



UNIVERSITI PUTRA MALAYSIA

***EFFICIENT XML QUERIES AND UPDATES ENCODING SCHEME IN THE
PRESENCE OF ACCESS CONTROL USING FRACTIONAL NUMBERS***

MEGHDAD MIRABI NOOSHABADI

FSKTM 2013 5



**EFFICIENT XML QUERIES AND UPDATES ENCODING SCHEME IN THE
PRESENCE OF ACCESS CONTROL USING FRACTIONAL NUMBERS**

By

MEGHDAD MIRABI NOOSHABADI

**Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia,
in Fulfillment of the Requirements for the Degree of Doctor of Philosophy**

March 2013

DEDICATION

This thesis is dedicated to my lovely wife and my darling parents for their endless support, encouragement, and patience.



Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment
of the requirement for the degree of Doctor of Philosophy

**EFFICIENT XML QUERIES AND UPDATES ENCODING SCHEME IN THE
PRESENCE OF ACCESS CONTROL USING FRACTIONAL NUMBERS**

By

MEGHDAD MIRABI NOOSHABADI

March 2013

Chairman: Hamidah Ibrahim, PhD

Faculty: Computer Science and Information Technology

As XML is emerging as a de facto standard for sharing and exchanging data over the internet, access control for XML data has become an important research topic. Several XML access control mechanisms have been proposed to enforce a fine-grained access control for querying XML data. However, only a few researches are proposed to update the structure of XML data in the presence of a fine-grained access control.

A way to accelerate the process of XML querying is to label the XML nodes in such a way that the structural relationships between two arbitrary XML nodes in the XML tree can be efficiently computed. However, if there is frequent demand for the XML data to be updated, most of the existing XML labeling schemes need to re-label pre-existing XML nodes in order to keep the order of XML nodes in the XML tree, which is rather time consuming.

In order to query and update XML data in the presence of a fine-grained access control, an accessibility map is required to determine the accessibility of XML nodes at runtime. Several researches have been done to compress the accessibility map in such a way that the accessibility of XML nodes can be rapidly determined but there is no effort to compress the accessibility map in dynamic XML environment where the accessibility of XML nodes can be updated frequently.

In this thesis, first we propose a novel XML encoding and labeling scheme based on fractional numbers to encode and label the XML nodes in the XML tree. Our proposed XML encoding and labeling scheme is able to determine the structural relationships between two arbitrary XML nodes in the XML tree and to eliminate the process of re-labeling pre-existing nodes during the process of XML updating. The experimental results demonstrate that our proposed XML encoding and labeling scheme is more efficient than existing XML encoding and labeling schemes for XML querying and updating.

Secondly, we propose a Dynamic Compressed Accessibility Map called DCAM to compress the accessibility map with rapid determination of accessibility of XML nodes at runtime in such a way that it needs minimum maintenance cost to be used in dynamic XML environment. In order to determine the structural relationships between XML nodes in the DCAM, we label the XML nodes in the XML tree as well as the XML nodes in the DCAM with our proposed XML encoding and labeling scheme. The experimental results demonstrate that the DCAM is more efficient than the CAM in compressing the accessibility map.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

SKEMA PENGKODAN PERTANYAAN DAN KEMASKINI XML YANG CEKAP DENGAN KEHADIRAN KAWALAN CAPAIAN MENGGUNAKAN NOMBOR PECAHAN

Oleh

MEGHDAD MIRABI NOOSHABADI

Mac 2013

Pengerusi: Hamidah Ibrahim, PhD

Fakulti: Sains Komputer dan Teknologi Maklumat

Disebabkan XML muncul sebagai suatu piawai *de facto* untuk perkongsian dan penukaran data melalui Internet, kawalan capaian untuk data XML telah menjadi satu topik penyelidikan yang penting. Beberapa mekanisme kawalan capaian XML telah dicadangkan untuk menguatkuasakan kawalan capaian terperinci untuk pertanyaan data XML. Walau bagaimanapun, hanya beberapa penyelidikan yang dicadangkan untuk mengemaskini struktur data XML dalam kehadiran kawalan capaian terperinci.

Satu cara untuk mempercepatkan proses pertanyaan XML adalah dengan melabelkan nod XML supaya hubungan berstruktur antara sebarang dua nod XML dalam pokok XML dapat dikira dengan cekap. Walau bagaimanapun, jika terdapat permintaan yang kerap untuk mengemaskini data XML, kebanyakan skim pelabelan XML sedia ada perlu melabelkan semula nod XML tersedia ada untuk memelihara susunan nod XML dalam pokok XML, di mana agak memakan masa.

Untuk menanya dan mengemaskini data XML dalam kehadiran kawalan capaian terperinci, suatu peta kebolehcapaian adalah diperlukan bagi menentukan kebolehcapaian nod XML pada masa larian. Beberapa penyelidikan telah dilakukan untuk memampatkan peta kebolehcapaian dengan cara kebolehcapaian nod XML dapat ditentukan dengan cepat tetapi tidak ada usaha untuk memampatkan peta kebolehcapaian di dalam persekitaran XML yang dinamik di mana kebolehcapaian nod XML boleh dikemaskini dengan kerap.

Dalam tesis ini, pertama kami mencadangkan satu skim pengkodan dan pelabelan XML yang *novel* berdasarkan nombor pecahan untuk mengkod dan melabelkan nod XML dalam pokok XML. Skim pengkodan dan pelabelan XML yang kami cadangkan mampu untuk menentukan struktur perhubungan antara sebarang dua nod XML dalam pokok XML dan untuk menghapuskan proses pelabelan semula nod sedia ada semasa proses pengemaskinian XML. Keputusan eksperimen menunjukkan bahawa skim pengkodan dan pelabelan XML yang kami cadangkan adalah lebih cekap daripada skim pengkodan dan pelabelan XML sedia ada untuk pertanyaan dan pengemaskinian XML.

Keduanya, kami mencadangkan *Dynamic Compressed Accessibility Map* dipanggil DCAM untuk memampatkan peta kebolehcapaian dengan penentuan cepat kebolehcapaian nod XML pada masa larian supaya ia memerlukan kos penyelenggaraan minimum untuk digunakan dalam persekitaran XML yang dinamik. Untuk menentukan struktur perhubungan antara nod XML dalam DCAM, kami melabelkan nod XML dalam pokok XML dan juga nod XML dalam DACM dengan skim pengkodan dan pelabelan XML yang kami cadangkan. Keputusan eksperimen

menunjukkan bahwa DCAM lebih cepat daripada CAM dalam memampatkan peta kebolehcapaian.



ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest praise to Allah who gave me the strength, faith, confidence, and patience to successfully complete this thesis.

Next, my deepest gratitude and admiration are to Prof. Dr. Hamidah Ibrahim, my thesis supervisor. Without her tremendous help and guidance this thesis would never have been completed. The respect I have for her as a researcher, supervisor, teacher, and person cannot be expressed deeply enough in words.

Next, I would also like to express my sincere thanks and appreciation to the supervisory committee members, Assoc. Prof. Dr. Ali Mamat and Assoc. Prof. Dr. Nur Izura Udzir for their guidance, valuable suggestions and advice throughout this thesis.

Finally, and most importantly, I would like to express my deepest gratitude and admiration to my darling wife, Leila Fathi, and my parents, Ali Mirabi and Zahra Poursina, for their support and encouragement, which enabled me to complete this thesis.

I certify that a Thesis Examination Committee has met on 04 March 2013 to conduct the final examination of Meghdad Mirabi Nooshabadi on his thesis entitled “Efficient XML Queries and Updates Encoding Scheme in the Presence of Access Control Using Fractional Numbers” in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Rodziah Atan, PhD

Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Rusli Abdullah, PhD

Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Mustafa Mat Deris, PhD

Professor
Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia
(Internal Examiner)

David Taniar, PhD

Associate Professor
Clayton School of Information Technology
Monash University
Australia
(External Examiner)

SEOW HENG FONG, PhD

Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 30 April 2013

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of Doctor of Philosophy. The Members of the Supervisory Committee were as follows:

Hamidah Ibrahim, PhD

Professor

Faculty of Computer Science and Information Technology

Universiti Putra Malaysia

(Chairman)

Ali Mamat, PhD

Associate Professor

Faculty of Computer Science and Information Technology

Universiti Putra Malaysia

(Member)

Nur Izura Udzir, PhD

Associate Professor

Faculty of Computer Science and Information Technology

Universiti Putra Malaysia

(Member)

BUJANG BIN KIM HUAT, PhD

Professor and Dean

School of Graduate Studies

Universiti Putra Malaysia

Date:

DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously, and is not concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institution.

MEGHDAD MIRABI NOOSHABADI

Date: 04 March 2013

TABLE OF CONTENTS

	Page
DEDICATION	ii
ABSTRACT	iii
ABSTRAK	v
ACKNOWLEDGEMENTS	viii
APPROVAL	ix
DECLARATION	xi
LIST OF TABLES	xv
LIST OF FIGURES	xvi
LIST OF ABBREVIATIONS	xix
CHAPTER	
1 INTRODUCTION	1
1.1 Research Motivation	1
1.2 Research Problems	5
1.3 Research Objectives	7
1.4 Research Scope	7
1.5 Research Contributions	8
1.6 Thesis Organization	10
2 LITERATURE REVIEW	12
2.1 Existing XML Labeling and Encoding Schemes	12
2.1.1 Classification of XML Labeling Schemes	12
2.1.1.1 Range-Based Labeling Scheme	12
2.1.1.2 Prefix Labeling Scheme	19
2.1.1.3 Multiplicative Labeling Scheme	29
2.1.2 Classification of XML Encoding Schemes	35
2.1.2.1 Binary Number Encoding Scheme	35
2.1.2.2 UTF8 Encoding Scheme	36
2.1.2.3 ORDPATH Encoding Scheme	37
2.1.2.4 Quaternary String Encoding Scheme	39
2.1.2.5 Bit String Encoding Scheme	40
2.1.3 Summary of Existing XML Labeling Schemes	41
2.2 Existing XML Access Control Models and Mechanisms	47
2.2.1 XML Access Control Model	48
2.2.1.1 Subject Specification	48
2.2.1.2 Object Specification	49
2.2.1.3 Access Privilege Specification	50
2.2.1.4 Access Authorization	50
2.2.2 XML Access Control Mechanism	53
2.2.3 Summary of Existing Access Control Models and Mechanisms	60
2.3 Summary	63

3	RESEARCH METHODOLOGY	64
3.1	Research Overview	64
3.2	Research Framework	66
3.3	Experimental Evaluation	72
3.3.1	Performance Evaluation on XML Encoding and Labeling Scheme	72
3.3.2	Performance Evaluation on Compressing the Accessibility Map	74
3.4	Summary	76
4	PROPOSED METHOD	77
4.1	XML Encoding and Labeling Based on Fractional Numbers	77
4.1.1	Fractional Number Based Encoding Scheme	77
4.1.2	Bit String Based Encoding Scheme	82
4.1.3	Applying the Proposed Bit String Based Encoding Scheme to Different Labeling Schemes	87
4.1.4	The Process of XML Querying	91
4.1.5	The Process of XML Updating	93
4.1.5.1	The Process of XML Updating in the BS-Containment Labeling Scheme	93
4.1.5.2	The Process of XML Updating in the BS-Prefix Labeling Scheme	102
4.1.6	The Process of Skewed Node Insertion	110
4.2	Accessibility Map Compression	114
4.2.1	Dynamic Compressed Accessibility Map (DCAM)	114
4.2.2	Efficient Storage and Lookup Methods for the DCAM	116
4.2.3	Accelerating the Process of Access Authorization Checking	120
4.2.4	Maintaining the DCAM in Dynamic Environment	124
4.2.5	Integrating Multiple DCAMs into an Integrated DCAM	128
4.3	Summary	132
5	RESULTS AND DISCUSSION	133
5.1	Performance Study on XML Encoding and Labeling Scheme	133
5.1.1	Experiment on Storage Space	134
5.1.2	Experiment on XML Querying	136
5.1.3	Experiment on XML Updating	138
5.1.4	Experiment on Frequent Node Deletion and Insertion	142
5.1.5	Experiment on Frequent Skewed Node Insertions	144
5.1.6	Discussion	144
5.2	Performance Study on Compressing the Accessibility Map	146
5.2.1	Experiment on Storage Space	146
5.2.1.1	Storage Cost with Low Access Locality	146
5.2.1.2	Storage Cost with High Access Locality	148
5.2.1.3	Effect of Diversity of Access Locality on Storage Cost	150
5.2.1.4	Effect of Number of Subjects on Storage Cost	151
5.2.2	Experiment on Construction Process	151
5.2.3	Experiment on Accessibility Lookup	153
5.2.4	Experiment on Access Authorization Checking Process	154

5.2.4.1	Effect of Different Methods on Access Authorization Checking Time	154
5.2.4.2	Effect of Diversity of Access Locality on Access Authorization Checking Time	156
5.2.5	Experiment on Dynamic XML Updating	157
5.2.5.1	Effect of Node Deletion on the Maintenance Cost	157
5.2.5.2	Effect of Updating the Accessibility of XML Nodes on the Maintenance Cost	158
5.2.6	Discussion	159
5.3	Summary	159
6	CONCLUSION AND FUTURE WORKS	160
6.1	Conclusion	160
6.2	Future Works	162
	REFERENCES	164
	BIODATA OF STUDENT	171
	LIST OF PUBLICATIONS	172