Managing Data Dynamics, Streams and Sharing in the Internet of Things

Yongrui (Louie) Qin

School of Computer Science

The University of Adelaide

This dissertation is submitted for the degree of

Doctor of Philosophy

Supervisors: A/Prof. Michael Sheng, Dr. Nickolas J.G. Falkner

and Prof. Hua Wang

August 2015

© Copyright by

Yongrui (Louie) Qin

August 2015

All rights reserved.

No part of the publication may be reproduced in any form by print, photoprint,
microfilm or any other means without written permission from the author.

To my mother and father,

my wife and my little princess,

my brother,

who made all of this possible,

for their endless encouragement and patience.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, 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. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

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 University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Yongrui (Louie) Qin

August 2015

Acknowledgements

This thesis would not have been possible without the support and help from some important people in my life. I would like to take the opportunity to thank all those who have helped me during my PhD journey.

First of all, my sincere thanks will go to my principal supervisor Prof. Michael Sheng, who has taught me how to do good research as well as how to be a person with better personality. He has always been patient, passionate, encouraging throughout the whole journey of my PhD study. His many insightful suggestions and comments on my research have significantly improved the work in this thesis.

Second, I am very thankful to my co-supervisors, Dr. Nickolas J.G. Falkner and Prof. Hua Wang. I want to thank Dr. Falkner for his insightful suggestions on improving my research work and drafts of my papers. I want to thank Prof. Wang for his guidance on my research and for his encouragement to do high-quality research.

I would also like to express my gratitude to Dr. Edward Curry at the Insight Centre for Data Analytics at the National University of Ireland, Galway for providing me the opportunity to work with him and his team as a research intern. My internship experience there has significantly contributed to the work in this thesis.

It has been a great pleasure working with the faculty, staff, students at the University of Adelaide, during my PhD study, and I would like to thank them all for such a great graduate school experience.

I express my sincere appreciation to the University of Adelaide, who provided the Adelaide Scholarship International (ASI), to financially support my work in this dissertation.

No need to mention the great love and support from my family. I am thankful to my parents, who have given up so much and worked so hard to earn money for my education and better future. I am forever indebted to them. I would like to thank my lovely younger brother for his company with me during my childhood and primary and secondary education. I sincerely wish that he will be able to escape from schizophrenia soon and live a better life in future. I am also deeply grateful to my wife and little princess for their constant love and support.

Abstract

Recently, the Internet of Things (IoT) has gained momentum in connecting everyday objects to the Internet and facilitating machine-to-human and machine-to-machine communication with the physical world. IoT offers the capability to connect and integrate both digital and physical entities, enabling a whole new class of applications and services.

This thesis firstly reviews the state-of-the-art research efforts in IoT from data-centric perspectives, including data stream processing, data storage models, complex event processing, and searching in IoT by identifying an IoT data taxonomy, which includes ten key data elements of IoT data under three categorizations. In this thesis, we focus ourselves on three aspects of data management in IoT: data dynamics, data velocity, and data incompleteness. More specifically, we study data dynamics in dynamic graphs, handle data velocity in streams, and tackle data incompleteness via sharing.

In IoT, connections and relations between things are universal and highly dynamic. It is natural to model these connections and relations using dynamic graphs. Meanwhile, shortest path computation is one of the most fundamental operations for managing and analyzing graphs. In this thesis, we focus on the problem of computing the shortest path distance in graphs subject to edge failures. We propose SIEF, a Supplemental Index for Edge Failures in a dynamic graph, which is based on distance labeling, to support distance queries in dynamic graphs with edge failures efficiently.

In IoT, one challenging issue is how to disseminate streaming data to relevant consumers efficiently. Semantic technologies aim to facilitate machine-to-machine (M2M) communication and are attracting more and more interest from both academia and industry, especially in the emerging IoT. This thesis leverages semantic technologies, such as Linked Data, which can facilitate M2M communications to build an efficient information dissemination system for semantic IoT. The system integrates Linked Data streams generated from various data collectors and disseminates matched data to relevant data consumers based on triple pattern queries registered in the system by the consumers. We also design new data structures, *TP-automata* and *CTP-automata*, to meet the high performance needs of Linked Data dissemination.

To tackle data incompleteness, we consider large-scale information sharing scenarios among mobile objects in IoT. By leveraging semantic techniques, we propose broadcasting Linked Data on-air to allow simultaneous access to the information and to achieve better scalability. We introduce a novel air indexing method to reduce the information access latency and energy consumption. We also study the data placement problem of periodic XML data broadcast in IoT environments to facilitate data sharing in IoT. Taking advantage of the structured characteristics of XML data, we present a theoretical analysis on the XML data placement on a wireless channel, which forms the basis of our novel data placement algorithm.

This thesis also discusses on-going and emerging IoT applications, and open research issues for processing and managing IoT data. Several representative domains where IoT can make profound changes are explored, and some key directions for future research and development from a data-centric perspective are identified.

Table of contents

List of figures

List of tables