Abstract

Graphs are powerful tools for a wide range of real applications, from Biological and Chemical Databases, Social Networks, Citation Networks to Knowledge Bases. Large graph data repositories have been consistently found in recent applications. Due to the high complexity of graph queries, e.g., NP-Completeness of subgraph query, and the lack of IT expertise, hosting efficient graph query services for the owners of graph data has been a technically challenging task. And hence, they may prefer to outsource their services to third-party service providers (SPs) for scalability, elasticity and efficiency.

Unfortunately, SPs may not always be trusted. Security, typically the integrity and confidentiality, of the data, has been recognized as one of the critical attributes of Quality of Services (QoS). This directly influences the willingness of both data owners and query clients to use SP’s services. To address these concerns, this thesis proposes novel techniques to solve both authentication-aware and privacy-aware subgraph query.

Firstly, we study authenticated subgraph query services (Chapter 3). To support the service, we propose Merkle IFTree (MIFTree) where Merkle hash trees are applied into our Intersection-aware Feature-subgraph Tree (IFTree). IFTree aims to minimize I/O in a well-received subgraph query paradigm namely the filtering-and-verification framework. The structures required to be introduced to verification objects (VOs) and the authentication time are minimized. Subsequently, the overall response time is minimized. For optimizations, we propose an enhanced authentication method on MIFTree.

Secondly, we propose structure-preserving subgraph query services (Chapter 4). A crucial step of this part is to transform the seminal subgraph isomorphism algorithm (the Ullmann’s algorithm) into a series of matrix operations. We propose a novel cyclic group based encryption (CGBE) method for private matrix operations. We propose a protocol that involves the query client and static indexes for optimizations. We prove that the
structural information of both query graph and data graph are preserved under CGBE and analyze the privacy preservation in the presence of the optimizations.

Thirdly, we propose asymmetric structure-preserving subgraph query processing (Chapter 5), where the data graph is publicly known and the query structure/topology is kept secret. Unlike other previous methods for subgraph queries, this part proposes a series of novel optimizations that only exploit graph structures, not the queries. Further, we propose a robust query encoding and adopt our proposed cyclic group based encryption method, so that the query processing can be transformed into a series of private matrix operations and performed securely.

The effectiveness and efficiency of all the techniques presented in this thesis are experimentally evaluated with both real-world and synthetic dataset.

**Keywords:** Outsourced Databases, Graph Databases, Subgraph Query, Authenticated Query Processing, Privacy-preserving Query Processing.
Acknowledgments

I would like to express my deep gratitude to my supervisor, Dr. Byron CHOI, for his great patience, inspiring guidance and constructive suggestions in my studies and research works in these years. He has brought me into this challenging research area and shared insightful experiences with me. I would also like to thank my co-supervisor, Dr. Jian-liang XU, for his continuous encouragement and help. Moreover, I would like to thank Prof. Wenfei FAN, for his great support and advice when I was visiting the University of Edinburgh.

I would like to thank my colleagues and friends for their direct and indirect help. In particular, I should mention Dr. Yun PENG, Mr. Qian CHEN, Mr. Peipei YI, Mr. Lei CHEN, Mr. Chuojie XIAN, Mr. Xiaojing XIE, Mr. Shuxiang YIN, Dr. Xin LIN, Dr. Dingming WU, Mr. Zhuo CHEN, Mr. Jingjing CHEN, Mr. Jintian DENG, Mr. Chao TIAN, Mr. Ruizhe HUANG, Mr. Shen GAO, Dr. Xin LIU, among many others.

Finally, I take this special occasion to thank my father Xiaohui FAN and my mother Lihui WANG for raising me and supporting me for so many years. Without them, I would never go so far.
Table of Contents

Declaration i

Abstract ii

Acknowledgments iv

Table of Contents v

List of Figures ix

Chapter 1 Introduction 1

1.1 Graph Databases .............................................. 1
  1.1.1 Modest Size Graphs ........................................ 1
  1.1.2 Large Graph Repositories ................................ 2
  1.1.3 Graph Queries ............................................ 3
  1.2 Outsourced Graph Databases and Their Security ............ 3
  1.3 Authenticated Subgraph Query Services ...................... 5
  1.4 Structure-preserving Subgraph Query Services .............. 7
  1.5 Asymmetric Structure-preserving Subgraph Query Service for Large Graphs 11
  1.6 Thesis Organization ......................................... 13

Chapter 2 Related Work 14

  2.1 Security-aware Graph Queries and Management ............. 14
    2.1.1 Authenticated Graph Query ............................... 14
    2.1.2 Privacy-preserving Graph Query ......................... 15
2.1.3 Privacy-preserving Graph Publication ................. 16

2.2 Graph Query Processing and Indexing .................. 16
  2.2.1 Indexing Techniques for Graph Query Processing on Transac-
    tional Graph Databases .............................. 16
  2.2.2 Subgraph Isomorphism Algorithm .................. 17

Chapter 3 Towards Efficient Authenticated Subgraph Query Service in Out-
    sourced Graph Databases 19
  3.1 Backgrounds and Preliminaries ....................... 19
    3.1.1 Data Model .................................. 19
    3.1.2 Subgraph Query ............................... 20
    3.1.3 Query Authentication .......................... 22
  3.2 Problem Formulation and Overview .................... 24
    3.2.1 Problem Formulation ............................ 24
    3.2.2 Baseline Authentication — MgIndex ............... 25
    3.2.3 Overview of our Approach ....................... 30
  3.3 Partially Overlapping Features ....................... 32
    3.3.1 Types of Overlapping Features .................. 32
  3.4 Intersection-aware Feature-subgraph Tree (IFTree) .... 34
    3.4.1 IFTree .................................... 34
    3.4.2 Query Processing on IFTree ..................... 35
  3.5 Merkle IFTree (MIFTree) ............................ 38
    3.5.1 Signing MIFTree ................................ 39
    3.5.2 Basic Authentication Method ..................... 40
  3.6 Enhanced Authentication .............................. 48
    3.6.1 Compact Representation of Graph IDs ............... 48
    3.6.2 Clustering Intersect-able Graphs ................ 51
  3.7 Experimental Evaluation ............................... 53
    3.7.1 Experimental Setup ............................. 53
    3.7.2 Experiments on AIDS ............................ 54