Patricia Seybold Group

E-Business Thought Leaders

Network Caching Guide

Optimizing Web Content Delivery

By Michael A. Goulde
March 1999

Inktomi®

Prepared for Inktomi Corporation

Table of Contents

Illustrations

Network Caching Guide

Optimizing Web Content Delivery

Executive Summary

Importance of the Web

The creation of the World Wide Web was one of the most important events in the history of information technology. Whether measured by amount of content reachable over the Web or by the number of users accessing that content, the Web continues growing at an incredible rate ten years after it was first invented. Estimates are that by 2001, the number of Web users will be greater than 175 million. One of the largest Internet Service Providers estimates that network traffic is doubling every 90 days as graphics, audio, and video become more common. The Web has penetrated households throughout the United States faster than any other technology in history, including the telephone, television, and VCR.

The Web's explosive growth has placed enormous stress on the capacity of the Internet, spurring massive investment on the part of carriers and network service providers in order to expand Internet capacity from the backbone all the way to homes and businesses. At the same time the Internet is becoming the network of choice for many types of business applications, spurring additional growth and increased load. The Internet is becoming a major, albeit distributed, center for commerce of all types. An increasing proportion of Internet traffic is in the form of business-to-business and business-to-consumer E-commerce transactions. Estimates of the economic impact of E-commerce over the next few years range from tens of billions to hundreds of billions of dollars. In any event, the Web is well on its way to becoming a major tool to people who conduct business all over the world. The expanding importance of the Web as a stimulant for world economic growth means that it must be a safe, reliable, predictable, and affordable place to do business.

Problems Caused by the Web's Success

The enormous success of the Web as a source of information and as a platform for E-commerce has led to serious problems. Users' expectations for a fruitful and enjoyable experience browsing the Web are often met with frustration and

annoyance because of slow performance and unreliable connections. Web sites take a long time to access and pages seemingly take forever to download. Cynics claim that WWW stands for "World Wide Wait." Many people despair as to the prospects for improvement in the foreseeable future. Companies using the Internet to conduct business need more reliable performance as well.

One of the major problems lies in the inherently inefficient nature of the Web. Every user seeking to view specific content must obtain it directly from the server that is the point of origin for that content. This is like everyone having to fly to Hollywood to see the latest movie. There is no distribution mechanism designed into the Web that is analogous to the system of movie theaters that place first-run movies in your hometown. Internet bandwidth is not an unlimited resource. All too often, users consume all available capacity. Since it is not possible to have dedicated, point-to-point bandwidth allocated to users, congestion is bound to occur. Special events, whether planned or unplanned, cause unpredictable peaks in bandwidth usage. As broadband access to home and work becomes more widespread, even greater pressure is placed on total available bandwidth. Added to bandwidth constraints is the fact that, as Internet use increases, servers are becoming overloaded and can't handle the load.

Simply adding more bandwidth to the Internet backbone isn't the solution. First, all the bandwidth added to date has barely kept pace with Internet growth. Second, whether the bandwidth is being added to the backbone that spans the globe or the last mile between an Internet Service Provider and a household or office, the cost of delivering more bandwidth is extremely high. Adding bandwidth also just tends to move the bottleneck, rather than eliminating it. Other solutions must be sought to provide improved performance and to make the medium more predictable for users. Otherwise the Web will not fully realize its potential as an information resource and a vehicle for global trade.

Caching Improves Performance

One way to improve performance and make it more predictable is to minimize the variability introduced by long trips across the Internet from the content source to the user's browser. By storing frequently accessed content at a location closer to the user, a great deal of latency and unpredictable delay in the Internet can be eliminated. The technique for doing this is called caching. Caching is a means of storing content objects from a Web server closer to the user, where they can be retrieved more quickly. The storehouse of objects, including text pages, images, and other content, is called a Web cache (pronounced cash).

Caching has long been used in computer design. Data that are frequently used by the central processing unit (CPU) are stored in very fast memory, sometimes right on the CPU chip, thereby reducing the latency for reading those data and reducing the frequency with which a CPU has to read data from a slower disk drive.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.