香港浸會大學
HONG KONG BAPTIST UNIVERSITY

DEPARTMENT OF
COMPUTER SCIENCE
計算機科學系

Privacy-preserving Linear Algebra Framework for Graph Query Algorithms for Massive Networks

PI : Dr. CHOI Byron Koon Kau

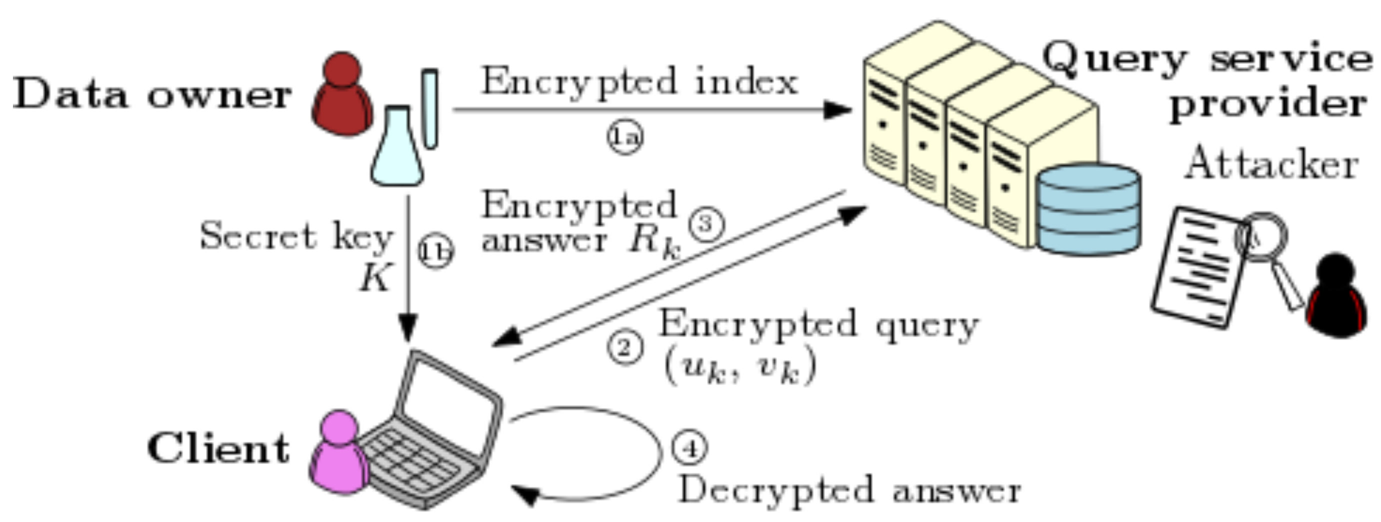
Funding Scheme: **General Research Fund**
Project Ref. No.: **12232716**
Amount Awarded (to HKBU): **HK$ 675,647**
Project Period: **Jan 2017 – Jun 2020**

OBJECTIVES

1. To study a set of linear algebra operators such as set intersection/union, scalar product, matrix multiplication/addition, and propose the encoding and encryption for graph queries
2. To apply privacy-preserving optimizations for the specific algebra operations
3. To unify the operations and develop a publicly available tool (API)

HIGHLIGHTS

1. System Model of Database Outsourcing



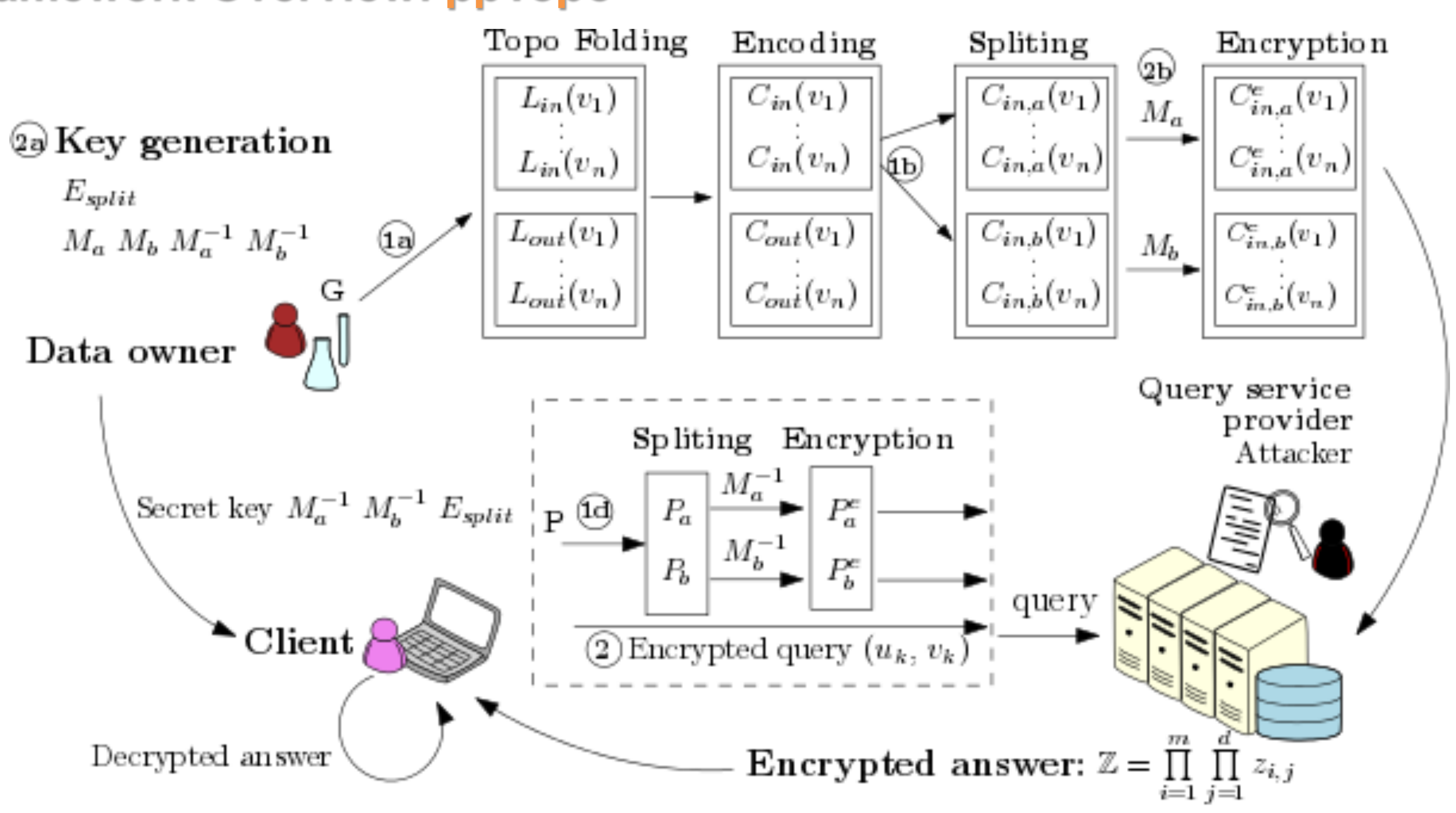
1. **Data Owner** owns the data graphs, cannot host query services to their data, and outsource the data to a service provider.
2. **Clients** submit the graph queries to SP to obtain the answer.
3. **Service Provider (SP)** receives the queries, process them and return the answers to clients. The SP is semi-honest.

Problem. How to efficiently process the queries and protect sensitive data from the graph?

2. Framework for Solving Reachability Query -- ppTopo

- Protect both query privacy and index privacy under the system model
- Use the framework to protect against the ciphertext only attack (COA) and the known plaintext attack (KPA).

3. Framework Overview: ppTopo



1. Data Owner
- Encodes *2-hop* labels (*Lin* and *Lout*) as vectors such that the sum of the plaintexts of *Lin* and *Lout* is 0 modulo 3.
- Splits the encoded vectors by a secret configuration bit-vector.

2. Client
- Encodes a random query permutation matrix.
- Splits the query permutation matrix by the same bit-vector as data owner does for applying the asymmetric scalar-product preserving encryption.
- Decrypts the query result by a secret inverse matrix of query.

3. SP
- The *SP* retrieves the encrypted labels for both *Lin* and *Lout*, then conducts the addition operation for intersection and aggregates the results for communication cost reduction by a serial of multiplications.

4. Discussions of results

Table 1. The query time at the SP side

Graph G	ASPE3	Paillier	CGBE
p2p-30	1.18s	152.2ms	11.97s
p2p-31	2.25s	266.6ms	23.16s
Cit-HepPh	4.26s	497.1ms	42.72s
Amazon0302	406ms	57.8ms	4.16s
Wiki-Vote	334ms	38ms	3.23s
WikiTalk	11.007s	*DNF*	2min2s
LiveJournal	14.180s	*DNF*	2min29s
web-BerkStan	11.072s	1.37s	1min55s

Table 2. The query time at the client side

Graph G	ASPE3	Paillier	CGBE
p2p-30	18.9ms	3min27s	55.5ms
p2p-31	34.6ms	6min2ss	56.4ms
Cit-HepPh	61.1ms	11min16s	56.7ms
Amazon0302	2.7ms	1min18s	59.3ms
Wiki-Vote	2.6ms	51.68s	55.8ms
WikiTalk	129.5ms	*DNF*	133ms
LiveJournal	174.4ms	*DNF*	153ms
web-BerkStan	128.7ms	30min19s	116ms

SELECTED PUBLICATIONS

1. L. Xu, J. Jiang, B. Choi, J. Xu and S. S. Bhowmick. Asymmetric Structure-Preserving Pattern Query Processing for Graph Data. Work in Progress, 2019.
2. Z Fan, B Choi, J Xu, S Bhowmick. Asymmetric structure-preserving subgraph queries for large graphs. ICDE 2015: 339-350.
3. Z Fan, B Choi, Q Chen, J Xu, H Hu, S Bhowmick. Structure-preserving subgraph query services. ICDE 2016: 1532-1533.
4. J. Jiang, P. Yi, B. Choi, Z. Zhang and X. Yu. Privacy-preserving Reachability Query Services for Massive Networks. CIKM 2016, Pages 145-154.