Algorithms for Data and Computation Privacy

Alex X. Liu • Rui Li

Algorithms for Data and Computation Privacy

Springer

Alex X. Liu
Chief Scientist
Ant Group
Hangzhou, Zhejiang, China

Rui Li
School of Cyberspace Security
Dongguan University of Technology
Dongguan, Guangdong, China

ISBN 978-3-030-58895-3 ISBN 978-3-030-58896-0 (eBook)
https://doi.org/10.1007/978-3-030-58896-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Dedicated with love and respect
to my parents
Yuhai Liu (God rest his soul) and Shuxiang Wang,
to my wife
Chenyang Li,
to my twin sons
Max Boyang and Louis Boyang,
to whom I owe
all that I am and all that I have accomplished.

– Alex X. Liu

To my dearest wife Jian Zhou, my son Zixuan Li, and my parents Jixiang Li and Yuqiong Li. Thanks for your support and understanding.

–Rui Li

Contents

Part I Privacy Preserving Queries

1 Range Queries over Encrypted Data 3
1.1 Introduction.. 3
1.1.1 Background and Motivation............................... 3
1.1.2 Threat Model.. 5
1.1.3 Security Model... 5
1.1.4 Summary and Limitation of Prior Art 6
1.1.5 Proposed Approach....................................... 7
1.1.6 Technical Challenges and Solutions 7
1.1.7 Key Contributions 8
1.2 Related Work 8
1.3 PBtree Construction... 8
1.3.1 Prefix Encoding.. 9
1.3.2 Tree Construction.. 9
1.3.3 Node Randomization Using Bloom Filters............... 12
1.3.4 Trapdoor Computation 13
1.3.5 Query Processing 14
1.3.6 False Positive Analysis 14
1.4 PBtree Search Optimization 15
1.4.1 Traversal Width Optimization............................ 16
1.4.2 Traversal Depth Optimization............................ 19
1.5 PBtree Update 21
1.5.1 PBtree Insertion Algorithm 21
1.5.2 PBtree Modification Algorithm.......................... 22
1.5.3 PBtree Deletion Algorithm............................... 23
1.6 Security Analysis.. 24
1.6.1 Security Model... 24
1.6.2 Security Proof.. 25
1.7 Experimental Evaluation.. 27
1.7.1 Experimental Methodology 27

1.7.2 Evaluation of PBtree Construction 29
1.7.3 Query Evaluation Performance 30
1.7.4 Experimental Results on Updating 32
1.8 Conclusions 34
References 34

2 Fast and Scalable Range and Keyword Query Processing Over Encrypted Data with Provable Adaptive Security 37
2.1 Introduction 37
2.1.1 Motivation and Problem Statement 37
2.1.2 Threat Model 38
2.1.3 Security Model 38
2.1.4 Limitation of Prior Art 39
2.1.5 Proposed Approach 40
2.1.6 Novelty and Advantages Over Prior Art 41
2.2 Related Work 41
2.3 Basic IBtree Algorithms 42
2.3.1 Index Element Encoding 42
2.3.2 IBF Construction 43
2.3.3 IBtree Construction 44
2.3.4 Trapdoor Computation 45
2.3.5 Query Processing 47
2.4 Optimized IBtree Algorithms 48
2.4.1 IBtree Traversal Width Minimization 48
2.4.2 IBtree Traversal Depth Minimization 51
2.4.3 IBtree Compression 53
2.5 Security Analysis 56
2.6 Experimental Evaluation 58
2.6.1 Experimental Methodology 58
2.6.2 Index Size 60
2.6.3 Index Construction Time 61
2.6.4 Query Processing Time 62
2.6.5 Compared with PBtree and KRB 63
2.7 Conclusions 65
References 65

3 Nearest Neighbor Queries over Encrypted Data 69
3.1 Introduction 69
3.2 Insecurity of ASPE 70
3.2.1 ASPE I and II 71
3.2.2 Attack Method 71
3.2.3 Experimental Results 73
3.3 Hardness Analysis 75
3.4 Conclusions 76
References 77

4 K-Nearest Neighbor Queries Over Encrypted Data ... 79
4.1 Introduction ... 79
4.1.1 Motivations ... 79
4.1.2 Problem Formulation ... 80
4.1.3 Service Model and Design Goals ... 80
4.1.4 Comparison with Prior Arts ... 82
4.1.5 Technical Challenges and Proposed Solutions ... 83
4.1.6 SecEQP Scheme Overview ... 84
4.1.7 Main Contributions ... 84
4.2 Space Encoding ... 85
4.2.1 Projection Function Introduction ... 85
4.2.2 Space Encoding via a Single Primitive Projection Function ... 86
4.2.3 Projection Function Composition Introduction ... 87
4.2.4 Space Encoding via Projection Function Composition ... 88
4.3 kNN Protocol for Plaintext Domain ... 90
4.3.1 kNN Protocol Design ... 90
4.3.2 Analysis of kNN Protocol Parameters ... 92
4.4 Transforming kNN to Secure kNN ... 95
4.4.1 Prefix-Free Encoding ... 95
4.4.2 Operation Transformation ... 96
4.4.3 Indistinguishable Bloom Filter Tree Based Secure Index 96
4.4.4 SkNN Protocol (SecEQP) Design ... 98
4.4.5 Security Analysis ... 99
4.5 Performance Evaluation ... 101
4.5.1 Parameters Settings ... 101
4.5.2 Datasets, Metrics, and Implementation ... 101
4.5.3 Experiment Results ... 102
4.5.4 Improve Result Accuracy ... 104
4.6 Related Work ... 107
4.7 Conclusions ... 107
References ... 107

5 Top-k Queries for Two-Tiered Sensor Networks ... 109
5.1 Introduction ... 109
5.1.1 Motivation ... 109
5.1.2 Problem Statement ... 110
5.1.3 Adversary and Security Model ... 111
5.1.4 Limitations of Prior Art ... 111
5.1.5 Technical Challenges and Proposed Approach ... 112
5.1.6 Key Contributions ... 113
5.2 Related Work ... 113
5.3 System Model and Assumptions ... 114
5.4 Sensor Data Pre-Processing: Mapping and Partitioning ... 115
5.4.1 Approximating Uniform Distribution ... 115

5.4.2 Data Partitioning for Integrity Verification 117
5.4.3 Embedding Intervals with Data 118
5.4.4 Index Selection 119
5.5 Privacy Preserving Index Generation 119
5.5.1 Prefix Encoding and Bloom Filter Indexing 119
5.5.2 Randomizing Bloom Filter Indexes 120
5.6 Trapdoor Computation and Query Processing 122
5.6.1 Top-*k* to Top-Range Query 122
5.6.2 Trapdoor Computation 124
5.6.3 Query Execution 124
5.6.4 Integrity Verification for Query Results 125
5.6.5 False Positive Rate Analysis 125
5.7 Security Analysis 126
5.8 Performance Evaluation 129
5.8.1 Experimental Setup 129
5.8.2 Summary for Experimental Results 130
5.8.3 Comparison with Prior Art 131
5.9 Conclusions 133
References 133

Part II Privacy Preserving Computation

6 Collaborative Enforcement of Firewall Policies in Virtual Private Networks 139
6.1 Introduction 139
6.1.1 Background and Motivation 139
6.1.2 Technical Challenges 140
6.1.3 Limitations of Prior Art 141
6.1.4 Our Solution 141
6.1.5 Key Contributions 142
6.2 Threat Model 142
6.3 Background 142
6.4 Oblivious Comparison 143
6.5 Bootstrapping Protocol 146
6.5.1 FDD Construction 146
6.5.2 Range Conversion 148
6.5.3 Prefix Numericalization 148
6.5.4 Applying XOR by MSU 149
6.5.5 Applying XOR and HMAC by IBM 149
6.6 Filtering Protocol 150
6.6.1 Address Translation 151
6.6.2 Prefix Membership Verification 151
6.6.3 Packet Preprocessing by IBM 152
6.6.4 Packet Preprocessing by The Third Party 153
6.6.5 Packet Processing by MSU 153

6.7 VGuard for Deep Packet Inspection 154
6.7.1 The Bootstrapping Protocol 154
6.7.2 The Filtering Protocol 155
6.8 Discussion 157
6.8.1 Firewall Updates 157
6.8.2 Decision Caching 157
6.8.3 Decision Obfuscation vs. Decision Encryption 158
6.8.4 Special Treatment of IP Addresses 158
6.8.5 Securing Keys of MSU 159
6.8.6 Stateful Firewalls 160
6.8.7 Statistical Analysis Attack and Countermeasures 161
6.8.8 Hash Collision 161
6.9 Related Work 162
6.9.1 Secure Function Evaluation 162
6.9.2 CDCF Framework 163
6.9.3 Secure Queries 164
6.10 Experimental Results 165
6.10.1 Efficiency on Real-Life Firewall Policies 165
6.10.2 Efficiency on Synthetic Firewall Policies 167
6.11 Concluding Remarks 169
References 169

7 Privacy Preserving Quantification of Cross-Domain Network Reachability 171
7.1 Introduction 171
7.1.1 Background and Motivation 171
7.1.2 Limitation of Prior Art 172
7.1.3 Cross-Domain Quantification of Reachability 173
7.1.4 Technical Challenges 174
7.1.5 Our Approach 175
7.1.6 Summary of Experimental Results 176
7.1.7 Key Contributions 176
7.2 Related Work 176
7.2.1 Network Reachability 176
7.2.2 Privacy Preserving Set Operation 178
7.2.3 Privacy Preserving Collaborative Firewall Enforcement in VPN 178
7.3 Problem Statement and Threat Model 179
7.3.1 Access Control Lists (ACLs) 179
7.3.2 Problem Statement 179
7.3.3 Threat Model 180
7.4 Privacy-Preserving Quantification of Network Reachability 181
7.4.1 Privacy-Preserving Range Intersection 181
7.4.2 ACL Preprocessing 183
7.4.3 ACL Encoding and Encryption 185

7.4.4 ACL Comparison 187
7.5 Incremental Updates of ACLs 189
7.5.1 Addition of Rules with Accept Decision 190
7.5.2 Addition of Rules with Discard Decision 190
7.5.3 Addition of New Routers 191
7.6 Stateful Firewalls 191
7.7 Security and Complexity Analysis 192
7.7.1 Security Analysis 192
7.7.2 Complexity Analysis 193
7.8 Protocol Optimization 194
7.9 Experimental Results 194
7.9.1 Efficiency on Real ACLs 195
7.9.2 Efficiency on Synthetic ACLs 196
7.9.3 Efficiency of Incremental Updates of ACLs 198
7.10 Conclusions 199
References 200

8 Cross-Domain Privacy-Preserving Cooperative Firewall Optimization 203
8.1 Introduction 203
8.1.1 Background and Motivation 203
8.1.2 Limitation of Prior Work 204
8.1.3 Cross-Domain Inter-Firewall Optimization 204
8.1.4 Technical Challenges and Our Approach 205
8.1.5 Key Contributions 206
8.2 Related Work 207
8.2.1 Firewall Redundancy Removal 207
8.2.2 Collaborative Firewall Enforcement in VPN 207
8.3 System and Threat Models 207
8.3.1 System Model 207
8.3.2 Threat Model 208
8.4 Privacy-Preserving Inter-Firewall Redundancy Removal 209
8.4.1 Privacy-Preserving Range Comparison 209
8.4.2 Processing Firewall FW_1 210
8.4.3 Processing Firewall FW_2 213
8.4.4 Single-Rule Coverage Redundancy Detection 215
8.4.5 Multi-Rule Coverage Redundancy Detection 216
8.4.6 Identification and Removal of Redundant Rules 219
8.5 Firewall Update After Optimization 220
8.6 Security and Complexity Analysis 221
8.6.1 Security Analysis 221
8.6.2 Complexity Analysis 222
8.7 Experimental Results 223
8.7.1 Evaluation Setup 223
8.7.2 Methodology 223

8.7.3 Effectiveness and Efficiency on Real Policies 224
8.7.4 Efficiency on Synthetic Policies 226
8.8 Conclusions and Future Work 228
References 229

9 Privacy Preserving String Matching for Cloud Computing 231
9.1 Introduction 231
9.1.1 Motivation 231
9.1.2 Problem Statement 232
9.1.3 Adversary and Security Model 232
9.1.4 Limitation of Prior Art 233
9.1.5 Proposed Approach 233
9.1.6 Technical Challenges and Solutions 234
9.1.7 Key Contributions 234
9.2 Related Work 235
9.3 Pattern Aware Secure Search Tree 235
9.3.1 String Pattern Matching 236
9.3.2 PASStree Structure 236
9.3.3 Preserving Privacy of Bloom Filters 238
9.3.4 Query Trapdoor Generation and Processing 239
9.4 PASStree+ 240
9.4.1 Challenge in Search Optimization 240
9.4.2 Optimizing PASStree 240
9.5 Ranking Search Results 242
9.5.1 Recording Matching Positions 242
9.5.2 Ranking Algorithm 243
9.6 Security Analysis 243
9.6.1 Security Model 243
9.6.2 Security Proof 244
9.7 Performance Evaluation 246
9.7.1 Experimental Methodology 246
9.7.2 PASStree Construction and Size 248
9.7.3 Query Processing Speed and Accuracy 249
9.7.4 Ranking Precision 249
9.8 Conclusion and Future Work 250
References 250

10 Privacy Preserving Information Hub Identification in Social Networks 253
10.1 Introduction 253
10.1.1 Background and Motivation 253
10.1.2 Limitations of Prior Art 254
10.1.3 Proposed Solution 254
10.1.4 Results and Findings 256
10.1.5 Key Contributions 256
10.2 Related Work 257

10.3 Proposed Solution 258
10.3.1 Eigenvector Centrality 258
10.3.2 Motivation for Principal Component Centrality 259
10.3.3 Definition of PCC 259
10.3.4 Generalized PCC 260
10.3.5 Selection of Number of Eigenvectors 263
10.3.6 Decentralized Eigendecomposition Algorithm 263
10.4 Performance Evaluation 265
10.4.1 Data Sets 265
10.4.2 Selection of PCC Parameter 268
10.4.3 Comparison with Ground Truth 268
10.5 Conclusions 274
References 274

Part III Differential Privacy

11 Publishing Social Network Data with Privacy Guarantees 279
11.1 Introduction 279
11.1.1 Background and Motivation 279
11.1.2 Problem Statement 280
11.1.3 Limitations of Prior Art 280
11.1.4 Proposed Approach 281
11.1.5 Technical Challenges 281
11.1.6 Key Contributions 282
11.2 Related Work 283
11.2.1 Differential Privacy 283
11.2.2 Differential Privacy in Data Publishing 283
11.3 Random Matrix Approach 284
11.3.1 Theoretical Guarantee on Differential Privacy 286
11.3.2 Theoretical Guarantee on Eigenvector Approximation ... 289
11.4 Experimental Results 292
11.4.1 Dataset 293
11.4.2 Node Clustering 293
11.4.3 Node Ranking 300
11.5 Utility Comparison 303
11.6 Conclusions 308
References 308

12 Predictable Privacy-Preserving Mobile Crowd Sensing 313
12.1 Introduction 313
12.2 Related Work 314
12.3 Privacy of MCS 315
12.3.1 Threat Model 315
12.3.2 Data Reconstruction Attack 316
12.4 Differentially Private Mechanisms for Privacy-Preserving MCS .. 316
12.4.1 Models and Definitions 316

12.4.2 The Basic Laplacian Mechanism 318
12.4.3 The Salus Algorithm 320
12.5 Role User: The Privacy Quantification 322
12.5.1 Data Reconstruction Error: A Quantitative Analysis 322
12.5.2 Data Reconstruction Error: A Lower Bound 325
12.6 Role Application Publisher: The Utility Prediction 328
12.6.1 Average (AVG) 328
12.6.2 Histogram (HIST) 331
12.6.3 Classifiers (CLS) 333
12.7 The P^3 Framework for Predictable Privacy-Preserving MCS 337
12.8 Performance Evaluation 338
12.8.1 Privacy Protection 339
12.8.2 System Overhead 339
12.8.3 Case Studies 340
12.9 Conclusions 344
References 344

13 Differentially Private and Budget Limited Bandit Learning over Matroids 347
13.1 Introduction 347
13.1.1 Limitations of Prior Art 349
13.1.2 Proposed Approach 349
13.1.3 Advantages over Prior Art 350
13.2 Related Work 351
13.2.1 MAB Algorithms Without Budgets 351
13.2.2 MAB Algorithms with Budgets 352
13.2.3 Privacy-Aware Online Learning 353
13.3 Problem Statement 353
13.3.1 Matroids 354
13.3.2 DP-Aware BMAB Over Matroids 354
13.3.3 An Example in Crowdsourcing 355
13.4 Bounding the Optimal Policy 356
13.5 Algorithm Design 357
13.5.1 Ensuring Differential Privacy 358
13.5.2 The OPBM Algorithm 358
13.6 Regret Analysis 364
13.7 Performance Evaluation 367
13.7.1 Experimental Setup 367
13.7.2 Metrics 368
13.7.3 Regret Performance 368
13.7.4 Time Efficiency 370
13.8 Conclusion 371
References 381

Part IV Breaking Privacy

14 Breaching Privacy in Encrypted Instant Messaging Networks 385
14.1 Introduction 385
14.1.1 Chapter Organization 387
14.2 Related Work 387
14.2.1 Mix Network De-anonymization 387
14.2.2 Social Network De-anonymization 388
14.3 Problem Description and Attack Scenarios 388
14.3.1 IM Service Architecture 388
14.3.2 Attack Scenarios 390
14.4 COLD: COmmunication Link De-anonymization 390
14.4.1 Architecture 391
14.4.2 Details 392
14.4.3 Example 394
14.5 Experimental Results 395
14.5.1 Data Set 396
14.5.2 Evaluation Metrics 398
14.5.3 Results 398
14.5.4 Discussions 400
14.6 Evasion and Countermeasures 402
14.7 Conclusions 403
References 403

Acronyms

ACL	Access control list
AM	Anonymization module
ASPE	Asymmetric scalar-product-preserving encryption
BDD	Binary decision diagram
BMAB	Budget-limited multi-armed bandit
CDCF	Cross-domain cooperative firewall
COLD	COmmunication Link De-anonymization
CTL	Computation tree logic
DP	Differential privacy
EVC	Eigenvector centrality
FDD	Firewall decision diagram
GSR	Galvanic skin response
IBF	Indistinguishable Bloom filter
IBtree	Indistinguishable binary tree
ICA	Independent component analysis
IDD	Interval decision diagram
IM	Instant messaging
IND-CKA	Indistinguishability against chosen keyword attack
IND-CPA	Indistinguishability against chosen plain-text attack
IND-OCPA	Indistinguishable under ordered chosen plain-text attack
IPSes	Intrusion prevention systems
ISP	Internet service provider
kNN	k-nearest neighbor
LSH	Locality sensitive hashing
MAB	Multi-armed bandit
MCS	Mobile crowd sensing
MSE	Average mean squared error
MSU	Michigan State University
NAT	Network address translation
OAR	Overall approximation ratio
OSN	Online social network

PASStree	Pattern aware secure search tree
PBtree	Privacy Bloom filter tree
PCA	Principal component analysis
PCC	Principal component centrality
PPT	Probabilistic polynomial time
QoS	Quality of service
SANE	Secure architecture for the networked enterprise
SecEQP	Secure and efficient query processing
SNN	Secure nearest neighbor
SP	Sub-string prefix
SRA	Secure RPC authentication
SSE	Searchable symmetric encryption
TF-IDF	Term-frequency inverse-document-frequency
VPN	Virtual private network