W H I T E P A P E R

Accelerating the Delivery of Web Content

New Service Delivery Platform Options for Hosters, ISPs and e-Business

- **WHY TRAFFIC MANAGEMENT**
- **WEB TRAFFIC MANAGEMENT PLATFORMS**
- **ALTEON'S INTEGRATED SERVICES MODEL**
- **ALTEON iSD DETAILS**
- **SPECIFIC APPLICATIONS**

Alteon WebSystems, Inc.

50 Great Oaks Boulevard

San Jose, California 95119

408-360-5500

408-360-5501 fax

http://www.alteon.com

WHY TRAFFIC MANAGEMENT

Internet content and application growth has produced an insatiable appetite for server compute cycles. With rapidly expanding server farms, Web site operators are faced with declining system scalability, increasing provisioning and maintenance complexity and rising hardware costs.

This creates an opportunity for front-end devices that regulate and load-balance traffic across Web servers as well as offloading repetitive functions from them. These server front-end functions are collectively categorized as Web traffic management.

Two types of Web traffic management systems dominate the market today: traffic management appliances and Web switches. Appliances are typically host-based products packaged to perform a dedicated function such as load balancing, proxy caching, content distribution and bandwidth management. Web switches, on the other hand, are networking devices, optimized for traffic filtering and forwarding, that are equipped with one or more embedded processors to offer value-added traffic management services.

Alteon's Integrated ServiceDirector (iSD) is a service delivery platform that combines Web switching and appliance technologies, taking full advantage of the strengths in both architectures to support a broad range of integrated, coherent Web traffic management capabilities.

This whitepaper outlines the strengths and limitations of Web switch and appliance architectures, describes Alteon's integration of both technologies with the iSD, and present two examples of iSD implementation—for SSL acceleration and Akaimization.

WEB TRAFFIC MANAGEMENT PLATFORMS

Web traffic management services are complex, stateful functions that provide Web site operators control over the unpredictable dynamics of Internet traffic, allowing applications to run smoothly in a Web data center. Traditionally, these services have been implemented directly within the Web server. More recently, however, some vendors have moved individual traffic management functions into dedicated, host-based platforms called appliances. Appliances provide leverage by offloading many Web servers with a single, streamlined device. Alteon pioneered a third alternative—Web switching. Web switches integrate multiple traffic management services onto a switching platform optimized for session traffic classification and forwarding functions.

Server-based Traffic Management

Early models put the burden of traffic management – such as clustering or load balancing, packet filtering and content distribution – on the Web server directly. While this approach allows traffic management functions to operate intimately with each application, the main problems are operational scaling, cost, and server management.

Scaling the system is a challenge as the number of applications grow. Each application adds traffic management complexity and correspondingly, the amount of server cycles expended on it. Subsequently, buying more servers just produces diminishing returns.

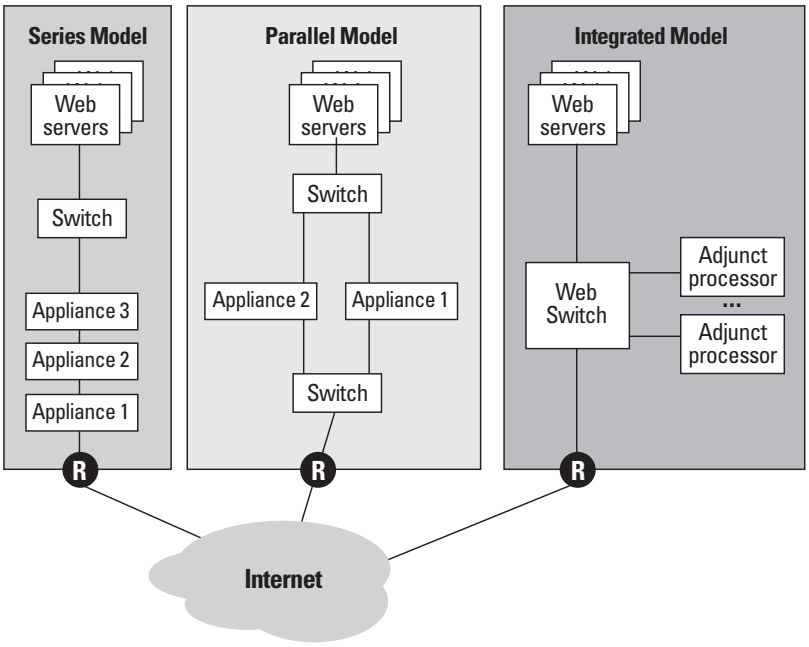
Server cost increases as each server must replicate all traffic management functions.

The system becomes less stable any time servers are modified to add new traffic management functions. The traffic management functions must be reconfigured with each operating system or application upgrade. Such management complexity increases exponentially as the system grows.

Appliance-based Traffic Management

A natural optimization of the server-based model is to move the majority of traffic management functions out of individual Web servers into a set of dedicated traffic management servers. The operating environment on these servers can be further streamlined since they do not have to perform general-purpose computing, while each server is fine-tuned for a particular traffic management task. This concept launched the traffic management appliance market.

FIGURE ONE – Series, Parallel and Integrated Deployment Models



The appliance model is well suited for compute-intensive or storage-intensive traffic management functions such as caching and firewalling, but it presents its own set of challenges as Web traffic grows.

Latency increases as each appliance adds a hop to the data path. Every piece of traffic to and from the Web servers must filter through each appliance, regardless of its relevance to the appliance's function.

System throughput is limited to that of the slowest appliance on the path. Since these appliances are really repackaged servers, their throughput is constrained by their operating systems, I/O buses, and other host-based overhead. Some appliance vendors have tried to overcome this problem by using network adapters that are equipped with co-processors to offload I/O overhead. This can alleviate the problem somewhat but eventually, system throughput is still limited by the PCI backplane in each appliance.

System resilience becomes more complex as configurations expand. Appliances offer two failure protection methods: cut-through and active-standby redundancy. Cut-though allows an appliance to behave like a transparent link upon a soft failure, allowing traffic to flow through to the device downstream. Clearly, this does not work in the event of a hardware failure.

Active-standby redundancy implements a watchdog function between two identical appliances, allowing the standby appliance to take over automatically when the active unit fails. Since appliances support limited network connectivity, external hubs or switches must be used to concentrate traffic into the redundant appliances, both upstream and downstream. Redundancy considerations must also be given to these networking devices. The topology complexity multiplies with each set of appliances in the data path.

Switch-based Traffic Management

Melding the appliance's software-processing facility with the performance and connectivity of a networking switch is the natural next-step in traffic management evolution. The Web switch is designed to do just that.

Web switching architectures can range from a layer 2/3 LAN switch with one or more turbo-charged processors and even a disk glued on to a purpose-built switch fabric that integrates network processors on every switch port *(Refer to Alteon's Content Switching Whitepaper for a detailed discussion of various Web switching architectures).*

This network-oriented approach works well for most traffic management functions such as load balancing of TCP, UDP and HTTP sessions, which require only examination of well-defined protocol headers. However, this solution is not ideal for every traffic management function.

Traffic processing tasks with long duty-cycles, such as parsing entire Web pages and logging all client-server transactions, can slow down the switch's ability to filter and forward traffic. Storage-intensive operations, such as content caching, content insertion and content copying, are best performed on a host-based device with both massive memory and disk capacity. Equipping a switch with a large amount of networking-class, high-speed memory translates to very expensive hardware for customers; while integrating moving parts such as a disk into the switch is too risky for a device that is the artery to a business-critical Web site.

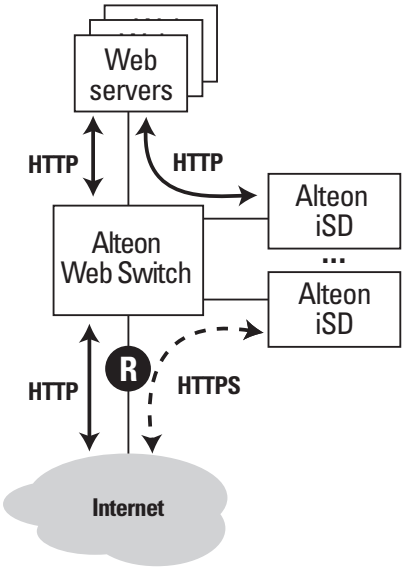
ALTEON'S INTEGRATED SERVICES MODEL

Alteon's integrated services model leverages the inherent personalized content switching features of the company's Web switches and the rich, host-based processing capabilities of appliances to deliver unprecedented breadth of traffic management services with the scaling, reliability, and transparency of operations required for next-generation Web sites. Central to this model is a new service delivery platform, the Alteon iSD, that provides a seamless, scalable processing extension to Alteon's high performance Web switches (see Figure 2).

Combining the strengths of the Web switch and iSD platforms, the integrated services model offers the following advantages:

The iSD platform enables the integration of extremely compute- or storage-intensive traffic management functions normally unsuitable for switch-based implementations into Alteon's Webworking solutions. The first two examples of such functions delivered by Alteon are

FIGURE TWO – Alteon's Integrated Services Deployment Model



Traffic Management Platforms	Appliance Model	Switch Model	Integrated Services Model
Throughput	Limited by appliance's PCI backplane and operating system. Tens to a few hundred megabits per second.	Depends on the complexity of the traffic management functions and limited by the processing and processing architecture of the Web switch. From a few hundred megabits to several gigabits per second	Scalable to many gigabits per second of aggregate throughput.
Functional Scalability	Poor. Single traffic management function per device.	Good. Multiple traffic management functions, except for extremely compute and storage-intensive tasks, can be implemented on a single system.	Excellent. Virtually unlimited possibilities.
Reliability	Adequate with active-standby redundancy; difficult to design large configurations without single-point-of failure.	Excellent with active-active redundant configurations.	Excellent with active-active redundant switch and iSD configurations.
Cost	High cost per function.	Depends on switch implementation. Typically low cost per function.	Low cost per function.
Management	Poor if multiple types of appliances are deployed. Each appliance has its own management interface.	Good; single, integrated management system.	Good; single, integrated management system.
Optimum use of capacity	Poor; processing resources consumed by filtering irrelevant traffic.	Depends on switch architecture.	Good; deploy only as much processing as necessary.

SSL acceleration and transparent Web page "Akamaization." Future applications that can be accelerated by the Web switch-iSD combination include content insertion, content filtering, and firewalling.

Unlike traditional appliances, the iSD is not situated in the data path; instead, it attaches only to the Web switch and is controlled by the Web switch. As a result, iSDs can be added without any service interruption, providing transparent operations and deployment. Once in place, the operator simply enables the iSD through the Web switch's management interfaces and the new service provided by the iSD is gracefully introduced into the network.

The integrated services model offers limitless capacity scaling. Because iSDs can be added to the Web switch in much the same way as load balancing Web servers, its traffic management capacity can scale as the needs arise. For example, using the iSD to deliver SSL co-processing allows performance to scale up to 12,000 SSL connections per second. Similarly, high availability is achieved transparently as multiple iSDs are clustered by the Web switch.

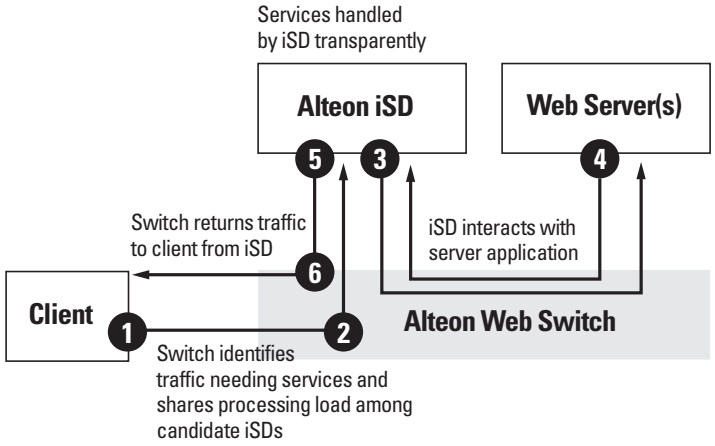
The iSD leverages existing content-intelligent traffic redirection capabilities on the Web switch for maximum processing efficiency. Unlike traditional appliances that need to filter all traffic, the Web switch pre-filters traffic for the iSD based on pre-set parameters such as TCP/UDP port number or URLs. As a result, processing resources on the iSD are completely dedicated to the specific purpose for which it is deployed.

ALTEON iSD DETAILS

Alteon's iSD is an integrated service delivery platform designed for the flexible delivery of complex, compute- and storage-intensive traffic management functions. Together with one or more Web switches, a cluster of iSDs offers reliable, high performance optimization of Web servers and applications. It is tightly integrated with Alteon Web switches and a key component of Alteon's Webworking Framework.

The iSD is pre-loaded with software for a variety of service functions, including SSL co-processing and dynamic URL rewriting ("Akamaization"). The iSD hardware is a 1u high package containing a high performance solid-state processor with massive amounts of volatile and non-volatile memory. A built-in, multi-speed Ethernet connection attaches the iSD to the Web switch.

FIGURE THREE– How Alteon's iSD Works



The iSD connects to the Web switch like a normal server, either directly or across a local broadcast domain. All iSDs performing the same function are on a shared VLAN with private addressing, and can be monitored through SNMP services on the Web switch. The iSDs are configured via the switch console or its WebUI interface.

The Web switch applies rich filters – based on address, session, and content information – and identifies traffic that requires iSD processing. This traffic is directed to the iSD cluster using the Web switch's inherent traffic redirection and server load balancing capabilities. iSD cluster support and other WebOS services can coexist cleanly on the same Web switch.

Leverages deep content awareness on Alteon Web Switches

The iSD model is possible because of the rich content-level parsing features inherent in Alteon's Web switches. iSDs are an ideal complement to the Alteon's Web switches – not only do they build on this deep content visibility for transparent integration and deployment, they also enable greater content-aware features within the switch by pre-processing and decrypting SSL traffic.

SPECIFIC APPLICATIONS

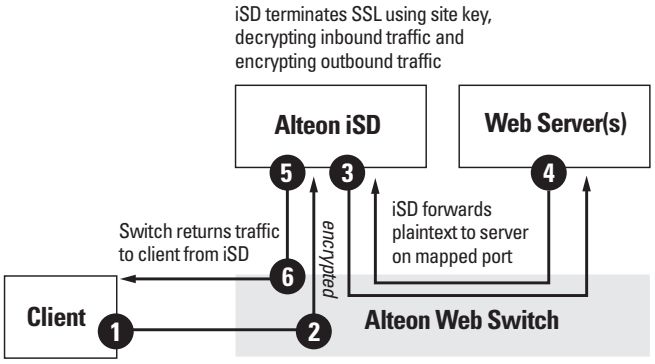
Alteon's iSD is a strategic platform for the deployment of complex traffic management services using an integrated services model. Two iSD services are available from Alteon.

SSL Acceleration

Secure Sockets Layer, or SSL, is the de facto means of secure communication for Web-based applications. It is increasingly common in both business-to-consumer and business-to-business applications, since it protects sensitive information such as passwords and financial transactions from prying eyes.

On the other hand, SSL presents several challenges for the site operator. It consumes an inordinate amount of server processor capacity, since it relies on "heavy" mathematics with large prime numbers – something for which today's general-purpose processors are poorly suited. It also hides critical information such as cookies or URLs from

FIGURE FOUR – SSL Acceleration



load-balancers and stateful firewalls, making it hard to correctly direct secure traffic. In fact, SSL often hides the very information needed to complete online transactions such as shopping carts.

How the iSD supports SSL

The iSD adds hardware-assisted cryptographic co-processing to the Web switch family. The Web switch identifies SSL-encrypted traffic – by looking for queries to the TCP port configured for SSL – and forwards it to an iSD according to a service-scheduling algorithm. The iSD then terminates the SSL connection using dedicated hardware acceleration that makes short work of the cryptography. Then, within a secure data center, the iSD forwards the plaintext query on to the chosen Web server.

On the return path, server traffic is forwarded to the iSD for encryption, and then returned to the client.

Benefits and advantages

The most obvious benefit of SSL co-processing is performance improvement. By using purpose-built processing hardware and software, sites can experience a dramatic increase in the number of SSL connections they can handle-and each connection will complete in less time, giving visitors a more responsive site. Unlike SSL co-processing appliances, only SSL traffic is passed to the iSD. Non-SSL traffic is not penalized by an extra hop.

Equally important, once the iSD has decrypted the query, all content information–including cookies, URLs, and other header information in the HTTP GET–is visible to the Web switch. The switch can apply content rules to the decrypted traffic in order to select an optimal server, support personalization services such as cookie-based QoS and persistence, as well as browser-based traffic redirection, and so on.

There are operational benefits to this model as well. By having SSL on a dedicated device that is not accessible to server users or system administrators, sensitive information such as root keys and certificates are far less likely to be compromised. New servers can also be added transparently, without the addition of new certificates.

The iSD is the first service platform to offer SSL services at gigabit speeds. Up to 32 iSD SSL Accelerators can be clustered to increase performance as traffic load increases.

Akamaization

Many media-heavy sites are using content delivery networks (CDNs), such as Akamai's, to proactively distribute content such as movie clips, audio files, and large images to the edge of the Internet. When end users ask for such content, they are transparently directed to the closest, most responsive available CDN site. To take advantage of Akamai's CDN service, Web site operator must convert specific URLs to Akamai Resource Locators (ARLs) a process known as "Akamaization." ARLs direct the end user's browser back to the Internet, to an Akamai DNS server that finds the closest content source through a set of proprietary algorithms.

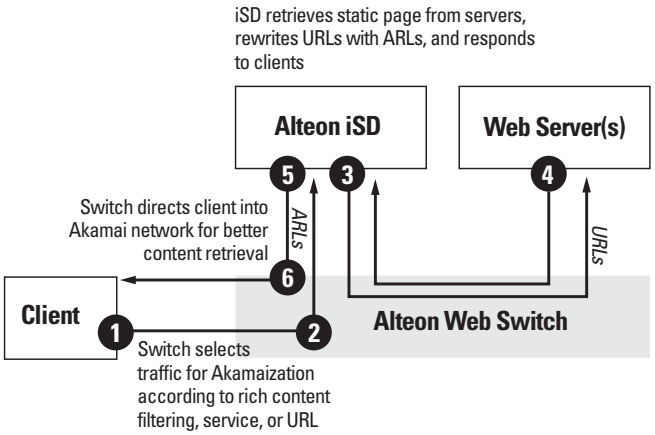
Today, to deploy Akamai's CDN, site builders must put software agents, which handle the URL-to-ARL transformation, on their servers. Administering and maintaining this agent on production servers is increasingly difficult and reduces server stability. With the introduction of the iSD Akamaizer, sites get the benefits of tight Akamai CDN integration without the administrative burden and processing overhead of manual or server-resident Akamaization.

How it works

The Web switch identifies requests for Web pages that must be Akamaized according to a set of content-level filters (such as *.htm). These requests are sent to a chosen iSD-Akamizer for processing. The iSD retrieves the requested content page from the origin Web server on behalf of the client, performs the URL-to-ARL translation on the links in the content page, and returns the Akamaized page to the client.

The client's browser then performs the typical process of parsing the Web page and extracts the ARLs which lead the client to a close, responsive copy of the content residing on Akamai's content distribution network. The Akamai network then fulfills the request for the content.

FIGURE FIVE – How Alteon's iSD-Akamaizer Works



Benefits and advantages

By moving the Akamaization service to a cluster of iSD processors, the CDN integration function is separated from the Web application. This ensures that Web servers aren't handling processor-intensive HTML rewriting in addition to their application duties thereby preserving the origin Web content on the server.

Having the Akamaization function on the iSD greatly simplifies the operation of an Akamaized site, and makes moves, adds, and changes trivial. Upgrades to server platforms no longer require reinstallation of Akamaization agents. In addition, origin content is kept on the source servers, simplifying content management.

CONCLUSIONS

Today's data centers demand quick forwarding of traffic, robust control of flows and transactions, and rich services to enable business-grade Web sites

Alteon's industry-leading integrated service model builds the most advanced, scalable, robust data centers with a clear migration path for integrating current and future traffic management requirements. The combination of Alteon Web switches and iSDs provides the best of both switch-based and appliance-based traffic management vehicles. Lightning-fast traffic filtering and forwarding comes from the dual embedded network processors on every Web switch port; aggregation of distributed processing and memory resources for optimum content processing comes from the Web switch's Virtual Matrix Architecture; and flexible, complex application acceleration services come from the iSD. The iSD also provides an open development platform for integrating third party Web applications that can benefit from Alteon's unique integrated services traffic management model.