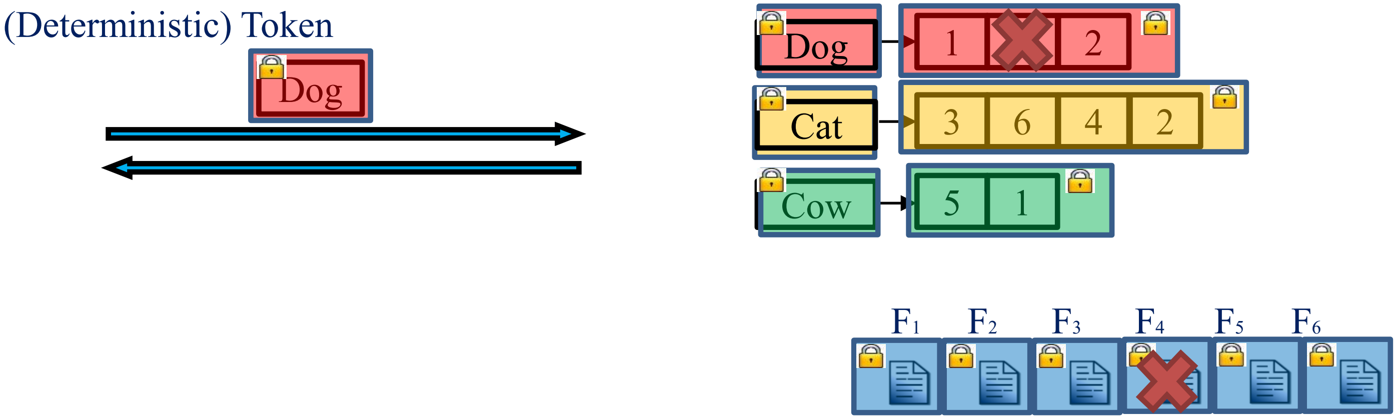
Not enough in Security!

- Vulnerable to volume attack.
  - Attackers can know how many files corresponding to a keyword.
  - Defense: append files to ensure that the number of files corresponding to all keys is the same.
- Update operations introduce additional privacy concerns, e.g., vulnerable to injection attacks:
  - The data addition can reveal the associations between newly added data and previous search results.
  - Defense: User maintains a counter for each key and updates the counter after each addition operation.

No ***Forward Privacy***: old search tokens can be used on new files!

Not enough in Security!

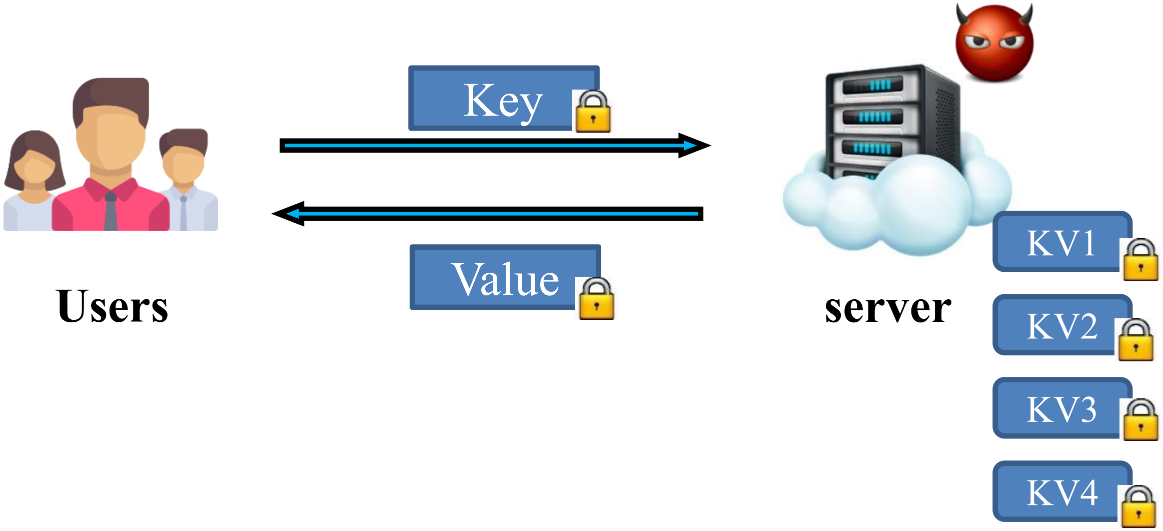
- The server may provide unfaithful query result.
  - Attackers can know how many files corresponding to a keyword.
  - Defense: the data owners to maintain the digests for pre-defined search results and conduct result verification locally.



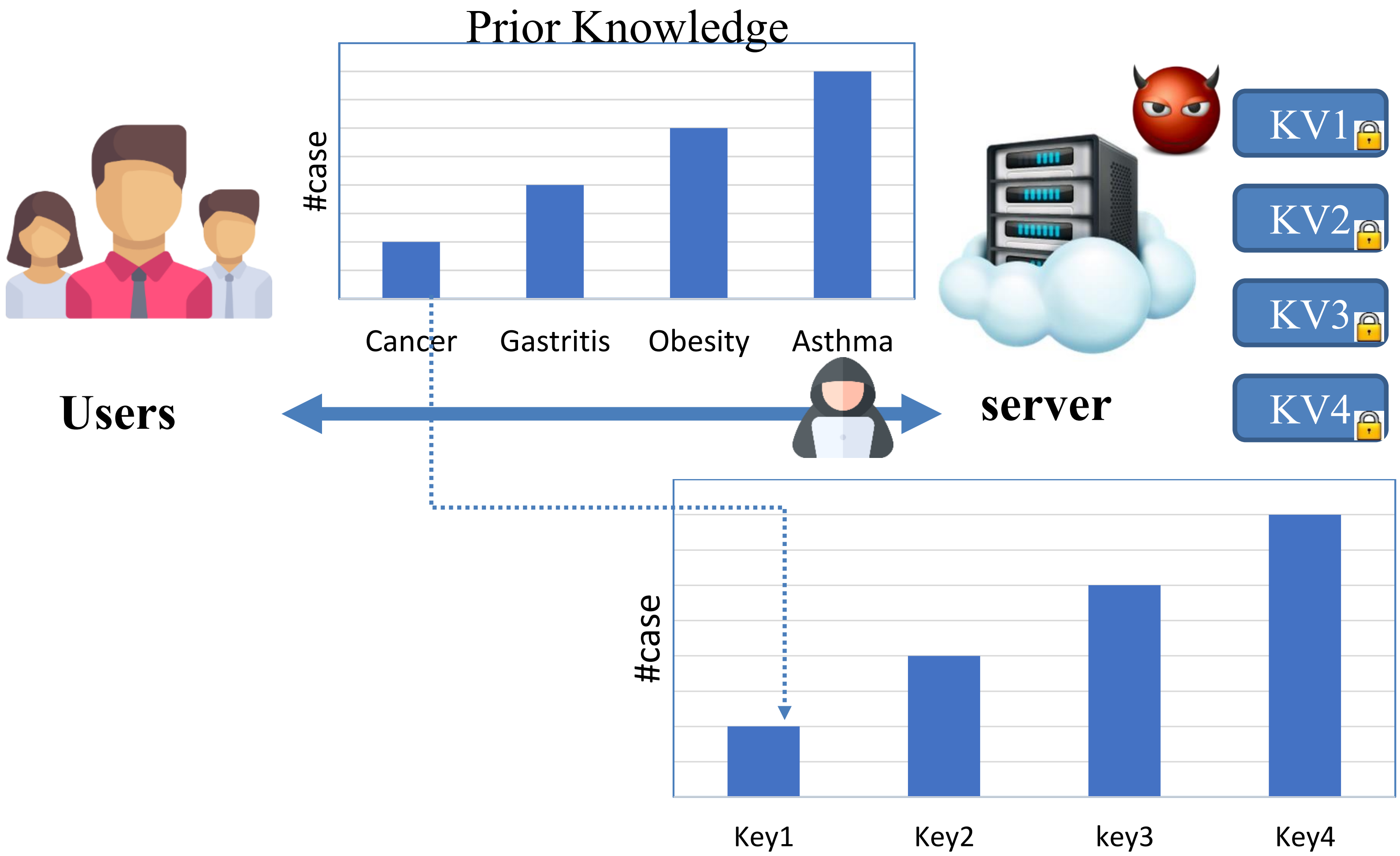
Adversary servers provide unfaithful query execution for saving computational cost

Not enough in Security!

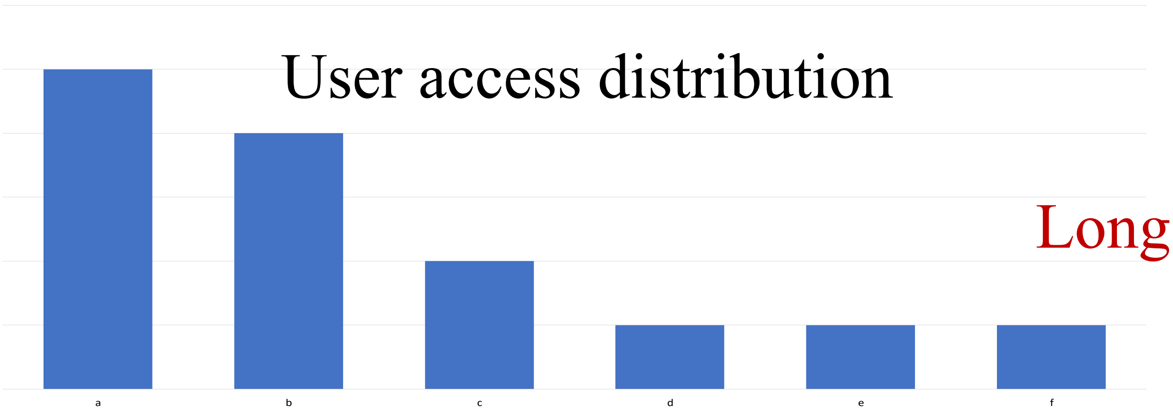
- Vulnerable to frequency-analysis attack.
  - Attackers can know how many files corresponding to a keyword.
  - Defense: add fake queries in the sequence of real queries to disrupt the original access pattern



Example of Frequency-Analysis Attacks

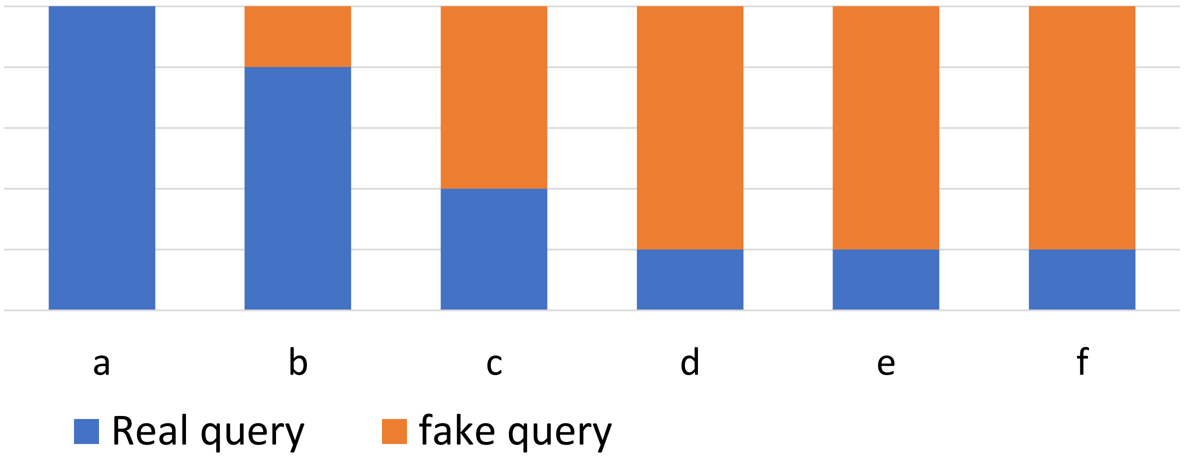


Resisting Frequency-Analysis Attack



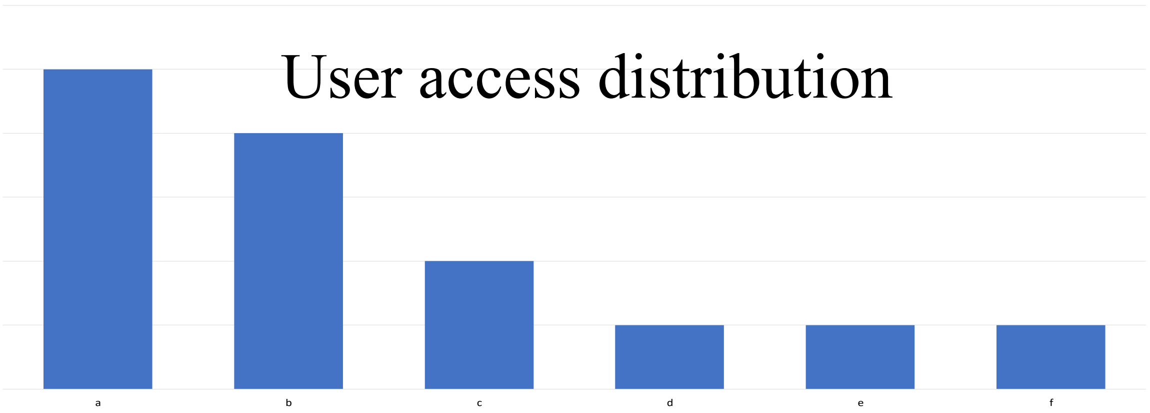
Long-tail law

Add fake queries



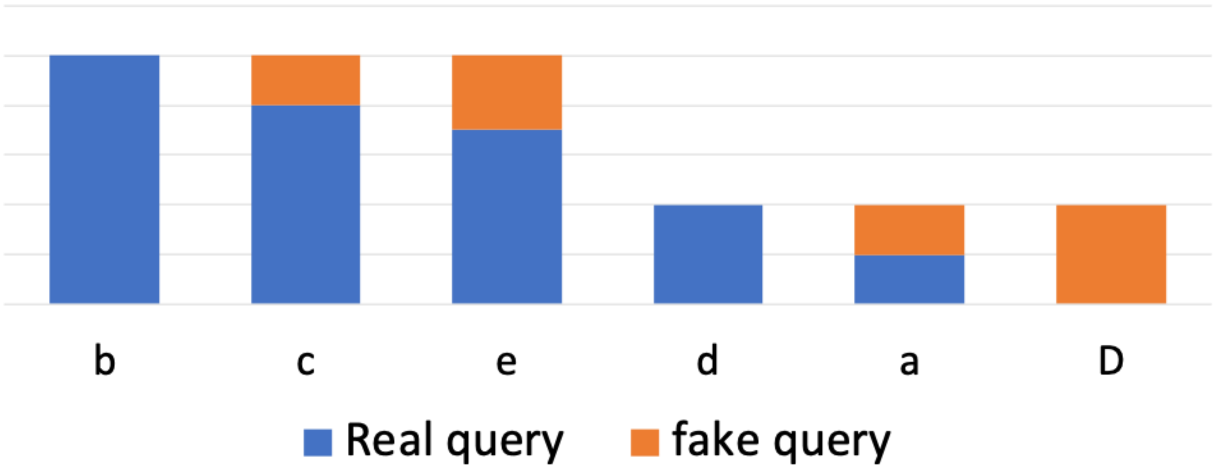
Problem: May incur high bandwidth overhead

Resisting Frequency-Analysis Attack



User access distribution

3-indistinguishable



Trade-off between security overhead and bandwidth overhead

Not Enough in Terms of Functionality!

- Only support put/get requests to access single encrypted value is not enough.
- Supporting rich queries (such as range query, Boolean query) is important.
- Conflict to the initial idea of encryption.
- More storage, bandwidth, and computational overheads are needed.

To ensure security, more complex functions incur more overhead.

Other Trade-Off Examples

- Data backup & data protection: Data backup is an important measure for data protection, but managing and storing backup data also increases storage overhead and system maintenance overhead.
- Access control & data sharing: Access control ensures that only authorized users can access sensitive data, but they can also limit data sharing and collaboration.
- Authentication & user experience: Strong authentication measures (such as multi-factor authentication) can enhance account security, but they can also make it more difficult for users to access their accounts.
- Firewall & data transfer speed: Firewalls are essential for protecting networks from cyber threats, but they can also slow down data transfer speeds.