



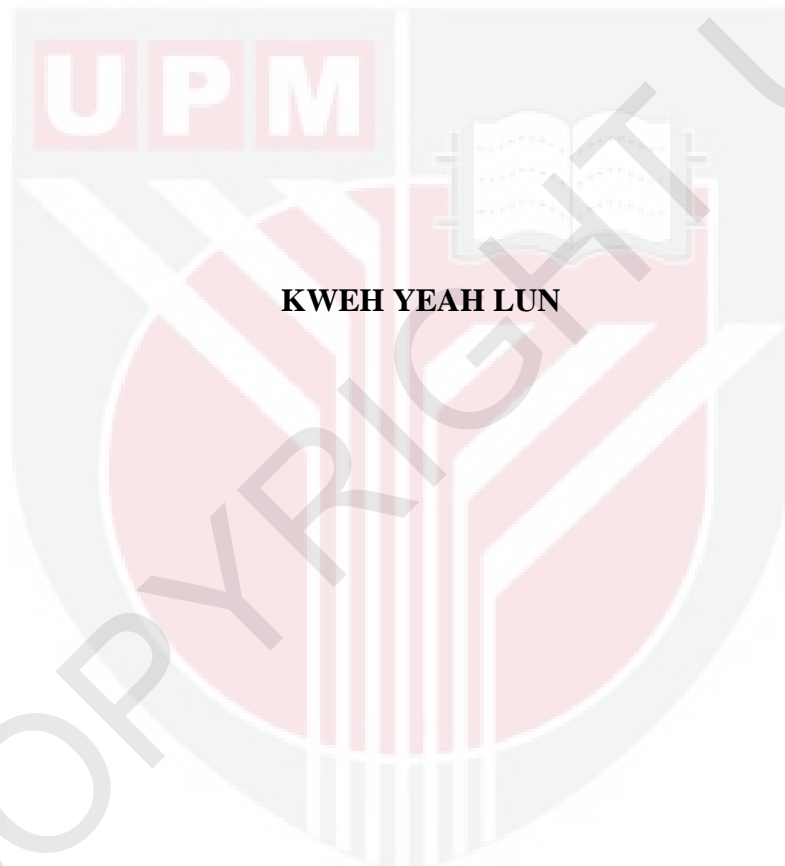
**UNIVERSITI PUTRA MALAYSIA**

**STATIC AND SELF-SCALABLE FILTER RANGE SELECTION  
ALGORITHMS FOR PEER-TO-PEER NETWORKS**

**KWEH YEAH LUN**

**FSKTM 2011 19**

**STATIC AND SELF-SCALABLE FILTER RANGE SELECTION  
ALGORITHMS FOR PEER-TO-PEER NETWORKS**



**KWEH YEAH LUN**

**DOCTOR OF PHILOSOPHY**

**UNIVERSITI PUTRA MALAYSIA**

**2011**

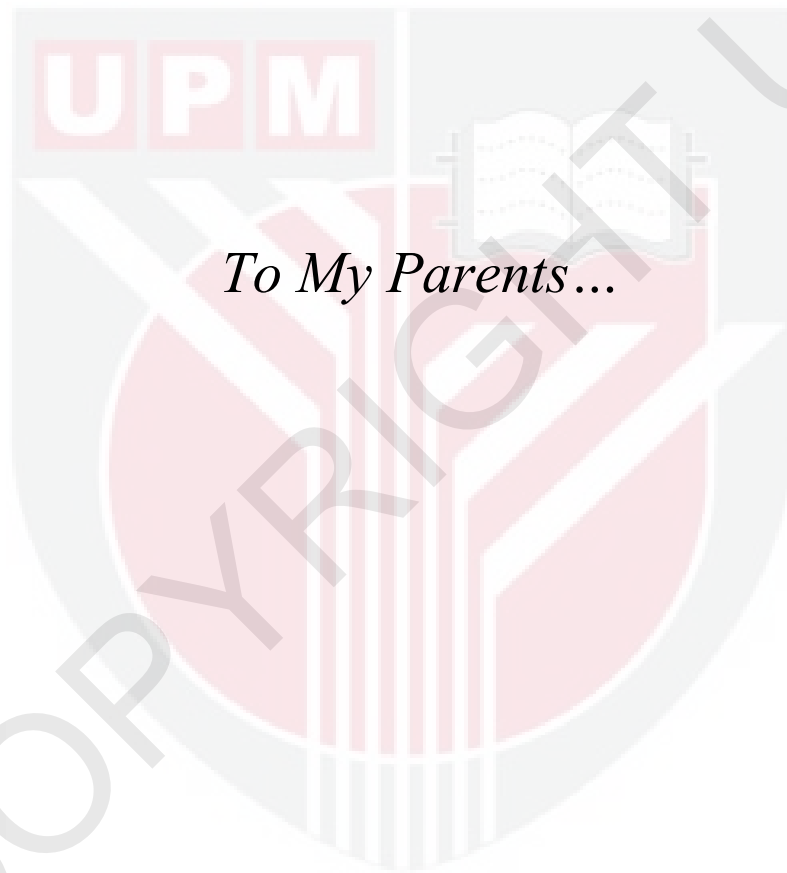
**STATIC AND SELF-SCALABLE FILTER RANGE SELECTION  
ALGORITHMS FOR PEER-TO-PEER NETWORKS**

By

**KWEH YEAH LUN**

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

**July 2011**



*To My Parents...*

© COPYRIGHT UPM

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

**STATIC AND SELF-SCALABLE FILTER RANGE SELECTION  
ALGORITHMS FOR PEER-TO-PEER NETWORKS**

By

**KWEH YEAH LUN**

**July 2011**

**Chairman: Professor Mohamed Othman, PhD**

**Faculty: Computer Science and Information Technology**

In this research, problems on the selection of keys in peer-to-peer networks are investigated. The key selection is about finding the target key with  $k^{\text{th}}$  rank in a global file with  $n$  keys that are distributed evenly among  $p$  nodes with each node holding  $n/p$  keys in the peer-to-peer network. In the literature, there are many selection algorithms proposed for different network topologies. For peer-to-peer networks, Loo (2005) selection algorithm has been selected as the benchmark as it is an established algorithm that claimed to be the best proposed for this network. The research works were implemented by simulation in which it was used to identify the selection problem, implementation of the proposed algorithms and the measurement of the results. Two multiple selection algorithm, which are known as “static filter range selection algorithm” and “self-scalable selection algorithm” are proposed. These algorithms are based on the statistical knowledge about the uniform distribution nature of the data and arranged in certain order in the file. The selection algorithms can perform multiple selections concurrently to find multiple target keys with different predefined target

ranks. The static filter range selection algorithm uses a fixed filter range approach to locate the target key, in which the filter range is preset at the beginning of the searching process. The range will be adjusted and becomes narrower while ensuring the target keys are still within it as the process iterates until the keys have been found. The self-scalable selection algorithm uses dynamic range where the filter range is not preset and is determined by the algorithm itself based on the distribution and the values of the data in the global and local file. After that, the ranges are made smaller and smaller until the target keys are found. Four parameters have been applied in this research to measure the performance of the algorithm. These are number of rounds needed, number of messages needed, success rate and execution time. The static filter range selection algorithm and the self-scalable selection algorithm are able to reduce the number of rounds and the number of messages needed, increase the success rate but longer execution time compared to the Loo (2005) selection algorithm. The self-scalable selection algorithm is also able to reduce the number of rounds and the number of messages needed, increase success rate and shorten the execution time compared to static filter range selection algorithm with filter range 15000, 20000 and 25000.

Abstrak tesis dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan ijazah Doctor Falsafah

**ALGORITMA STATIK DAN BOLEH-SKALA SENDIRI JULAT PENAPIS  
UNTUK RANGKAIAN PERANGKAI PADAN**

Oleh

**KWEH YEAH LUN**

**Julai 2011**

**Pengerusi: Profesor Mohamed Othman, PhD**

**Fakulti: Sains Komputer dan Teknologi Maklumat**

Dalam penyelidikan ini, masalah untuk pemilihan kunci dalam rangkaian perangkai padan telah disiasatkan. Pemilihan kunci adalah berkaitan dengan pencarian kunci-kunci sasaran dengan kedudukan sasaran ke- $k$  dalam sebuah fail global dengan  $n$  kunci yang ditaburkan secara sama rata di antara  $p$  nod dengan setiap nod memegang  $n/p$  kunci dalam rangkaian perangkai padan. Terdapat banyak algoritma pemilihan telah dicadangkan untuk topologi rangkaian yang berlainan dalam kepustakaan. Untuk rangkaian perangkai padan, algoritma pemilihan Loo (2005) telah dipilih sebagai sebab ia merupakan satu algoritma yang mantap dan didakwa yang terbaik dalam rangkaian ini. Kerja penyelidikan telah dilaksanakan dengan penyelakuan di mana ia digunakan untuk mengenalpasti masalah dalam pemilihan, pelaksanaan cadangan algoritma-algoritma dan pengukuran keputusan-keputusan yang didapati. Dua algoritma pemilihan berbilang, iaitu “*static filter range selection algorithm*” dan “*self-scalable selection algorithm*” telah dicadangkan. Algoritma-algoritma ini adalah berdasarkan pengetahuan statistik mengenai data dalam taburan seragam dan disusunkan dalam tertib tertentu

dalam fail. Algoritma pemilihan dapat melaksanakan pemilihan berbilang secara serempak untuk mencari kunci sasaran berbilang dengan pratakrif kedudukan yang berbeza. “*Static filter range selection algorithm*” menggunakan julat penapis tetap untuk menentukan lokasi kunci sasaran, di mana julat penapis telah dipreset pada permulaan proses pencarian. Proses pencarian berulang dengan julat dilaraskan dan menjadi semakin sempit dengan kunci-kunci sasaran masih berada dalam julat tersebut sehingga kunci-kunci tersebut didapati. “*Self-scalable selection algorithm*” menggunakan julat dinamik di mana julat penapis tidak dipreset dan ditentukan oleh algoritma sendiri berdasarkan taburan dan nilai kunci-kunci dalam fail global dan setempat. Selepas itu, julat itu dicecilkan sehingga kunci-kunci sasaran didapati. Empat parameter telah diaplikasikan dalam penyelidikan ini untuk mengukur prestasi algoritma-algoritma tersebut. Mereka adalah bilangan pusingan yang diperlukan, bilangan mesej yang diperlukan, kadar kejayaan dan masa perlakuan. “*Static filter range selection algorithm*” dan “*self-scalable selection algorithm*” berupaya mengurangkan bilangan pusingan dan mesej yang diperlukan, meningkatkan kadar kejayaan dengan masa perlakuan yang lebih panjang berbanding dengan algoritma pemilihan Loo (2005). “*Self-scalable selection algorithm*” juga dapat mengurangkan bilangan pusingan dan bilangan mesej yang diperlukan, meningkatkan kadar kejayaan dan memerlukan masa perlakuan yang lebih singkat berbanding kepada “*static filter range selection algorithm*” yang menggunakan julat penapis 15000, 20000 dan 25000.



## ACKNOWLEDGEMENTS

I would like to thank my main supervisor, Professor Dr. Mohamed Othman for giving me guidance in conducting this research. He is not only showering me with wisdoms and ideas for my work but also the way to manage the time well so that everything runs according to the plan. This gives me a lot of courage to solve the problems that I face and to finish the project on time. Professor Dr. Fatimah binti Dato Ahmad and Associate Professor Dr. Hamidah Ibrahim have gave me a lot of useful advices when I'm stuck in the research. Thanks for the understanding and giving me full support on my work.

Finally, I would like to thank my families who provided me with continual encouragement and those who are not mentioned here giving me endless support on my work.

**Kweh Yeah Lun**

July 2011

I certify that a Thesis Examination Committee has met on 7 July 2011 to conduct the final examination of Kweh Yeah Lun on his thesis entitled “Static and Self-Scalable Selection Algorithms for Peer-to-Peer Network” in accordance with the University and University College 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:

**NORWATI MUSTAPHA, PhD**

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

**RUSLI BIN HAJI ABDULLAH, PhD**

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

**MD NASIR BIN SULAIMAN, PhD**

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

**NICOLA SANTORO, PhD**

Professor  
School of Computer Science  
Carleton University  
Canada  
(External Examiner)

---

**NORITAH OMAR, PhD**

Associate Professor and Deputy Dean  
School of Graduate Studies  
Universiti Putra Malaysia.

Date:

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

**Mohamed Othman, PhD**

Professor

Faculty of Computer Science and Information Technology

Universiti Putra Malaysia

(Chairman)

**Fatimah bt. Dato Ahmad, PhD**

Professor

Faculty of Science and Defense Technology

Universiti Pertahanan Nasional Malaysia

(Member)

**Hamidah Ibrahim, PhD**

Associate Professor

Faculty of Computer Science and Information Technology

Universiti Putra Malaysia

(Member)

---

**HASANAH MOHD GHAZALI, 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 currently, submitted for any other degree at Universiti Putra Malaysia or other institutions.

---

**KWEH YEAH LUN**

Date:



# TABLE OF CONTENTS

	<b>Page</b>
<b>DEDICATION</b>	ii
<b>ABSTRACT</b>	iii
<b>ABSTRAK</b>	v
<b>ACKNOWLEDGEMENT</b>	vii
<b>APPROVAL SHEET</b>	viii
<b>DECLARATION</b>	x
<b>LIST OF TABLES</b>	xvi
<b>LIST OF FIGURES</b>	xx
<b>LIST OF ABBREVIATION</b>	xxii

## CHAPTERS

<b>1. INTRODUCTION</b>	1
1.1 Background	1
1.2 Statement of Problem	2
1.3 Objective	4
1.4 Scope	5
1.5 Contributions	5
1.6 Organization of Thesis	7
<b>2. PROGRESS IN SELECTION ALGORITHMS AND ITS PLATFORMS</b>	9
2.1 Introduction	9
2.2 Various Kind of Selection Algorithms	10
2.2.1 Median Searching Algorithm	10
2.2.2 Types of $k^{\text{th}}$ Selection Algorithms	12
2.2.3 Distributed Selection Algorithm on Different Topologies and Architectures	13
2.2.4 Parallel Selection	21
2.3 Related Works on Multiple Selection Algorithms in Peer-to-Peer Networks	28
2.4 Summary	30
<b>3. METHODOLOGY</b>	31
3.1 Introduction	31
3.2 Research Framework	31
3.2.1 Problem Identification	33
3.2.2 Objective Determination	33
3.2.3 Design	34
3.2.4 Implementation	35

3.2.5	Performance Parameters Identification	35
3.2.6	Analysis and Discussion	36
3.3	Peer-to-Peer Model and Selection Model	37
3.3.1	Peer-to-Peer Model	37
3.3.2	Selection Algorithm Model	38
3.3.3	Simulation Environment	41
3.4	Performance Parameters Measurement	46
3.4.1	Success Rate	46
3.4.2	Number of Rounds Needed	48
3.4.3	Number of Messages Needed	50
3.4.4	Execution Time	51
3.5	Summary	53
4.	<b>THEORETICAL FORMULATION AND ITS ANALYSIS FOR STATIC AND SELF-SCALABLE SELECTION ALGORITHMS</b>	54
4.1	Introduction	54
4.2	Selection Processes of Static Filter Range Selection Algorithm	54
4.2.1	Hash Value and Expected Actual Rank	55
4.2.2	Theoretical Expectation of Number of Rounds, Number of Messages with Different Filter Range and Number of Nodes Involved	61
4.3	Filter Determination Mechanism of Self-Scalable Selection Algorithm	68
4.3.1	Assumptions for Filter Determination	68
4.3.2	Upper Filter Determination	70
4.3.3	Lower Filter Determination	71
4.4	Summary	72
5.	<b>STATIC FILTER RANGE SELECTION ALGORITHM</b>	73
5.1	Introduction	73
5.2	Preprocessing Stage	73
5.2.1	Hashing Stage	74
5.2.2	Identification of Lower and Upper Filter	76
5.2.3	Determination of the Global Ranks for the Filters	77
5.3	Selection Processes	79
5.3.1	Pivot Sending	81
5.3.2	Pivot Receiving	82
5.4	Termination Process	83
5.5	Node Identification Mechanism	85
5.6	Summary	88
6.	<b>SELF-SCALABLE SELECTION ALGORITHM</b>	89
6.1	Introduction	89
6.2	Preliminary Stage	89
6.2.1	Hashing Process	90
6.2.2	First Filter Candidate	90
6.2.3	Second Filter Candidate and its Role Determination	91

6.3 Selection Processes of Self-Scalable Selection Algorithm	93
6.3.1 First Round: Sending	95
6.3.2 First Round: Receiving	95
6.3.3 Subsequent Rounds of Selection and Termination	96
6.4 The Strength of the Algorithm	97
6.5 Summary	99
7. <b>RESULTS AND DISCUSSION</b>	100
7.1 Introduction	100
7.2 Results of the Number of Rounds Needed	100
7.2.1 Results Analysis for the Static Filter Range Selection Algorithm with Loo (2005) Algorithm	100
7.2.2 Relationship between the Number of Nodes and the Number of Rounds Needed	106
7.2.3 Relationship between the Filter Range Applied and the Number of Rounds Needed	107
7.2.4 Decrement and Increment in the Difference in the Number of Rounds Needed Compared to Loo (2005) Algorithm	108
7.2.5 Results Analysis for the Self-Scalable Selection Algorithm with Loo (2005) Algorithm	109
7.2.6 Results Analysis for the Self-Scalable Selection Algorithm with the Static Filter Range 15000, 20000 and 25000 Selection Algorithms	111
7.2.7 Reduction in the Number of Rounds as the Number of Nodes Increases for the Self-Scalable Selection Algorithm	114
7.2.8 Results Analysis for the Self-Scalable Selection Algorithm and the Static Filter Range 5000 and 10000 Selection Algorithm	115
7.3 Results of the Number of Messages Needed	119
7.3.1 Results Analysis for the Static Filter Range Selection Algorithm with Loo (2005) Algorithm	119
7.3.2 Relationship between the Number of Nodes and the Number of Messages Needed	123
7.3.3 Relationship between the Filter Range Applied and the Number of Messages Needed	124
7.3.4 Relationship between the Number of Rounds and the Number of Messages Needed	124
7.3.5 Increment in the Difference in the Number of Messages Needed as the Number of Nodes Increase	124
7.3.6 Result Analysis for the Self-Scalable Selection Algorithm with Loo (2005) Algorithm	126
7.3.7 Results Analysis for the Self-Scalable Selection Algorithm with the Static Filter Range 15000, 20000 and 25000 Selection Algorithms	127
7.3.8 Relationship between the Number of Messages and the Number of Nodes for the Self-Scalable Selection Algorithm	130

7.3.9	Result Analysis for the Self-Scalable Selection Algorithm with the Static Filter Range 5000 and 10000 Selection Algorithms	132
7.3.10	Reduction in Improvement by the Self-Scalable Selection Algorithm Compared to the Static Filter Range 5000 and 10000 Selection Algorithms	135
7.3.11	Significance of Improvement of the Self-Scalable Selection Algorithm and the Number of Nodes	136
7.4	Results of the Success Rate	136
7.4.1	Results Analysis for the Static Filter Range Selection Algorithm with Loo (2005) Algorithm	137
7.4.2	Inability of Loo (2005) Algorithm to a Locate Certain Target Key	140
7.4.3	Increasing Success Rate as the Number of Nodes Increases	141
7.4.4	Consistent Success Rate for a Particular Static Filter Range Approach	142
7.4.5	Reducing the Success Rate as the Filter Ranges become Narrower for the Static Filter Range Selection Algorithm	142
7.4.6	Accuracy of the Hash Approach Applied in the Algorithm	143
7.4.7	Relationship between the Number of Rounds, Number of Messages and the Success Rate for the Static Filter Range Selection Algorithm	144
7.4.8	Results Analysis for the Self-Scalable Selection Algorithm with Loo (2005) Algorithm	146
7.4.9	Results Analysis for the Self-Scalable Selection Algorithm with the Static Filter Range Selection Algorithm	148
7.5	Results of the Execution Time	152
7.5.1	Results Analysis for the Static Filter Range Selection Algorithm with Loo (2005) Algorithm	152
7.5.2	Relationship between Filter Range and Execution Time	155
7.5.3	Results Analysis for the Self-Scalable Selection Algorithm with Loo (2005) Algorithm	156
7.5.4	Relationship between Execution Time and the Number of Nodes for the Self-Scalable Selection Algorithm	158
7.5.5	Longer Execution Time Factor for Self-Scalable Selection Algorithm	159
7.5.6	Results Analysis for the Self-Scalable Selection Algorithm with the Static Filter Range 15000, 20000 and 25000 Selection Algorithms	160
7.5.7	Significance of Improvement by the Self-Scalable Selection Algorithm as the Number of Nodes Increases	163
7.5.8	Results Analysis for the Self-Scalable Selection Algorithm with the Static Filter Range 5000 and 10000 Selection Algorithms	163
7.5.9	Greater Execution Time for the Self-Scalable Selection Algorithm Compared to the Static Filter Range 5000 and 10000 Selection Algorithm	166
7.6	Performance Parameters and Filter Range in General	166



7.7 Tradeoff between the Static Range Selection Algorithm and the Self-Scalable Selection Algorithm	168
7.7.1 Tradeoff for the Number of Rounds	169
7.7.2 Tradeoff for the Number of Messages	169
7.7.3 Tradeoff for the Execution Time	170
7.7.4 Tradeoff for the Success Rate	170
7.8 Summary	171
<b>8. CONCLUSION AND FUTURE WORKS</b>	<b>174</b>
8.1 Conclusion	174
8.2 Future Works	177
<b>REFERENCES</b>	<b>178</b>
<b>APPENDICES</b>	<b>186</b>
<b>BIODATA OF STUDENT</b>	<b>210</b>
<b>LIST OF PUBLICATIONS, AWARDS AND COPYRIGHT</b>	<b>211</b>

