

Enabling Traceability in Large-Scale RFID Networks

A dissertation submitted in fulfillment
of the requirements for the degree of

Doctor of Philosophy
in
Computer Science

Yanbo Wu

Supervisor: Prof. Hong Shen, Dr. Quanzheng Sheng

School of Computer Science
The University of Adelaide

December, 2011

TABLE OF CONTENTS

1	Introduction	2
1.1	RFID Enabled Traceability	4
1.2	Supply Chain Management System: an Example	6
1.3	Research Issues	8
1.4	Contribution Overview	10
1.4.1	Peer-to-Peer Traceability Model and Architecture	10
1.4.2	Traceability Mining over Distributed RFID Streams	11
1.4.3	Traceability as a Service	12
1.4.4	Implementation and Performance Study	13
1.5	Dissertation Organization	13
2	Background	16
2.1	RFID Traceable Networks: Preliminaries	17
2.1.1	RFID Systems	17
2.1.2	Traceable RFID Networks	21
2.2	Traceability in RFID Networks: Applications, Queries and Requirements	26
2.2.1	RFID Enabled Traceability Applications	26
2.2.1.1	Eliminating Inventory Inaccuracies	28
2.2.1.2	Inventory Shrinkage	29
2.2.1.3	Eliminating Wastage and Damage	29

2.2.1.4	Fine Grain Product Recalls	30
2.2.1.5	Anti-counterfeiting	30
2.2.2	Traceability Queries	31
2.2.3	Requirements of Traceability Applications	35
2.2.3.1	System Development Requirements	35
2.2.3.2	Data Model Requirements	39
2.3	Overview of Traceability Models for RFID Data	42
2.3.1	DRER Model	42
2.3.2	RFID Cuboid	44
2.3.3	KAIST Trace Model	46
2.3.4	SPIRE Model	47
2.3.5	Comparison and Open Issues	49
2.4	Overview of Architectures for Traceable RFID Networks	50
2.4.1	EPCglobal Architecture Framework	52
2.4.2	BRIDGE	55
2.4.3	IBM Theseos	58
2.4.4	DIALOG	59
2.4.5	Hierarchical P2P-based RFID Code Resolution Network	60
2.4.6	Comparison and Open Issues	62
2.5	Summary	66
3	MOODS	68
3.1	The MOODS Model	70

3.1.1	The Key Traceability Functions	70
3.1.2	The Design of the MOODS Model	71
3.2	A P2P Traceable RFID Network Architecture	74
3.3	An Enhanced Model Maintenance Algorithm	77
3.3.1	The Overview of the Design	77
3.3.2	Determining the Width of the Sliding Window	79
3.3.3	The Key Factor : The Length of Prefixes	81
3.3.4	Prefix Triangle	83
3.3.5	The Group-based Indexing Algorithm on Prefix Triangle	87
3.3.6	Algorithm Analysis	89
3.4	Traceability Query Processing Algorithms	90
3.4.1	Item Level Queries	90
3.4.2	Statistical Queries	91
3.4.3	Algorithm Analysis	92
3.4.3.1	Item Level Queries	92
3.4.3.2	Statistical Queries	93
3.5	Replication and Fault Recovery	93
3.5.0.3	Hardware Faults	93
3.5.0.4	Network Faults	94
3.6	Related Work	96
3.7	Summary	99
4	Mining Moving Patterns	102

4.1	Problem Definition	104
4.2	The Architecture for Distributed Stream Mining	106
4.3	The TISH Model	110
4.3.1	The Overview of TISH Design	110
4.3.2	Algorithm : RFID Stream Sampling	114
4.3.3	Algorithm : Update for the Current Slot	116
4.3.4	Algorithm : Merging with the Next Slot	117
4.4	Building TISH in a P2P Fashion	117
4.4.1	Tracing and Tracking Objects	118
4.4.2	Building the Flow Synopsis	120
4.4.3	The Business Neighbor Tree	124
4.4.4	Determining w_s and w_e	126
4.5	Performance Analysis	128
4.5.1	Model Maintenance Cost	128
4.5.2	Performance of Building Flow Synopsis	131
4.6	Related Works	132
4.6.1	Data Structures and Data Transformation	133
4.6.2	Knowledge Discovery	134
4.6.3	Distributed Modeling and Query Processing	135
4.7	Summary	137
5	Traceability as a Service	138
5.1	Motivations and Challenges	142

5.2	The Architecture of PeerTrack Cloud	147
5.2.1	Modules of the Architecture	147
5.2.2	PT-T2S Data Model	150
5.2.3	PT-S2 Data Model	153
5.3	Data Partition and Replication	156
5.3.1	Point-Based Data Partition and Replication	158
5.3.2	Path-Based and Graph-based Data Partition and Replication	159
5.4	Performance Analysis and Comparison	162
5.5	Related Work	164
5.6	Summary	166
6	Implementation and Performance Study	168
6.1	PeerTrack Platform: An Overview	169
6.2	Implementation Details	173
6.2.1	Rule Engine	174
6.2.2	Tracking Engine	176
6.3	PeerTrack AMS: A Demonstration	178
6.4	Performance Study	181
6.4.1	Performance Study on the P2P Architecture and MOODS	181
6.4.1.1	Performance Study on Scalability	181
6.4.1.2	Performance Study on Bandwidth Efficiency	184
6.4.1.3	Performance Study on Load Balancing	186
6.4.2	Performance Study on the TISH Model	188

6.4.2.1	Accuracy of TISH	190
6.4.2.2	The Cost of Model Maintenance	193
6.4.3	Performance Study on the PeerTrack Cloud	195
6.5	Summary	197
7	Conclusions	200
7.1	Summary	200
7.2	Future Directions	203
A	Curriculum Vitae	206
	Bibliography	216

LIST OF FIGURES

1.1	Supply Chain Management Scenarios	7
2.1	Overview of an RFID System	17
2.2	Reference Model of Traceable RFID Networks	22
2.3	RFID Data Model Overview : DRER Model	43
2.4	RFID Data Model Overview : RFID Cuboid	44
2.5	RFID Data Model Overview : Gateway-based RFID Cuboid	45
2.6	RFID Data Model Overview : KAIST Trace Model	46
2.7	RFID Data Model Overview : SPIRE Model	48
2.8	EPCglobal Architecture Framework	53
3.1	MOODS Model Design Overview	72
3.2	P2P Traceable RFID Network Workflow	75
3.3	Group-based P2P Traceable RFID Network Workflow	78
3.4	Prefix Triangle Example	84
3.5	Algorithm for Indexing a Group of Objects	88
3.6	An Example of Missing Readings	94
3.7	Replication of Indices	95
3.8	Replication of MOODS	95
4.1	The Architecture for Distributed Stream Mining	107
4.2	The Logarithmic Tilted Time Frame	111
4.3	An Example of Tilted Time Frame Series of Histograms	112

4.4	The Structure of a Slot in TISH	113
4.5	Algorithm to Update the LTF Model	117
4.6	Algorithm to Merge the LTF Model	118
4.7	Algorithm to Trace an Object	119
4.8	Algorithm to Build Flow Synopsis	123
4.9	An Example of Sideway Problem	124
4.10	An Example of Business Neighbor Tree	125
4.11	An Example of Overlapped Sliding Windows	127
4.12	Example of Performance in Modeling Accuracy	129
4.13	Examples of Tracing Efficiency	132
5.1	The Waste of Resources in Existing Deployment Scheme	143
5.2	Cloud Computing Architecture	144
5.3	PeerTrack Cloud Architecture	147
5.4	Algorithm to Update Index in PT-T2S	151
5.5	Algorithm to Update Index in PT-T2S	153
5.6	PT-S2 Graph Model	154
5.7	Example of PT-S2 Graph Model	155
5.8	Example of Partitioning Point-based Data	158
5.9	Example of Partitioning Path-based Data	160
5.10	Algorithm to Replicate the Graph-based Data	161
6.1	The Architecture of PeerTrack Platform	170
6.2	Screenshot of the Rule Editor	175

6.3	Screenshot of the PeerTrack AMS Client	178
6.4	Scalability on Network Size	183
6.5	Scalability on Data Volume	184
6.6	Bandwidth Cost in Different Scenarios	185
6.7	Load Balancing with Different Schemes	187
6.8	Default Settings of Experiments for TISH	189
6.9	Patterns in Experiments for TISH	190
6.10	Accuracy of the Model for Different Patterns	191
6.11	Accuracy of the Model with Mixed and Random Pattern	192
6.12	Number of Network Calls vs. Time	194
6.13	Distribution of Number of Network Calls for Model Maintenance .	194
6.14	Query Processing Performance of PeerTrack Cloud	196

LIST OF TABLES

2.1	Notations in the Reference Model of Traceable Networks	25
2.2	Comparison: Data Models vs. Data Model Requirements	51
2.3	Comparison: Data Models vs. Supporting Traceability Queries	51
2.4	Comparison: System Architectures vs. System Development Re- quirements	65
4.1	Symbols in The Overview of TISH	112
5.1	Comparison of Performance in Different Architectures	164
6.1	Enabling Technologies in PeerTrack Platform	173
6.2	Testing Queries for Scalability of PeerTack and MOODS	182

ABSTRACT OF THE DISSERTATION

Enabling Traceability in Large-Scale RFID Networks

by

Yanbo Wu

Doctor of Philosophy in Computer Science

The University of Adelaide, 2011

The emergence of radio frequency identification (RFID) technology brings significant social and economic benefits. As a non line of sight technology, RFID provides an effective way to record movements of objects within a networked system formed by a set of distributed and collaborating parties. A trail of such recorded movements is the foundation for enabling traceability applications. While traceability is a critical aspect of the majority of RFID applications, realizing traceability for these applications brings many fundamental research and development issues, including storage efficiency, query processing complexity, privacy etc.

In this dissertation, we present a novel approach to realize RFID-based traceability in large, autonomous and heterogeneous distributed networks. We first propose a Peer-to-Peer (P2P) architecture, namely PeerTrack. PeerTrack does not require any kind of centralized database for the RFID data or their index, neither it requires RFID data to be fully shared to partners. In PeerTrack, only a specific portion of data is requested by partners, when the access is necessary.

We introduce a distributed model, namely MOODS (a **M**odel for **m**Oving **O**bjects in **D**iscrete **S**pace), for the essential data structures of traceability.

MOODS is maintained by a distributed index on the top of a structured Peer-to-Peer overlay. We then propose efficient algorithms for the maintenance of MOODS. The algorithms are optimized to consume statistically minimal cost of bandwidth. Based on this model, we propose algorithms for efficient item-level and statistical traceability query processing.

We also propose a traceability mining model for distributed RFID streams, namely TISH (**T**ilted **T**ime **F**rame of **H**istogram). TISH takes advantages of two important data mining tools, namely *Tilted Time Series* and *Histogram*, and combines them to describe the patterns of RFID streams in the dimensions of both time and space, and capture the dynamicity of the patterns. We propose efficient algorithms to maintain TISH and algorithms that use it for traceability query processing and RFID stream mining.

We present a platform, namely PeerTrack Cloud, to bring the aforementioned RFID data modeling and traceability query processing techniques to the Cloud Environments. The platform features specific traceability-oriented modules for real-time query processing and efficient data storage.

The techniques proposed in this dissertation are implemented in “Asset Management System”, which is a collaborative project with a local company. Finally, we conduct extensive performance studies of the proposed techniques. The experimental results reveal that our system i) is more scalable and outperforms the centralized approach when the data volume or the network becomes larger; ii) provides powerful programming interfaces for query processing; iii) is economy in both storage and bandwidth; and iv) can be easily adopted in cloud computing platforms.

ORIGINALITY STATEMENT

“Yanbo Wu certifies that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the Universitys digital research repository, the Library catalogue, also through web search engines, unless permission has been granted by the University to restrict access for a period of time.”

Yanbo Wu

December 1st, 2011

To my mother and father,
who made all of this possible,
for their endless encouragement and support.

ACKNOWLEDGMENTS

It has been a great pleasure working with the faculty, staff, students at the University of Adelaide, during my tenure as a doctoral student, and I would like to thank them all for such a great graduate school experience. My foremost thank goes to my thesis supervisor Dr. Quanzheng Sheng, a talented teacher and passionate scientist. Dr. Sheng instilled a thirst for excellence in me, taught me how to do high-quality research, and helped me think independently and creatively. He not only guided my research, but also served as a mentor and role model as I embarked on my academic career. I will forever cherish his full support during my study. I also would like to thank Prof. Hong Shen, who gave me many valuable suggestions to my research.

I thank my co-authors: Quanzheng Sheng, Hong Shen, Sherali Zeadally, Jian Yu, Damith Ranasinghe and Jun Han, for their productive and enjoyable collaborations. I would like also to thank anonymous reviewers for their valuable comments on earlier drafts of my papers.

I would like to thank my mother and father for their constant support and encouragement. They have been always being there whenever I needed them. I also would like to thank my dearest friend, Chenke Yang, for not only giving me valuable suggestions to my papers, but also supporting me spiritually.

Finally, I express my sincere appreciation to the University of Adelaide, who provided the Adelaide University Fee Scholarship (AFSI) and ARC-grant Funded Scholarship and to Google Inc., who provided the Google PhD Top-up Scholarship, to financially support my work in this dissertation.