



**UNIVERSITI PUTRA MALAYSIA**

***HYBRIDFLOOD ALGORITHMS MINIMIZING REDUNDANT  
MESSAGES AND MAXIMIZING EFFICIENCY OF SEARCH IN  
UNSTRUCTURED P2P NETWORKS***

**HASSAN BARJINI**

**FSKTM 2012 12**

**HYBRIDFLOOD ALGORITHMS MINIMIZING  
REDUNDANT MESSAGES AND MAXIMIZING  
EFFICIENCY OF SEARCH IN  
UNSTRUCTURED P2P NETWORKS**

The logo of Universiti Putra Malaysia (UPM) is a shield-shaped emblem. It features a red and white design with a book in the upper right corner and the letters 'UPM' in a red box at the top left. The shield is set against a light grey background.

**HASSAN BARJINI**

**DOCTOR OF PHILOSOPHY  
UNIVERSITI PUTRA MALAYSIA**

**2012**

**HYBRIDFLOOD ALGORITHMS MINIMIZING  
REDUNDANT MESSAGES AND MAXIMIZING  
EFFICIENCY OF SEARCH IN UNSTRUCTURED  
P2P NETWORKS**



By

**HASSAN BARJINI**

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

July 2012

## DEDICATIONS

*To My dear wife for her encouragement,  
and My sweet daughters*



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

**HYBRIDFLOOD ALGORITHMS MINIMIZING REDUNDANT MESSAGES AND MAXIMIZING EFFICIENCY OF SEARCH IN UNSTRUCTURED P2P NETWORKS**

By

**HASSAN BARJINI**

**July 2012**

**Chairman: Professor Mohamed Othman, PhD**

**Faculty: Computer Science and Information Technology**

Unstructured peer-to-peer (P2P) networks, aggregate the slack resources on each peer, which may include bandwidth, storage space, and computing power. As a peer joins this P2P network, the total demand and total capacity of the system simultaneously increase. However, in a typical client-server network, as a client joins the network it only shares its demands, not its resources. Thus as more clients join the client-server network fewer resources are available to serve each client. Besides, the decentralized structure of a P2P network increases its robustness because it removes the single point of failure that can be inherent in a client-server based system.

Therefore the unstructured model of the P2P network has attracted the greatest attention from both users and the researcher communities. Searching is an essential and basic activity for all P2P applications. Thus, there are a large number of research works have focused on unstructured P2P search facilities. There are two main reasons driving the research interest in this area.

First, the upward trend of digital information production, such as HTML, music and image files, requires a scalable information infrastructure that is capable of indexing and searching. Recent studies have shown that more than 97% of information produced worldwide is in a digital form. The amount of digital information is expected to grow exponentially. There are many challenges posed by such a huge amount of information for existing search systems.

Second, compared to traditional centralized networks, unstructured P2P networks are particularly attractive and promising due to their scalability, availability, low cost, easy deployment, and data freshness. Meanwhile the fundamental property of existing and scalable unstructured P2P networks is the high heterogeneity of the peers that participate in the network. The heterogeneity of peers in unstructured P2Ps introduces both challenges and opportunities when designing a P2P network. Flooding is a basic file search procedure in unstructured P2P file-sharing systems. In flooding a peer initiates the file search operation by broadcasting a query to its neighbors, who continue to propagate it to their neighbors. Flooding has no knowledge about network topology nor files or resources distribution, so it offers an attractive method for file discovery in dynamic and developing networks. In the meantime, flooding produces exponentially redundant messages at each hop. Consequently, the growth of redundant messages limits the system's scalability and causes unnecessary traffic in networks.

In this thesis, we combine two search techniques to tackle this issue and improve P2P search performance in terms of search efficiency and the quality of the search results. We proposed two novel search algorithms named QuickFlood and HybridFlood. QuickFlood combines two flood-based searches; flooding and teeming. QuickFlood is performed in two steps, in a first step the algorithm performs flooding in a limited number of hops. In the second step the algorithm follows a teeming search. QuickFlood compared with blocking expanding ring search decreased re-

dundant messages by 70%, increased two-time success rate and decreased latency by 30%. Therefore, the algorithm enhanced unstructured P2P search by increasing scalability, efficiency and reliability of search.

HybridFlood is also performed in two steps. The first step performs flooding with a limited number of hops. In the second step, nosey nodes are selected in each searching horizon. The nosey nodes are nodes which have the most links to other nodes. These nodes maintain the data index of all client nodes. HybridFlood in comparison to the blocking expanding ring search decreased redundant messages by 80%, increased the success rate by 2.5, and decreased of latency by 80%. In the other word the algorithm improved unstructured P2P search's scalability, efficiency and reliability.

We provided analytical studies for flooding, QuickFlood and HybridFlood. The analytical results provided the best hop threshold point for the optimum growth rate coverage and redundant messages from the three systems. It also proved that in HybridFlood, broadcasting messages are reduced by at least an order of magnitude. Thus, the proposed algorithms enhance the performance of the search by reducing redundant messages, increasing the success rate and decreasing latency. The simulation experiments validated the analytical results.

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

**ALGORITMA HYBRIDFLOOD UNTUK MEMINIMAKAN  
KEBERULANGAN MESEJ DAN MEMAKSIMAKAN  
KECEKAPAN DALAM RANGKAIAN TIDAK BERSTRUKTUR  
P2P**

Oleh

**HASSAN BARJINI**

**Julai 2012**

**Pengerusi: Profesor Mohamed Othman, PhD**

**Fakulti: Sains Komputer dan Teknologi Maklumat**

Rangkaian tidak berstruktur rakan-kepada-rakan (P2P), mengumpulkan sumber-sumber yang bekerja perlahan pada setiap rakan, yang mungkin merangkumi jalur lebar, ruang simpanan, dan kuasa pengkomputan. Apabila rakan menyertai rangkaian P2P ini, jumlah permintaan dan jumlah kapasiti sistem meningkat secara serentak. Walau bagaimanapun, dalam rangkaian pelanggan-pelayan biasa, apabila pelanggan menyertai rangkaian ia hanya berkongsi permintaan, bukan sumber. Jadi apabila lebih banyak pelanggan menyertai rangkaian pelanggan-pelayan kurang sumber-sumber yang ada untuk melayani setiap pelanggan. Selain itu, struktur tidak berpusat rangkaian P2P meningkatkan keteguhannya kerana ia membuang titik tunggal kegagalan yang boleh wujud dalam sistem berasaskan pelanggan-pelayan.

Oleh itu, model tidak berstruktur rangkaian P2P telah menarik perhatian yang paling besar daripada kedua-dua pengguna dan komuniti penyelidik. Carian meru-



pakan aktiviti asas dan penting untuk semua aplikasi P2P. Oleh itu, terdapat sebilangan besar kerja-kerja penyelidikan tertumpu kepada kemudahan carian tidak berstruktur P2P. Terdapat dua sebab utama yang menggalakkan minat penyelidikan dalam bidang ini.

Pertama, aliran menaik penghasilan maklumat digital, seperti HTML fail-fail muzik dan imej, memerlukan infrastruktur maklumat berskala yang mampu mengindeks dan mencari. Kajian terkini menunjukkan bahawa lebih daripada 97% maklumat yang dihasilkan di seluruh dunia adalah dalam bentuk digital. Bilangan maklumat digital dijangkakan akan bertambah secara eksponen. Terdapat banyak cabaran ditimbulkan oleh maklumat sebegitu banyak bagi sistem carian yang sedia ada.

Kedua, daripada maklumat yang dihasilkan di seluruh dunia adalah dalam bentuk digital. Jumlah maklumat digital dijangka berkembang dengan pesat. Terdapat banyak cabaran yang ditimbulkan seperti jumlah maklumat yang besar untuk sistem carian yang sedia ada. Kedua, dibandingkan dengan rangkaian berpusat tradisional, rangkaian P2P tidak berstruktur khususnya menarik dan menjanjikan kebolehskalaan, ketersediaan, kos rendah, penggunaan mudah, dan kesegaran data. Sementara itu, perkara asas rangkaian tidak berstruktur dan berskala P2P yang sedia ada adalah kepelbagaian yang tinggi daripada rakan-rakan sebaya yang mengambil bahagian dalam rangkaian. Kepelbagaian rakan-rakan dalam P2P tidak berstruktur mendedahkan kedua-dua cabaran dan peluang-peluang apabila mereka mereka bentuk bentuk satu rangkaian P2P.

Pembanjiran adalah prosedur asas carian fail dalam sistem perkongsian fail P2P yang tidak berstruktur. Dalam pembanjiran rakan sebaya memulakan operasi fail carian dengan menyebarkan permintaan kepada jiran-jiran, yang terus menyebarkan kepada jiran-jiran mereka yang lain. Pembanjiran tidak mempunyai pengetahuan mengenai topologi rangkaian mahu pun fail atau pengagihan sumber, jadi ia menawarkan satu kaedah yang menarik untuk penemuan fail dalam rangkaian

yang dinamik dan membangun. Sementara itu, pembanjiran menghasilkan mesej-mesej lewah di setiap loncatan masing-masing. Akibatnya, penambahan mesej berlebihan menghadkan kebolehskalaan sistem dan menyebabkan lalu lintas yang tidak perlu dalam rangkaian.

Dalam tesis ini, kami menggabungkan dua teknik carian untuk menangani isu ini dan meningkatkan prestasi carian P2P dari segi kecekapan carian dan kualiti hasil carian. Kami mencadangkan dua algorithma carian baharu yang dinamakan QuickFlood dan HybridFlood. QuickFlood menggabungkan dua carian berasaskan pembanjiran; pembanjiran dan penuh sesak. QuickFlood dilaksanakan dalam dua langkah: dalam langkah pertama algorithma ini melaksanakan pembanjiran di beberapa loncatan terhad. Dalam langkah kedua prosedur pengiraan mengikut satu carian penuh sesak. QuickFlood dibandingkan dengan carian cincin sekatan berkembang mangurangkan mesej lewah sebanyak 70%, meningkat kan kadar kejayaan dua-kali dan pendaman menurun kan sebanyak 30%. Oleh itu, prosedur pengiraan dipertingkatkan carian P2P tidak berstruktur dengan meningkatkan kebolehan untuk diskala, kecekapan dan kebolehpercayaan carian.

HybridFlood juga dilaksanakan dalam dua langkah. Langkah pertama melaksanakan pembanjiran dengan bilangan loncatan yang terhad. Pada langkah kedua, nod-nod yang ingin tahu dipilih dalam setiap lapisan carian. Nod-nod ingin tahu adalah nod yang mempunyai pautan kepada nod-nod lain. Nod ini mengekalkan indeks data semua nod pelanggan. HybridFlood, jika dibandingkan dengan carian cincin sekatan berkembang menurunkan mesej lewah sebanyak 80%, meningkatkan kadar kejayaan sebanyak 2.5, dan menurunkan pendaman sebanyak 80%. Dengan kata lain, algorithma ini meningkatkan kebolehskalaan, kecekapan dan kebolehpercayaan carian tidak berstruktur P2P.

Kami menyediakan kajian analisis untuk pembanjiran, QuickFlood dan HybridFlood. Keputusan analisis memberikan titik ambang loncatan yang terbaik untuk

pertumbuhan kadar liputan yang optima dan mesej lewah dari ketiga-tiga sistem. Ia juga membuktikan bahawa dalam HybridFlood, penyebaran mesej dapat dikurangkan sekurang-kurangnya satu tertib magnitud. Oleh itu, algoritma yang dicadangkan meningkatkan prestasi carian dengan mengurangkan mesej lewah, meningkatkan kadar kejayaan dan mengurangkan pendaman. Eksperimen simulasi mengesahkan keputusan analisis.



## ACKNOWLEDGEMENTS

First all, praise is for *Allah Subhanahu Wa Taala* for giving me the strength, guidance and patience to complete this thesis.

I would like to express my sincere gratitude to my supervisor Professor Dr. Mohamed Othman and also my supervisory committee members Associate Professor Dr. Hamidah Ibrahim and Dr. Nur Izura Udzir for their guidance and advice throughout this work in making this a success.

I certify that a Thesis Examination Committee has met on **19 July 2012** to conduct the final examination of **Hassan Barjini** on his thesis entitled "**Hybrid-Flood Algorithms Minimizing Redundant Messages and Maximizing Efficiency of Search in Unstructured P2P Networks**" 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:

**Ali b Mamat, PhD**

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

**Azizol b Hj Abdullah, PhD**

Senior Lecturer  
Faculty of Computer Science and Information Technology  
Universiti Putra Malaysia  
(Internal Examiner)

**Kasmiran Jumari, PhD**

Professor  
Faculty of Engineering and Built Environment  
Universiti Kebangsaan Malaysia  
(External Examiner)

**Xiaodong Zhang, PhD**

Professor  
Faculty of Computer Science and Engineering  
College of Engineering  
The Ohio State University United State  
(External Examiner)

**ZULKARNAIN ZAINAL, PhD**

Professor and Deputy Dean  
School of Graduate Studies  
Universiti Putra Malaysia

Date: 27 August 2012

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

**Mohamed Othman, PhD**

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

**Hamidah Ibrahim, PhD**

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

**Nur Izura Udzir, PhD**

Lecturer  
Faculty of Computer Science and Information Technology  
Universiti Putra Malaysia  
(Member)

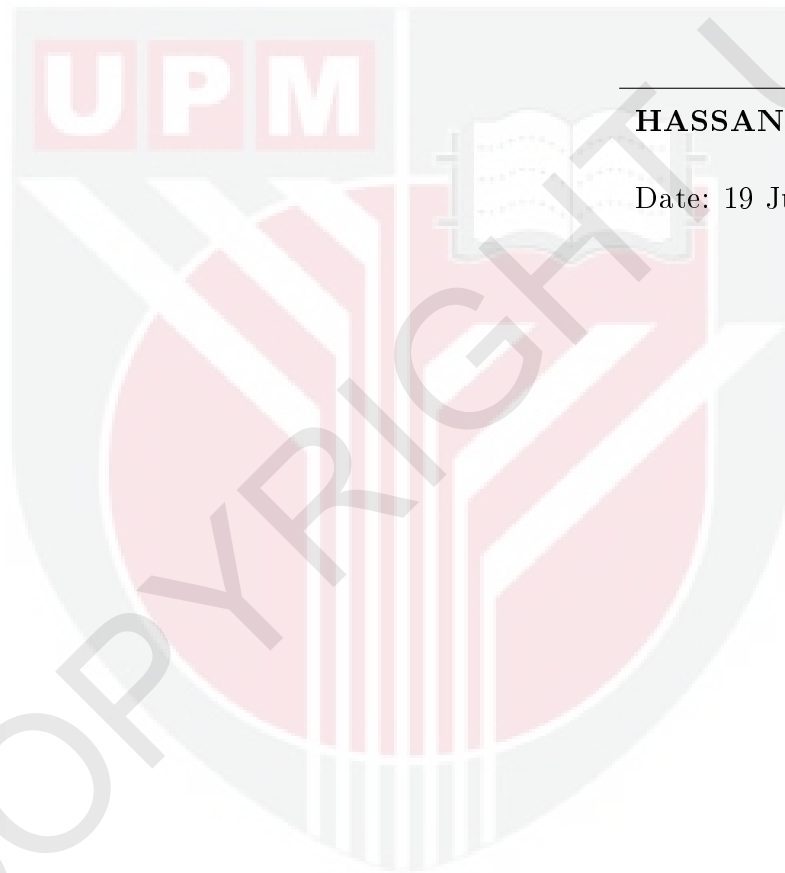
**BUJANG BIN KIM HUAT, PhD**

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.



---

**HASSAN BARJINI**

Date: 19 July 2012



## TABLE OF CONTENTS

|  | <b>Page</b> |
|--|-------------|
| <b>DEDICATIONS</b>                                 | i           |
| <b>ABSTRACT</b>                                    | ii          |
| <b>ABSTRAK</b>                                     | v           |
| <b>ACKNOWLEDGEMENTS</b>                            | ix          |
| <b>APPROVAL</b>                                    | x           |
| <b>DECLARATION</b>                                 | xii         |
| <b>LIST OF TABLES</b>                              | xvi         |
| <b>LIST OF ABBREVIATIONS</b>                       | xx          |
| <br>   |             |
| <b>CHAPTER</b>                                     |             |
| <b>1 INTRODUCTION</b>                              | 1           |
| 1.1 Background                                     | 1           |
| 1.2 P2P Search Architecture                        | 1           |
| 1.2.1 P2P Structured                               | 1           |
| 1.2.2 P2P Unstructured                             | 2           |
| 1.2.3 P2P Hybrid                                   | 4           |
| 1.3 Significant of the Research                    | 5           |
| 1.4 Problem Statement                              | 7           |
| 1.5 Objectives of the Research                     | 8           |
| 1.6 Research Scope                                 | 8           |
| 1.7 Contributions of the Research                  | 9           |
| 1.8 Thesis Organization                            | 10          |
| <br>   |             |
| <b>2 LITERATURE REVIEW</b>                         | 11          |
| 2.1 Introduction                                   | 11          |
| 2.2 Fundamental Terms and Concepts                 | 12          |
| 2.3 Overlay Network                                | 12          |
| 2.3.1 Flat P2P Network Topology                    | 12          |
| 2.3.2 Tree-based P2P Network Topology              | 14          |
| 2.3.3 Hybrid P2P Network Topology                  | 15          |
| 2.4 Search Techniques in Unstructured P2P Networks | 16          |
| 2.4.1 Blind Search Algorithms                      | 17          |
| 2.4.2 Informed Search Algorithms                   | 27          |
| 2.5 Super-Peer Selection and Searching Algorithms  | 30          |
| 2.5.1 OceanStore                                   | 31          |
| 2.5.2 Brocade                                      | 32          |
| 2.5.3 Gnutella                                     | 34          |
| 2.5.4 KaZaa  | 35          |



|          |  |           |
|----------|--|-----------|
| 2.5.5    | Gnutella 0.6   | 36        |
| 2.5.6    | eDonkey  | 37        |
| 2.5.7    | Supernode Based P2P File-Sharing Networks  | 37        |
| 2.6      | Related Works on Search Mechanisms   | 38        |
| 2.6.1    | Blind Search Algorithms Review   | 38        |
| 2.6.2    | Informed Search Algorithms Review  | 45        |
| 2.6.3    | Super-peer Search Algorithms Review  | 45        |
| 2.7      | Summary  | 48        |
| <b>3</b> | <b>RESEARCH METHODOLOGY</b>  | <b>50</b> |
| 3.1      | Introduction   | 50        |
| 3.2      | An Overview  | 50        |
| 3.3      | Research Framework   | 51        |
| 3.3.1    | Problem Formulation  | 51        |
| 3.3.2    | Previous Flood-Based Search Algorithm  | 52        |
| 3.3.3    | Proposed Schemes   | 54        |
| 3.3.4    | Perform Analytical Studies and Simulation Experiments                                    | 55        |
| 3.4      | Experiment Environment   | 55        |
| 3.4.1    | Data Collection  | 55        |
| 3.4.2    | Network Model Assumptions and Parameters   | 56        |
| 3.4.3    | Experimental Setup   | 63        |
| 3.5      | Performance Metrics  | 63        |
| 3.5.1    | Query Success Rate   | 63        |
| 3.5.2    | Number of Redundant Messages   | 64        |
| 3.5.3    | Number of Latency  | 64        |
| 3.6      | Simulator Overview   | 64        |
| 3.7      | Summary  | 65        |
| <b>4</b> | <b>ANALYTICAL STUDIES AND EXPERIMENTAL EXAMINES FOR FLOODING-BASED SEARCH ALGORITHMS</b> | <b>67</b> |
| 4.1      | Introduction   | 67        |
| 4.2      | Flooding Algorithm across the Hops   | 67        |
| 4.2.1    | Trend of Coverage Growth Rate  | 68        |
| 4.2.2    | Trend of Redundant Messages  | 71        |
| 4.3      | Proposed Performance Metric  | 75        |
| 4.3.1    | Critical Metric  | 76        |
| 4.3.2    | Critical Metrics in Flood-Based Algorithms   | 77        |
| 4.3.3    | Evaluation of Critical Metrics in Flood-Based Algorithms                                 | 83        |
| 4.4      | Analytical Results   | 86        |
| 4.5      | Experimental Results   | 87        |
| 4.6      | Summary  | 89        |

|          |   |            |
|----------|---|------------|
| <b>5</b> | <b>PROPOSED QUICKFLOOD SEARCH ALGORITHM</b>   | <b>93</b>  |
| 5.1      | Introduction                                  | 93         |
| 5.2      | Proposed QuickFlood Model                     | 94         |
| 5.3      | First Step of Algorithm                       | 94         |
| 5.4      | Second Step Description                       | 96         |
| 5.5      | Second Step Algorithm                         | 97         |
| 5.6      | Analytical Studies                            | 98         |
| 5.7      | Experimental Results                          | 101        |
| 5.8      | Summary                                       | 104        |
| <b>6</b> | <b>PROPOSED HYBRIDFLOOD SEARCH ALGORITHM</b>  | <b>106</b> |
| 6.1      | Introduction                                  | 106        |
| 6.2      | Proposed HybridFlood Search Model             | 106        |
| 6.3      | First Step of Algorithm                       | 107        |
| 6.4      | Second Step Description                       | 108        |
| 6.4.1    | Nosey Node Selection                          | 108        |
| 6.4.2    | Indexing Procedure in Nosey Nodes             | 110        |
| 6.4.3    | Nosey Node Overlay Network                    | 111        |
| 6.4.4    | Nosey Node Redundancy                         | 112        |
| 6.4.5    | Discussion on Nosey Nodes                     | 112        |
| 6.5      | Second Step of Algorithm                      | 113        |
| 6.6      | Analytical Study                              | 114        |
| 6.6.1    | Optimum Hop Count to Switch between Two Steps | 114        |
| 6.6.2    | Analytical Study in the Second Step           | 116        |
| 6.7      | Experimental Results                          | 118        |
| 6.8      | Summary                                       | 122        |
| <b>7</b> | <b>CONCLUSION AND FUTURE WORKS</b>            | <b>123</b> |
| 7.1      | Conclusion                                    | 123        |
| 7.2      | Future Works                                  | 124        |
|          | <b>REFERENCES</b>                             | <b>126</b> |
|          | <b>APPENDIX I</b>                             | <b>134</b> |
|          | <b>APPENDIX II</b>                            | <b>150</b> |
|          | <b>BIODATA OF STUDENT</b>                     | <b>165</b> |
|          | <b>LIST OF PUBLICATIONS</b>                   | <b>166</b> |