



Chapter XIII

Web Caching: A Survey

Antonis Danalis and Evangelos Markatos
Institute of Computer Science—FORTH, Greece

World Wide Web traffic increases at exponential rates saturating network links and web servers. By replicating popular web pages in strategic places on the Internet, web caching reduces core network traffic, reduces web server load, and improves the end-users' perceived quality of service. In this paper we survey the area of web caching. We identify major research challenges and their solutions, as well as several commercial products that are being widely used.

CACHING METHODOLOGY

Introduction

Web caches should be placed at strategic positions on the Web in order to maximize their effectiveness. Depending on the desired result, a web cache should be deployed close to the server, on the gateway of a local area network, or at the client side.

A cache deployed close to the web server (or even in the same host) would accelerate the server's performance. A cache deployed in a strategic place in a network (LAN or WAN), such as a gateway, would serve clients by bringing data closer to them and would reduce network traffic. Finally, a cache may be at the client's host as a part of a browser, serving a single end-user.

Client-Side Caching

Most users tend to access pages they have accessed in the near past. For example, they use the "back" button of browsers quite often to access a page that has been loaded a few minutes ago. In addition, most users frequently access some

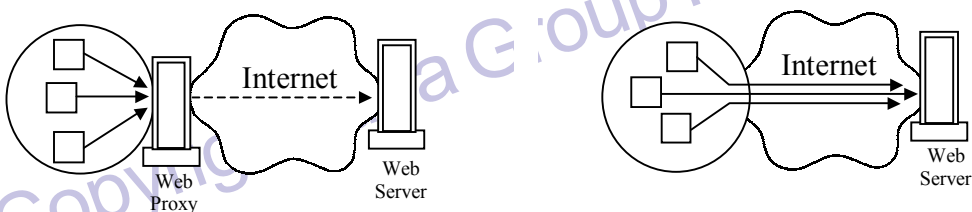
specific websites (like the default home page they have set on the browser's settings) very often. To exploit this temporal locality in users' requests, most browsers keep recently accessed URLs in main memory and possibly in the local disk. Thus, the browser serves some requests from its local cache, saving network bandwidth and giving the user low-latency responses. Although it is effective in reducing client latency, client-side caching has the least benefits in reducing bandwidth costs because the amount of available disk space used as a cache is usually quite small, and there is no sharing between different user caches.

Proxy Caches

In addition to client-side caches that serve single users, enterprise caching systems may be installed to serve an entire local or wide area network. Non-transparent caches, also called proxies, are dedicated computer systems that filter user requests within a network. All the web requests of the clients are served by the proxy instead of the original content provider. When a client requests an object that is in the cache (hit), the proxy serves it without the intervention of the original server. When a miss occurs, the proxy forwards the request to the server, serves the client with the reply and possibly stores the returned objects. Web proxies can be either a software application (like, CERN and Squid) running on top of a traditional operating system (like, UNIX or Windows) or a caching appliance (like, NetCache and Imimic DataReactor) that integrates caching software with a hardware platform and a proprietary operating system. Traditional proxy caches serve only the users that have explicitly set their browser to communicate with the proxy. These non-transparent proxies are difficult to deploy since all the end-users must change the appropriate attributes of their browsers to take advantage of the proxy benefits.

To eliminate the necessity of manual browser setup, required by traditional caches, a proxy can operate transparently. To do so, a transparent proxy cooperates closely with the gateway that intercepts all the outgoing TCP/IP packets of the local area network. The gateway recognizes web traffic (i.e., TCP/IP packets directed to port 80) by examining all the outgoing TCP/IP packets. Every network flow recognized as a web transaction is forwarded to a traditional cache, and the reply is forwarded back to the client. Because the client (who ignores the presence of the proxy) expects a reply from the original content provider, the transparent proxy should make the packets of the reply appear to be sent by the original content provider. To be able to examine and manipulate the TCP packets, the proxy should

Figure 1: LAN with proxy (left) and LAN without proxy (right)



13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/web-caching-survey/18424

Related Content

Benefits of Information Technology Implementations for Supply Chain Management: An Explorative Study of Progressive Indian Companies

Prashant R. Nair (2011). *Enterprise Information Systems: Concepts, Methodologies, Tools and Applications* (pp. 596-616).

www.irma-international.org/chapter/benefits-information-technology-implementations-supply/48568

Using Knowledge-Based Intelligent Reasoning to Support Dynamic Equipment Diagnosis and Maintenance

Yin-Ho Yao, Gilbert Y.P. Lin Linand Amy J.C. Trappey (2006). *International Journal of Enterprise Information Systems* (pp. 17-29).

www.irma-international.org/article/using-knowledge-based-intelligent-reasoning/2094

Dynamic Performance and Stability Research of VMI-APIOBPCS in Apparel Industry Based on Control Theory

Xueli Zhan, Qin Zhangand Wenfeng Xie (2018). *International Journal of Enterprise Information Systems* (pp. 56-76).

www.irma-international.org/article/dynamic-performance-and-stability-research-of-vmi-apiobpcs-in-apparel-industry-based-on-control-theory/203039

Conducting Multi-Project Business Operations in SMEs and IS Support

Igor Vrecko, Anton Hauc, Vesna Cancerand Igor Perko (2011). *Enterprise Information Systems: Concepts, Methodologies, Tools and Applications* (pp. 1871-1901).

www.irma-international.org/chapter/conducting-multi-project-business-operations/48649

Maintaining Meaning of Information when it is Shared amongst Information Systems

Janet Aisbett, Shuxin Zhao, Greg Gibbonand C. N.G. Dampney (2006). *International Journal of Enterprise Information Systems* (pp. 63-77).

www.irma-international.org/article/maintaining-meaning-information-when-shared/2102