

**A HYBRID PREFETCHING APPROACH FOR IMPROVEMENT OF
WEB LATENCY AND NETWORK TRAFFIC**

BY

Doliente, Bernil Ed G.

Mendoza, Francis Lou P.

Tolentino, Jasmine P.

**SCHOOL OF ARTS AND SCIENCES
ATENEO DE DAVAO UNIVERSITY**

MARCH 2002

**A HYBRID PREFETCHING APPROACH FOR IMPROVEMENT OF
WEB LATENCY AND NETWORK TRAFFIC**

A Independent Research

Presented to

The Faculty of the Computer Science Division

Ateneo de Davao University

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Science major in Computer Science

by

Doliente, Bernil Ed G.

Mendoza, Francis Lou P.

Tolentino, Jasmine P.

MARCH 2002

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF TABLES	viii
LIST OF FIGURES	ix
ABSTRACT	x
CHAPTER	
I. INTRODUCTION	1
1.1 Background of the Study	1
1.2 Statement of the Problem	2
1.3 Objectives	2
1.4 Scope and Limitation	3
1.5 Significance of the Study	3
II. REVIEW OF RELATED LITERATURE	4
2.1 Software Systems	4
2.2 Algorithms and Simulations	6
III. METHODOLOGY	11
IV. THEORETICAL BACKGROUND	14
4.1 Basic Prefetching Concepts	14
4.2 Basic Prefetching Techniques	18
4.3 Client-Proxy-Server Framework	22
4.4 Prediction Method	24
4.5 Discovery Techniques on Web Transactions	33
V. RESULTS AND DISCUSSION	34

5.1 Definition of a Selection Criteria	37
5.2 Selection of Two Prefetching Schemes	40
5.3 The Hybrid Prefetching Approach	42
5.4 Analysis of Unenhanced and Enhanced Implementation. .	45
VI. CONCLUSION AND RECOMMENDATIONS	48
APPENDICES	49
A-1. Prefetching Features Summary	49
A-2. Performance Evaluation of Prefetching Schemes ...	50
A-3. Advantages and Disadvantages of Prefetching Schemes. ...	52
B-1. Installation Guide	53
BIBLIOGRAPHY	54

ABSTRACT

The growth of the World Wide Web has emphasized the need for improved Web latency. Primarily two techniques, caching and prefetching are being used for improving Web latency. Caching is a mechanism used to increase download speeds by keeping frequently accessed information in a location close to the requester. But several studies show that the maximum hit ratio achievable by any caching algorithm is only 30% - 50%. Therefore, another effort preferred on further increasing the hit ratio is prefetching. Prefetching predicts future documents before clients request for them. The only drawback with prefetching is that it inherently impose heavy network traffic.

In this paper, the proponents developed a hybrid prefetching approach for improvement of Web latency and network traffic. It combined the strengths of the two existing prefetching on the web namely: Interactive Prefetching and Prefetching Using Association Rules. Trace driven simulation based on access logs from both the unenhanced proxy server (caching only) and the enhanced proxy server (with prefetching) were used to evaluate the new prefetching scheme. Performance results suggested that the hybrid prefetching approach improved Web performance by reducing access latency without increasing network traffic significantly.

CHAPTER I

INTRODUCTION

1.1 Background of the Study

The Internet has continued to grow since its popularity during the early 90's. It has become the primary source of information for different kinds of people. Students, businessmen and professionals from various fields all share this new revolution in computing for their several needs. But the Internet has its weakness from the very start: speed. Web latency has become a serious problem to Internet users that they referred to the Web as the *World Wide Wait*. There have been various efforts performed to improve the Web performance and reduce network traffic.

One is through caching. *Caching* is a proven technique for reducing the latency perceived by Web clients. The fundamental concept of caching is the intermediate storage of copies of popular Web documents close to Web clients. Caching is effective because many Web documents are requested more than once, and it can also reduce network traffic and server load. But several papers reported that the hit-rate of the caching proxy server is only 30%-50%.

Another is through prefetching. *Prefetching* is an alternative way to reduce Web latency where it complements the work of caching as well as further increasing the hit ratio. It does so by anticipating future document requests and then prefetches these documents in the local cache before actual users request for them. But prefetching has its drawback. It increases hit ratio up to 60%-70% but it also adds up to the network load and therefore induce network traffic.

In this research, the proponents presented a new approach to prefetching where it compromised the strengths of two existing prefetching methods on the Web, namely: Interactive Prefetching by Ken-ichi Chinen and Suguru Yamaguchi; and Prefetching Using Association Rules by Ajay Bhushan Pandey et al. The result was a hybrid prefetching approach where it increased the hit ratio as well as maintained a light traffic on the network.

1.2 Statement of the Problem

With the prolific popularity of the Web, Internet users are being faced with a serious problem: Web latency. Caching is a popular solution used in WWW proxy servers but its hit-rate is only 30%-50%. Prefetching is an emerging technique that complements the work of caching but still has its drawback. There have been prefetching methods developed that would increase the hit ratio but would impose a heavy traffic on the network. Therefore, this research aimed to answer the problem: *“How can Web latency be addressed by using prefetching but without causing heavy network traffic?”*

1.3 Objectives of the Study

The general objective of this research was to develop a hybrid prefetching approach for improvement of Web latency and network traffic.

In order to realize this general objective, the following specific objectives was first addressed:

- To identify previous prefetching techniques
- To evaluate the advantages and disadvantages of each technique

- To create a framework for the hybrid prefetching approach
- To implement the said hybrid prefetching approach on a proxy server
- To test and compare the performance of the enhanced proxy server with hybrid prefetching against the unenhanced proxy server without hybrid prefetching

1.4 Scope and Limitation of the Study

The study generally focused on two prefetching mechanisms that existed on the Internet, namely: Interactive Prefetching and Prefetching Using Association Rules. The result was a new prefetching approach implemented on a proxy server that better boosted its prefetching performance.

The limitation of the study was that it was only implemented on the Ateneo de Davao University (AdDU) Computer Laboratory A (Lab A). The trace-driven-simulation that was used to come up with results of the new prefetching method was only limited within the scope of the trace logs of the said domain.

1.5 Significance of the Study

The proponents strongly believed that it was relevant to conduct this study since it will greatly benefit the whole Internet population. By developing a new prefetching approach that will minimize Web latency and network traffic, Internet users could access and view more pages without getting into the hassle of waiting several seconds or even minutes for a Web page to download.