Alteon Web OS

APPLICATION AND CONTENT-INTELLIGENT TRAFFIC CONTROL SERVICES

Alteon WebSystems' Web OS provides Layer 2 to 7 Web traffic control services integrated with Alteon's Web switch product lines. Web OS gives administrators unprecedented control of their IP networks by providing local and global server load balancing, application redirection, non-server (e.g., firewall and router) load balancing, bandwidth management and Quality of Service, multi-level high availability and server security services.

FEATURES

- TCP, UDP and IP server load balancing, including HTTP (persistent and non-persistent), FTP and passive FTP, SSL, SMTP, LDAP, DNS, Radius, Telnet and NNTP
- Virtual Matrix Architecture enables dynamic utilization of all processors and memory
- Gigabit-class, content-intelligent bandwidth management enables SLAs and usage-based services
- Full inspection of URLs, cookies and host headers, up to 4,500 bytes across multiple packets.
- High-performance content-intelligent switching using HTTP headers, cookies and URLs allows infrastructure optimization
- Global server load balancing to distributed servers based on health, user proximity, server weights and response times
- Persistent connections using source source IP addresses, HTTP cookies or SSL IDs
- User-scriptable server health checks enable content verification
- Complete site redundancy and failover
- Unlimited number of virtual IP (VIP) addresses per switch via address masking
- Application redirection for any traffic type
- Firewall and WAN router load balancing

www.alteonwebsystems.com

Local Server Load Balancing (SLB) Increases Application Performance and Availability

Web OS local server load balancing enables virtually unbounded server capacity by transparently distributing client traffic across a server farm. Web OS ensures high application availability through the use of continuous server, application and content health checking, server load policing, support for backup and overflow servers and redundant switch configurations.

With server load balancing enabled, Alteon's Web switches appear to the network as one or many "virtual servers," each represented by a Virtual IP (VIP) address or address range. Web OS supports a virtually unlimited number of VIP addresses with up to 8,192 services allowed per VIP address. Real servers can be associated with any number of virtual servers for maximum flexibility.

The Web switches distribute load by assigning each application request to the most available real server associated with the destination VIP address and tracking each session from start to finish.

Effective load balancing algorithms tailor request distribution based on application and server types. These algorithms include least connections and round robin with per server weighting and maximum load thresholds. Excess load is directed to designated overflow servers which can be located anywhere in the network. In addition, Alteon's Web OS API allows users to design custom server-based scripts to dynamically control load distribution on Web switches.

Web OS also supports the inspection and parsing of URLs and HTTP headers and cookies letting Web administrators optimize their server investment. In each of these applications, Web OS directs user requests to the appropriate server or server group based upon URL or HTTP host header matches. By using URL-based load balancing, Web servers can be optimized by storing different types of objects on a server optimized for managing that object type. And, Web hosters can create a "virtual host" by configuring a single VIP to represent multiple domains. This simplifies the host infrastructure configuration and transparently minimizes the number of public IP addresses required.

Flexible Persistence Support Enables Effective E-Commerce Infrastructure

Web OS provides a variety of persistent methods to facilitate server assignment based on specific user information. These methods include SSL session ID tracking, HTTP cookie tracking and source IP address-based persistence using hashing or inactivity timers. With these persistence methods, Web OS looks deep into the HTTP connection – up to 4,500 bytes across multiple packets – to parse user data. Requests from new users are load balanced to the best available server based on the configured load-balancing algorithm. Subsequent requests from the same users are automatically tracked and forwarded to the same server.

Alteon Web OS

MORE FEATURES

- Up to 32 million concurrent sessions (700 series)
- Up to 12 million session setups and teardowns per second (700 series)
- Up to 64 geographically distributed sites
- Unlimited number of virtual IP (VIP) addresses per switch via address masking
- Up to 8192 services per virtual IP address
- Up to 224 packet filtering rules per port
- Full network address translation enables multi-site load balancing and traffic redirection
- Active-active and hot-standby for high availability with in-band keep-alives and configuration updates
- Advanced server security features and secure switch administration
- Support for direct server return

www.alteonwebsystems.com

Global Server Load Balancing (GSLB) Directs Client Requests to the Nearest, Most Responsive Server

Global server load balancing allows content and servers to be distributed to up to 64 sites around the world and directs client requests to the best site based on server health, proximity to the client, server weights and response times.

Each Web switch monitors the health, response time and performance of other sites and exchanges this information with all other GSLB Web switches. With a global view of each site's health and performance, Alteon Web switches direct traffic to sites in proportion to their performance measurements. As a result, the best performing sites receive more user requests than other sites, in proportion to their ability to handle additional requests. This gives users consistently good response times regardless of where their requests are directed.

Web OS directs users to the best site using DNS or static client-to-site proximity tables. Additionally, granular client proximity discovery can be implemented by allowing the user's DNS requests to be resolved based upon Border Gateway Protocol (BGP) autonomous system hops. If all servers at a GSLB site fail or are overloaded, the Web switch at the site immediately notifies its peer switches and redirects all incoming requests to the best peer site using HTTP redirect or IP Proxy. This ensures no service disruption. The prefer local option causes user requests to be redirected to the same site as their DNS request – in effect, directing client requests to the site in closest proximity to their DNS server.

Application Redirection Provides High-Performance Policy Routing and Load Balancing

With application redirection, powerful filtering rules and IP Type of Service settings can be used to intercept any type of IP traffic and redirect it to a designated server, server farm or IP interface. Filters can be set by source and destination MAC and IP addresses, protocols, layer 4 source and destination port numbers, as well as URLs and cookies. Any or all filters can be applied dynamically to each switch port to allow, deny or redirect packets. Dozens of unique filters can be executed simultaneously with minimal performance impact. Application redirection offloads policy routing from expensive routers and increases system performance.

Application redirection also allows DNS requests to be intercepted and transparently redirected to a local DNS server or group of DNS servers regardless of the destination DNS address in the request. This minimizes bandwidth usage and increases user performance.

Content-Intelligent Web Cache Redirection and Load Balancing Optimizes Infrastructure

Application redirection also enables the use of transparent Web caches to reduce Internet access response time and WAN bandwidth consumption without placing caches in the data path. All, or selected, requests for Web pages can be redirected to local Web caches transparently.

By parsing the URL and HTTP commands in each request, Web OS-equipped switches further offload caches by redirecting to them only HTTP GET requests for cacheable objects and forwarding other requests directly to the origin servers. For example, non-GET requests, such as HEAD, POST, PUT, as well as CGI or ASP calls, can be automatically identified by the Web switch and bypassed from the cache – optimizing Web cache operations.

Cache server farm processing can be further offloaded through URL parsing. Specific URLs or URL substrings are matched and transparently redirected to specific cache servers. If multiple caches are used, the Web switch automatically distributes client requests using an algorithm that maximizes cache hits.

Firewall Load Balancing Speeds User Response Times, Increases Site Availability

Network managers can use redirection filters to add the benefits of scaling and resiliency to inline packet processing devices such as firewalls.

For instance, application redirection allows traffic to be distributed among multiple (up to 256), fully active firewalls to improve performance with no changes to firewall software. Health checks are performed on the entire data path through load balanced firewalls increasing resilience and availability.

TRAFFIC CONTROL SERVICES

Bandwidth Management and Quality of Service Enable New Services and Improve User Experience

Web OS bandwidth management provides granular control of bandwidth and traffic into and out of servers while enabling a high level of security. Because Web OS-based switches are designed to connect directly to servers, and have multi-processor computing horsepower, they are uniquely positioned to meter server output and enable usage-based services.

Web site managers can meter, control and account for bandwidth use – by client, server farm, virtual service, application, user class, content type, and many other traffic classes – using a variety of Layer 2 through 7 attributes.

Additionally, Web hosters and service providers can implement bandwidth contracts to guarantee service rates, offer preferential services, and improve the transmission quality of jitter- or loss-sensitive applications.

For each traffic class, the Web site manager assigns a bandwidth policy composed of three bandwidth rates: a *Committed Information Rate* which is the guaranteed minimum bandwidth, a *Soft Limit* which is the metered rate used when bandwidth is available, and a *Hard Limit* which is the maximum burst rate. Web OS logs a variety of flow statistics that can be used to track usage by bandwidth contracts, traffic classes and bandwidth policies, providing data for accounting applications and capacity planning.

Web OS-equipped switches provide last-mile QoS by tagging traffic with IP ToS settings, allowing upstream routers and switches to transport traffic at the appropriate service levels.

Multi-level Resiliency Services Ensure Application Availability

Web OS enables non-stop content access in the event of network, switch, server, application or even entire site failures. Alteon Web switches inherently support full-meshed topologies – eliminating all system-wide single points of failure. Redundant switches can be deployed in an active-standby or active-active configuration with both switches simultaneously servicing users. Active-active redundancy improves availability, resource scalability, and performance and minimizes user impact when failures occur.

Web OS-equipped switches constantly monitor server, application and content availability, bypassing unhealthy servers and services when distributing new sessions. Intelligent application health checking ensures integrity of the entire data path, including content retrieval for HTTP, SSL, POP3, SMTP, IMAP, NNTP, FTP, Radius and DNS services. In addition, Alteon's Web OS allows users to dynamically verify applications and content availability using "send/expect" script-based health checking.

Effective Server Attack Protection

Web OS server security protects Web sites against malicious attacks and unwanted intrusion while providing continuous service for legitimate connection requests. Web OS-equipped switches can thwart TCP SYN attacks without blocking legitimate session requests by terminating all TCP connection requests at the switch.

Also, Web OS bandwidth management can limit the amount of bandwidth allowed for suspect traffic, such as broadcasts or TCP SYN packets. This is superior to the traditional prevention approach, which forces a threshold for total connections at which the switch stops accepting TCP SYN requests, including those from legitimate users.

Other Web OS server security measures include network address translation to allow implementation of private Internet addresses, access list filtering and secure switch administration and management using the Secure Shell Protocol (SSH), Secure Copy Protocol (SCP), and Radius authentication.

APPLICATION EXAMPLES

- Traffic distribution across multiple Web servers
- Scaling firewall and Web server performance for secure e-commerce sites
- Transparent caching infrastructure for ISPs
- Web, Email, News, FTP, DNS and Radius load balancing for ISP online services
- Host header parsing for simplified Virtual Hosting
- High performance bandwidth management for usage-based and preferential services
- Multi-site load distribution for content publishers
- Enables high availability services for Web hosters