

DOCTORAL THESIS

Query authentication in data outsourcing and integration services

Chen, Qian

Date of Award:
2015

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Abstract

Owing to the explosive growth of data driven by e-commerce, social media, and mobile apps, data outsourcing and integration have become two popular Internet services. These services involve one or more data owners (DOs), many requesting clients, and a service provider (SP). The DOs outsource/synchronize their data to the SP, and the SP will provide query services to the requesting clients on behalf of DOs. However, as a third-party server, the SP might alter (leave out or forge) the outsourced/integrated data and query results, intentionally or not. To address this trustworthy issue, the SP is expected to deliver their services in an authenticatable manner, so that the correctness of the service results can be verified by the clients. Unfortunately, existing work on query authentication cannot preserve the privacy of the data being queried. Furthermore, almost all previous studies assume only a single data source/owner, while data integration services usually combine data from multiple sources. In this dissertation, we take the first step to study the authentication of location-based queries with confidentiality and investigate authenticated online data integration services. Cost models, security analysis, and experimental results consistently show the effectiveness and robustness of our proposed schemes under various system settings and query workloads.

Keywords: Query Authentication, Privacy Preservation, Location-based Services, Data Integrity, Data Integration

Acknowledgements

I would like to express my special appreciation and thanks to my supervisor, Prof. Jianliang Xu, for his expertise on 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-supervisors, Dr. Haibo Hu and Dr. Byron Choi, for their great patience in guiding me shaping good research topics, and improving skills of presentation and paper writing. Without their supervision and constant help, this dissertation would not have been possible.

I would like to thank my colleagues for their direct and indirect help. In particular, I should mention Mr. Zhe Fan, Mr. Yafei Li, Mr. Lei Chen, Mr. Peipei Yi, Mr. Jingjing Chen, Mr. Cheng Xu, Dr. Yun Peng, Dr. Xin Lin, Dr. Rui Chen, Dr. Qijun Zhu, Dr. Dingming Wu, Mr. Shen Gao, Mr. Jintian Deng, Mr. Ziwei Yang, Dr. Qilong Han, Mr. Xiaojing Xie, Ms. Shenshen Peng, Mr. Zhuo Chen among many others.

Finally, I take this special occasion to thank my father Guoliang Chen and my mother Yunxian Wu for raising me and supporting me for so many years. I also wish to thank my wife Chengcheng Dai for her love, kindness and understanding. Without them, I would never go so far.

Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Figures	viii
Chapter 1 Introduction	1
1.1 Authenticating Location-based Range Queries with Confidentiality . . .	2
1.2 Authenticating Location-based Top- k Queries with Confidentiality . . .	5
1.3 Authenticating Online Data Integration Services	6
1.4 Dissertation Outline	8
Chapter 2 Literature Review	9
2.1 Data Integration	9
2.2 Query Authentication in Outsourced Databases	10
2.3 Authenticated In-Network Aggregation in Distributed Systems	12
Chapter 3 Authenticating Location-based Range Queries With Confidentiality	15
3.1 Problem Formulation	15
3.1.1 Security Model	16
3.2 Preliminary: Privacy-Preserving Authentication for Single-Dimensional Range Queries	17
3.3 Privacy-Preserving Authentication for Multi-Dimensional Range Queries	20
3.3.1 Authentication on R -tree Index	20
3.3.2 Authentication on Grid-File Index	22

3.3.3	Accumulative Digest for Grid-File Index	26
3.3.4	Cost Models of Three Authentication Schemes	29
3.3.5	Security Analysis	32
3.4	Optimizations	34
3.4.1	Turbo Digest Function (TDF)	34
3.4.2	Linear Ordering and Embedding	37
3.5	Performance Analysis	39
3.5.1	Performance of Turbo Digest Function	42
3.5.2	Basic Query Authentication Performance	42
3.5.3	Performance with Optimizations	43
3.5.4	Update Costs of User Locations	44
3.6	Chapter Summary	46

Chapter 4 Authenticating Location-based Top- k Queries with Confidentiality **48**

4.1	Problem Formulation	48
4.1.1	Security Model	50
4.2	Private Ranking Comparison	51
4.2.1	Private-Paillier based (PPB) Method	51
4.2.2	Pre-signed Lines based (PLB) Method	54
4.3	Authenticating top- k Queries without Compromising Privacy	56
4.3.1	Authentication on MR-tree	57
4.3.2	Authentication on the Power Diagram	63
4.4	Security Analysis	66
4.4.1	Security of PPB and PLB methods	67
4.4.2	MR-tree Based Authentication Scheme	67
4.4.3	Power Diagram based Scheme	69
4.4.4	Security Model for Continuous Top- k Queries	70
4.5	Offline and Online Strategy on Pre-signed Lines	71
4.5.1	DO Offline Strategy on Pre-signed Lines	72

4.5.2	SP Online Strategy on Pre-signed Lines	72
4.6	Performance Evaluation	73
4.6.1	Basic Query Authentication Performance	75
4.6.2	Performance with Pre-signed Line Optimization	76
4.6.3	Impact of Non-spatial Scores	78
4.7	Chapter Summary	79
Chapter 5 Authenticating Online Data Integration Services		80
5.1	Problem Definition and System Model	80
5.2	Homomorphic Secret Sharing Seal (HS ³)	82
5.2.1	Preliminaries	82
5.2.2	HS ³ Design and Query Authentication	84
5.2.3	Security Analysis	88
5.3	Authenticating Multi-Dimensional Data	91
5.3.1	HS ³ -G-tree	91
5.3.2	HS ³ -R-tree	94
5.3.3	Other Types of Queries	96
5.4	Analytical Models	97
5.4.1	Cost Model for Seal	98
5.4.2	Cost Model for HS ³ -G-tree	98
5.4.3	Cost Model for HS ³ -R-tree	99
5.5	Handling Data Updates	100
5.5.1	Basic Update Scheme	102
5.5.2	Lazy Update in HS ³ -G-tree	103
5.5.3	Loosely-Bounded HS ³ -R-tree	104
5.6	Performance Evaluation	105
5.6.1	Authenticated Data Structure Construction	107
5.6.2	Query Authentication Performance	108
5.6.3	Data Update Performance	110
5.7	Chapter Summary	113

Chapter 6	Conclusions	114
6.1	Contributions	114
6.2	Possibilities for Future Work	115
Appendix A	Discussion on Ranking Function	117
Appendix B	Cost Analysis of Private Ranking Comparison	121
Appendix C	Extension of PLB Method to 3D Space	123
Appendix D	Encryption Enhancement of Area A_2 Computation in PLB Method	124
Bibliography		126
Curriculum Vitae		133

List of Figures

1.1	Authenticatable Location-Based Service	3
1.2	Multi-Source Data Integration and Query Processing	6
2.1	Basic Authentication Tools	11
3.1	Verification Object for 1D Range Query	19
3.2	Query Authentication on R-tree Index	21
3.3	Query Authentication on Grid-File Index	23
3.4	Accumulative Digest	28
3.5	Linear Ordering	38
3.6	Linear Embedding	39
3.7	$g()$ v.s. $g_t()$	41
3.8	Basic Query Authentication Performance	41
3.9	Optimized Query Authentication Performance	45
3.10	Data Owner Update Cost	46
4.1	Top-k Query Example	50
4.2	Ranking Comparison in Different Dimensions	55
4.3	Pre-signed Line Based Method	56
4.4	Nodes, Objects, and Query	58
4.5	Query Authentication on MR-tree	59
4.6	Ranking Comparison between an MBB and an Object	63
4.7	Power Diagram and Query	63
4.8	VO and Authentication on Power Diagram	66
4.9	Privacy Circles before and after Query q	71
4.10	Online Strategy on Pre-signed Lines. (a) the pre-signed lines in the top-3 query running example. (b) the spanning trees and comparison chains for both strategies.	73
4.11	Basic Query Authentication Performance	75

4.13	Impact of Scores	76
4.12	Performance with Pre-signed Lines ($k = 1$ or 128)	77
5.1	Query Authentication Example in Data Integration	81
5.2	Prefix Tree Index	84
5.3	Content of Seal S_i	86
5.4	HS ³ -G-tree	92
5.5	HS ³ -R-tree	96
5.6	Complex Queries	97
5.7	Lazy Update	101
5.8	Server Construction Cost vs. DO Population	107
5.9	Basic Query Authentication Performance	108
5.10	Performance of Varying Query Ranges	109
5.11	Performance vs. Data Population	109
5.12	Performance vs. Data Dimensionality	110
5.13	Index Update Performance	111
5.14	Impact of Query Workload on Server	111
5.15	Impact of Hashtable Size on Server	112
5.16	Impact of Bounding-Square Size	113
5.17	Performance under Mixed Workload	113
1	Top- k in Sum Ranking vs. Top- k in Euclidean Ranking	119
2	Cumulative Probability Distribution of k' (as multiple of k , $k=16$)	119
3	Presigned Line in 3D Space	123