W H I T E P A P E R

Scaling Next Generation Web Infrastructure with Content-Intelligent Switching

- **WEB SWITCHING EVOLUTION**
- **WHAT DRIVES CONTENT-INTELLIGENT TRAFFIC MANAGEMENT**
- **THE BENEFITS OF CONTENT-INTELLIGENT SWITCHING**
- **CONTENT-INTELLIGENT SWITCHING —WHAT'S INVOLVED?**
- **CONTENT SWITCH DESIGN CONSIDERATIONS**
- **PURPOSE-BUILT WEB SWITCHING**
- **CONCLUSION**

Alteon WebSystems, Inc.
50 Great Oaks Boulevard
San Jose, California 95119
408-360-5500
408-360-5501 fax

http://www.alteon.com

WEB SWITCHING EVOLUTION

The Internet is in a constant state of change. To respond to increased load and new application demands, e-business innovators are deploying new architectures that will scale with their business growth. In turn, the Web switch has emerged to manage, route and load balance traffic across data center devices including Internet servers, firewalls, caches and gateways. In addition to connectivity and packet routing services offered by traditional layer 2/3 switches, Web switches use sophisticated policies that are unknown to traditional LAN switches and routers to incorporate traffic management capabilities such as local and global server load balancing, access control, quality-of-service and bandwidth management. By enabling e-businesses to increase performance, security, availability and scalability at the application level, the Web switch has become an indispensable component of new Web data center infrastructures.

Today, Web switching technologies have already advanced beyond simple transport (layer-4) routing to encompass content (layer-7) intelligence. What drives this rapid technology evolution? What does content switching involve? What applications can content switching enable? What are the key differentiations among content switches? This white paper will answer these questions and conclude with a discussion of purpose-built, content-intelligent Web switch design.

WHAT'S DRIVING CONTENT-INTELLIGENT TRAFFIC MANAGEMENT?

The Internet's hyper-growth has given rise to a number of issues that Web site infrastructure designers must address:

1. Virtualization of Web servers and content

The days of a single host serving a Web site are gone. More often than not, multiple servers, represented by a virtual IP address (VIP), work in tandem to support a Web site. Load balancers are deployed to direct traffic across the server farm by intercepting client requests destined for the VIP and distributing each request to the best performing, healthiest server. Server load balancing allows a service to continue in the event of a server failure. It also enables processing capacity to grow smoothly through simple addition of more servers.

The virtualization of Web servers has led to content management challenges. If all servers in a Web site were to be able to service any request, then they must all have access to the entire Web site's content. This was feasible when Web content was smaller and Web sites were hosted in a single data center. But with Web content growing at a blazing pace of 60% per year1 from 2000 to 2003, replicating entire content on every server in multiple data centers will no longer be economically viable.

Furthermore, the use of content- and function-specific server appliances also drives the need for content segregation. For example, dynamic Web content is best hosted on high-performance servers optimized for executing scripts and applets. Conversely, static content such as logos, templates, videos clips, and the like can be hosted on low-end servers with large storage capacity to reduce costs. Likewise, deployment of specialized servers for streaming video, compressing graphics for mobile devices, etc, has become commonplace.

In short, content segregation mandates intelligent routing of user requests to the correct content locations.

2. Virtualization of Web Users

With the use of DHCP servers and proxy firewalls, IP users are no longer uniquely identifiable by IP address. In fact, with widespread use of proxies at Internet access points, continuous requests generated by a user to the same Web site may actually carry different source IP addresses.

This presents a challenge for organizations looking to differentiate their services based on user categories. This also makes it difficult for Web site designers to have multiple requests from each individual user be processed by the same server - a requirement known as persistent- or sticky-connections that helps maximize server efficiency and enables stateful transactions such as e-shopping and multi-page forms.

To solve this problem, many Web sites insert an electronic "cookie," representing a unique user identifier, into a new visitor's browser. The browser will automatically transmit the cookie in subsequent visits to the same Web site. The ability for a device to recognize these cookies is mandatory to providing any user-aware traffic management services.

3. Multimedia Applications

Multimedia content is growing fast. Real Time Streaming Protocol (RTSP) and Voice-over-IP (VoIP) protocols use separate channels for transmitting control and data traffic between a client and a server. The specific sockets (TCP or UDP port numbers) used for the data transmission channels are generated dynamically and communicated between clients and servers over pre-established control channels. To properly route these applications to the right servers, a traffic management device must parse the control channels to extract the dynamic socket numbers for the data channels, so related control and data channels can be processed as a single, logical session.

CONTENT SWITCHING BENEFITS

Traditionally, redirecting Web requests using content or user classification has been a function of Web servers. However, Internet traffic and business growth is fast outpacing that of computing power. Offloading content classification to Web switches provides leverage for the entire Web site infrastructure in the following ways:

Flexible Content Location

By examining the URL in a Web request, the Web switch can determine the type of content being requested and direct the request to servers hosting the requested URL. With content switching, Web site content can be segregated with no change to the applications. This allows partial instead of entire content mirroring on each server and makes it easy for e-businesses to deploy servers optimized for specific content types or processing functions.

Persistent Application Support

Applications such as shopping cart, payment transactions, search display and multi-page forms require persistent connections. This means a client must constantly talk to the same real server for the duration of the transaction, which typically spans multiple TCP connections. If a client-server association is not persistent, it may result in broken shopping carts and disgruntled users. In a mega-proxy environment, the only reliable way to match multiple connections to the same user is by matching the cookies embedded in non-secure HTTP connections or the SSL session identifiers embedded in secure HTTP-S sessions. The ability to parse content enables the Web switch to accurately associate consecutive requests from a user with the same server, ensuring transaction integrity.

Increased Server Efficiency

Even if an application doesn't break when visitors are sent to different servers during the course of a transaction, there are other reasons for persistent sessions. Servers store recently accessed information in memory. Retrieving information from local memory is many times faster than retrieving it from a back-end database or hard drive. A content-intelligent Web switch can send successive requests with the same cookie to the same server, taking advantage of server cache memory to improve server efficiency and performance. Where cache servers are used, the content switch can intelligently filter incoming client requests to avoid passing irrelevant requests to the cache servers. For example, requests for dynamic content, requests with embedded cookies, requests other than HTTP GET, etc, can be forwarded directly to the origin server to reduce unnecessary load on the cache servers.

Improved Web Site Performance

HTTP version 1.1 allows multiple HTTP transactions to be transported over a single TCP connection to reduce TCP processing overhead. A layer-4 Web switch with no content intelligence will forward all HTTP 1.1 requests on each TCP connection to a single server. In contrast, a content switch can forward each request within the TCP connection to a different server, increasing load distribution granularity. This optimizes resource utilization and speeds overall Web site performance.

Enhanced Differentiated Services and Bandwidth Management

In order to provide preferential services based on user categories, (frequent shoppers versus frequent browsers, for example,) a Web switch needs to be cookie-aware. Likewise, to enforce the appropriate bandwidth and jitter characteristics for transporting different content types, the Web switch must be URL-aware. Without content awareness, traffic classification and hence, quality of service, can only be applied at gross levels such as per IP address or application port.

Virtual Hosting Support

Virtual hosting conserves IP addresses by allowing multiple domains to be represented by a single public IP address. When a content-intelligent Web switch receives a client request for the shared IP address, it can extract the requested domain name from the “Host Header” portion of the HTTP header, concatenate it with the IP address to obtain the unique host identifier, and redirect the request to the appropriate server or server farm.

Flexible Content-based Server Health Checks

Content-intelligent Web switches allow Webmasters to customize server health checks to verify content accessibility in large Web sites. As the amount of content grows and information is distributed across different server farms, flexible, customizable content health checks are critical to ensuring end-to-end availability.

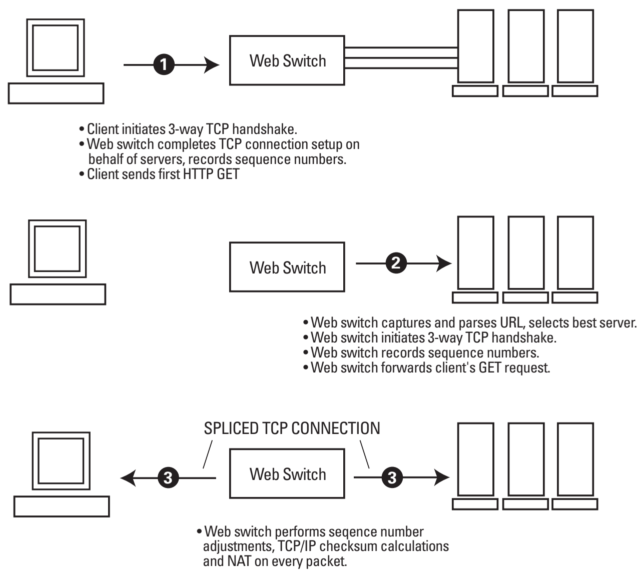
CONTENT SWITCHING —WHAT'S INVOLVED?

Working with session content is much more demanding than examining TCP/IP protocol headers because:

- Content is non-deterministic. Content identifiers such as URLs and cookies can be of varying lengths and can appear at unpredictable locations within a request. Scanning through session traffic for a specific string is far more processor intensive than looking at a known location in a session for a specific number of bytes.

- Parsing content requests means temporarily terminating the TCP connection from a client. In other words, the Web switch must first pretend that it is the server, ask the client what it wants, examine the request, and then open a connection to an appropriate server. While this is happening, the Web switch must temporarily buffer the request, which consumes system memory. This temporary termination is called a “delayed binding" (see Figure 1).

FIGURE 1
How "TCP Connection Splicing" Works



- With delayed binding, two independent TCP connections span a Web session: one from the client to the Web switch and the second from the Web switch to the selected server. The Web switch must modify the TCP header, including performing TCP sequence number translation and recalculating checksums on every packet that travels between the client and the server, for the duration of the session. This function, known as "TCP connection splicing," heavily tasks a Web switch, particularly when the switch must process thousands of these sessions simultaneously.

- In addition to real-time traffic and connection processing, a content switch needs to monitor the servers to ensure that requests are forwarded to the best performing and healthy servers. This monitoring involves more than simple ICMP or TCP connection tests as servers continue to process network protocols while failing to retrieve any content. Furthermore, if content is segregated in different servers or server farms, the Web switch must provide a flexible, user-customizable mechanism allowing a relevant set of application and content tests to be applied to each server or server farm.

What Content Switching is...

To fulfill the requirements described above, a Web switch needs to perform numerous processing tasks for each incoming session, including connection setup, traffic parsing, applying server selection algorithms, splicing connections and translating session addresses, metering and controlling server bandwidth usage, processing traffic filters, collecting statistics and so on. Not only are these functions CPU intensive, they are executed whenever a new request arrives. In addition, the switch must also perform background functions such as updating network topology, health-checking servers, applications and server sites, measuring server performance, etc., on a periodic basis.

Imagine the processing load that the Web switch has to bear during a flash crowd when millions of requests flood a site within a 15 minute period! Consequently, a content switch is fundamentally different from a conventional packet switch in that high performance and availability is dependent on powerful software processing capabilities, in addition to massive switch fabric capacity.

What Content Switching isn't...

With the ever-present threat of unpredictable Internet flash traffic, it is also extremely important to NOT burden the Web switch with every content-related feature to avoid overloading the Web switch. In general, tasks that must occur inline are ideal for integration into the switch's small but tightly optimized processing environment, while complex background work that can bog down real-time traffic processing is best left to external host machines.

For example, the job of content replication is a function ideally performed by content management systems designed to process and compare large electronic files, provide version control, execute data compression and security encryption, interface with Web event logging systems and perform bulk data copies. The Web switch is not the right platform for content staging and replication - a non-realtime process with long duty cycles. If replication of a large file occurs at the same time as a sustained traffic burst, it will interfere with session processing and adversely affect site performance.

CONTENT SWITCH DESIGN CONSIDERATIONS

Since content classification involves parsing information of variable sizes, often at non-deterministic locations within a Web session, it is impractical, if not impossible, to integrate the entire function directly in hardware. Hence the first and foremost criterion in designing a content switch is to insert ultra high performance processors into the switching path. Several designs can be found in Web switch products today:

Centralized Software Processing on an L2/3 Switch Fabric

The easiest but perhaps the sloppiest approach is to force-fit content switching software into the management processor on a L2/3 packet switch. From a design perspective, this involves little more than software programming and is the fastest way for hardware vendors to add value to their existing LAN switch offerings.

But when traffic is heavy, every packet requiring layer 4 to 7 processing must pass through the central processor, which easily becomes a bottleneck. As described in the preceding section, content classification is extremely processing intensive. Adding to that the need to support concurrent traffic management services, such as load balancing and bandwidth management, performance degrades as a function of traffic and processing load if enough processing capacity is not readily available. The scalability limitation of this design under heavy traffic and processing load was demonstrated when first generation multi-protocol routers were obsoleted by routers with distributed processing functions across all line cards. This architecture is only suitable for sites with low traffic expectation and simple traffic management requirements.

Distributed Software Processing on an L2/3 Switch Fabric

A better approach is to integrate content classification software in multiple processors within the switch, as parallel processing improves performance. However, this model generates more intra-switch communications, (between the distributed processors and the switch ports,) which either exerts higher load on the switch fabric or requires a separate, out-of-band link between the CPUs and the switch fabric.

Hence, the merit of this design is highly dependent on the "distance" between the distributed processors and the switch fabric, as every packet within a session must depart the switch fabric for one or more side trips to a processor. Distance is a function of the architectural proximity (such as the speed of the shared memory) between the processors and the switch, as well as the speed of the medium over which they communicate (such as the bus rate). For example, if the switching fabric operates at multiple gigabits per second but the internal link between the fabric and each distributed processor operates at one-tenth the speed, the internal link can become the performance bottleneck.

Integrated Network Processors in an L2-7 Switch Fabric

To eliminate the distance between the processors and the switch fabric, many companies are working on new switch designs based on "network processor" ASICs. This state-of-the-art approach embeds network processors directly into a high speed switch fabric so that software processing can occur inline at any stage of the switching function, without packets ever leaving the fabric.

Other performance-boosting advances in Web switching include new switching fabrics that integrate L2/3 packet switching and L4 session switching performed directly in hardware. Certain L7 content-switching tasks, such as TCP connection splicing, can also be embedded into hardware to minimize software processing load.

PURPOSE-BUILT WEB SWITCH DESIGNS

Alteon's Web switches combine custom developed L2-7 network processor ASICs, called WebICs, and a parallel-processing operating environment, called WebOS Virtual Matrix Architecture (VMA), to offer maximum performance and flexibility at the same time.

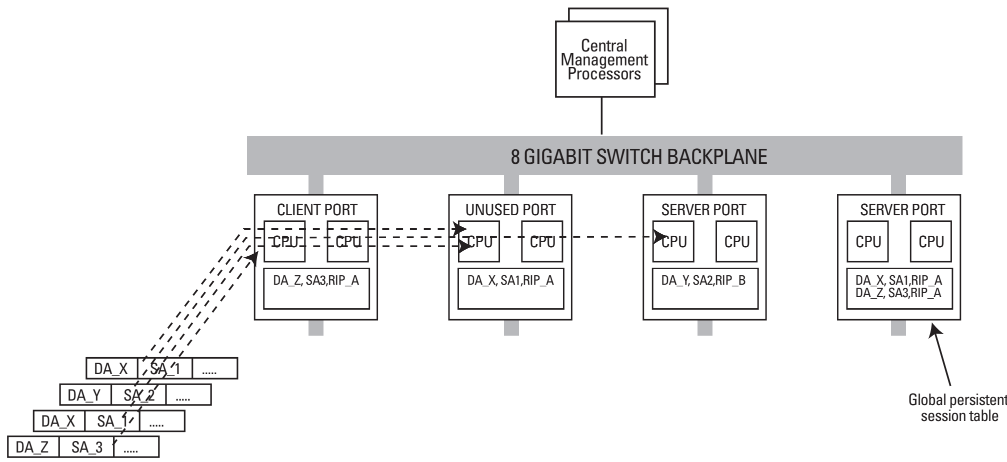
Alteon 180 and ACEdirector Architecture

The Alteon 180 and ACEdirector line is the industry's first Web switch implementation based on advanced network processor technology and a distributed processing architecture. Each port on the Web switch features a WebIC network processing ASIC that combines a L2 packet engine with two RISC processors onto a single chip. Up to 10 WebICs are interconnected over an 8 Gbps switch backplane. The packet engine in each WebIC switches L2 packets in hardware while the network processors support L3-7 switching in software. An additional WebIC handles all switch-wide background processing, including server healthchecks, routing updates and switch management.

With this architecture, 20 RISC network procesors support Web switching operations in parallel inside a multi-gigabit switch fabric. With the addition of Alteon's Virtual Matrix Architecture (VMA) software, every network processor across all switch ports can process traffic simultaneously regardless of the physical ports through which session traffic traverses. As a result, VMA creates a virtual matrix of memory and processor resources that can be used to process traffic from any port at any time within the switch.

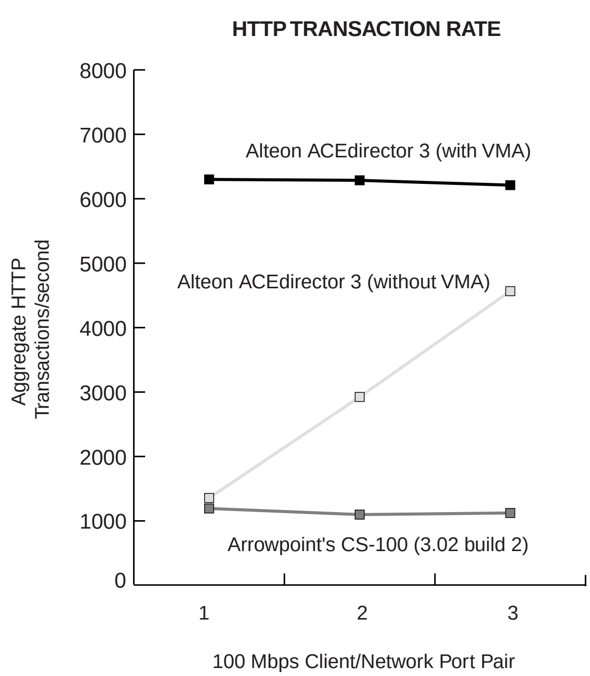
VMA optimizes performance for Web sites with heavy traffic load and sophisticated traffic management requirements. With processor and memory pooling, the Alteon Web switch supports a rich set of Web switching services that can be enabled concurrently with minimal performance degradation. Massive parallel processing increases total system performance while memory aggregation supports large switching tables and extensive buffering for deep content parsing. Each port has access to all content and flow information it needs with no topology and connectivity constraints. Figure 2 shows a diagram of the hardware architecture.

FIGURE 2
Functional View of Alteon Virtual Matrix Architecture



Layer 7, URL-switching testing, conducted by the Tolly Group, demonstrates the performance superiority of the Alteon network processor-based architecture over conventional centralized processor approaches (see Figure 3).

FIGURE 3
Layer 7 URL Switching Performance Tests



CONFIGURATION

- Traffic generation tool: Ziff Davis WebBench 3.0
- 50 Pentium II-class clients, 11 NT 4.0 servers
- Threads per client: 12
- Protocol: HTTP 1.0
- URL parsing rules: 11

SOURCE
The Tolly Group, March 2000

Alteon Series 700 Architecture

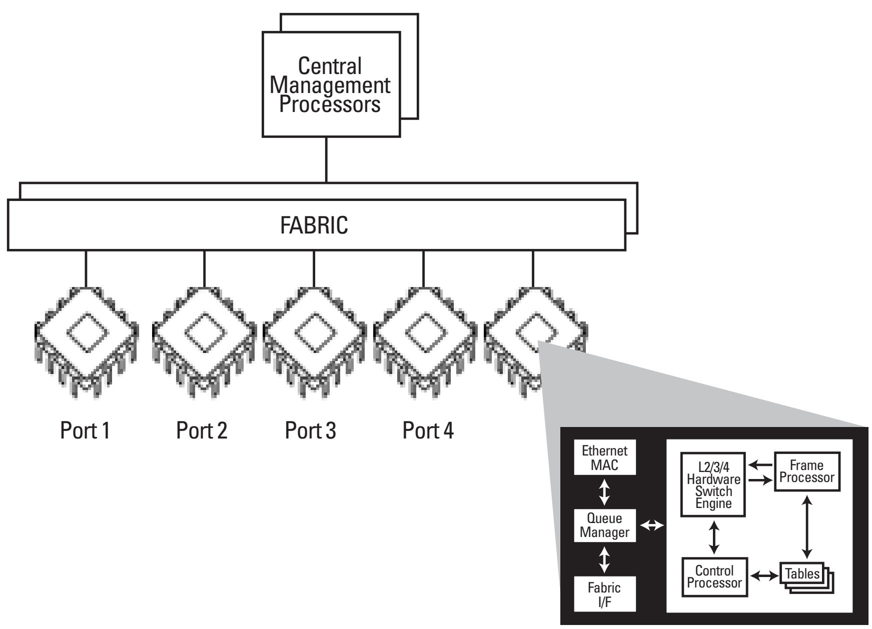
With many times the port density and traffic capacity, the Alteon 700 Series features a next generation WebIC that further optimizes performance for multi-service content switching at large and dense traffic Web sites.

The WebIC in the Series 700 integrates dual network processors, an advanced L2-4 packet and session forwarding engine, L7 hardware-assist functions, and a QoS manager, all within a single chip. Up to 32 WebICs are interconnected by a cross-bar fabric that scales to 180 Gbps of switching capacity.

The Series 700 features the most advanced level of hardware integration, in combination with network processors, in a high capacity switch fabric (see Figure 4). L4 sessions are switched in the hardware engine, and hardware-assist is added to accelerate L7 content switching. The increased level of hardware acceleration is mandatory in order to achieve wire-speed content switching at gigabit rates on all ports.

Network processors continue to be a vital part of the new WebIC architecture. The embedded processors provide the 700 Series with feature flexibility and scalability, allowing more content-intelligence and traffic management services to be integrated into the Series 700 without hardware changes. To ensure performance scalability, the new WebIC is designed to allow the processors to insert hundreds of instructions at any stage of the hardware switch path without impacting wire-speed throughput.

FIGURE 4
Functional View of Alteon 700 Series and Next Generation WebIC Network Processing ASIC



CONCLUSIONS

Content switching is becoming a mandatory traffic management service in new Web data center infrastructures in order for e-businesses to scale their server and application architectures and respond quickly to new business demands.

Content classification is extremely processing intensive. It calls for an advanced, purpose-built platform optimized for both performance and flexibility simultaneously.

Alteon WebSystems, the inventor of Web switching, has developed the most sophisticated Web switch platform that is uniquely suited for high performance content switching. Alteon's ACEdirector and Series 700 lines are purpose-built Web switches with network processor cores integrated directly into their high-speed switch fabrics to deliver an architecture that combines performance with feature scalability and flexibility. ■

[1] Source: IDC, "The Global Market Forecast for Internet Usage and Commerce: Based on Internet Commerce Market Model. Version 5", June 1999