
**UNIVERSITI PUTRA MALAYSIA**

# *NETWORK BANDWIDTH UTILIZATION BASED ON COLLABORATIVE WEB CACHING USING MACHINE LEARNING ALGORITHMS IN PEERTO-PEER SYSTEMS FOR MEDIA WEB OBJECTS*

**WAHEED YASIN MOHAMMED**

**FSKTM 2018 65**

# NETWORK BANDWIDTH UTILIZATION BASED ON COLLABORATIVE WEB CACHING USING MACHINE LEARNING ALGORITHMS IN PEER-TO-PEER SYSTEMS FOR MEDIA WEB OBJECTS


**By**

**WAHEED YASIN MOHAMMED**


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

**May 2018**


© COPYRIGHT UPM

**COPYRIGHT**

All material contained within the thesis, including without limitation text, logos, icons, photographs and all other artwork, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial use of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright © Universiti Putra Malaysia



© COPYRIGHT UPM

**DEDICATION**

Dedicated to the human beings



© COPYRIGHT UPM

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

**NETWORK BANDWIDTH UTILIZATION BASED ON COLLABORATIVE WEB CACHING USING MACHINE LEARNING ALGORITHMS IN PEER-TO-PEER SYSTEMS FOR MEDIA WEB OBJECTS**

By

**WAHEED YASIN MOHAMMED**

**May 2018**

**Chairman: Professor Hamidah Binti Ibrahim, PhD**
**Faculty: Computer Science and Information Technology**

Web caching plays a key role in delivering web items to end users in World Wide Web (WWW). Many benefits can be gathered from caching such as improving the hit rates, alleviating loads on origin servers, and reducing network traffic.

Cache size is considered as a limitation of web caching. Furthermore, retrieving the same media object from the origin server many times consumes the network bandwidth. On the other hand, full caching for media objects is not a practical solution and consumes cache storage in keeping few media objects because of its limited capacity. Moreover, traditional web caching policies such as Least Recently Used (LRU) and Least Frequently Used (LFU) suffer from caching pollution (i.e. media objects that are stored in the cache are not frequently visited, which negatively affects on the performance of web proxy caching). This problem has been addressed in the works of Ali et al. (2012a), Ali et al. (2012b), Julian et al. (2014), and Julian and Sagayaraj (2015). For example, the average improvement of Hit Ratio (HR) in the works of Ali et al. (2012a) and Ali et al. (2012b) achieved by NB-LRU approach over LRU increased by 7.68%. In terms of Byte Hit Ratio (BHR), the average improvement achieved by NB-LRU, NB-LFU approaches over LRU and LFU are 11.65%, 2.88%, respectively. On the other hand, they do not consider the advantages that can be given by applying these approaches in peer-to-peer systems.

In this work, intelligent collaborative web caching approaches based on C4.5 decision tree and Naïve Bayes (NB) supervised machine learning algorithms are presented. The proposed approaches take the advantage of structured peer-to-peer systems where peers' caches contents are shared in order to enhance the performance of the web caching policy.



© COPYRIGHT UPM

The performance of the proposed approaches is evaluated by running simulations on a two datasets that are collected from YemenNet which is the Internet Service Provider (ISP) in Yemen, and IRCache network which is used as a source for dataset in many researches. The results demonstrate that the new proposed approaches improve the performance of LFU and LRU traditional web caching policies in terms of HR, BHR, and Cost Throughput (CT), the results are compared with the most relevant and state-of-the-art web proxy caching policies.



© COPYRIGHT UPM

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

**PENGGUNAAN JALUR LEBAR RANGKAIAN BERDASARKAN KERJASAMA *WEB CACHING* MENGGUNAKAN ALGORITMA PEMBELAJARAN MESIN DALAM SISTEM *PEER-TO-PEER* UNTUK OBJEK WEB MEDIA**

Oleh

**WAHEED YASIN MOHAMMED**

**Mei 2018**

**Pengerusi: Professor Hamidah Binti Ibrahim, PhD**
**Fakulti: Sains Komputer dan Teknologi Maklumat**

*Web caching* memainkan peranan yang penting dalam menghantar item Web kepada pengguna akhir di *World Wide Web (WWW)*. Banyak kebaikan boleh diperolehi daripada *caching* seperti meningkatkan kadar hit, mengurangkan beban kepada pelayan asal, dan mengurangkan trafik rangkaian.

Saiz *cache* dianggap sebagai kekangan bagi *Web caching*. Tambahan pula, mencapai objek media yang sama dari pelayan asal beberapa kali mengambil sejumlah rangkaian jalur lebar. Sebaliknya, *caching* objek media secara keseluruhan bukan jalan penyelesaian yang praktikal dan mengambil storan *cache* dalam menyimpan beberapa objek media disebabkan kekurangan kapasiti. Tambahan pula, polisi *Web caching* tradisional seperti *Least Recently Used (LRU)* dan *Least Frequently Used (LFU)* mengalami pencemaran *caching* (iaitu objek media yang disimpan di dalam *cache* tidak selalunya dilawati dengan kerap yang memberi kesan negatif kepada prestasi *Web Proxy Caching*). Masalah ini telah ditangani dalam karya-karya Ali et al. (2012a), Ali et al. (2012b), Julian et al. (2014), dan Julian and Sagayaraj (2015). Contohnya, peningkatan purata *Hit Ratio (HR)* dalam karya-karya Ali et al. (2012a) dan Ali et al. (2012b) yang dicapai oleh pendekatan *NB-LRU* terhadap *LRU* meningkat sebanyak 7.68%. Dari segi Nisbah *Byte Hit Ratio (BHR)*, peningkatan purata yang dicapai oleh *NB-LRU*, pendekatan *NB-LFU* terhadap *LRU* dan *LFU* masing-masing adalah 11.65%, 2.88%. Sebaliknya, mereka mengabaikan kelebihan yang boleh diberikan dengan menggunakan pendekatan ini dalam sistem *peer-to-peer*.

Di dalam kerja ini, pendekatan *Web caching* kerjasama cerdas berdasarkan pepohon keputusan *C4.5* dan algoritma pembelajaran mesin terselia *Naïve Bayes (NB)* dibentangkan. Pendekatan yang dicadangkan mengambil peluang yang ada di dalam



© COPYRIGHT UPM

struktur sistem *peer-to-peer*, di mana kandungan *cache peer* dikongsi bagi meningkatkan prestasi polisi *Web caching*.

Prestasi bagi pendekatan yang dicadangkan dinilai dengan simulasi ke atas set data yang dikumpul dari rangkaian YemenNet dan IRCache. Keputusan menunjukkan bahawa pendekatan baharu yang dicadangkan meningkatkan prestasi dasar cache web tradisional LFU dan LRU dari segi *Hit Ratio (HR)*, *Byte Hit Ratio (BHR)* dan *Cost Throughput (CT)*. Tambahan pula hasil dibandingkan dengan polisi *Web proxy caching* yang paling relevan dan *state-of-the-art*.



© COPYRIGHT UPM

**ACKNOWLEDGEMENTS**

First of all, I want to thank ALLAH for giving me the opportunity to increase my knowledge by being a Doctor of Philosophy candidate. I have learned a lot of things during this period of time. I improve my skills in many fields of life.

Moreover, I want to thank my supervisory committee members whom have helped me by guiding me and providing the advice and comments to fulfil this work. Professor Dr. Hamidah Ibrahim was really patient throughout my PhD study. Also, I am grateful to Associated Professor Dr. Nur Izura Udzir, and Dr. Nor Asilah Wati Abdul Hamid.

Also, the first dataset which is used in this work is provided by the National Science Foundation (grants NCR-9616602 and NCR-9521745), and the National Laboratory for Applied Network Research (NLANR). Furthermore, the second dataset which is used in this work is provided by YemenNet ISP.

This work is partly sponsored by the Yemeni Ministry of Higher Education, the Malaysian Ministry of Higher Education, Yemen Telecom, Yemen Mobile, and Yemen Net.

Finally, I want to thank my parents, wife, brother, sisters, colleagues, and friends whom always encourage me and pray for me.

© COPYRIGHT UPM

I certify that a Thesis Examination Committee has met on 4 May 2018 to conduct the final examination of Waheed Yasin Mohammed on his thesis entitled "Network Bandwidth Utilization Based on Collaborative Web Caching using Machine Learning Algorithms in Peer-to-Peer Systems for Media Web Objects" 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.

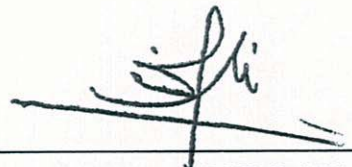
Members of the Thesis Examination Committee were as follows:

**Azizol bin Hj. Abdullah, PhD**
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

**Lilly Suriani Affendey, PhD**
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

**Datin Norwati Mustapha, PhD**
Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

**Xiaofang Zhou, PhD**
Professor
University of Queensland
Australia
(External Examiner)



**RUSLI HAJI ABDULLAH, PhD**
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 30 July 2018

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 Binti Ibrahim, PhD**
Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

**Nur Izura Binti Udzir, PhD**
Associated Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

**Nor Asilah Wati Binti Abdul Hamid, PhD**
Senior Lecturer,
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

**ROBIAH BINTI YUNUS, PhD**
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

© COPYRIGHT UPM

**Declaration by Graduate Student**

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other
- degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned
- by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia
- (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia(Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: _____ Date: _____

Name and Matric No.: Waheed Yasin Mohammed GS26835

© COPYRIGHT UPM

**Declaration by Members of Supervisory Committee**

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: _____
Name of Chairman of Supervisory Committee: Prof. Dr. Hamidah Binti Ibrahim

Signature: _____
Name of Member of Supervisory Committee: Assoc. Prof. Dr. Nur Izura Binti Udzir

Signature: _____
Name of Member of Supervisory Committee: Dr. Nor Asilah Wati Binti Abdul Hamid

© COPYRIGHT UPM

**TABLE OF CONTENTS**





© COPYRIGHT UPM

© COPYRIGHT UPM

© COPYRIGHT UPM

## LIST OF TABLES





© COPYRIGHT UPM

© COPYRIGHT UPM

## LIST OF FIGURES



© COPYRIGHT UPM

© COPYRIGHT UPM

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ABC | Artificial Bee Colony |
| ACHP | Average Cache Hit Probability |
| AODE | Averaged One-Dependence Estimators |
| AP | Access Point |
| APS | Adaptive Path Selection |
| BHR | Byte Hit Ratio |
| BPNN | Back Propagation Neural Network |
| BSR | Bandwidth to Space Ratio |
| CART | Classification and Regression Trees |
| CBC | Content-Based Centrality |
| CBR | Case-Based Reasoning |
| CCN | Content Centric Network |
| CCR | Correct Classification Rate |
| CDN | Content Delivery/Distribution Network |
| CR | Communication Rate |
| CT | Cost Throughput |
| CPU | Central Processing Unit |
| DA | Dynamic Aging |
| DDoS | Distributed Denial of Service |
| DDSB | Data Distribution Service Bridge |
| DHT | Distributed Hash Table |
| DWT | Discrete wavelet transform |
| ECHO | Efficient Complex Query over DHT |
| EPC | Evolved Packet Core |
| FIFO | First In First Out |
| GA | Genetic Algorithm |
| GDS | Greedy Dual Size |
| GDSF | Greedy Dual Size Frequency |
| GFRT-Chord | Grouped Flexible Routing Tables Chord |
| HR | Hit Ratio |
| HRR | Hop Reduction Ratio |
| HTML | HyperText Markup Language |
| HTTP | Hypertext Transfer Protocol |
| IC-C4.5 | Intelligent Collaborative C4.5 |
| IC-C4.5-LFU | Intelligent Collaborative C4.5 Least Frequently Used |
| IC-C4.5-LRU | Intelligent Collaborative C4.5 Least Recently Used |
| ICN | Information Centric Networking |
| IC-NB | Intelligent Collaborative NB |
| IC-NB-LFU | Intelligent Collaborative Naïve Bayes Least Frequently Used |
| IC-NB-LRU | Intelligent Collaborative Naïve Bayes Least Recently Used |
| IP | Internet Protocol |
| ISO | International Organization for Standardization |
| ISP | Internet Service Provider |
| LAN | Local Area Network |
| LFU | Least Frequently Used |
| LRU | Least Recently Used |
| LTE | Long-Term Evolution |



© COPYRIGHT UPM

| | |
|---|---|
| M-DART | Multipath Dynamic Address Routing |
| MANET | Mobile Ad hoc Network |
| MARS | Multivariate Adaptive Regression Splines |
| MDC | Multiple Description Coding |
| MLP | Multilayer Perceptron |
| mSCTP | Mobile Stream Control Transmission Protocol |
| NB | Naïve Bayes |
| NNPCR | Neural Network Proxy Cache Replacement |
| OSI | Open System Interconnection model |
| POD | Peer Connection over Data |
| PRISM | Portal Infrastructure for Media objects caching Media |
| PSO | Particle Swarm Optimization |
| PST | Path Selectivity Table |
| QoE | Quality of Experience |
| QoS | Quality of Service |
| RF | Random Forest |
| RAM | Random Access Memory |
| RAN | Radio Access Network |
| RAND | Random |
| SIEMS | Security Information Event Management System |
| SPRON | Scalable Private Reappearing Overlay Network |
| SVM | Support Vector Machine |
| TN | TreeNet |
| TANB | Tree Augmented Naive Bayes |
| UDN | Ultra Dense Network |
| URL | Uniform Resource Locator |
| VCR | Video Cassette Recording |
| VFDT | Very Fast Decision Tree |
| WAN | Wide Area Network |
| WebRTC | Web Real Time Communication |
| WORM | Write-Once-Read-Many |
| WWW | World Wide Web |
| WS-Chord | Web Service-Chord |

© COPYRIGHT UPM

# CHAPTER 1

# INTRODUCTION

## 1.1 Overview

Web caching is a technique where local copies of a web page are stored in places close to the end-users. Web caching plays a key role in delivering web items to end users in World Wide Web (WWW). Web caches are used to improve the performance of the WWW. Many benefits can be gathered from caching such as improving the hit rates, alleviating loads on origin servers, and reducing network traffic.

Caching replacement policies is the core of web caching which is the procedure that has to be taken when the cache is full while there are web objects that have to be cached. Thus, it is common that a web caching policy is defined according to the cache replacement algorithm. The web object is divided into two types, namely: text object and media object. This work focuses on media objects because of their huge sizes compared to text objects. In this work, the term object and the term item are used interchangeably.

Collaborative caching offers an efficient utilization of cache storage of neighbour peers that are joining a network by sharing web objects. Furthermore, collaborative caching reduces the latency that is required to answer a query, because the query is answered from a neighbour cache rather than fetching it from the origin server. Also, collaborative caching increases Hit Ratio (HR) and Byte Hit Ratio (BHR) (Hara et al., 2010).

Many challenges have to be considered when collaborative caching is applied in the system such as the cache replacement decision that must to be taken when there is no enough space for new coming objects. This replacement is based on entry time, location, expiration time of the objects in peers' caches, frequency, and last access. Furthermore, patterns of user requests, heterogeneous cache sizes, and network topology are other challenges of collaborative caching (Xiuhua et al., 2017).

The collaborative approaches that are presented in this work are based on machine learning that provide algorithms to deal with information from data, where humans are not able to deal with because of its complexity level or the size of data to be treated per time unit. Furthermore, the proposed collaborative approaches share information about peers caches' contents in order to enhance the traditional web caching policies that are Least Recently Used (LRU) and Least Frequently Used (LFU). Many variants of LRU and LFU have been presented in the literature (Podlipnig, 2003) such as SVM-LRU, NB-LRU, (Ali et al., 2012a; Ali et al., 2012b), LFU-Aging and LFU-DA (ElAarag,



© COPYRIGHT UPM

2013; Zink & Shenoy, 2005). However, these approaches do not consider the contents of neighbours' caches.

In this work, two supervised machine learning algorithms are implemented, namely: the C4.5 decision tree and Naïve Bayes (NB) supervised machine learning algorithms because they are popular supervised machine learning algorithms that have been applied successfully in many domains such as medical, military, forecasting, modelling, control, and computer science (Ali et al., 2011; Sulaiman et al., 2008; Darwiche, 2010; de Melo & Sanchez, 2008; Goubanova & King, 2008). Moreover, NB and C4.5 machine learning algorithms are considered simple classifiers because of the independent assumptions among features. Also, they are more effective compared to other more sophisticated classifiers. In this work, NB and C4.5 classifiers are incorporated effectively with traditional web caching policies. More details are presented in Section 4.3.

A peer-to-peer system is a system in which interconnected peers share resources amongst each other without the use of a centralized administrative system. There are two different ways for configuring a peer-to-peer system that are structured peer-to-peer systems and unstructured peer-to-peer systems.

Many challenges have to be taken into account when a peer-to-peer system has to be applied. For example, the communication overhead between peers, searching mechanism, and the impact of peer's joining/leaving the overly network (Guoqiang et al., 2014). Furthermore, the Quality of Service (QoS) of the cached objects in terms of the delay which is taken to answer the query from peers' caches has to be considered. Moreover, updating and maintaining the cached objects in peers is considered as a challenge of caching in peer-to-peer system. Also, if a problem happened in a signal power, it might lead to packet loss or errors.
Mobility is one of the common characteristics in peer-to-peer system, where a peer can move at any time during its joining the overly network. This movement might cause unpredictable results. This issue has been addressed in (Elfaki et al., 2014).

## 1.2 Problem Statement

Utilizing the bandwidth of a network is considered as an important aim for network administrators (Jelenković & Radovanović, 2009). Caching policies such as LRU and LFU play a key role in utilizing the network bandwidth by delivering web items to end users in WWW. However, these caching policies have some limitations such as making a cache replacement decision by a node without depending on other nodes, which results in many nodes cache the same data that means wastes the cache storage in caching duplicated web items.

Recent studies show that media objects are the dominant form of the traffic on the Internet. For example, YouTube and Netflix consume 20-30% of the traffic on the



© COPYRIGHT UPM

Internet (Pouya et al., 2016). Thus, a cache size has to be considered when caching is required. Furthermore, retrieving the same media object from the origin server many times consumes the network bandwidth (Kyounghwan et al., 2004; Radhamani et al., 2010; Shoushou et al., 2015; Xiuhua et al., 2017), which might lead to the following:

- More delays for web users due to the increase in traffic on the Internet (Kumar & Norris, 2008; Kumar, 2009).
- Push network administrators to extend network uplinks (outgoing communication links) at a rate which is out of proportion to the subscribers' growth, or push them to increase the cache storage size, which costs a lot of money (Abhari et al., 2006). These extensions of bandwidth and cache size are immediately consumed without reaching the expected improvement of network performance, which results in web user dissatisfaction.

Moreover, traditional web caching policies such as Least Recently Used (LRU) and Least Frequently Used (LFU) suffer from caching pollution (i.e. media objects that are stored in the cache are not frequently visited) which negatively affects on the performance of web proxy caching (Arlitt et al., 2000; Cherkasova & Ciardo, 2001; Karlsson, 2005; Koskela et al., 2003). Many variants of traditional web caching policies have performed investigations on the problem of caching pollution such as Support Vector Machine-Least Recently Used (SVM-LRU), Naïve Bayes-Least Recently Used (NB-LRU), Support Vector Machine-Least Frequently Used (SVM-LFU), Naïve Bayes- Least Frequently Used (NB-LFU), Very Fast Decision Tree-Least Recently Used (VFDT-LRU), Very Fast Decision Tree-Greedy Dual Size Frequency (VFDT-GDSF), Tree Augmented Nave Bayes- Least Recently Used (TANB-LRU), and Tree Augmented Naïve Bayes-Greedy Dual Size Frequency (TANB-GDSF) (Ali et al., 2012a,b; Julian et al., 2014; Julian & Sagayaraj, 2015); however, they do not take into account the contents of neighbours' caches in the network. As a result of that, repeated data in peers' caches that wastes the cache storage. On other hand, in the work of Raghee and Sajeev (2015), the caching pollution in peer-to-peer systems has been addressed where an approach based on C4.5 machine learning algorithm has been presented that is called Intelligent Greedy Dual Size (IGDS).

Collaborative caching is a kind of caching where a cache replacement decision is made by a node depending on other nodes which enables resources utilization by caching web objects and sharing them with other neighbour nodes. Thus, it does not result in caching the same data in nodes' caches, thus wasting the cache storage in caching duplicated media objects.

There are a lot of works that are based on collaborative caching to investigate the above issues such as the work that has been presented in (Elfaki et al., 2014) and the work of Mushtaq & Ahmed(2008). The cache space allocation is the main issue in collaborative caching. Furthermore, in heterogeneous environment, caching multiple bitrate versions of the media objects incurs high overhead in terms of cache storage (Tuyen et al., 2017). However, the policies that are presented in this work take the advantages of collaborative approach and machine learning algorithms as presented in more details in Chapter 4. The performance metrics including HR, BHR, and Cost Throughput (CT) are presented in Section 3.3.



© COPYRIGHT UPM

The problem which is tackled in this work can be divided in two main problems that are: (i) The problem of caching pollution which leads in wasting cache storage in storing infrequent objects. (ii) The problem of wasting network bandwidth in retrieving media objects which leads in insufficient utilizing of network bandwidth. The problem statement is illustrated in Figure 1.1.

## 1.3 Research Objectives

The research objective of this work is as follows:

- To propose collaborative web caching policies that have better performance compared to existing web caching policies in terms of HR, BHR that reduce caching pollution.
- To reduce the network bandwidth consumption by reducing the CT to fetch media objects based on NB and C4.5 supervised machine learning algorithms in peer-to-peer systems. The performance of the proposed approaches are evaluated and compared to the traditional and state-of-the-art web proxy caching policies.

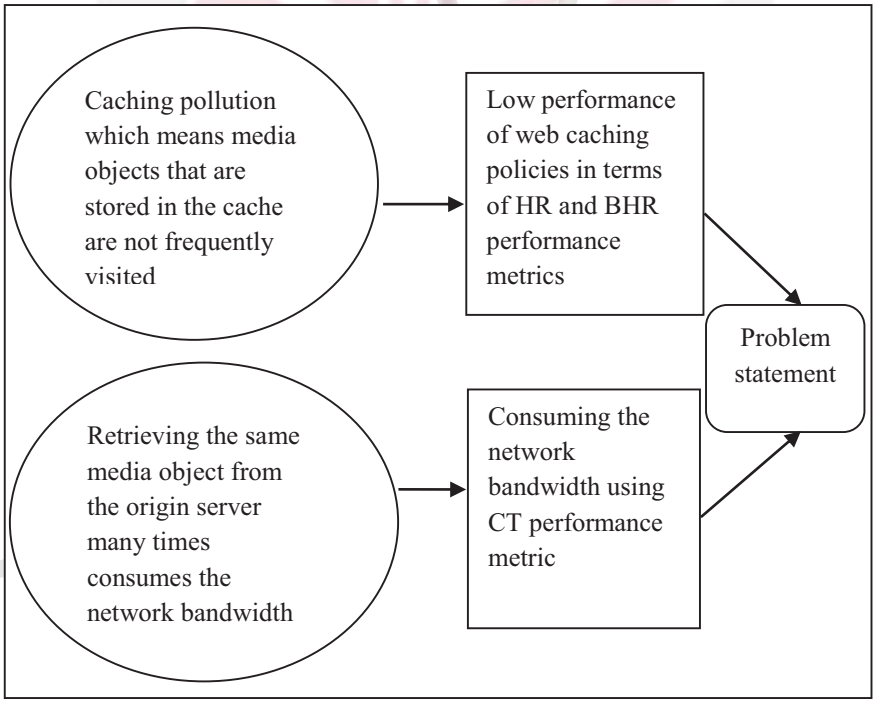



**Figure 1.1: Problem statement**



© COPYRIGHT UPM

## 1.4 Scope of the Research

A web page may include, among other things: texts, images, sounds, videos, data, applications, e-services, and animations. Web pages are normally updated and the problem of data consistency appears, however; media objects follow the principle of Write-Once-Read-Many (WORM) which makes cache consistency issues more resilient than those considered in caching other web objects. The scope of this research is limited to web caching for media objects because they have huge sizes compared to other objects such as text objects. Thus many researches have focused on the media object.

## 1.5 Contribution of the Thesis

This research has made the following contributions:

- It proposes collaborative web caching approaches for LRU and LFU based on C4.5 decision tree and NB supervised machine learning algorithms in order to utilize the network bandwidth by reducing network traffic caused by cache pollution. The proposed approaches are listed below:
    - Intelligent Collaborative Naïve Bayes Least Recently Used (IC-NB-LRU).
    - Intelligent Collaborative Naïve Bayes Least Frequently Used (IC-NB-LFU).
    - Intelligent Collaborative C4.5 Least Recently Used (IC-C4.5-LRU).
    - Intelligent Collaborative C4.5 Least Frequently Used (IC-C4.5-LFU).
- The proposed approaches take the advantages of structured peer-to-peer systems where peers caches' contents are shared using Distributed Hash Table (DHT) in order to enhance the performance of the web caching policy. The proposed approaches are evaluated using simulations.

## 1.6 Thesis Outline

Chapter 1 introduces the research study and the significance of this study. Furthermore, it states the problem that is tackled in this work. Also, this chapter presents the research objectives, scope of the research, contribution of thesis, and thesis outline.
The literature review is presented in Chapter 2 including a background on web caching including benefits of caching, caching limitations, caching stages, and traditional web caching policies. Also, in this chapter, media object caching is presented including media object characteristics, media object systems, and the advantages of caching media object. Also, this chapter introduces peer-to-peer systems, their categories, and collaborative caching based on peer-to-peer systems. Finally, this chapter provides the latest related works of the current state of the art techniques and comment on them.



© COPYRIGHT UPM

Research methodology is a systematic methods analyzed to accomplish the objectives of any research. Thus, in this work, Chapter 3 illustrates the research methodology design. Moreover, it presents the research methodology including the stages that are performed to conduct this work.

Chapter 4 presents the infrastructure of the media object caching in peer-to-peer systems. Furthermore, the proposed collaborative caching approaches using machine learning algorithms in peer-to-peer systems for media objects are presented in Chapter 4. In this work, two supervised machine learning algorithms are adopted namely NB and C4.5 are presented in this chapter.

Chapter 5 illustrates the results and discussion where the proposed approaches are compared to the conventional caching policies namely LRU and LFU caching policies. Furthermore, they are compared to the existing intelligent web caching approaches.

Chapter 6 provides the conclusion of this work by summarizing the findings and contributions of this work. It also provides a specific guidance of the directions of future work suggestions that are recommended in order to follow on from this work.

© COPYRIGHT UPM

## REFERENCES


Abdallah Khreishah & Jacob Chakareski (2015). Collaborative caching for multicell-coordinated systems, *Proceedings of the 2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Hong Kong, China, pp. 257-262. doi: 10.1109/INFCOMW.2015.7179394

Abhari, A., Dandamudi, S. P., & Majumdar, S. (2006). Web object-based storage management in proxy caches. *Future Generation Computer Systems, 22*: 16-31. doi:10.1016/j.future.2005.08.003

Acharya, Soam, Smith, & Brian (2000). MiddleMan: A video caching proxy server. *Proceedings of the 10th International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV 2000),* Chapel Hill, NC, USA. pp. 1-10. doi:10.1.1.113.5574

Aizaz Ullah (2015). mSCTP P2P (2-layered Chord) supported decentralized seamless mobility framework. *Proceedings of the International Conference on Emerging Technologies (ICET2015)*, Peshawar, Pakistan, pp. 1-5. doi: 10.1109/ICET.2015.7389214

Ali, W., and Shamsuddin, S. M. (2009). Intelligent client-side web caching scheme based on Least Recently Used algorithm and Neuro-Fuzzy system. In W. Yu, H. He & N. Zhang (Eds.), *Advances in Neural Networks* (pp. 70-79). Springer, Berlin / Heidelberg.

Ali, W., Shamsuddin, S. M., & Ismail, A. S. (2011). Patterns analysis and classification for web proxy cache. *Proceedings of the 11th International Conference on Hybrid Intelligent Systems (HIS),* Malacca, Malaysia. pp. 97-102. doi:10.1109/HIS.2011.6122087

Ali, W., Shamsuddin, S. M., & Ismail, A. S. (2012a). Intelligent naïve bayes-based approaches for web proxy caching. *Knowledge-Based Systems,* 31: 162-175. doi:10.1016/j.knosys.2012.02.015"

Ali, W., Shamsuddin, S. M., & Ismail, A. S. (2012b). Intelligent web proxy caching approaches based on machine learning techniques. *Decision Support Systems,* 53: 565-579. doi:10.1016/j.dss.2012.04.011"

Andreas Disterhoft & K. Graffi (2015). Protected chords in the web: secure P2P framework for decentralized online social networks. *Proceedings of the IEEE International Conference on Peer-to-Peer Computing (P2P2015)*, MA, USA, pp. 1-5. doi: 10.1109/P2P.2015.7328520

Arlitt, M., Friedrich, R., & Jin, T. (2000). Performance evaluation of web proxy cache replacement policies. *Performance Evaluation,* 39: 149-164. doi:10.1016/S0166-5316(99)00062-0




© COPYRIGHT UPM

Balaji E. & G. Gunasekaran (2016). Efficient range query processing, load balancing and fault tolerance with popular web cache in DHT. P*roceedings of the International Conference on Information Communication and Embedded Systems (ICICES2016)*, Chennai, India, pp. 1-4. doi: 10.1109/ICICES.2016.7518891

Benter M., M. Divband, S. Kniesburges, A. Koutsopoulos, & K. Graffi (2013). Ca-Re-Chord: A Churn Resistant Self-Stabilizing Chord Overlay Network. *Proceedings of the 2013 Conference on Networked Systems (NetSys2013)*, Stuttgart, Germany, pp. 27-34. doi: 10.1109/NetSys.2013.11

Bo Chen, Liang Liu, Zhao Zhang, Wenbo Yang, & Huadong Ma (2016). BRR-CVR: a collaborative caching strategy for information-centric wireless sensor networks, *Proceedings of the 12th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)*, Hefei, China, pp. 31-38. doi: 10.1109/MSN.2016.013

Candan, K., Akca, Y., & Lit, W. (2005). Policy-based resource sharing in streaming overlay networks. In X. Tang, J. Xu, S. T. Chanson & Y. Zhang (Eds.), *Web content delivery.* (pp. 215-244). Springer US. doi:10.1007/0-387-27727-7_10

Cao, P. & Irani, S. (1997). Cost-aware WWW proxy caching algorithms. *Proceedings of the USENIX Symposium on Internet Technologies and Systems on USENIX Symposium on Internet Technologies and Systems,* Monterey, California, USA. pp. 193-206. doi:10.1.1.30.7285

Carmela Comito, Domenico Talia, & Paolo Trunfio (2016). A distributed selectivity-driven search strategy for semi-structured data over DHT-based networks. *Journal of Parallel and Distributed Computing*, 93: 10-29. doi: 10.1016/j.jpdc.2016.03.015

Chae Youngsu, Guo, K., Buddhikot, M. M., Suri, S., & Zegura, E. W. (2002). Silo, rainbow, and caching token: Schemes for scalable, fault tolerant stream caching. *IEEE Journal on Selected Areas in Communications,* 20: 1328-1344. doi:10.1109/JSAC.2002.802062

Chen, R., & Hsieh, C. (2006). Web page classification based on a support vector machine using a weighted vote schema. *Expert Systems with Applications,* 31: 427-435. doi:10.1016/j.eswa.2005.09.079

Chen T. Min & L. T. Ming (2015). Investigate SPRON Convergence Time Using Aggressive Chord and Aggressive AP-Chord. *Proceedings of the 12th International Conference on Information Technology - New Generations (ITNG2015)*, NV, USA, pp. 61-66 doi: 10.1109/ITNG.2015.16

Chen Yan (2014). Distributed Hash Table Based Routing Algorithm for Wireless Sensor Networks. *Proceedings of the Fifth International Conference on Intelligent Systems Design and Engineering Applications (ISDEA2104)*, Hunan, China, pp. 430-433. doi: 10.1109/ISDEA.2014.103

Cherbal S., A. Boukerram, & A. Boubetra (2015). A survey of locality-awareness solutions in mobile DHT systems. *Proceedings of the 12th International*



© COPYRIGHT UPM

*Symposium on Programming and Systems (ISPS2015)*, Algiers, Algeria, pp. 1-7. doi: 10.1109/ISPS.2015.7244963

Cherbal, A. Boukerram, & A. Boubetra (2017). An improvement of mobile Chord protocol using locality awareness on top of cellular networks, *Proceedings of the 5th International Conference on Electrical Engineering - Boumerdes (ICEE-B)*, Boumerdes, Algeria, pp. 1-6 doi: 10.1109/ICEE-B.2017.8192071

Cherkasova, L., & Ciardo, G. (2001). Role of aging, frequency, and size in web cache replacement policies. In B. Hertzberger, A. Hoekstra, & R. Williams (Eds.), *High-Performance Computing and Networking* (pp. 114-123). Springer Berlin Heidelberg. doi:10.1007/3-540-48228-8_12

Cobb, J. & ElAarag, H. (2008). Web proxy cache replacement scheme based on back-propagation neural network. *Journal of Systems and Software,* 81: 1539-1558. doi:10.1016/j.jss.2007.10.024

Cranor, C. D., Green, M., Kalmanek, C., Shur, D., Sibal, S., Van der Merwe, J. E., & Sreenan, C. J. (2001). Enhanced streaming services in a content distribution network. *IEEE Internet Computing,* 5: 66-75. doi:10.1109/4236.939452

Dan, A. & Sitaram, D. (1995). An online video placement policy based on bandwidth to space ratio (BSR). *Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data,* San Jose, California, USA. pp. 376-385. doi:10.1145/223784.223853

Darwiche, A. (2010). Bayesian networks. *Communications of the ACM,* 53: 80-90. doi:10.1145/1859204.1859227

de Melo, A. C. V., & Sanchez, A. J. (2008). Software maintenance project delays prediction using bayesian networks. *Expert Systems with Applications,* 34: 908-919. doi:10.1016/j.eswa.2006.10.040

De-gan Zhang, W. b. Li, & Z. Ma (2015). New service discovery algorithm based on DHT for mobile application. *Proceedings of the 12th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON2015)*, WA, USA, pp. 145-147. doi: 10.1109/SAHCN.2015.7338301

Dhib E., N. Tabbane, N. Zangar, & K. Boussetta (2014). C-Chord: Hierarchical peer-to-peer protocol over a fully decentralized IaaS Cloud. *Proceedings of the International Conference and Workshop on the Network of the Future (NOF2014)*, Paris, France, pp. 1-5. doi: 10.1109/NOF.2014.7119776

Ding Z. m. & Q. Qian (2014). A chord-based load balancing algorithm for P2P network. *Proceedings of the 2014 International Conference on Information and Network Security (ICINS2014)*, Beijing, China, pp. 91-96. doi: 10.1049/cp.2014.1271

ElAarag, H. (2013). A quantitative study of web cache replacement strategies using simulation. In Setephan, Romano (Eds.), *Web Proxy Cache Replacement Strategies* (pp. 17-60). Springer London. doi:10.1007/978-1-4471-4893-7_4



© COPYRIGHT UPM

ElAarag, H., & Romano, S. (2009). Training of NNPCR-2: An improved neural network proxy cache replacement strategy. *Proceedings of the International Symposium on Performance Evaluation of Computer & Telecommunication Systems (SPECTS 2009)*, Istanbul, Turkey, pp. 260-267. doi:978-1-4244-4165-5

Elfaki Ahmed, M., Ibrahim, H., Mamat, A., Othman, M., & Safa, H. (2014). Collaborative caching priority for processing requests in MANETs. *Journal of Network and Computer Applications*, 40: 85-96. doi:http://dx.doi.org/10.1016/j.jnca.2013.08.013

Fan, L., Poh, K., & Zhou, P. (2009). A sequential feature extraction approach for naïve bayes classification of microarray data. *Expert Systems with Applications,* 36: 9919-9923. doi:10.1016/j.eswa.2009.01.075

Fengkai Yuan, Zhenzhou Ji, & Zhongchuan Fu (2017). RACMan: Replication-aware cache management for manycore CMPs with private LLCs. *Microprocessors and Microsystems*, 51: 165 – 175. doi: 10.1016/j.micpro.2017.04.015

Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning,* 29: 131-163. doi:10.1023/A:1007465528199

Gao Shen, Li Pei, Pan Zhiwen, Liu Nan, & You Xiaohu (2017). Machine Learning based Small Cell Cache Strategy for Ultra Dense Networks, *Proceedings of the 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, Nanjing, China, pp. 1-6. doi:10.1109/WCSP.2017.8170936

Gianluigi Folino, F. Pisani, & P. Trunfio (2014). Efficient discovery of data mining services over DHT-based overlays. *Proceedings of the International Conference on High Performance Computing Simulation (HPCS2014)*, Bologna, Italy, pp. 484-490. doi: 10.1109/HPCSim.2014.6903724

Gkantsidis, C. & Rodriguez, P. R. (2005). Network coding for large scale content distribution. *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies,* Miami, Florida, USA. pp. 2235-2245. doi:10.1109/TDSC.2011.32

Goubanova, O. & King, S. (2008). Bayesian networks for phone duration prediction. *Speech Communication,* 50: 301-311. doi:10.1016/j.specom.2007.10.002

Gunjae Yoon, Hoon Choi, & Won-Tae Kim (2014). A load balancing scheme using a chord algorithm for DDSB service. *Proceedings of the International Conference on Big Data and Smart Computing (BIGCOMP2014)*, Bangkok, Thailand, pp. 291-295. doi: 10.1109/BIGCOMP.2014.6741455

Guo, L., Chen, S., Ren, S., Chen, X., & Jiang, S. (2004). PROP: A scalable and reliable P2P assisted proxy streaming system. *Proceedings of the 24th International Conference on Distributed Computing Systems,* Tokyo, Japan. pp. 778-786. doi:10.1109/ICDCS.2004.1281646



© COPYRIGHT UPM

Guo, Y., Ge, Z., Urgaonkar, B., Shenoy, P., & Towsley, D. (2006). Dynamic cache reconfiguration strategies for cluster-based streaming proxy. *Computer Communications,* 29: 1710-1721. doi:10.1016/j.comcom.2005.09.017

Guoqiang Gao, Ruixuan Li, Heng He & Zhiyong Xu (2014). Distributed caching in unstructured peer-to-peer file sharing networks. *Computers & Electrical Engineering*,40: 688 - 703. doi: 10.1016/j.compeleceng.2013.12.001

Hara, T., Maeda, K., Ishi, Y., Uchida, W., & Nishio, S. (2010). Cooperative caching by clients constructing a peer-to-peer network for push-based broadcast. *Data & Knowledge Engineering,* 69: 229-247. doi:10.1016/j.datak.2009.10.011

Hiroya Nagao & S. Kazuyuki (2014). GFRT-chord: Flexible Structured Overlay Using Node Groups. *Proceedings of the IEEE 11th International Conference on Ubiquitous Intelligence and Computing, the IEEE 11th International Conference on Autonomic and Trusted Computing, and the IEEE 14th International Conference on Scalable Computing and Communications and Its Associated Workshops (UTC-ATC-ScalCom2014)*, Bali, Indonesia, pp. 187-195. doi: 10.1109/UIC-ATC-ScalCom.2014.69

Hu B., X. Zhang, & X. Zhang (2013). Automatic Balancing Chord: A Dynamic Load Balancing Mechanism for High Efficient. *Proceedings of the Fourth International Conference on Emerging Intelligent Data and Web Technologies (EIDWT2013)*, Xi'an, China, pp. 137-144. doi: 10.1109/EIDWT.2013.29

Isaac Woungang, F. Tseng, Y. Lin, L. Chou, H. Chao, & M. Obaidat (2015). MR-Chord: Improved Chord Lookup Performance in Structured Mobile P2P Networks. *IEEE Systems Journal*, 9: 743-751. doi: 10.1109/JSYST.2014.2306147

Jelenković P. R. & Radovanović A. (2009). Asymptotic optimality of the static frequency caching in the presence of correlated requests. *Operations Research Letters,* 37: 307-311. doi:10.1016/j.orl.2009.03.011

Jin, S. & Jiang, H. (2007). Novel approaches to efficient flooding search in peer-to-peer networks. *Computer Networks,* 51: 2818-2832. doi:10.1016/j.comnet.2006.12.002

Josyl Mariela B. Rocamora & Jhoanna Rhodette I. Pedrasa (2016). Evaluation of hierarchical DHTs to mitigate churn effects in mobile networks. *Computer Communications*, 85: 41-57. doi: 10.1016/j.comcom.2016.02.003

Julian Benadit , Sagayaraj Francis and U. Muruganantham (2014). Improving the Performance of a Proxy Cache Using Tree Augmented Naive Bayes Classifier. *Proceedings of the International Conference on Information and Communication Technologies (ICICT 2014)*, Kochi, India, pp. 184-193. doi: 10.1016/j.procs.2015.02.010

Julian Benadit & Sagayaraj Francis (2015). Improving the Performance of a Proxy Cache Using Very Fast Decision Tree Classifier. *Proceedings of the International*



© COPYRIGHT UPM

*Conference on Computer, Communication and Convergence (ICCC 2015)*, Odisha, India. pp. 304-312. doi: 10.1016/j.procs.2015.04.186

Junaid Ahmed Khan, Cedric Westphal, & Yacine Ghamri-Doudane (2017). A content-based centrality metric for collaborative caching in information-centric fogs, *Proceedings of the 2017 IFIP Networking Conference (IFIP Networking) and Workshops*, Stockholm, Sweden, pp. 1-6. doi: 10.23919/IFIPNetworking.2017.8264869

Jung Paik & Dong Lee (2015). Scalable signaling protocol for Web real-time communication based on a distributed hash table. *Computer Communications*, 70: 28-39. doi: 10.1016/j.comcom.2015.05.013

Karlsson, M. (2005). Replica placement and request routing. In Y. Zhang, X. Tang, J. Xu & S. T. Chanson (Eds.), *Web content delivery* (pp. 23-43). Springer USA. doi:10.1007/0-387-27727-7_2

Katti, S., Rahul, H., Wenjun Hu, Katabi, D., Medard, M., & Crowcroft, J. (2008). XORs in the air: Practical wireless network coding. *IEEE/ACM Transactions on Networking*, 16: 497-510. doi:10.1109/TNET.2008.923722

Kaur R., A. L. Sangal, & K. Kumar (2014). Analysis of different churn models in chord based overlay networks. *Proceedings of the Recent Advances in Engineering and Computational Sciences (RAECS2014)*, Chandigarh, India, pp. 1-6. doi: 10.1109/RAECS.2014.6799651

Kellerer, W., Kunzmann, G., Schollmeier, R., & Zöls, S. (2006). Structured peer-to-peer systems for telecommunications and mobile environments. *AEU - International Journal of Electronics and Communications,* 60: 25-29. doi:10.1016/j.aeue.2005.10.005

Koskela, T., Heikkonen, J., & Kaski, K. (2003). Web cache optimization with nonlinear model using object features. *Computer Networks,* 43: 805-817. doi:10.1016/S1389-1286(03)00334-7

Kumar, C. (2009). Performance evaluation for implementations of a network of proxy caches. *Decision Support Systems,* 46: 492-500. doi:10.1016/j.dss.2008.09.002

Kumar, C. & Norris, J. B. (2008). A new approach for a proxy-level web caching mechanism. *Decision Support Systems,* 46: 52-60. doi:10.1016/j.dss.2008.05.001

Kyounghwan An, Bonggi JunJietae, ChaBonghee Hong (2004). A log-based cache consistency control of spatial databases in mobile computing environments. *Proceedings of the International Conference on Database Systems for Advanced Applications (DASFAA 2004)*, Suzhou, China, pp. 630-641. doi: 10.1007/978-3-540-24571-1_57

Lei Shi, Jing Zhou, & Qi Huang (2013). A Chord-based super-node selection algorithm for load balancing in hybrid P2P networks. *Proceedings of the International Conference on Mechatronic Sciences, Electric Engineering and Computer*



© COPYRIGHT UPM

*(MEC2013)*, Shengyang, China, pp. 2090-2094. doi: 10.1109/MEC.2013.6885395

Li, S. R., Yeung, R. W., & Ning Cai. (2003). Linear network coding. *IEEE Transactions on Information Theory,* 49: 371-381. doi:10.1109/TIT.2002.807285

Li, X. & Ji, H. (2009). Scalable peer-to-peer resource discovering scheme for wireless self-organized networks. *The Journal of China Universities of Posts and Telecommunications,* 16: 47-52. doi:10.1016/S1005-8885(08)60200-9

Lu, S., Chiang, D., Keh, H., & Huang, H. (2010). Chinese text classification by the naïve bayes classifier and the associative classifier with multiple confidence threshold values. *Knowledge-Based Systems,* 23: 598-604. doi:10.1016/j.knosys.2010.04.004

Ming Wan, Ying Liu, Huachun Zhou, & Hongke Zhang (2010). A chord-based handoff authentication scheme under ID/locator separation architecture. *Proceedings of the International Conference on Advanced Intelligence and Awarenss Internet (AIAI2010)*, Beijing, China, pp. 309-314. doi: 10.1049/cp.2010.0776

Muhammad Ibrar, M. Ahmad, M. Umar, M. Habib, & M. Iqbal (2016). Stability analysis of DHT based multi-path routing protocol under group-based mobility models and entity-based mobility models in mobile ad-hoc networks. *Proceedings of the IEEE Information Technology, Networking, Electronic and Automation Control Conference*, Chongqing, China, pp. 777-783. doi: 10.1109/ITNEC.2016.7560467

Mushtaq, M. & Ahmed, T. (2008). P2P-based collaborative media streaming for heterogeneous network terminals. *Proceedings of the IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR 2008),* Cluj-Napoca, Romania. pp. 330-335. doi:10.1109/AQTR.2008.4588763

Nicolas Hidalgo, Luciana Arantes, Pierre Sens, & Xavier Bonnaire (2016). ECHO: Efficient complex query over DHT overlays. *Journal of Parallel and Distributed Computing*, 88: 31-45. doi: 10.1016/j.jpdc.2015.10.007

Nitin Varyani, N. S., & V. S. Shekhawat (2016). Latency and Routing Efficiency Based Metric for Performance Comparison of DHT Overlay Networks. *Proceedings of the 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA2016)*, Crans-Montana, Switzerland, pp.337-342. doi: 10.1109/WAINA.2016.165

Peng Wang, Julong Lan, Yuxiang Hu, & Shuqiao Chen (2015). Towards locality-aware DHT for fast mapping service in future Internet. *Computer Communications*, 66: 14-24. doi: 10.1016/j.comcom.2015.04.003

Podlipnig Stefan, L. B. (2003). A survey of web cache replacement strategies. *ACM Computing Surveys,* 35: 374-398. doi:10.1145/954339.954341

Poo Kuan Hoong, Ong Yen Shyang, & Tan, I. K. T. (2013). Red-black tree architecture for P2P media streaming. *Proceedings of the 10th IEEE 10*



© COPYRIGHT UPM

*Conference Region (TENCON 2013 -31194),* Xi'an, China. pp. 1-4. doi:10.1109/TENCON.2013.6719004

Pouya Ostovari, Jie Wu, & Abdallah Khreishah (2016). ). Scalable video streaming with helper nodes using random linear network coding. *Proceedings of the IEEE 13th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, Brasilia, Brazil, pp. 136-144. doi: 10.1109/MASS.2016.027

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. ISBN: 1-55860-238-0

Radhamani G., Umamaheswari S. (2010). Comparison of cooperative caching strategies in Mobile Ad-Hoc Network (MANET). *Recent Trends in Networks and Communications*, 90: 439-446. doi: org/10.1007/978-3-642-14493-6_45

Raghee Chandran M & G P Sajeev (2015). Intelligent pollution controlling mechanism for peer to peer caches, *Proceedings of the 7th International Conference on Computational Intelligence, Modelling and Simulation (CIMSim)*, Kuantan, Malaysia, pp. 141-146 doi: 10.1109/CIMSim.2015.24

Sam Romano, H. E. (2008). A quantitative study of recency and frequency based web cache replacement strategies. *Proceedings of the 11th Communications and Networking Simulation Symposium,* Ottawa, Canada. pp. 70-78. doi:10.1145/1400713.1400725

Shahrear Tanzil , William Hoiles, & Vikram Krishnamurthy (2017). Adaptive scheme for caching YouTube content in a cellular Network: machine learning approach. *IEEE Access*, 5: 5870-5881. doi: 10.1109/ACCESS.2017.2678990

Shoushou Ren, Tao Lin, Wei An, Yang Li, Yu Zhang, & Zhen Xu (2015). Collaborative EPC and RAN caching algorithms for LTE mobile networks. *Proceedings of the 2015 IEEE Global Communications Conference (GLOBECOM)*, San Diego, USA, pp. 1-6. doi: 10.1109/GLOCOM.2015.7417681

Song, J., Lei, G., & Xiaodong, Z. (2003). LightFlood: An efficient flooding scheme for file search in unstructured peer-to-peer system. *Proceedings of the International Conference on Parallel Processing,* Taiwan, China. pp. 627-635. doi:10.1109/ICPP.2003.1240631

Stephen V. Stehman (1997). Selecting and interpreting measures of thematic classification accuracy. Remote Sensing of Environment, 62: 77-89. doi: 10.1016/S0034-4257(97)00083-7

Stoica, I., Morris, R., Karger, D., Kaashoek, M. F., & Balakrishnan, H. (2001). Chord: A scalable peer-to-peer lookup service for internet applications. *ACM SIGCOMM Computer Communication Review,* 31: 149-160. doi:10.1145/964723.383071

Stojnic N., L. Probst, & H. Schuldt (2013). COMPASS - Optimized routing for efficient data access in mobile Chord-based P2P systems. *Proceedings of the IEEE 14th International Conference on Mobile Data Management (MDM2013)*, Milan, Italy, pp. 46-55. doi: 10.1109/MDM.2013.15



© COPYRIGHT UPM

Sulaiman, S., Shamsuddin, S. M., Forkan, F., & Abraham, A. (2008). Intelligent web caching using neurocomputing and particle swarm optimization algorithm. *Proceedings of the Second Asia International Conference on Modeling & Simulation (AICMS 08),* Kuala Lumpur, Malaysia. pp. 642-647. doi:10.1109/AMS.2008.40

Sulaiman S., S. M. Shamsuddin, and A. Abraham (2011). Intelligent Web caching using Adaptive Regression Trees, Splines, Random Forests and Tree Net. *Proceedings of the 3rd Conference on Data Mining and Optimization (DMO2011)*, Selangor, Malaysia, pp. 108-114. doi: 10.1109/DMO.2011.5976513

Sunho Lim, Yumin Lee, Jongpil Cheon, Manki Min, & Wei Wang (2014). User-defined consistency sensitive cache invalidation strategies for wireless data access. *Computer Communications*, 41: 55-66. doi: 10.1016/j.comcom.2013.12.006

Ting Li, Hong Ji, Jingqing Mei, Yi Li, & Chao Hu. (2009). Topology mismatch avoidable cross-layer protocol for P2P file discovery in MANETs. *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC 2009),* Budapest, Hungary. pp. 1-5. doi:10.1109/WCNC.2009.4917995

Tuyen X. Tran, Parul Pandey, Abolfazl Hajisami, & Dario Pompili (2017). Collaborative multi-bitrate video caching and processing in mobile-edge computing networks. . *Proceedings of the 13th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*, Jackson, USA, pp. 165-172. doi: 10.1109/WONS.2017.7888772

Weiyuan Li, Yang Li, Wei Wang, Yonghui Xin, & Yuemei Xu (2016). A collaborative caching scheme with network clustering and hash-routing in CCN, *Proceedings of the 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Valencia, Spain, pp. 1-7. doi: 10.1109/PIMRC.2016.7794975

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems,* 14: 1-37. doi:10.1007/s10115-007-0114-2

Xiong Wang, Jing Ren, Tong Tong, Rui Dai, Shizhong Xu, & Sheng Wang (2016). Towards Efficient and Lightweight Collaborative In-Network Caching for Content Centric Networks, *Proceedings of the 2016 IEEE Global Communications Conference (GLOBECOM)*, Washington, USA, pp. 1-7. doi: 10.1109/GLOCOM.2016.7842342

Xiuhua Li, Xiaofei Wang, Keqiu Li, & Victor C. M. Leung (2017). Collaborative hierarchical caching for traffic offloading in heterogeneous networks. *Proceedings of the 2017 IEEE International Conference on Communications (ICC)*, Paris, France, pp.1-6. doi: 10.1109/ICC.2017.7996705

Yasuhiro Ando, H. Nagao, T. Miyao, &K. Shudo (2014). FRT-2-Chord: A DHT supporting seamless transition between one-hop and multi-hop lookups with symmetric Routing Table. *Proceedings of the International Conference on*



© COPYRIGHT UPM

*Information Networking (ICOIN2014)*, Phuket, Thailand, pp. 170-175. doi: 10.1109/ICOIN.2014.6799686

Zink, M. & Shenoy, P. (2005). Caching and distribution issues for streaming content distribution networks. In Y. Zhang, X. Tang, J. Xu, & S. T. Chanson (Eds.), *Web content delivery* (pp. 246-263). Springer USA. doi:10.1007/0-387-27727-7_11



© COPYRIGHT UPM