

CSC2231 – Internet Systems and Services

Paper Review – On the Scale and Performance of Cooperative Web Proxy Caching

Name: Alex Wun

Date: Oct. 19th, 05

Using one week long traffic traces obtained at the University of Washington and the Microsoft Corporation, the authors perform simulations to determine the benefits of web caching. Wolman et al. found that cache hit rates rise quickly for the first ~2500 clients of a proxy cache before diminishing returns take effect. They also found that cooperative caching has significant benefits across medium organizations but not across very large organizations – since unpopular content remains globally unpopular. Finally, through extending their traces with models, the authors show that cache hits for very large client populations can dominate cache misses caused by rapidly changing (or updating) web content.

This paper provides a very useful study on the effects of web caching. By performing simulations using an infinite resource server and perfect cacheability, the authors are able to characterize the behaviour of “ideal” caching. Wolman et al. are thorough in their approach to provide answers to interesting questions: How does different groupings of clients affect hit rates? How much does “realistic” cacheability differ from perfect cacheability?

Although the analytic work in the second part of the paper is also interesting, it is less convincing due to the assumptions and simplifications that are made (but such is the nature of all analytic models). Particularly, the use of exponential distributions to model the time between client requests and the time between document changes is never justified. I suspect it is to make the distributions memoryless, although the authors don’t explicitly state this.

The most interesting result from this paper is the identification of the “knee” in the Request Rate vs. Population curve, which occurs at roughly 2500 clients. This is a surprisingly low number of clients for diminishing returns to start taking effect. As the authors note, a single proxy cache can easily handle this many clients. This *turning point* is likely the result of the Zipf-like distribution of web page popularity - since requests for unpopular content become significant as population size increases. The implications of this low *turning point* seem to be that while caching is the de facto solution for improving performance, there are significant limits to its efficacy. It would be interesting to perform similar studies on the limits of caching solutions (such as in computer hardware).