

Explaining ISA Server 2006 Web server Load Balancing

Abstract

In this article, I will show you how to load balance multiple web servers with ISA Server 2006 Web Server load balancing capabilities. I will also cover some NLB basics from ISA Server and Windows Server 2003 to complete the overview of the load balancing capabilities of ISA Server 2006 / Windows Server 2003.

Let's begin

ISA Server 2006 can distribute Web traffic to identical configured web servers that are normally a special function of a Hardware load balancer. Web server load balancing distributes network traffic to different hosts in the internal network without using classic NLB functions of the Windows operating system.

It is possible to publish a hardware load balancing device to balance web traffic to internal web server but ISA Server Web farm load balancing has a number of advantages (but also disadvantages):

Some Hardware load balancer uses source IP addresses to balance web requests, but this solution might only be suitable in environments where these servers are not behind a NAT device. ISA Server 2006 doesn't forward the original IP address in a standard Web server publishing scenario. The IP address from the external client will always be masked with the IP address of the ISA Server. If you want to be able to forward the original client IP from the external requesting client, the published web servers have to set its Default Gateway to ISA Server which is not suitable in some environments.

Another way to distribute traffic to web servers is to use the Windows integrated Network Load Balancing (NLB) mechanism. NLB allows distributing network traffic based on port rules. All nodes in an NLB cluster use one Virtual IP address (VIP) which is used by ISA Server to forward traffic. The NLB algorithm distributes traffic across the NLB cluster members.

Network Load balancing basics

Very briefly; NLB is a kind of cluster technology which is not exclusive to Microsoft Windows. NLB is part of the Windows Server 200x operating system family and is used to distribute network traffic for up to 32 hosts in the network. NLB uses a distributed algorithm that load-balances incoming traffic to all nodes in a Windows NLB cluster. So, NLB can be used to provide failover and Load balancing capabilities

It is possible to enable the Network Load Balancing feature on every Windows Server 2003 version. The following figure shows the Windows Server 2003 Network Load Balancing Manager with only one NLB node.

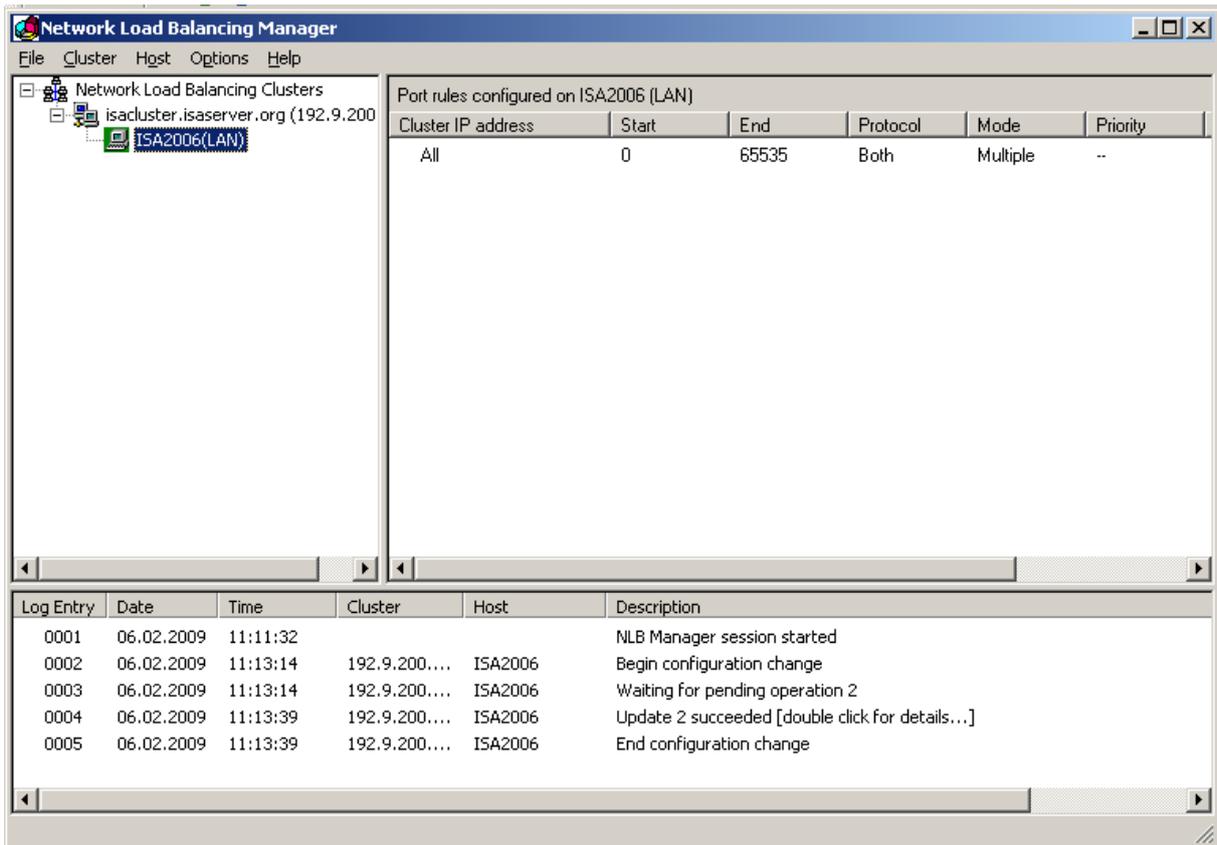


Figure 1: Windows Network Load Balancing Manager

As you can see in Figure 1, it is possible to create port rules which determine on which port the NLB Cluster is listening. For the configured port rule, the NLB mechanism distributes Web Server traffic to every NLB cluster node, but only based on IP addresses and port numbers.

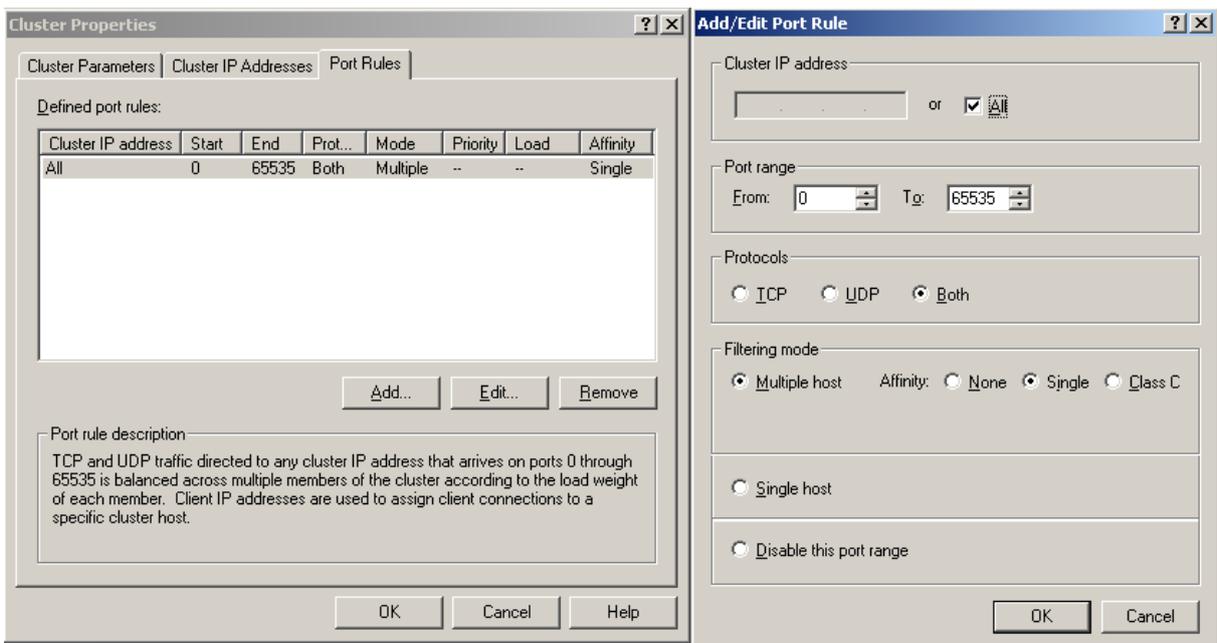


Figure 2: Network Load Balancing Cluster properties – Add/Edit port rules

It is possible to select the protocol TCP or UDP, HTTP is TCP based, so TCP is the best selection to publish Web Servers.

NLB with ISA Server 2006

If you plan to load balance internal Web Servers with the ISA Server Web Server Farm Load Balancing feature, you should also keep in mind, that ISA Server might be the single point of failure (SPOF) when ISA is not load balanced. ISA Server 2006 Enterprise uses NLB to load balance ISA Servers. It is possible to use NLB in integrated mode, the preferred and recommended mode in ISA Server 2006. It is also possible to use NLB with ISA Server 2006 but this is not official supported by Microsoft and has some limitations.

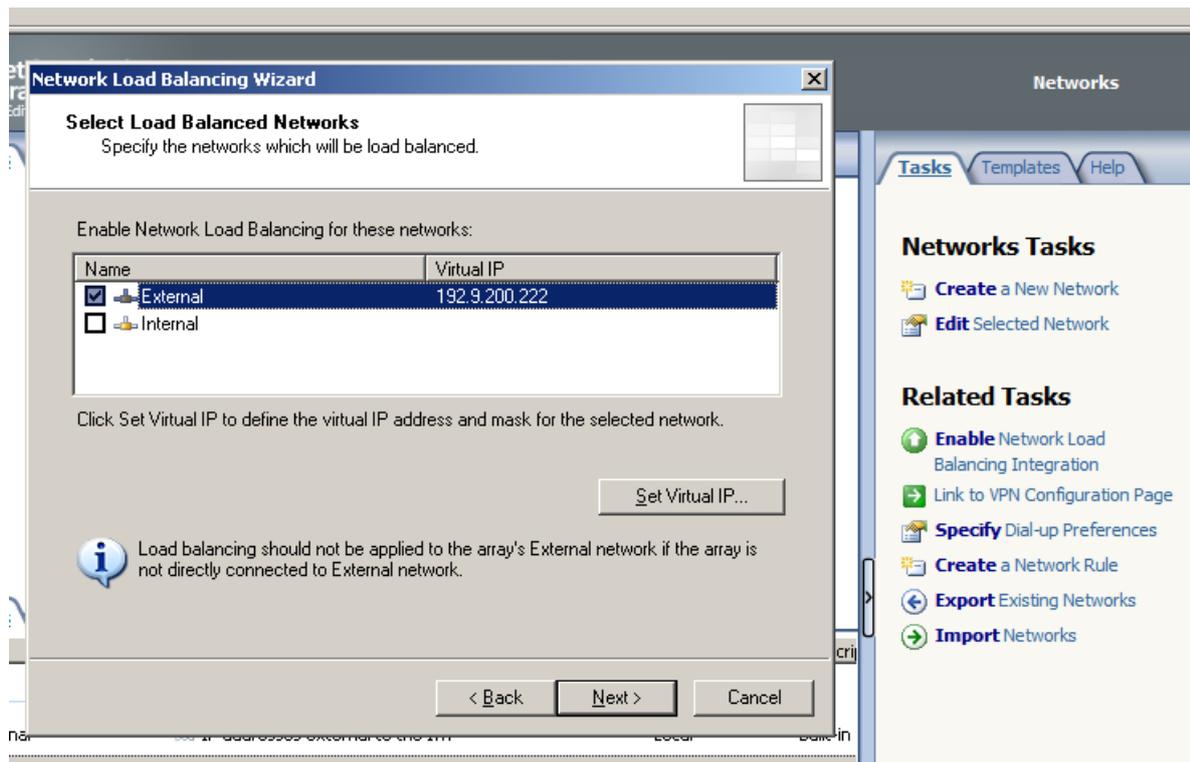


Figure 3: ISA Network Load Balancing configuration

Web Server Farm Load Balancing

Publishing a Web Server Farm is similar to a normal web server publishing scenario. Start the publishing wizard and select the radio button *Publish a Server Farm of load balanced Web servers*.



Figure 4: New Web Publishing Rule Wizard – Publish a server farm

As one of the next steps we must create a new web server farm. The web server farm contains all web servers you want to have to be part of the load balancing farm.



Figure 5: Welcome to the new Server Farm Wizard

Now we must select to publish a web server with Cookie-based Load Balancing or Source-IP based Load Balancing.

Cookie based affinity

Session (Cookie) based affinity is normally used to publish Outlook Web Access (OWA) from Exchange Server 200x or Microsoft SharePoint services/Servers sites.

You should not use Session affinity if you want to publish RPC over HTTP(S) services. RPC over HTTP(S) is used to give Outlook clients full access to Exchange Server from the Internet. RPC traffic will be tunneled through HTTPS. With Outlook it is not possible to use Cookie based affinity.

IP affinity

With IP affinity, the web server traffic is distributed based on IP to all members of the Web farm. If one Server fails to respond, the traffic will be send to another member of the Web farm.

You should not use IP based affinity, if remote clients are located behind a NAT server, because the web server farm will only see the IP address of the ISA Server. If this is the case you should use Session affinity, if it is possible.

IP affinity is useful in an Exchange RPC over HTTP(S) scenario, where session affinity cannot be used or in Exchange Active Sync publishing scenarios where the client does not fully understand HTTP 1.1 (which is needed for cookie based affinity).

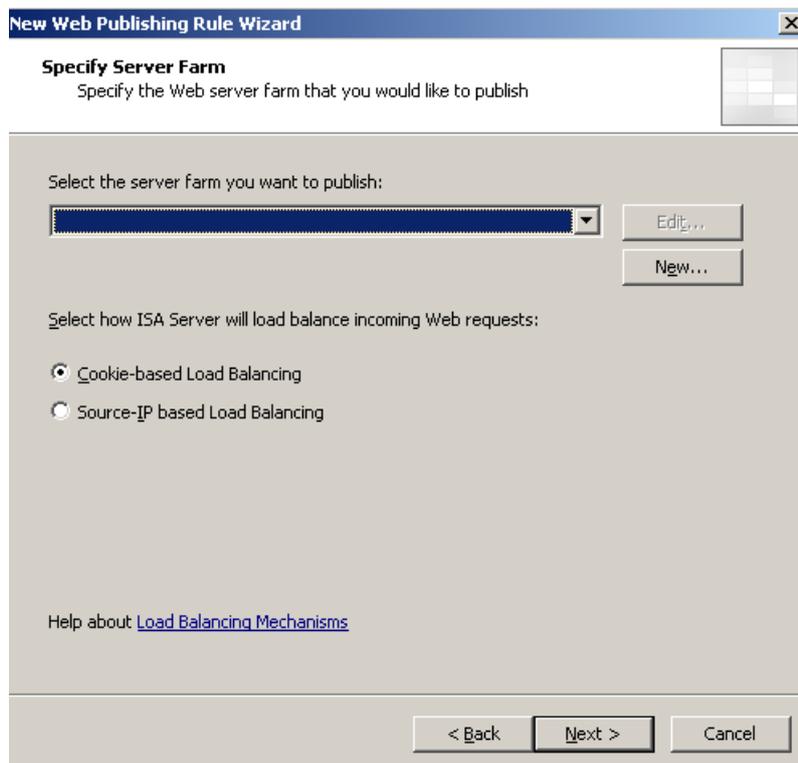


Figure 6: Select the Web Server Farm you want to publish

Now it is time to create the web server farm. Select the servers that should be part of the web server farm.

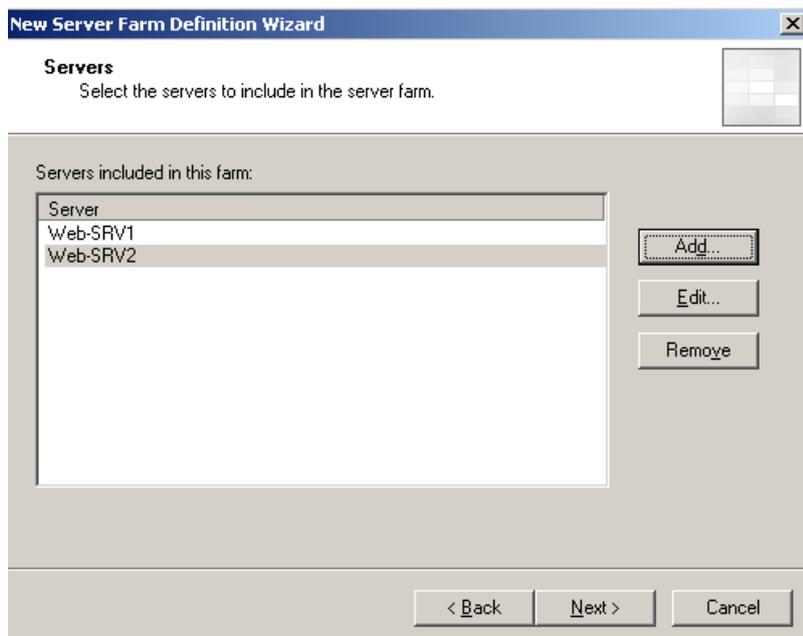


Figure 7: Select Servers for the Web Server farm

Monitoring Web Server Farm Status

If you want to know, which web server farm member is available and which is not, ISA Server automatically creates connection verifiers when you create the Web Server farm. A connection verifier detects the status of farm member and reports this event to the alert configuration in ISA Server, which creates notifications like e-mail messages, entries in the event log and many more.

Servers in a web server farm can have five different states:

Active

This is the normal state of a web server in the farm and indicates that the server is reachable and able to accept requests.

Out-of-service

This state indicates that the web server didn't respond to the connection verifier within the specified timeout. No requests are sent to this farm member.

Draining

This state indicates that the web server is in the process of being drained. Existing connections will be finished but new requests will not be sent to this server. This feature is useful if you want to place one Server of the Web Server Farm in maintenance mode.

Removed

This state indicates that the web server has been removed from the farm, and is not accepting requests.

Unable to verify

This indicates that the server state cannot be verified.

Depending on the type of published servers it is possible to create different connection verifiers. For session (cookie) based affinity, use a URL connection verifier, for IP based affinity use a PING request or a port connection verifier.

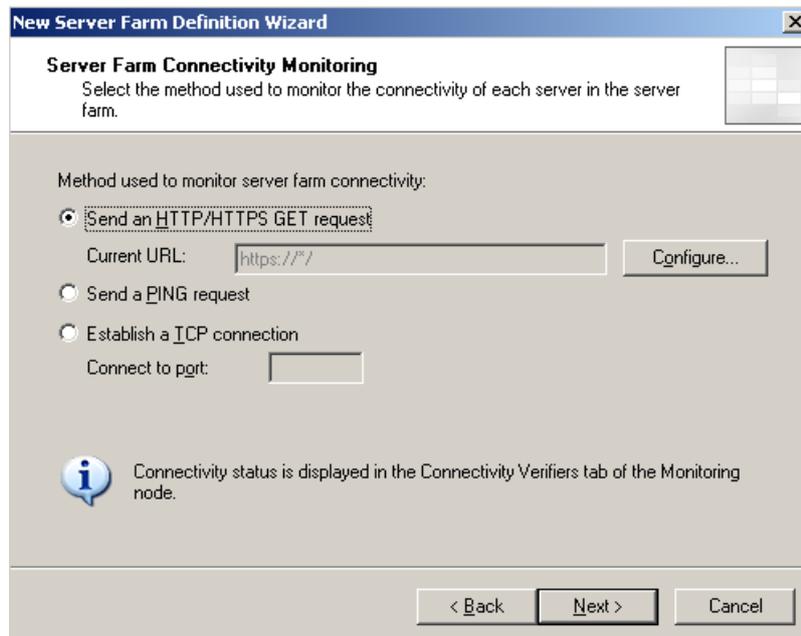


Figure 8: Server Farm Connectivity Monitoring

It is possible to check the status of farm members in the Connectivity Verifiers tab on the monitoring node in the ISA Server 2006 management console (MMC).

To use a connection verifier, a rule must be created which allows HTTP or HTTPS to the specified destination (the web server farm). ISA Server automatically creates this rule for you.



Figure 9: Enable HTTP Connectivity Verification

After the wizard has finished creating the web server publishing rule, you will find the new publishing rule in the Firewall Policy tab in the ISA Server MMC. A special symbol in the publishing rule indicates that a web server farm has been published.

Firewall Policy					
Order	Policy	Name	Action	Protocols	From / Listener
Enterprise Policy Rules Applied Before Array Firewall Policy					
Firewall Policy Rules					
1	Array	WebserverFarm	Allow	HTTP	WWW

Figure 10: The new Web Server Farm Firewall policy rule

If you open the web server publishing rule, you will find a new tab called *Web Farm* which is a new tab when you publish a web server farm. In this tab it is possible to change the configuration of the web server farm.

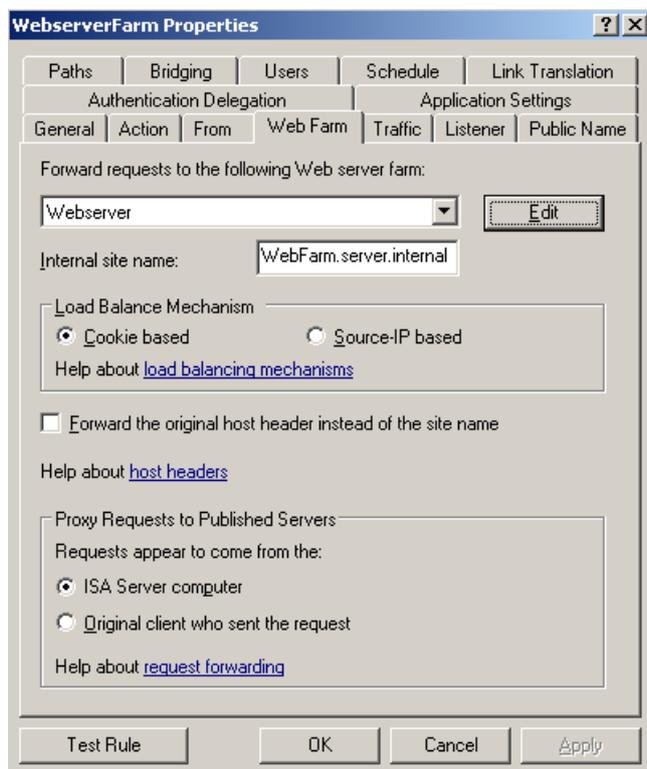


Figure 11: Web Farm configuration properties

Click *Edit* to display the properties of the Web Server Farm.

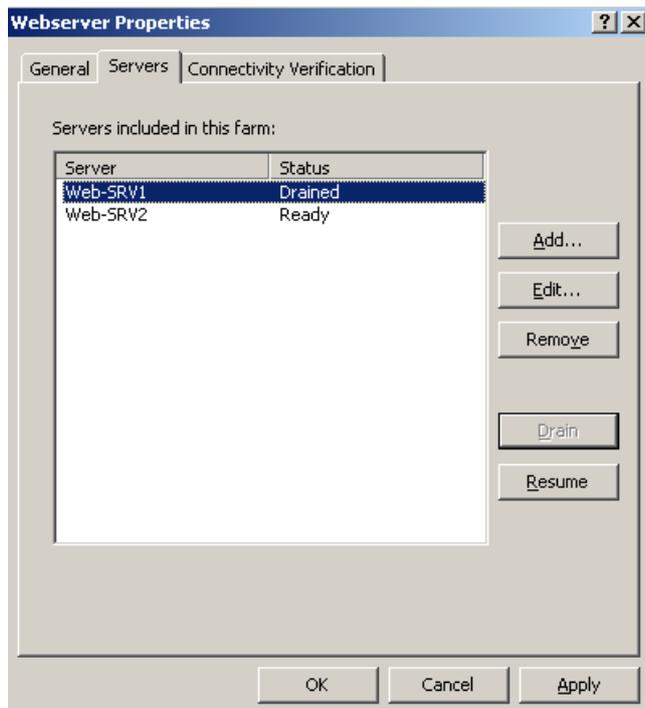


Figure 12: Web Server Farm connection properties

Navigate to the Servers tab and again click *Edit* to change the server properties or to place one Server manually into maintenance mode if you select the *Drain* button. For session based affinity, the server will continue to handle current sessions but will not accept new connections. If you are using IP based affinity a drained server stops receiving requests, but existing connections to that server are still handled.

Conclusion

In this article, I tried to give you a step by step guide solution how to enable ISA Server Web Server Load Balancing to load balance web traffic to several internal web servers without using a classic Hardware Load Balancer or a NLB (Network Load Balancing) solution based on Windows Server 2003. In my opinion, the Web Server Load Balancing feature of ISA Server 2006 is a nice feature for a limited number of Web Servers with basic functionality. A traditional Hardware Load Balancer might have a few more advanced features.

Related links

Web Server Farm Load Balancing in ISA Server 2006

<http://technet.microsoft.com/en-us/library/bb794841.aspx>

Network Load Balancing Integration Concepts for Microsoft Internet Security and Acceleration (ISA) Server 2006

<http://technet.microsoft.com/en-us/library/bb794741.aspx>

How to load balance a Web server farm by using one SSL certificate in IIS 6.0 and in IIS 5.0

<http://support.microsoft.com/kb/313299>