An Improvised Dynamic and Semantic based Web Cache Replacement Policy

K Geetha\textsuperscript{a} and N Ammasai Gounden\textsuperscript{b}

\textsuperscript{a}Department of Computer Science and Engineering, National Institute of Technology, Tiruchirappalli-15, Tamilnadu, India 620015, Contact: geethavalavan@yahoo.com

\textsuperscript{b}Department of Electrical and Electronics Engineering, National Institute of Technology, Tiruchirappalli-15, Tamilnadu, India 620015

This paper proposes a web cache replacement policy based on semantic content of the pages cached at the client side. Two models namely Clustered Model (CM) and Relational Model (RM) are proposed that focus on the Dynamicity which refers to the dynamic nature of the content and the Semantic content which exhibits the relation of information available among cached web pages and hence the name DynaSem. The proposed policy marks the page for eviction prioritized by Eviction Index (EI) in CM and Relation Index (RI) in RM. CM uses an interface with a web browser incorporated into it. The Trie data structure that enables the searching process to be more efficient has been framed to store the well-known categories of cached content as clusters. Pages with highest EI are marked for eviction. RM employs a technique to reveal the relation among cached documents. It evicts documents that are less related (minimum RI) to an incoming document which needs to be stored in the cache to ensure that only related documents are cached; hence the contents of the cache represent the documents of interest to the user and those which are of more static in nature. The proposed policy has been developed to incorporate two algorithms- one to find the dynamic count of the given web page \textquote{P} and the other to find semantic relation between the pages cached. Both the models (CM and RM) are used to establish the semantic relation. The policy has been simulated under model driven simulation with the help of an input set consisting of a few web pages. The parameters pertinent to cache replacement algorithms are computed and the result shows there is a factual improvement compared to the original semantic based policies.

\textbf{Keywords}: Web caching, Replacement Policies, Eviction, Semantic Relation, Dynamism.

1. INTRODUCTION

Web caching is the process of storing the frequently accessed web pages or documents. The increasing demand for web services insisted the need for web caching that can indubitably reduce the Internet traffic, download time, network bandwidth usage, server load, and perceived lag. Cache being a limited resource in terms of size, becomes saturated quite frequently and hence eviction has to be made often. Especially in wireless network, size of the client cache at mobile terminal is very small that demands frequent replacement. The state of art dictates multitude policies based on recency, frequency, size, and combination of the above parameters as some function.

A cache server stores web objects (e.g., HTML pages, images, and files) locally for the use of future requests to those objects. As cache size is finite, a cache replacement policy is needed to manage cache content. If a cache is full when an object needs to be stored, the policy will determine which object is to be evicted to make room for the new object.

However, in practical implementation, a replacement policy usually takes place before the cache is really full. The cache uses two water marks, high and low, to guide the replacement process. If the size of total cached objects exceeds the high watermark, the policy will evict objects until the low watermark is reached. The advantage of doing this is reducing the overhead of invoking the policy on demand. The goal of the replacement policy is to make the best use of available resources, including disk space, processing power, server load, and network bandwidth. The increasing use of the internet and its emerging appli-
naSem policy supersedes the other policies in all the above metrics.

4. CONCLUSIONS AND FUTURE WORK

This work aims at improvising the semantic based web cache replacement policy by considering the level of dynamism among the pages possessing the same relation. A modified policy termed ‘DynaSem’ has been developed. A formal framework for the DynaSem policy has been designed that incorporates two logics - dynamic and semantic relations. Rating the document based on the dynamic count leads to conservative improvement in most of the performance metrics used to evaluate the cache replacement policy. The semantic relation has been established using clustered model and relational model. Both models aim at identifying the dependencies among the cached document and the new incoming document. The request set is generated by a simulation process and the clustered model can be chosen, if the user access pattern is going to explore all the possible sub links provided in a web site.

Relational model can be preferred if the user switches between various web sites for related information. Though both models experience same time complexity, better results can be obtained in CM inspite of its increased space complexity. Using model driven simulation, the performance of DynaSem policy has been analyzed for related and unrelated request sets. Even if the user access pattern is not related, this policy has not deteriorated much from other policies namely LRU, SEMANTIC and SEMALRU. Ranking the documents based on their dynamism shows commendable results and hence it can be used as a vital parameter for tuning the performance of all prevalent replacement strategies that ignore file relations and communication overhead. An increasingly important technique to enhance the web caching performance is to prefetch web pages. Prefetching can happen in a predictive manner or in an interactive manner. For predictive prefetching, the proposed policy can be used to predict the reference probability of new requests after analyzing the user access pattern. The experiments conducted in the proposed work are restricted to only isolated cache. A possible extension to this work could be to experiment with a grid of cooperating caches. This policy tackles single user interest and can be extended to satisfy group of users. Instead of trie structure used in CM, standard vector model that is widely employed in search engines for information retrieval can also be adopted.

ACKNOWLEDGEMENT

The authors would like to thank Mr. N Rama-subramanian, Associate Professor, Department of Computer Science and Engineering, National Institute of Technology, Tiruchirappalli for permitting to conduct the simulation work in his lab and for his valuable suggestions in making this work possible.

REFERENCES

6. K Psounis and Balaji Prabhakar. A Randomized
An Improvised Dynamic and Semantic based Web Cache Replacement Policy


K Geetha was born in Tiruchirappalli, Tamilnadu, India, on June 24, 1972. She received the B.E degree in Computer Science and Engineering from Regional Engineering College, Tiruchirappalli, India in 1996 and M.Tech. in computer science from the National Institute of Technology(formerly REC), Tiruchirappalli, India in 2003. She is currently pursuing her doctoral programme in the area of Web caching and its applications. She worked as Research Associate in the National Institute of Technology, Tiruchirappalli, India from 2000 to 2006 and taught courses on Computer Architecture, Computer Networking and Digital Systems and Microprocessors for graduate and undergraduate students. Since 2009 she is working as an Assistant professor in the Department of Computer Science and Engineering, Bharathidasan University, Tiruchirappalli, India. She has also guided projects related to network security, web technology and computer architecture.

N Ammasai Gounden was born in Coimbatore, TamilNadu, India, on October 5, 1955. He received the B.E degree from the College of Engineering, Guindy India,(Madras University)in 1978 and the M.E. degree in control systems from P.S.G.College of Technology, Coimbatore, India (Madras University) in 1980. He received the Ph.D. degree from the Bharathidasan University, Tiruchirappalli, India, in 1990. Currently, he is a professor with the Department of Electrical and Electronics Engineering, National Institute of Technology, Tiruchirappalli where he has been working since 1982. His areas of interest are power electronic applications in renewable energy systems, energy conversion, use of programmable digital controllers in hybrid renewable systems, pattern recognition and memory architecture.