



Chapter VI

Web Switching

Vishal Sharma
Metanoia, Inc., USA

Rakhi Sharma
Harvard University Extension School, USA

INTRODUCTION

What is Web Switching?

Web switching may be viewed as an optimized combination of networking functions, such as load balancing, bandwidth and traffic management, cache switching, and site-level security, that are implemented on a single device and are specifically designed to address the unique requirements of World Wide Web (or simply, Web) traffic; for example, efficient management of Web traffic, simplified website management, and fast, reliable, and error-free website operation. More precisely, “Web switches are network devices that can be used to construct a ‘user- and content-aware’ network infrastructure, which has the explicit objective of seamless and high-performance routing of specific user requests for specific Web content to the best server containing that content, irrespective of its physical location (Johnson, 1999a).” A web switch, therefore, is an extremely fast and “intelligent” LAN switch that switches traffic based on content, rather than just addresses (Passmore, 1999), and integrates the traffic management and control functions that have traditionally run on a number of separate devices. These include sophisticated URL load balancing capabilities via local and global server load balancing, bandwidth control, network address translation (NAT), packet filtering, cache redirection, and policies to manage and speed the flow of web traffic.

The Need for Web Switching

Over the past few years, enterprise websites have become mission-critical resources, as corporations use them not only to market and sell products but also to support internal operations, communicate with partners, and conduct real-time

This chapter appears in the book, *Enterprise Networking: Multilayer Switching and Applications* by Vasilis Theoharakis and Dimitrios Serpanos.

Copyright © 2002, Idea Group Publishing.

business transactions (Alteon Web Systems, 1999a). This, together with the increasing use of the Web for e-commerce (business-to-business e-commerce, for example, was more than \$43 billion in 1998, and is expected to reach \$1,300 billion by 2003 [Flanagan, 1999]) has had a great impact on the architecture of today's web data centers, and on the requirements for the underlying server and network infrastructure. Building a high availability, low latency, secure, and scalable web data center, for example, requires a number of web traffic management functions, which (because of the times at which the need for these functions arose and the rate at which the associated technologies matured) today reside on a number of discrete platforms. For instance, traffic coming into a web data center may flow through Internet routers, bandwidth managers/rate shapers, global server load balancers, firewall and tunneling devices, cache servers, local server load balancers, and a number of Web servers, back-ended by multiple application and database servers, in that order. This brings about a need to: reduce the number of internal hops, decrease possible single points of failure, lower latency, increase reliability, and lower cost (Sevick, 1999), and thus motivates the need for an integrated, fast, intelligent, and highly robust next generation switching platform that can manage, route, and load balance traffic across data center devices (such as, servers, firewalls, caches, and gateways). In addition to connection and packet routing services offered by traditional routers and switches, web switches use policies to enable the traffic management capabilities mentioned earlier (server load balancing, access control, quality-of-service, and so on). The considerably more functionality demanded by these new requirements far exceeds that provided by ordinary LAN switches, and gives rise to the need for specialized web switches.

How Web Switching Works: The Basics

The World Wide Web is based on the TCP/IP protocol suite, which consists of five layers (not to be confused with the seven layer OSI reference model): Layer 1, the physical layer comprised of the physical medium; Layer 2, the data link layer comprised of a link-level transmission technology, like Ethernet; Layer 3, the network layer comprised of IP; Layer 4, the transport layer comprised of TCP; and Layer 5, the application layer comprised of the protocols used to request and deliver content on the Web, including such protocols as HTTP, FTP, and RTSP/RTCP (Real-time Streaming and Control Protocols).

To deliver high-performance routing based on content, a web switch must understand precisely what content is requested by a user, and must, therefore, be able to read and parse complete URLs and user cookies embedded within HTTP requests, and route the requests to the locations best capable of serving them. That is, a web switch must intelligently direct web traffic in real time, based on policies and on information contained in Layers 2 through 5 of the TCP/IP protocol stack (such as URLs, TCP/UDP port numbers, SYN/FIN bits that mark the start and end of TCP application sessions, and IP source/destination addresses). The use of URLs, in addition to IP addresses to make switching decisions is what allows web switches to optimize web traffic. Indeed, URLs provide a ubiquitous method of identifying

18 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-switching/18417

Related Content

Evaluating Information Systems Constructing a Model Processing Framework

João Duarte and André Vasconcelos (2010). *International Journal of Enterprise Information Systems* (pp. 17-32).

www.irma-international.org/article/evaluating-information-systems-constructing-model/46065

Systemic Enterprise Architecture as Future: Tackling Complexity in Governments in the Cusp of Change

Pallab Saha (2014). *A Systemic Perspective to Managing Complexity with Enterprise Architecture* (pp. 1-70).

www.irma-international.org/chapter/systemic-enterprise-architecture-as-future/80906

Marketing in SMEs: The Sales Process of SMEs on the Food and Drink Industry

Tony Douglas and Maktoba Omar (2013). *Enterprise Development in SMEs and Entrepreneurial Firms: Dynamic Processes* (pp. 163-196).

www.irma-international.org/chapter/marketing-smes-sales-process-smes/74466

The Impact of Dual-Fairness Concerns Under Different Power: Structures on Green-Supply-Chain Decisions

Tianjian Yang, Guangdong Liu, Yao Wei, Xuemei Zhang and Xinglin Dong (2019). *International Journal of Enterprise Information Systems* (pp. 1-26).

www.irma-international.org/article/the-impact-of-dual-fairness-concerns-under-different-power/232162

E-Government and EIS Change Management and Critical Success Factors: An Omani Success Story

Wafi Al-Karaghoul, Ahmed Al Azri and Zahran Al Salti (2013). *Cases on Enterprise Information Systems and Implementation Stages: Learning from the Gulf Region* (pp. 181-214).

www.irma-international.org/chapter/government-eis-change-management-critical/70310