



TRIBHUVAN UNIVERSITY
INSTITUTE OF SCIENCE AND TECHNOLOGY

**ANALYSIS OF QUERIES ROUTING IN SUPER-SUPER-PEER BASED P2P
ARCHITECTURE USING NBTree: THE HYBRID ALGORITHM**

Thesis

Submitted to:

Central Department of Computer Science and Information Technology
Kirtipur, Kathmandu, Nepal

In Partial Fulfillment of the Requirements for
Master's Degree in Computer Science & Information Technology

Submitted by:

Mr. Abhishek Dewan

September, 2014

Under the Guidance of

Prof. Dr. Subarna Shakya

Department of Electronics and Computer Engineering

Institute of Engineering, Pulchowk, Nepal

ACKNOWLEDGEMENT

ABSTRACT

The Internet is converging to a more dynamic, huge, fully distributed peer-to-peer (P2P) overlay networks containing millions of nodes typically for the purpose of information distribution and file sharing as the increase in the number of computers connected to the Internet are increasing rapidly. Because of which a challenging problem in unstructured P2P system is how to locate peers that are relevant with respect to a given query with minimum query processing and minimum answering time. Connected peers can leave the overlay network any time and new peers can join it any time. To achieve our goal we suggest an unstructured P2P system which is based on an organization of peers around super-peers that is connected to super-super-peer according to their semantic domains and also uses NBTree: The Hybrid Algorithm to extract Super-Peer that contains peers with relevant data respect to a given query.

Keywords: Decision Tree, Machine Learning, NBTree, P2P, P2P Queries Answering, P2P Queries Routing, Super-Super-Peer, Weka

TABLE OF CONTENTS

TITLE	PAGE
COVER PAGE	
ACKNOWLEDGEMENT	II
ABSTRACT	III
TABLE OF CONTENTS	IV-VI
LIST OF TABLE	VII
LIST OF FIGURE	VIII
LIST OF ABBRIVATION	IX
CHAPTER 1 INTRODUCTION	1 - 3
1.1. Background of the Study	1
1.2. Statement of the Problem	2
1.3. Objectives of the Study	3
1.4. Scope and Limitation of the Study	3
1.5. Structure of the Report	3
CHAPTER 2 LITERATURE REVIEW	4 - 20
2.1. Peer-to-Peer Networks	4
2.1.1. Importance of P2P Network	4
2.1.2. Dark Sides of P2P Network	5
2.2. P2P Architecture	5
2.3. P2P Overlay Networks	6
2.3.1. Unstructured P2P Overlay Network	6
A. A Centralized Approach: Napster	7
B. A Distributed Approach: Gnutella	8
C. A Hybrid Approach: Kazaa	9

	2.3.2. Structured P2P Overlay Network	10
2.4.	Machine Learning	10
	2.4.1. Supervised Learning	11
	2.4.2. Unsupervised Learning	12
2.5.	Classification	13
	2.5.1. Decision Tree	13
	2.5.2. Naive-Bayes	17
	2.5.3. NBTree	18
2.6.	Application Programming Interface (API) for Data Mining	19
	2.6.1. WEKA	19
2.7.	Related Works	19
CHAPTER 3	RESEARCH METHODOLOGY	22 - 24
3.1.	Background	22
3.2.	P2P System Architecture Model	22
3.3.	Algorithms	23
	3.3.1. Query Routing Algorithm	23
	3.3.2. NBTree	24
3.4.	Source of Data	24
3.5.	Experiments Protocol	24
	3.5.1. Experiment Setup and Evaluation	24
CHAPTER 4	EXPERIMENT AND RESULT ANALYSIS	25-36
4.1.	Background	25
4.2.	Tool	25
4.3.	Data Structure	26

4.4.	Data Samples	26
4.5.	Experiments and Results	27
4.5.1.	Experiment	27
4.5.2.	Results	30
4.6.	Context Analysis	35
CHAPTER 5	CONCLUSION AND FUTURE WORK	37
5.1.	Conclusion	37
5.2.	Future Works	38
REFERENCES		39-42

LIST OF TABLE

TABLE	TOPIC	PAGE
Table 3.1:	NBTree Experimental Parameters	24
Table 4.1:	Confusion Matrix	29
Table 4.2:	Result of NBTree	30
Table 4.3:	Result of NBTree taking Accuracy	31
Table 4.4:	Result of NBTree taking Precision, Recall and F-Measure	32
Table 4.5:	Result of NBTree taking Tree Size	33
Table 4.6:	Result of NBTree taking Time Taken to Build Model (TTBM)	34

LIST OF FIGURE

FIGURE	TOPIC	PAGE
Figure 2.1:	P2P Network	4
Figure 2.2:	P2P overlay networks distribution	6
Figure 2.3:	Search process in structured process in unstructured P2P networks a) Napster b) Gnutella c) Kazaa and d) BitTorrent	7
Figure 2.4:	Architecture of Napster	8
Figure 2.5:	Example of Kazaa network	10
Figure 2.6:	Decision tree example	14
Figure 2.7:	Possibility of attribute as node	15
Figure 3.1:	Semantic Network of Super-Peer organized by cluster with Super-Super-Peer	22
Figure 4.1:	Portion of input samples viewed in WEKA ARFF-Viewer	26
Figure 4.2:	Result of running NBTree: The Hybrid Algorithm	27
Figure 4.3:	Result of classified instances by using NBTree	28
Figure 4.4:	Graph of Table 4.2	31
Figure 4.5:	Graph of Table 4.3	32
Figure 4.6:	Graph of Table 4.4	33
Figure 4.7:	Graph of Table 4.5	34
Figure 4.8:	Graph of Table 4.6	35

LIST OF ABBREVIATION

API	:	Application Programming Interface
ARFF	:	Attribute-Relation File Format
ARPANET	:	Advanced Research Projects Agency Network
CART	:	Classification and Regression Tree
DHT	:	Distributed Hash Table
GAs	:	Genetic Algorithms
GATree	:	Genetically Evolved Decision Tree
GNU	:	General Public License
HTTP	:	Hyper Text Transfer Protocol
ID3	:	Iterative Dichotomiser
IP	:	Internet Protocol
MANETs	:	Mobile Ad hoc Networks
P2P	:	Peer-to-Peer
QoS	:	Quality of Service
SIP	:	Single Inline Package
TTBM	:	Time Tiken to Build Model
TTL	:	Time-to-Live
URL	:	Uniform Resource Locator
WEKA	:	Waikato Environment for Knowledge Analysis