

Layer 7 Switching + Load Balancing = Layer 7 Load Balancing

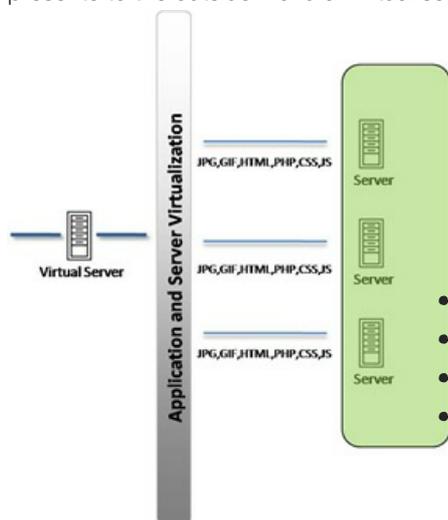


Lori MacVittie, 2008-12-08

Modern load balancers (application delivery controllers) blend traditional load-balancing capabilities with advanced, application aware layer 7 switching to support the design of a highly scalable, optimized application delivery network. Here's the difference between the two technologies, and the benefits of combining the two into a single application delivery controller.

LOAD BALANCING

Load balancing is the process of balancing load (application requests) across a number of servers. The load balancer presents to the outside world a "virtual server" that accepts requests on behalf of a pool (also called a cluster or farm) of servers and distributes those requests across all servers based on a load-balancing algorithm. All servers in the pool must contain the same content.



Load balancers generally use one of several industry standard algorithms to distribute request. Some of the most common standard load balancing algorithms are:

- *round-robin*
- *weighted round-robin*
- *least connections*
- *weighted least connections*

Load balancers are used to increase the capacity of a web site or application, ensure availability through failover capabilities, and to improve application

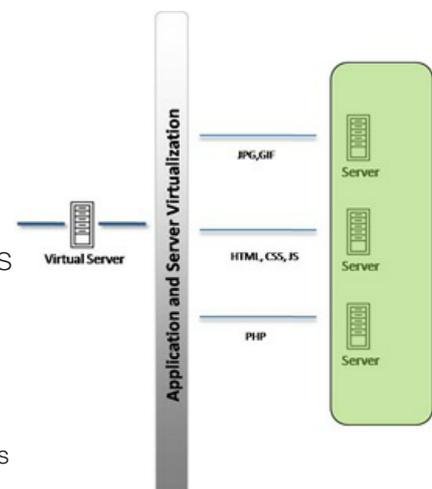
performance.

LAYER 7 SWITCHING

Layer 7 switching takes its name from the OSI model, indicating that the device switches requests based on layer 7 (application) data. Layer 7 switching is also known as "request switching", "application switching", and "content based routing".

A layer 7 switch presents to the outside world a "virtual server" that accepts requests on behalf of a number of servers and distributes those requests based on policies that use application data to determine which server should service which request. This allows for the application infrastructure to be specifically tuned/optimized to serve specific types of content. For example, one server can be tuned to serve only images, another for execution of server-side scripting languages like PHP and ASP, and another for static content such as HTML , CSS , and JavaScript.

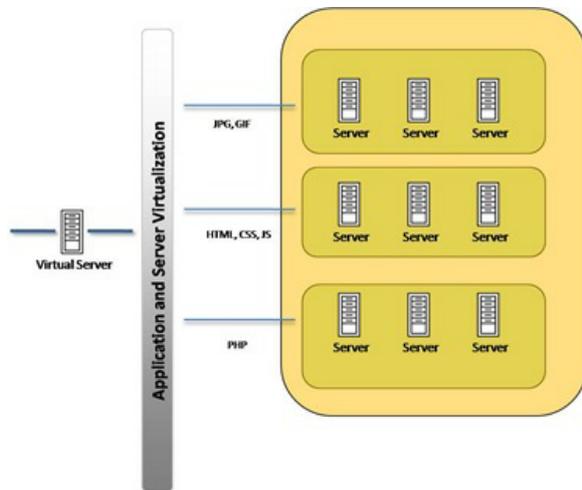
Unlike load balancing, layer 7 switching does not require that all servers in the pool (farm/cluster) have the same content. In fact, layer 7 switching expects that servers will have different content, thus the need to more deeply inspect requests before determining where they should be directed. Layer 7 switches are capable of **directing requests based on URI**, host, HTTP headers, and anything in the application message.



The latter capability is what gives layer 7 switches the ability to perform [content based routing for ESBs](#) and XML/SOAP services.

LAYER 7 LOAD BALANCING

By combining load balancing with layer 7 switching, we arrive at layer 7 load balancing, a core capability of all modern load balancers (a.k.a. application [delivery controllers](#)).



Layer 7 load balancing combines the standard load balancing features of a load balancing to provide failover and improved capacity for specific types of content. This allows the architect to design an application delivery network that is highly optimized to serve specific types of content but is also highly available.

Layer 7 load balancing allows for additional features offered by application delivery controllers to be applied based on content type, which further improves performance by executing only those policies that are applicable to the content. For example, data security in the form of [data scrubbing](#) is likely not necessary on JPG or GIF images, so it need only be applied to HTML and PHP.

Layer 7 load balancing also allows for increased efficiency of the application infrastructure. For example, only two highly tuned image servers may be required to meet application performance and user concurrency needs, while three or four optimized servers may be necessary to meet the same requirements for PHP or ASP scripting services. Being able to separate out content based on type, URI, or data allows for better allocation of physical resources in the application infrastructure.



F5 Networks, Inc. | 401 Elliot Avenue West, Seattle, WA 98119 | 888-882-4447 | [f5.com](#)

F5 Networks, Inc.
Corporate Headquarters
info@f5.com

F5 Networks
Asia-Pacific
apacinfo@f5.com

F5 Networks Ltd.
Europe/Middle-East/Africa
emeainfo@f5.com

F5 Networks
Japan K.K.
f5j-info@f5.com