An Analysis of Issues Facing World Wide Web Servers

A Thesis presented

by

Stephen Lee Manley

to

Computer Science

in partial fulfillment of the honors requirements

for the degree of

Bachelor of Arts

Harvard College

Cambridge, Massachusetts

April 7, 1997

Abstract

The World Wide Web has captured the public's interest like no other computer application or tool. In response, businesses have attempted to capitalize on the Web's popularity. As a result propaganda, assumption, and unfounded theories have taken the place of facts, scientific analysis, and well-reasoned theories. As with all things, the Web's popularity comes with a price for the first time the computer industry must satisfy exponentially increasing demand. As the World Wide Web becomes the "World Wide Wait" and the Internet changes from the "Information Superhighway" to a "Giant Traffic Jam," the public demands improvements in Web performance.

The lack of cogent scientific analysis prevents true improvement in Web conditions. Nobody knows the source of the bottlenecks. Some assert that the server must be the problem. Others blame Internet congestion. Still others place the blame on modems or slow Local Area Networks. The Web's massive size and growth have made research difficult, but those same factors make such work indispensable.

This thesis examines issues facing the Web by focusing on traffic patterns on a variety of servers. The thesis presents a method of categorizing different Web site growth patterns. It then disproves the theory that CGI has become an important and varied tool on most Web sites. Most importantly, however, the thesis focuses on the source of latency on the Web. An in-depth examination of the data, leads to the conclusion that the server cannot be a primary source of latency on the World Wide Web.

The thesis then details the creation of a new realistic, self-configuring, scaling Web server benchmark. By using a site's Web server logs, the benchmark can create a model of the site's traffic. The model can be reduced by a series of abstractions, and scaled to predict future behavior. Finally, the thesis shows the benchmark models realistic Web server traffic, and can serve as a tool for scientific analysis of developments on the Web, and their effects on the server.

Table of Contents

Chapter 1

Introduction

Computer scientists have been looking at exponential growth patterns for years. Exponential growth in computing power has extended computer functionality in ways that few people thought possible. Only four years ago, an Intel 486, 33MHz processor was considered the top-of-the-line. Now, the Intel Pentium Pro runs at 200 MHz -- 6 times as fast. Four years ago, machines ran with 4 MB of RAM. Now, retailers sell standard PCs with 16 MB of RAM. The advances in computer power have made a desktop machine more powerful than the supercomputers of the past. Advances in technology enable someone to send text, pictures, movies, or sounds to another person thousands of miles away.

After years of enjoying exponential growth in computing power, the computer community now faces exponential growth in demand. The Internet and World Wide Web (WWW or Web) have captured the world's imagination like no other computer product or service. The Internet, nicknamed “The Information Superhighway” enables any two computers in the world to send information to one another. Popular World Wide Web browsers such as Netscape or Microsoft Internet Explorer give users a simple to use Graphical User Interface (GUI). With this simple but powerful interface, a user can download a file from another computer with only a click of the mouse. The simplicity of the Web promotes the construction of Web sites, groups of Web pages, for purposes ranging from Kellogg's advertising their cereals to high school and college students setting up their resumes. The WWW has enabled everybody to become a publisher, with the HyperText Markup Lan-

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.