Knowledge Transfer Office
香港城市大學
City University of Hong Kong

Privacy-assured Similarity Joins over Encrypted Datasets

Communications & Information

Computer/AI/Data Processing and Information Technology

Digital Broadcasting, Telecommunication and Optoelectronics

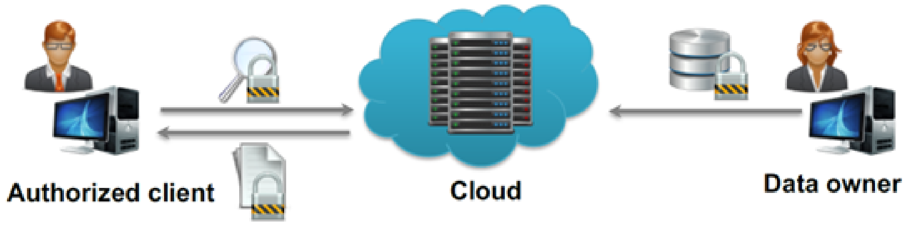

Figure 1. System architecture

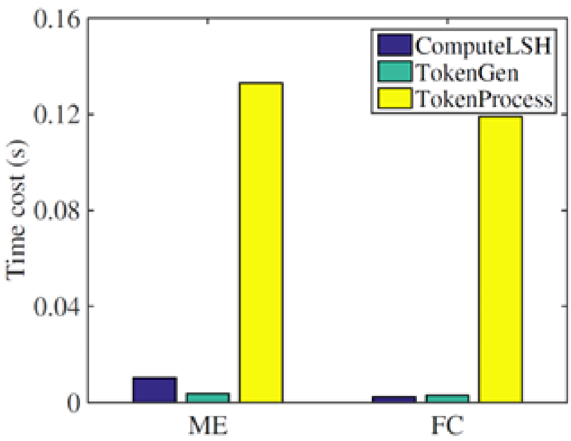

Figure 2. Per token processing time

IP Status

Patent granted

Technology Readiness Level (TRL) ?

4

Inventor(s)

Prof. WANG Cong
Mr. WANG Xinyu
Dr. YUAN Xingliang
Dr. YU Chenyun
Dr. NUTANONG Sarana
Enquiry: kto@cityu.edu.hk

Opportunity

A similarity join algorithm identifies strings in datasets which contain similar content, typically those where the differences fall below a predefined threshold level. With the growing use of “big data”, such algorithms are frequently used in data analysis and cleaning, for instance where data items (tokens) from different sources may be related – to give a simple example the same address may be written as “First Floor XX Building” in one record and “XX Bldg 1/F” in another.

Large databases – sometimes comprising millions of records – commonly contain sensitive material such as financial information, medical records or genetic documentation, and usually hold high-dimensional data, i.e. data records with multiple attributes. To maintain privacy, encryption is used, particularly where they are stored on public cloud facilities. Though encryption ensures data confidentiality, it restricts complex similarity join operations, preventing clouds from performing useful computations on the data. This invention proposes a privacy-assured similarity join algorithm over large-scale encrypted datasets, which enables the public cloud to answer the similarity join query without learning the content of the query dataset and the target dataset, thus ensuring secure query processing.

Technology

The algorithm is realized through three modules, functioning together in a security protocol:

- The module for the data owners transfers the dataset to ciphertext, indexed in an encrypted data structure.
- The module for the users can generate secure queries from their query dataset.
- The module on the cloud side can process those secure queries and return the encrypted candidates.

It is optimized based on the well-known fast and effective algorithm for similarity search, a practical cryptographic technique called searchable encryption, and a carefully designed “result sharing query scheme” that protects the query set distribution while greatly improving the query efficiency.

Advantages

- Efficient and simplified one-step gas-solid reaction
- Reduced the production cost
- High purity and controllability

Applications

- Catalysis (e.g., CO_2 reduction reaction and hydrogen evolution reaction)
- Energy storage and conversion (e.g., battery and supercapacitors)
- Electronic devices
- Condensed matter physics