

# LOAD BALANCING XML APPLICATIONS

## BUILDING NEXT-GENERATION APPLICATION INFRASTRUCTURE WITH XML INTELLIGENCE ON SERVERIRON LOAD BALANCERS



### Overview

XML (EXtensible Markup Language) has become the dominant technology and document format for exchanging information over the networks. Applications using XML are continuing to proliferate and will be ubiquitous in the not too distant future. Many new communication protocols like WAP (Wireless Application Protocol) and FIX (Financial Information Exchange) are using XML as the standard for defining message formats and exchanging application information. XML's simplicity and versatility are the two critical factors propelling it to be the technology of choice when building new Web-based applications.

IP and Web applications are at the heart of every business operation. Application performance, availability, scalability and security are the critical challenges that Enterprise and Service Provider customers face. Intelligent Layer 4-7 load balancing switches solve these challenges and create massively scalable and always-on applications. They allow businesses to leverage commodity servers to build highly sophisticated and differentiated application infrastructure to serve the businesses' mission-critical needs. Additionally, load balancers act the last line of defense to applications and servers by preventing and defeating DoS (Denial of Service), Virus and Worm attacks. To date, Layer 4-7 switching intelligence has centered around Web switching. Load balancing decisions are based on URL names, cookies and SSL session IDs embedded in Web transaction messages. As more applications use XML, load balancing based on XML content is a natural and critically important next-step in the evolution of load balancing. It dramatically increases the benefits of load balancing switches by providing advanced means to customize and optimize key performance metrics of the application infrastructure.

Foundry Networks' ServerIron products are the industry's first load balancing switches to feature comprehensive support for XML content switching. ServerIron products are used by the most demanding Enterprise and Service Provider customers, and are industry leaders in performance and security. XML switching capabilities further increase the benefits of ServerIron products to Enterprise and Service Provider application infrastructure, and help maximize productivity and profits.

### XML and SOAP

XML is a document format that is highly flexible, yet structured. It allows data to be structured and represented in a format that is easy to understand and is highly extensible. XML is not a mechanism to exchange information - it is simply a way to represent it. XML documents have keywords called Tags and Attributes that de-limit and highlight key information. Below is a sample XML Document:

```
<?xml version="1.0"?>
<note type="internal">
  <to> John </to>
  <from> Mary </from>
  <body>
    Text.....
  </body>
</note>
```

Each XML document has a declaration (e.g: conforms to XML 1.0 specification). The document has nested elements starting with the root element "note". Rest of the elements "to", "from" and "body" are all child elements. SOAP (Simple Object Access Protocol) is an XML-based protocol used to exchange messages containing XML documents. Each SOAP message consists of an envelope defining the contents of the message, an optional header, and the message body. HTTP is the most commonly used transport protocol to carry SOAP messages.

*NOTE: Reference material is available at [www.w3schools.com](http://www.w3schools.com) and [www.w3.org/TR/SOAP](http://www.w3.org/TR/SOAP).*

# LOAD BALANCING XML APPLICATIONS

## BUILDING NEXT-GENERATION APPLICATION INFRASTRUCTURE WITH XML INTELLIGENCE ON SERVERIRON LOAD BALANCERS

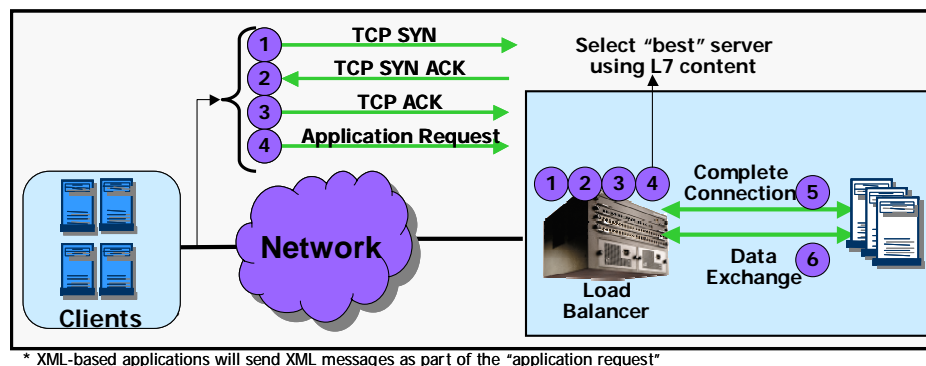


### Load Balancing using XML Switching

XML is a document format that is structured and simple with few rules. XML documents have tags and attributes (strings of alphanumeric text) that are like keywords and reveal highly valuable information about application transactions. Application messages formed using XML may contain tags and attributes that indicate application type, client class, message type and other valuable information that can be used to efficiently switch and service application traffic. Because XML documents are structured and simple, extremely flexible and very powerful XML content switching is possible to load balance XML applications for non-stop availability and massive scalability.

XML content switching involves searching the XML documents embedded in the SOAP/HTTP messages for user-defined tags and attributes, and applying user-defined rules. SOAP/HTTP messages are widely used as the transport mechanism for XML documents. Once the tags and attributes are identified, they are matched against the user-defined content switching rules to direct application requests to the “best” application server(s). The best server is determined by a combination of health checks and load balancing policies defined by the user. Messages belonging to different applications can be switched to appropriate server clusters, and further load balanced to individual servers within a cluster based on sophisticated load balancing methods.

Load balancing using Layer-7 content, including XML, requires that the load balancer delay selection of the “best” server (binding of client connections to a specific real server) until after the application messages are sent by the client and received by the load balancer. This mechanism is called “Delayed Binding”, and is required for intelligent XML content switching. The load balancer, acting as a proxy, completes the connection with the client, and awaits the first application message to arrive. After inspecting the content of the first message and matching with the user-configured content switching rules, the load balancer picks the “best” server for the client connection. After selecting the “best” server, the load balancer completes the connection to the server and binds the client-side connection to the new server-side connection. Figure-1 below shows the concept of delayed binding in a Layer-7 content switching application.



**Figure-1: Delayed Binding to Inspect Layer-7 XML Content Prior to Server Selection**

Foundry's ServerIron Layer 4-7 load balancing switches provide industry's highest performance and scalability, and are the first to feature comprehensive support for XML switching. ServerIron switches are based on Foundry's IronCore™ architecture and enable always-on applications and highly reliable server farms. They are also the last line of defense for applications, and protect against wire-speed Gigabit rate DoS and DDoS attacks.

Foundry's Layer 4-7 switching solution goes beyond the traditional URL and HTTP header switching, and inspects deep into the XML messages to intelligently pick the “best” application server. Support for XML switching allows IT managers to build application and network infrastructure that is highly flexible and customizable for individual applications and mission-critical business needs.

# LOAD BALANCING XML APPLICATIONS

## BUILDING NEXT-GENERATION APPLICATION INFRASTRUCTURE WITH XML INTELLIGENCE ON SERVERIRON LOAD BALANCERS



ServerIron switches are capable of inspecting application messages for XML tags and attributes in content up to 8 KBytes without adversely impacting application performance. Content inspection and matching may be configured based on prefix, suffix and pattern matching to suit a range of application needs. Content switching rules are used to ensure that the application requests are sent to the server (or the server pool) that delivers optimal performance. ServerIron XML switching combines the most advanced content rules engine with the best-in-class load balancing algorithms to deliver superior application response time.

Session persistence is critical for many applications where each session (group of transactions) spans multiple layer-4 connections. All client requests belonging to a given session must be load balanced to the same server. ServerIron switches feature many session persistence mechanisms like cookie, SSL session ID, and source IP. When XML switching is enabled, session persistence is also achieved using XML tag and attribute values.

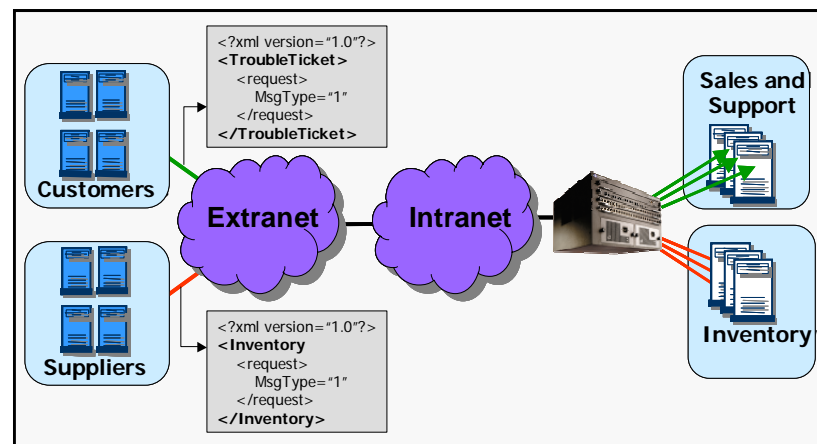
### Example XML Load Balancing Applications

Layer 4-7 load balancing switches with XML intelligence are critical for performance, availability, scalability and security of XML applications. As XML applications proliferate, load balancers with XML intelligence become a strategic business choice to build the next-generation application and network infrastructure. XML switching also has the potential to open the door to new application architectures and business models. This section presents some example uses of XML switching and their benefits.

#### Load Balancing Multiple XML Applications - Example #1

Enterprises are increasingly using XML to exchange information over the extranets with suppliers, customers and partners. These real-time information exchanges are the key drivers of business productivity and profitability. Availability and performance are the critical requirements to ensure timely and uninterrupted access to applications and information to support business operations. Using the XML intelligence, Layer 4-7 load balancers help Enterprises build highly reliable and massively scalable extranet application infrastructure that delivers highest availability and performance to application users.

Figure-2 below shows the use of XML switching to load balance multiple applications. Suppliers and customers access different applications and engage in various types of transactions. The Load Balancer inspects application packets to determine the application type and select the appropriate server cluster to direct client requests. After selecting the cluster, the load balancer can further inspect the content or use other user-defined load balancing methods to select the "best" server within a given cluster to process the client request. Using XML switching gives high flexibility in differentiating applications and transactions.



**Figure-2: Load Balancing Multiple XML Applications using XML Content Switching**

# LOAD BALANCING XML APPLICATIONS

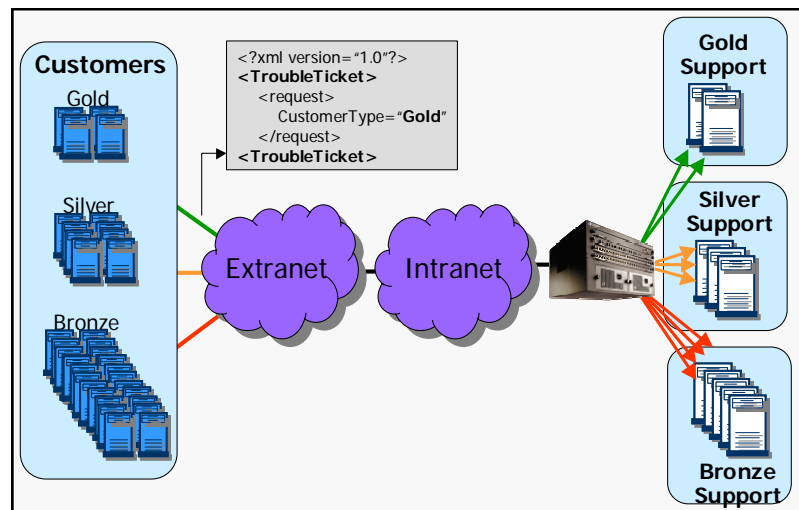
## BUILDING NEXT-GENERATION APPLICATION INFRASTRUCTURE WITH XML INTELLIGENCE ON SERVERIRON LOAD BALANCERS



### Differentiated Client Service - Example #2

Most applications serve a large number of clients' needs. But, not all clients need to be treated exactly the same. Businesses routinely segment their customers (or partners) into different classes and offer differentiated service to each class. For example, a company may sell tiered support contracts to its customers and categorize them as gold, silver and bronze. Gold customers are highly valued, and must be offered highest level of service and least response time.

With XML switching, network and application infrastructure can be designed to offer differentiated service to different segments of clients. The load balancer inspects XML messages to identify the class of client and then directs the requests to servers dedicated for a given client class. Figure-3 below shows an example customer support application with differentiated service. The load balancer uses its XML awareness to identify the type of client and directs requests to the cluster of dedicated servers. Load balancing to the individual servers within the cluster is performed using a variety of user-preferred methods.



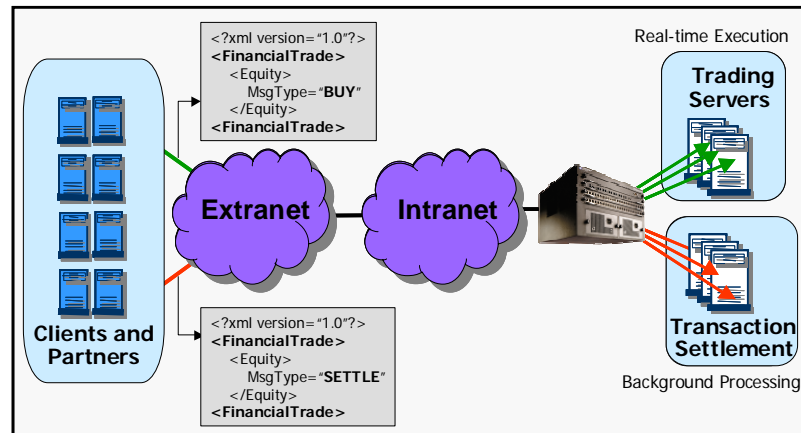
**Figure-3: Differentiated Client Service with XML Switching**

### Application Performance Optimization - Example #3

Above example showed client classification and prioritization. Even within a class of clients, not all transactions are equal. Some transactions may be very time sensitive, and have significantly higher performance and response time needs. Load balancers must be capable of differentiating and prioritizing application transactions to deliver the highest overall performance and availability. For example, financial transactions involving trade executions are extremely time sensitive and need non-stop uptime, whereas the ones involving settlement are less time critical. Using XML switching, the load balancer can distinguish between transaction types, prioritize requests, and optimize performance. Additionally, the load balancer can switch the high priority transactions to a different cluster of servers that are capable of delivering higher performance. Figure-4 below shows an example financial application with transaction level prioritization.

# LOAD BALANCING XML APPLICATIONS

BUILDING NEXT-GENERATION APPLICATION INFRASTRUCTURE  
WITH XML INTELLIGENCE ON SERVERIRON LOAD BALANCERS



**Figure-4: Transaction Prioritization for Application Performance Optimization using XML Switching**

## Key Benefits

### Application Performance

Foundry ServerIrons switches, with their intelligent application-aware load balancing and content switching, significantly improve overall application performance by optimally utilizing server resources. Using customizable load balancing methods and metrics, application performance can be tuned to achieve best response time and maximum throughput. XML switching provides added intelligence to maximize performance by directing client requests with finer grained content switching rules to servers and server clusters. Using the ability to classify and prioritize application transactions, performance optimization can be specifically targeted to the most productive and profitable client and transaction categories.

### Application Availability

ServerIron's high-performance load balancing ensures always-on applications by intelligently distributing application traffic among all available servers, and dynamically monitoring the ability of serves and applications running on them to deliver optimal performance. Using customizable health checks at various levels of granularity like host, port, application and transaction levels, the ServerIron switches instantaneously and transparently react to increases and decreases in server resources by re-directing client traffic as needed. To protect applications from catastrophic failures, the switches can be deployed in multiple high-availability modes with stateful session failover. Applications are completely transparent to ServerIron failures, and continue to function uninterrupted.

### Application and Server Farm Security

Security is a critical challenge for businesses, especially for the mission-critical applications where the stakes are very high. As reliance on the network to deliver the mission-critical applications increases, so does the threat posed by network-based attacks. ServerIron switches have many intelligent features and superior performance to reliably protect against many forms of DoS, Virus and Worm attacks. They protect application infrastructure and server farms against wire-speed Gigabit rate DoS attacks, which translates to 1.5 million attack messages in a one second duration. ServerIron product family features industry's most advanced security intelligence to provide high-performance IronShield™ security that meets the needs of even the most demanding networks and applications serving millions of clients.

# LOAD BALANCING XML APPLICATIONS

## BUILDING NEXT-GENERATION APPLICATION INFRASTRUCTURE WITH XML INTELLIGENCE ON SERVERIRON LOAD BALANCERS



### Application and Server Farm Scalability

Scaling applications and server farms is one of the most fundamental requirements for continued business growth, and is easily and permanently met by the ServerIron load balancers. The ServerIron switches provide unlimited scalability to any IP-based application, and allow businesses to leverage commodity servers to build highly sophisticated and secure application infrastructure. Massive scalability is achieved with complete transparency to existing clients and servers. Using the fine-grained intelligence of XML switching, application resource needs can be identified and targeted, allowing even higher levels of scalability. Additionally, existing resources can be more efficiently used by implementing content switching rules that maximize overall utilization.

### Higher Return on Investment (ROI)

Foundry Networks' ServerIron load balancers provide immediate ROI, and also improve the ROI of application and server infrastructure. They support significantly higher application traffic and clients on existing resources with efficient utilization. By reducing the overhead of connection management on the servers, and dedicating their resources to application processing, overall performance and capacity of the server farm is increased. Downtime associated with security breaches, and server and application maintenance is eliminated, resulting in improved availability. Load balancers also simplify management of applications and server farms, which improves productivity and helps conserve valuable capital to address other critical problems in the network.

### Summary

Performance, availability, scalability and security of applications and server farms are the most fundamental challenges faced by Enterprises and Service Providers. Layer 4-7 load balancing switches are a key strategic solution to these challenges. As increasing number of IP applications use XML, combining XML intelligence with load balancing dramatically improves the benefits of ServerIron switches by allowing customization of application performance and availability. It also opens the opportunity for new application services and infrastructure capabilities that mirror the high-level business models. Foundry's ServerIron products combine the industry's best XML intelligence with the high performance switching to meet the next-generation application infrastructure needs.