Load Balancing Microsoft Terminal Services

v1.2.1

Deployment Guide

NOTE: This guide has been archived and is no longer being maintained. While the content is still valid for the particular software versions mentioned, it may refer to outdated software that has now reached end-of-life. For more information please contact support@loadbalancer.org.

loadbalancer.org

Contents

1. About this Guide

This guide details the steps required to configure a load balanced Microsoft Terminal Services environment utilizing Loadbalancer.org appliances. It covers the configuration of the load balancers and also any Microsoft Terminal Services configuration changes that are required to enable load balancing.

For more information about initial appliance deployment, network configuration and using the Web User Interface (WebUI), please also refer to the relevant Administration Manual:

- v7 Administration Manual
- v8 Administration Manual

2. Loadbalancer.org Appliances Supported

All our products can be used with Terminal Services. The complete list of models is shown below:

Discontinued Models	Current Models *
Enterprise R16	Enterprise R20
Enterprise VA R16	Enterprise MAX
Enterprise VA	Enterprise 10G
Enterprise R320	Enterprise 40G
	Enterprise Ultra
	Enterprise VA R20
	Enterprise VA MAX
	Enterprise AWS
	Enterprise AZURE **
	Enterprise GCP **

* For full specifications of these models please refer to: http://www.loadbalancer.org/products/hardware

** Some features may not be supported, please check with Loadbalancer.org support

3. Loadbalancer.org Software Versions Supported

- v7.6.4 and later

4. Microsoft Windows Versions Supported

- Windows 2000 to Windows 2008 R1

5. Microsoft Terminal Services

Terminal Services is one of the components of Microsoft Windows that allows a user to access applications and data on a remote computer over a network. Terminal services is Microsoft's implementation of thin-client Terminal Server computing, where Windows applications, or even the entire desktop of the computer running terminal services, are made accessible to a remote client machine.

6. Load Balancing Terminal Services

> Note: It's highly recommended that you have a working Terminal Server environment first before implementing the load balancer.

The Basics

Session Load Balancing

The fundamental purpose of deploying a load balancer is to share the load from multiple clients between two or more back-end Terminal Servers. Typically, all Terminal Servers within the cluster/farm have the same applications installed to ensure all clients get the same applications irrespective of which server they are connected to.

Session Persistence (aka Server Affinity)

A critical aspect of load balancing terminal services is session persistence. Within a Terminal Server environment, this relates to the ability to reconnect to disconnected sessions that occur when a client session is closed rather than logged off. If this reconnection process is not handled correctly, users may not be able to return to their previous sessions.

Port Requirements

The following table shows the ports that must be load balanced.

Port	Description
3389	RDP Protocol

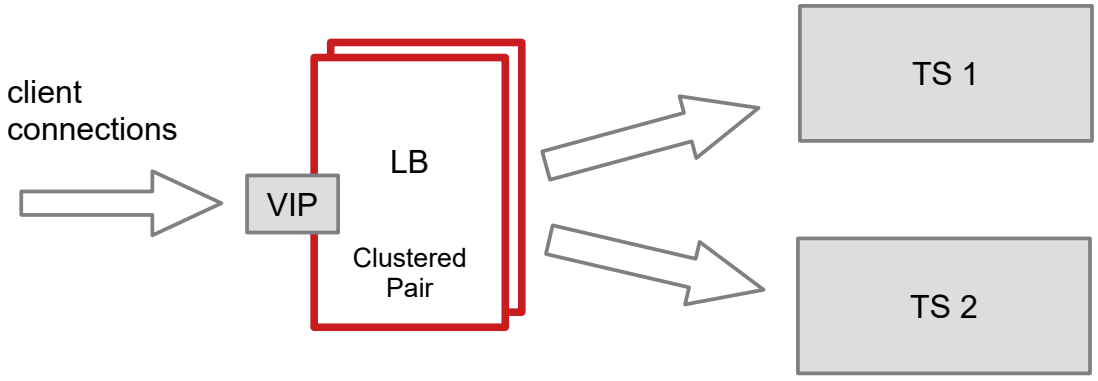
> Note: It is possible to change the port used, but the default is 3389.

Load Balancer Deployment

The load balancer is deployed in front of the Terminal Servers to provide load balancing and fail-over functionality.

Once deployed, clients then connect to the Virtual Service (VIP) on the load balancer rather than connecting directly to a one of the Terminal Servers. These connections are then load balanced across the Terminal Servers to distribute the

load according to the load balancing algorithm selected.



VIPs = Virtual IP Addresses

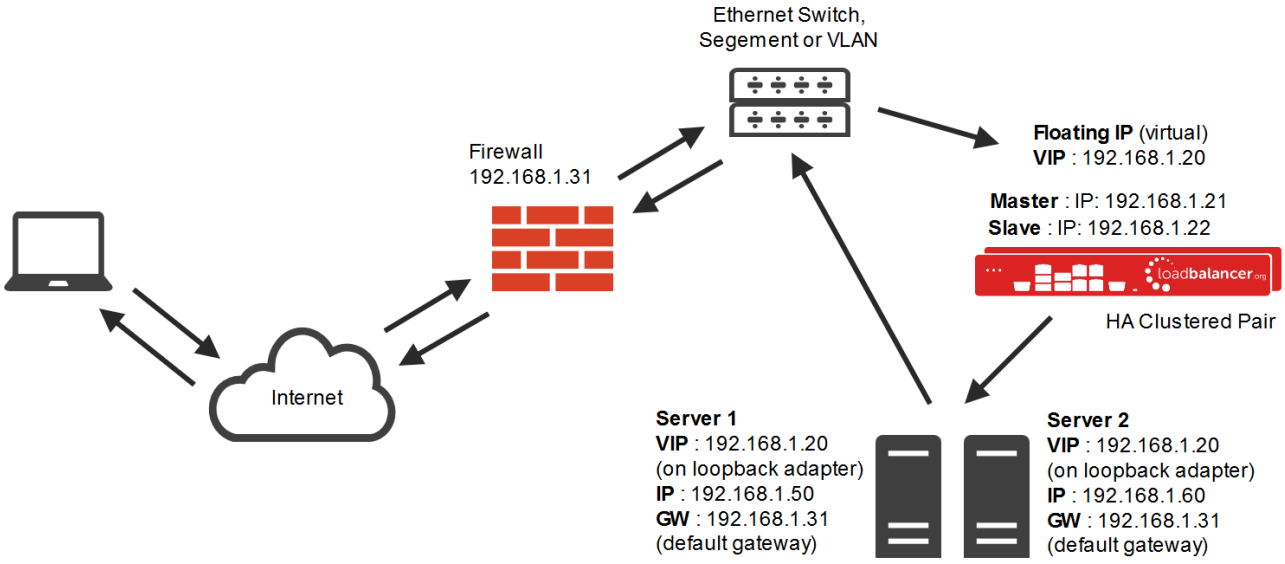
> Note: The load balancer can be deployed as a single unit, although Loadbalancer.org recommends a clustered pair for resilience & high availability. Please refer to section 1 in the appendix on page 30 for more details on configuring a clustered pair.

Load Balancer Deployment Modes

The load balancer can be deployed in 4 fundamental ways: *Layer 4 DR mode*, *Layer 4 NAT mode*, *Layer 4 SNAT mode* and *Layer 7 SNAT mode*. For Terminal Services, Layer 4 DR mode, Layer 4 NAT mode and Layer 7 SNAT are typically used. These are described below.

Layer 4 Direct Server Return (DR Mode)

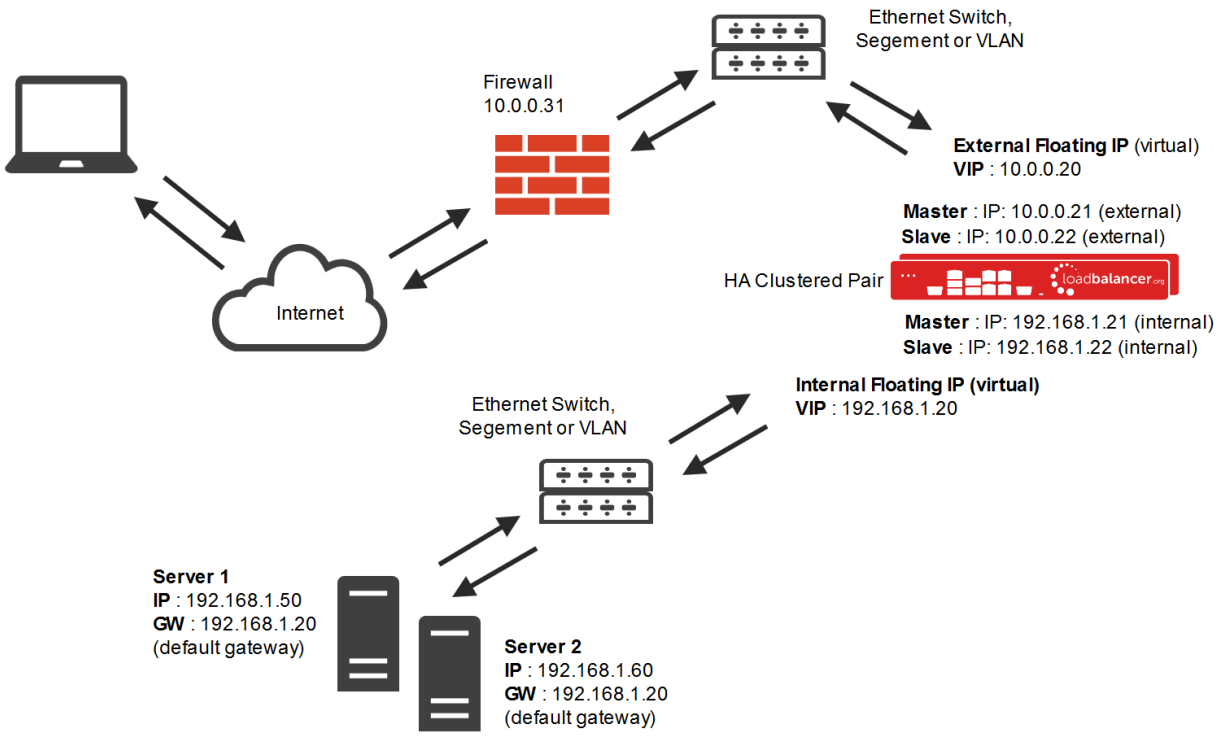
One-arm Direct Routing (DR) mode is a very high performance solution that requires little change to your existing infrastructure.



- Direct Routing mode works by changing the destination MAC address of the incoming packet to match the selected Real Server on the fly which is very fast
- When the packet reaches the Real Server it expects it to own the Virtual Services IP address (VIP). This means that you need to ensure that the Real Server (and the load balanced application) respond to both the Real Servers own IP address and the VIP
- The Real Server should not respond to ARP requests for the VIP. Only the load balancer should do this. Configuring the Real Servers in this way is referred to as *Solving the ARP Problem.* please refer to chapter 6 in the Administration Manual for more information
- On average, DR mode is 8 times quicker than NAT for HTTP, 50 times quicker for Terminal Services and much, much faster for streaming media or FTP
- The load balancer must have an Interface in the same subnet as the Real Servers to ensure layer 2 connectivity required for DR mode to work
- The VIP can be brought up on the same subnet as the Real Servers, or on a different subnet provided that the load balancer has an interface in that subnet
- Port translation is not possible in DR mode i.e. having a different RIP port than the VIP port
- DR mode is transparent, i.e. the Real Server will see the source IP address of the client

Layer 4 Network Address Translation (NAT Mode)

Sometimes it's not possible to use DR mode. The two most common reasons being: if the application cannot bind to the RIP & VIP at the same time; or if the host operating system cannot be modified to handle the ARP problem. The second choice is Network Address Translation (NAT) mode. This is also a high performance solution but it requires the implementation of a two arm infrastructure with an internal and external subnet to carry out the translation (the same way a firewall works).



- The load balancer translates all requests from the external Virtual Service to the internal Real Servers
- Normally eth0 is used for the *internal* network and eth1 is used for the *external* network although this is not

mandatory. If the Real Servers require Internet access, Autonat should be enabled using the WebUI option: *Cluster Configuration > Layer 4 – Advanced Configuration*, the external interface should be selected

- NAT mode can be deployed in the following ways:

  2-arm (using 2 Interfaces), 2 subnets (as shown above) - One interface on the load balancer is connected to subnet1 and the second interface and Real Servers are connected to subnet2. The VIP is brought up in subnet1. The default gateway on the Real Servers is set to be an IP address in subnet2 on the load balancer. Clients can be located in subnet1 or any remote subnet provided they can route to the VIP

  2-arm (using 1 Interface), 2 subnets - same as above except that a single interface on the load balancer is allocated 2 IP addresses, one in each subnet

  1-arm (using 1 Interface), 1 subnet - Here, the VIP is brought up in the same subnet as the Real Servers. For clients located in remote networks the default gateway on the Real Servers must be set to be an IP address on the load balancer. For clients located on the same subnet, return traffic would normally be sent directly to the client bypassing the load balancer which would break NAT mode. To address this, the routing table on the Real Servers must be modified to force return traffic to go via the load balancer - for more details on 'One-Arm NAT Mode' please refer to chapter 6 in the Administration Manual

- If you want Real Servers to be accessible on their own IP address for non-load balanced services, e.g. SMTP or RDP, you will need to setup individual SNAT and DNAT firewall script rules for each Real Server or add additional VIPs for this - please refer to chapter 6 in the Administration Manual
- NAT mode is transparent, i.e. the Real Server will see the source IP address of the client
- Port translation is possible in NAT mode, i.e. VIP:80 → RIP8080 is possible

Layer 4 Persistence Methods

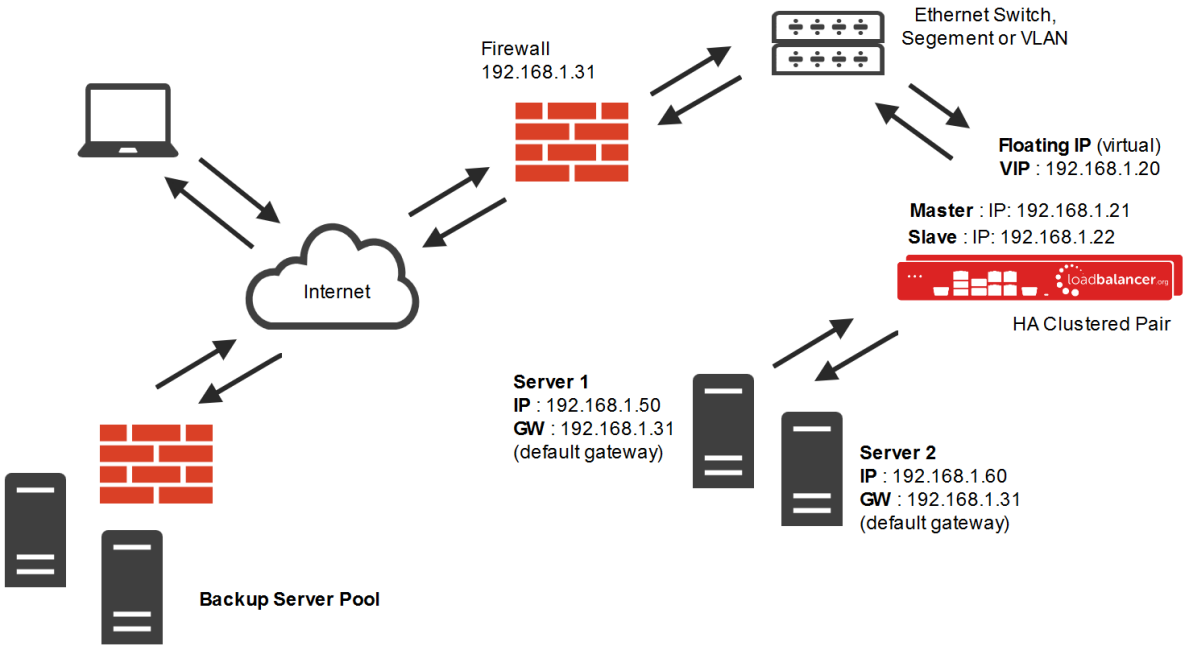
Source IP Persistence

Layer 4 methods only support source IP address based persistence (affinity) for reconnecting user sessions to the same backend server. This works very well in many situations, but in cases where clients connect via some form of NAT device, then these methods may not be appropriate because the source IP address for all clients would be the same. If this is the case, layer 7 methods can be used instead (see page 8).

Layer 7 SNAT Mode

Layer 7 load balancing uses a proxy (HAProxy) at the application layer. Inbound requests are terminated on the load balancer, and HAProxy generates a new request to the chosen real server. As a result, Layer 7 is a slower technique than DR or NAT mode at Layer 4. Layer 7 is generally chosen when the network topology prohibits the use of the layer 4 methods.

Single–arm and two–arm configurations are supported as shown below. In both cases return traffic passes via the load balancer. Since layer 7 works as a proxy, there is not need to set the appliance as the gateway.



This mode has the advantage of a one arm configuration and does not require any changes to the application servers. However, since the load balancer is acting as a full proxy it doesn't have the same raw throughput as the layer 4 methods.

The load balancer proxies the application traffic to the servers so that the source of all traffic becomes the load balancer.

- SNAT mode is a full proxy and therefore load balanced Real Servers do not need to be changed in any way
- Because SNAT mode is a full proxy any server in the cluster can be on any accessible subnet including across the Internet or WAN
- SNAT mode is not transparent by default, i.e. the Real Servers will not see the source IP address of the client, they will see the load balancers own IP address by default, or any other local appliance IP address if preferred (e.g. the VIP address), this can be configured per layer 7 VIP
- SNAT mode can be deployed using either a 1-arm or 2-arm configuration

Layer 7 Persistence Methods

Three persistence methods are supported to ensure that clients can reconnect to their sessions. These are '*Source IP Persistence*', '*Microsoft Session Directory Persistence*' and '*RDP Cookie Persistence*' and are described in the following sections.

Source IP Persistence

As at Layer 4, this method is appropriate when client PC's have unique IP addresses. This method is very straight forward to configure and requires no changes to the Terminal Servers.

Microsoft Session Directory/Broker Persistence

Session Directory/Broker provides functionality that allows a group of Terminal Servers to coordinate the reconnection of disconnected sessions. All sessions are stored as records in a central database. This database is updated and

queried by the Terminal Servers whenever users log on, log off, or disconnect their session, while leaving their applications active.

The load balancer is able to interact with Session Broker by enabling Routing Token Redirection mode. This mode allows the reconnection of disconnected sessions by utilizing a routing token to enable the load balancer to re-connect the client to the correct Terminal Server.

WIN 2000/2003/2008 R1: For these versions of Windows clustering must be used to provide HA for the session database. If this is not done the session database is vulnerable to failure and therefore data loss resulting in the inability for disconnected sessions to reconnected correctly.

> Note: Session Directory was renamed as Session Broker in Windows 2008 R1.

RDP Cookie Persistence

This method utilizes the cookie sent from the client in the Connection Request PDU. This cookie is created when the username is entered at the first client login prompt (mstsc.exe). If the username is not entered here, the cookie is not created.

The cookie only supports up to 9 characters, so this method may have limited use, especially in cases where users login using the domain\username format. In this case, if the domain name was 9 characters in length, the RDP cookie would be the same for all users, resulting in all sessions being sent to the same session host. If users login using the UPN format (User Principle Name), i.e. **username@domain**, it's more likely to be unique.

> Note: When RDP cookie persistence is selected, the load balancer will attempt to use RDP cookie persistence, but if a cookie is not found, source IP persistence will be used instead as a fallback.

> Note: In certain scenarios depending on client version as well as the specific client & server settings, the RDP cookie (mstshash) is not consistently sent. Please also refer to our blog post on this topic: http://www.loadbalancer.org/blog/microsoft-drops-support-for-mstshash-cookies
>
> *Update (October 2015)* – with the latest versions of Windows Servers & RDP Client, this problem appears to have been resolved.

Which Mode Should I Use?

Mode Summary

Layer 4 DR Mode offers the best performance and requires limited physical Real Server changes. The load balanced application must be able to bind to the Real Servers own IP address and the VIP at the same time. This mode requires the

"ARP Problem" to be solved.

Layer 4 NAT Mode is also a high performance solution but not as fast as DR mode. It requires the implementation of a two-arm infrastructure with an internal and external subnet to carry out the translation (the same way a firewall works). Also each Real Server must use the load balancer as the default gateway.

Layer 7 SNAT Mode offers greater flexibility but at lower performance levels. It supports HTTP cookie insertion, RDP cookies, Connection Broker integration and works very well with either Pound or STunnel when SSL termination is required. It does not require any changes to the Real Servers and can be deployed in one-arm or two-arm mode and. HAProxy is a high performance solution, but since it operates as a full proxy, it cannot perform as fast as the layer 4 solutions.

Our Recommendation

Where possible we recommend that Layer 4 Direct Routing (DR) mode with source IP persistence is used. This mode offers the best possible performance since replies go directly from the Real Servers to the client, not via the load balancer. It's also relatively simple to implement. Ultimately, the final choice does depend on your specific requirements and infrastructure.

7. Loadbalancer.org Appliance – the Basics

Virtual Appliance Download & Deployment

A fully featured, fully supported 30 day trial is available if you are conducting a PoC (Proof of Concept) deployment. The VA is currently available for VMware, Virtual Box, Hyper-V, KVM and XEN and has been optimized for each Hypervisor. By default, the VA is allocated 1 CPU, 2GB of RAM and has an 8GB virtual disk. The Virtual Appliance can be downloaded here.

> Note: The same download is used for the licensed product, the only difference is that a license key file (supplied by our sales team when the product is purchased) must be applied using the appliance's WebUI.

> Note: Please refer to the Administration Manual and the ReadMe.txt text file included in the VA download for more detailed information on deploying the VA using various Hypervisors.

Initial Network Configuration

The IP address, subnet mask, default gateway and DNS settings can be configured in several ways as detailed below:

Method 1 - Using the Network Setup Wizard at the console

After boot up, follow the instructions on the console to configure the IP address, subnet mask, default gateway and DNS settings.

Method 2 - Using the WebUI

Using a browser, connect to the WebUI on the default IP address/port: **https://192.168.2.21:9443**

To set the IP address & subnet mask, use: *Local Configuration > Network Interface Configuration*

To set the default gateway, use: *Local Configuration > Routing*

To configure DNS settings, use: *Local Configuration > Hostname & DNS*

Accessing the Web User Interface (WebUI)

1. Browse to the following URL: **https://192.168.2.21:9443/lbadmin/**

   (replace with your IP address if it's been changed)

   ** Note the port number → **9443***

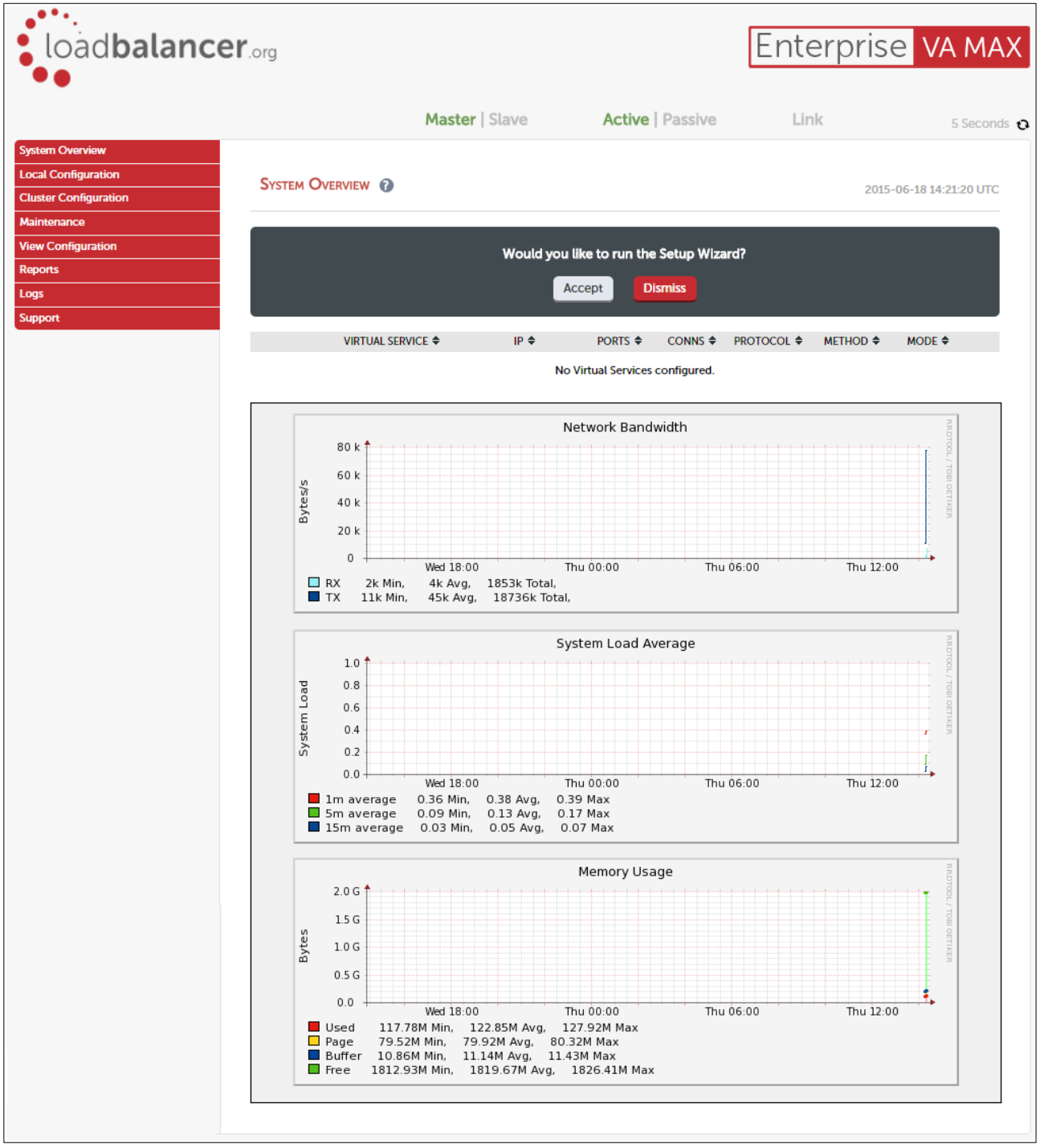
2. Login to the WebUI:

Username: loadbalancer

Password: loadbalancer

> Note: To change the password , use the WebUI menu option: *Maintenance > Passwords.*

Once logged in, the WebUI will be displayed as shown below:



HA Clustered Pair Configuration

Loadbalancer.org recommend that load balancer appliances are deployed in pairs for high availability. In this guide a single unit is deployed first, adding a secondary slave unit is covered in section 1 of the Appendix on page 30.

8. Load Balancing Terminal Servers

EXAMPLE 1 – Layer 4 DR Mode (Using Source IP Persistence)

Overview

- **Configure the network interface –** A single Interface is required, eth0 is normally used in one-arm deployments, however this is not mandatory
- **Configure the Virtual Service (VIP) –** This is created on the load balancer and is the cluster address through which all back-end Terminal Servers are accessed
- **Configure the Real Servers (RIPs)** – Define the Terminal Servers that make up the cluster
- **Configure the Terminal Servers** – In DR mode, the ARP issue must be solved on each Terminal Server

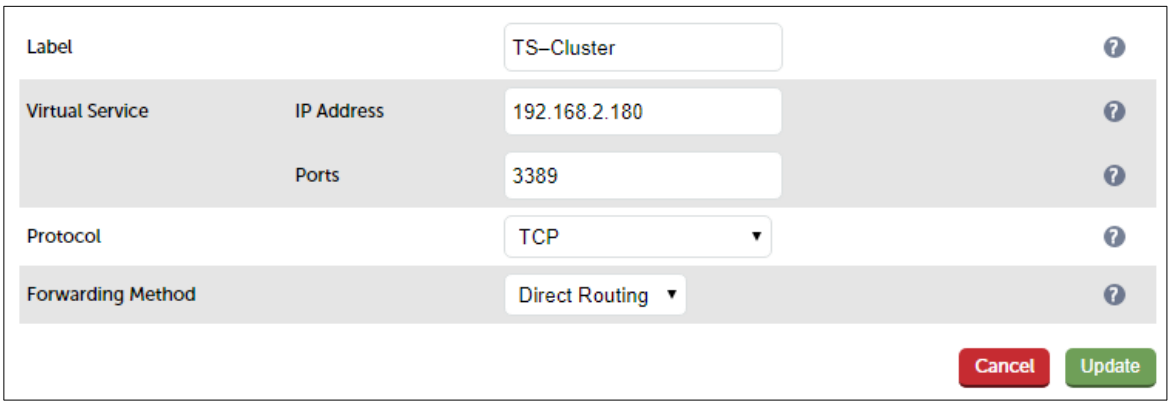
Appliance Configuration

Configure the Network Interface

1. One interface is required. Page 12 of this guide covers the various methods available to configure network settings.

Configure the Virtual Service

1. Using the WebUI, navigate to: *Cluster Configuration > Layer 4 – Virtual Services* and click **Add a New Virtual Service**
2. Enter the following details:



3. Enter an appropriate name (label) for the VIP, e.g. **TS-Cluster**
4. Set the *Virtual Service IP address* field to the required IP address, e.g. **192.168.2.180**
5. Set the *Virtual Service Ports* field to **3389**
6. Ensure that *Protocol* is set to **TCP**
7. Ensure that *Forwarding Method* is set to **Direct Routing**
8. Click **Update**

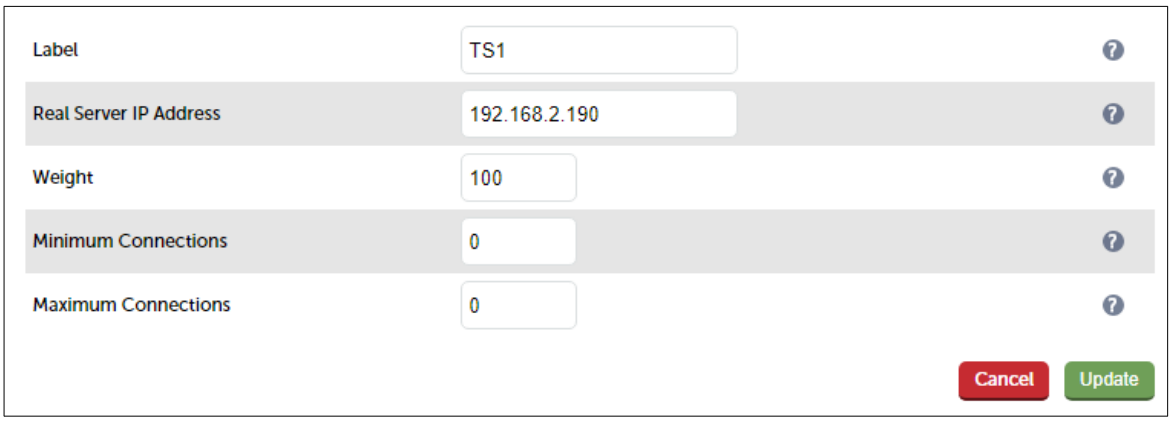
9. Now click **Modify** next to the newly created Virtual Service
10. Ensure *Persistent* is enabled
11. Set *Persistence Timeout* to an appropriate value, e.g. **3600** (i.e. 1 hour)

> Note: This is the time that the load balancer tracks the client IP to Terminal Server mapping and should typically be set to be the same as the RDP idle session timeout configured on the servers.

12. Click **Update**

Define the Real Servers

1. Using the WebUI, navigate to: *Cluster Configuration > Layer 4 – Real Servers* and click **Add a New Real Server** next to the newly created Virtual Service
2. Enter the following details:



3. Enter an appropriate label (name) for the RIP, e.g. **TS1**
4. Change the *Real Server IP Address* field to the required IP address, e.g. **192.168.2.190**
5. Click **Update**
6. Repeat for your remaining Terminal Server(s)

Terminal Server Configuration

Solve the 'ARP Problem'

For Windows 2000, 2003, 2008 & 2012 a Loopback adapter must be added to each connection server to enable them to accept traffic destined for the VIP. Also, for Windows 2008 a series of 3 netsh commands must also be run to configure the strong/weak host behavior.

For more details on solving the ARP problem, please refer to page 35

EXAMPLE 2 – Layer 4 NAT Mode (Using Source IP Persistence)

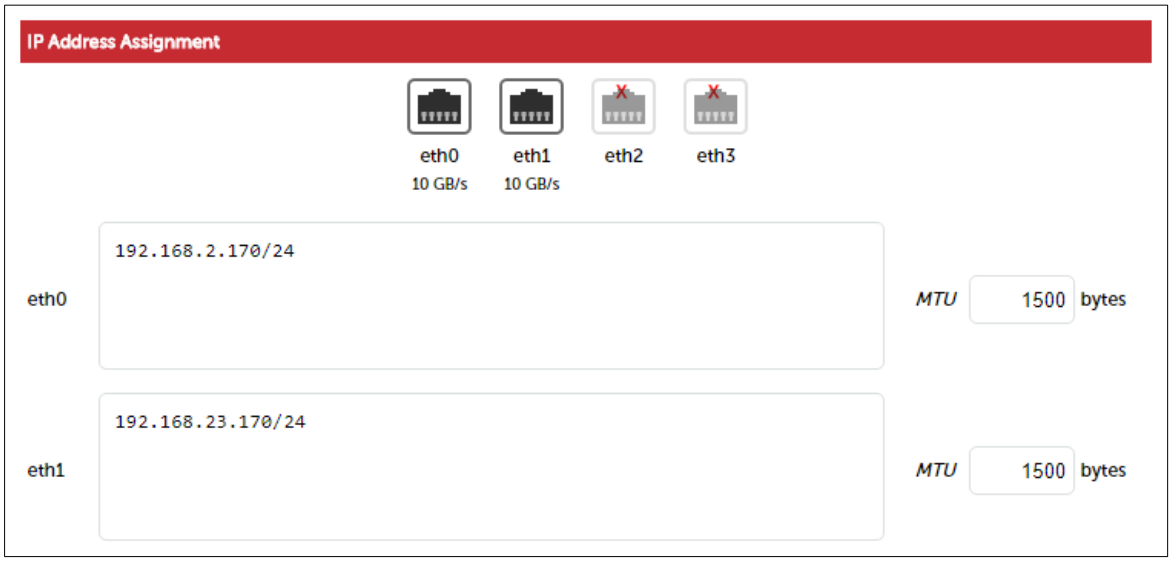
Overview

- **Configure the Network Interfaces –** Two Interfaces are needed, this can be achieved by using two network adapters, or by creating VLANs on a single adapter
- **Configure the Virtual Service (VIP) –** This is created on the load balancer and is the cluster address through which all back–end Terminal Servers are accessed
- **Configure the Real Servers (RIPs)** – Define the Terminal Servers that make up the cluster
- **Configure the Terminal Servers** – In NAT mode, the Terminal Servers default gateway must be the load balancer

Appliance Configuration

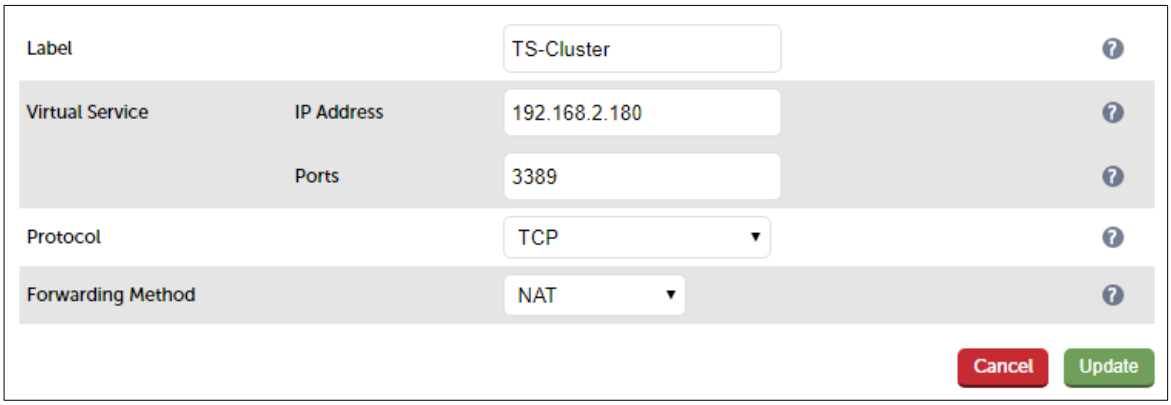
Configure the Network Interfaces

1. Set the first IP address for eth0 using one of the methods listed on page 12
2. Using the WebUI, navigate to: *Local Configuration > Network Interface Configuration*
3. Define an additional IP address for a second interface, e.g. *eth1* , in a different subnet as shown below:



Configure the Virtual Service

1. Using the WebUI, navigate to: *Cluster Configuration > Layer 4 – Virtual Services* and click **Add a New Virtual Service**
2. Enter the following details:



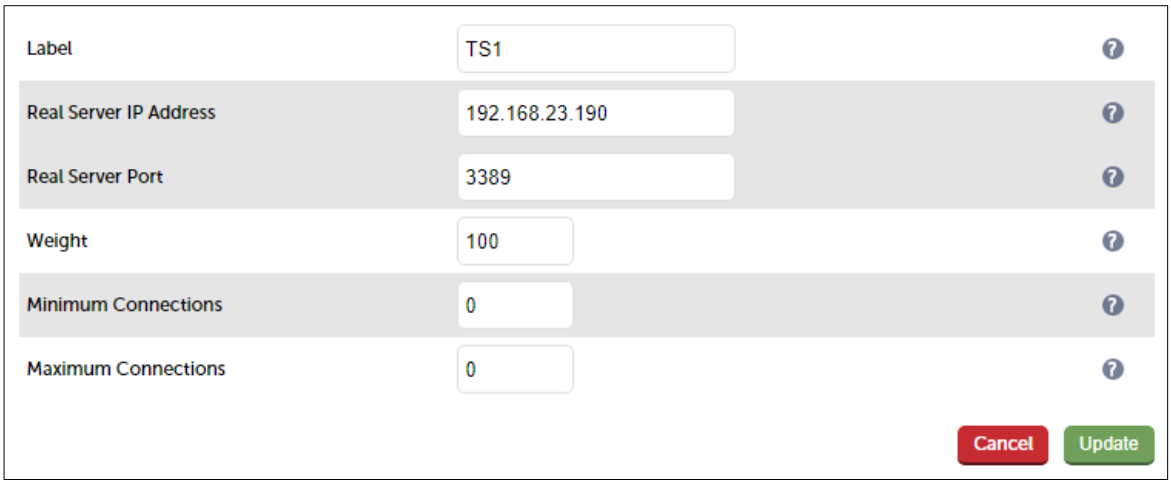
3. Enter an appropriate label (name) for the VIP, e.g. **TS–Cluster**
4. Set the *Virtual Service IP address* field to the required IP address, e.g. **192.168.2.180**
5. Set the *Virtual Service Ports* field to **3389**
6. Ensure that *Protocol* is set to **TCP**
7. Ensure that *Forwarding Method* is set to **NAT**
8. Click **Update**
9. Now click **Modify** next to the newly created Virtual Service
10. Ensure *Persistent* is enabled
11. Set *Persistence Timeout* to an appropriate value, e.g. **3600** (i.e. 1 hour)

> Note: This is the time that the load balancer tracks the client IP to Terminal Server mapping and should typically be set to be the same as the RDP idle session timeout configured on the servers.

12. Click Update

Configure the Real Servers

1. Using the WebUI, navigate to: *Cluster Configuration > Layer 4 – Real Servers* and click **Add a New Real Server** next to the newly created Virtual Service
2. Enter the following details:



3. Enter an appropriate label (name) for the RIP, e.g. **TS1**
4. Change the *Real Server IP Address* field to the required IP address, e.g. **192.168.23.190**
5. Change the *Real Server Port* to **3389**
6. Click **Update**
7. Repeat for your remaining Terminal Server(s)

Terminal Server Configuration

Default Gateway

It is possible to use the internal IP address on eth0 for the default gateway, although it's recommended that an additional floating IP is created for this purpose. This is required if two load balancers (our recommended configuration) are used. If the master unit fails, this will enable the floating IP to be brought up on the slave..

To create a floating IP address on the load balancer:

1. Using the WebUI, navigate to: *Cluster Configuration > Floating IP(s)*
2. Enter the required IP address to be used for the default gateway and click **Add Floating IP**
3. Once added, there will be two floating IP's, one for the Virtual Service (**192.168.2.180**) and one for the default gateway (e.g. **192.168.23.254**) as shown below:

4. Now configure your Terminal Servers to use this additional IP address as their default gateway

EXAMPLE 3 – Layer 7 SNAT Mode (Using Source IP Persistence)

Overview

- **Configure the Network Interface(s) –** HAProxy can be deployed in single–arm or two–arm mode. As with layer 4 NAT mode, with a two–arm Layer 7 configuration, this can be achieved by using two network adapters, or by creating VLANs on a single adapter
- **Configure the Virtual Service (VIP) –** This is created on the load balancer and is the cluster address through which all back–end Terminal Servers are accessed
- **Configure the Real Servers (RIPs)** – Define the Terminal Servers that make up the cluster
- **Configure the Terminal Servers** – No Terminal Server changes are required for SNAT mode

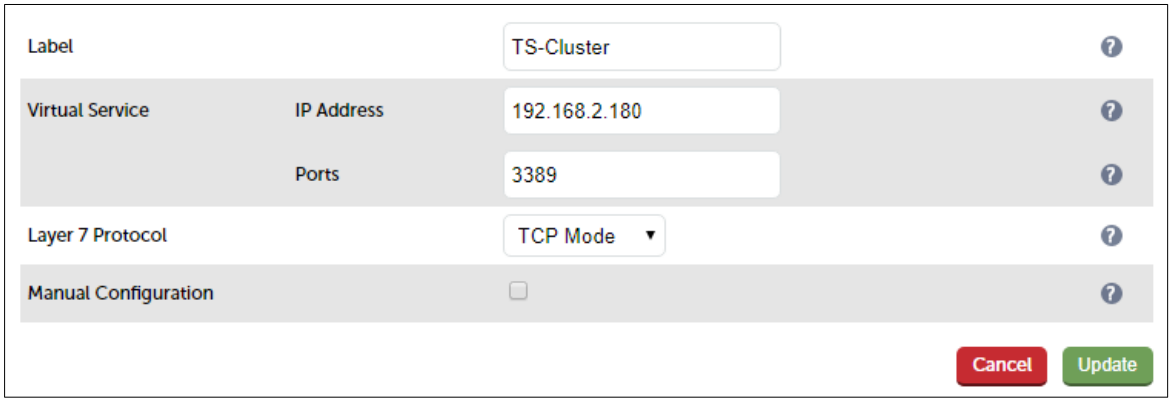
Appliance Configuration (single–arm Example)

Configure the Network Interface

1. One interface is required. Page 12 of this guide covers the various methods available to configure network settings.

Configure the Virtual Service (VIP)

1. Using the WebUI, navigate to: *Cluster Configuration > Layer 7 – Virtual Services* and click **Add a New Virtual Service**
2. Enter the following details:



3. Enter an appropriate name (Label) for the Virtual Service, e.g. **TS–Cluster**
4. Set the *Virtual Service IP address* field to the required IP address, e.g. **192.168.2.180**
5. Set the *Virtual Service Ports* field to **3389**
6. Click **Update**
7. Now click **Modify** next to the newly created Virtual Service

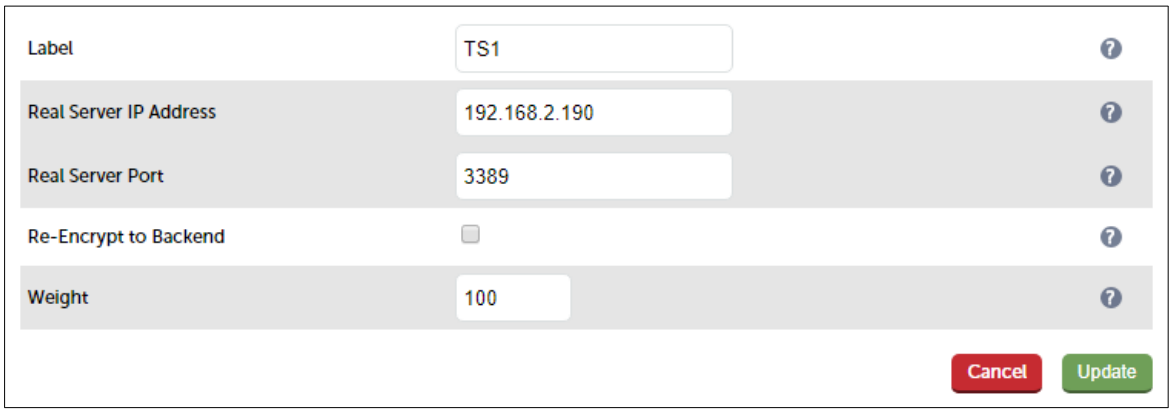
8. Under the *Persistence* section, click **Advanced** to show more options
9. Ensure *Persistence Mode* is set to **Source IP**
10. Set *Persistence Timeout* to an appropriate value, e.g. **3600** (i.e. 1 hour)

> Note: This is the time that the load balancer tracks the client IP to Terminal Server mapping and should typically be set to be the same as the RDP idle session timeout configured on the servers.

11. Click Update

Configure the Real Servers (RIPs)

1. Using the WebUI, navigate to: *Cluster Configuration* > *Layer 7 – Real Servers* and click **Add a New Real Server** next to the newly created Virtual Service
2. Enter the following details:



3. Enter an appropriate name (Label) for the first Terminal Server, e.g. **TS1**
4. Change the *Real Server IP Address* field to the required IP address, e.g. **192.168.2.190**
5. Set the *Real Server Port* field to **3389**
6. Click **Update**
7. Now repeat for your remaining Terminal Server(s)

EXAMPLE 4 – Layer 7 SNAT Mode (Using RDP Cookie Persistence)

Overview

- **Configure the Network Interface(s) –** HAProxy can be deployed in single–arm or two–arm mode. As with layer 4 NAT mode, with a two–arm Layer 7 configuration, this can be achieved by using two network adapters, or by creating VLANs on a single adapter
- **Configure the Virtual Service (VIP) –** This is created on the load balancer and is the cluster address through which all back–end Terminal Servers are accessed

- **Configure the Real Servers (RIPs)** – Define the Terminal Servers that make up the cluster
- **Configure the Terminal Servers** – No Terminal Server changes are required to support SNAT mode

Appliance Configuration (single–arm Example)

Configure the Network Interface

Please refer to the previous example.

Configure HAProxy Timeouts

Please refer to the previous example.

Configure the Virtual Service (VIP)

Please refer to the previous example.

> Note: When configuring persistence, choose *RDP Client Cookie* rather than *Source IP*.

Configure the Real Servers (RIPs)

Please refer to the previous example.

Terminal Server Configuration

No changes are required to the Terminal Servers.

EXAMPLE 5 – Layer 7 SNAT Mode (Using Connection Broker Persistence)

(In Windows 2008 R1 this is known as Session Broker, in Windows 2003 and earlier as Session Directory)

Overview

- **Configure the Network Interface(s) –** HAProxy can be deployed in single–arm or two–arm mode. As with layer 4 NAT mode, with a two–arm Layer 7 configuration, this can be achieved by using two network adapters, or by creating VLANs on a single adapter
- **Configure the Virtual Service (VIP) –** This is created on the load balancer and is the cluster address through which all back–end Terminal Servers are accessed
- **Configure the Real Servers (RIPs)** – Define the Terminal Servers that make up the cluster
- **Configure the Terminal Servers** – No Terminal Server changes are required to support SNAT mode although the back–end servers must be configured to use session broker in Routing token re–direction mode

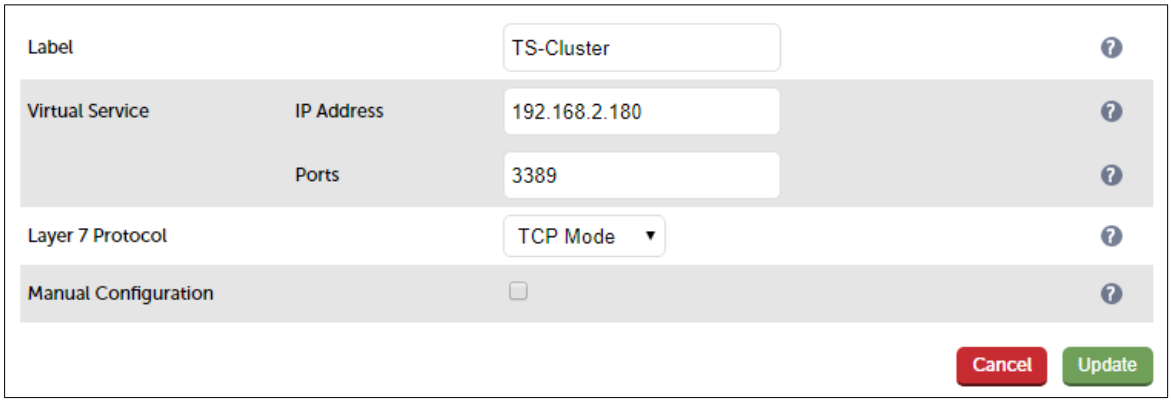
Appliance Configuration (single-arm Example)

Configure the Network Interface

1. One interface is required. Page 12 of this guide covers the various methods available to configure the network settings.

Configure the Virtual Service (VIP)

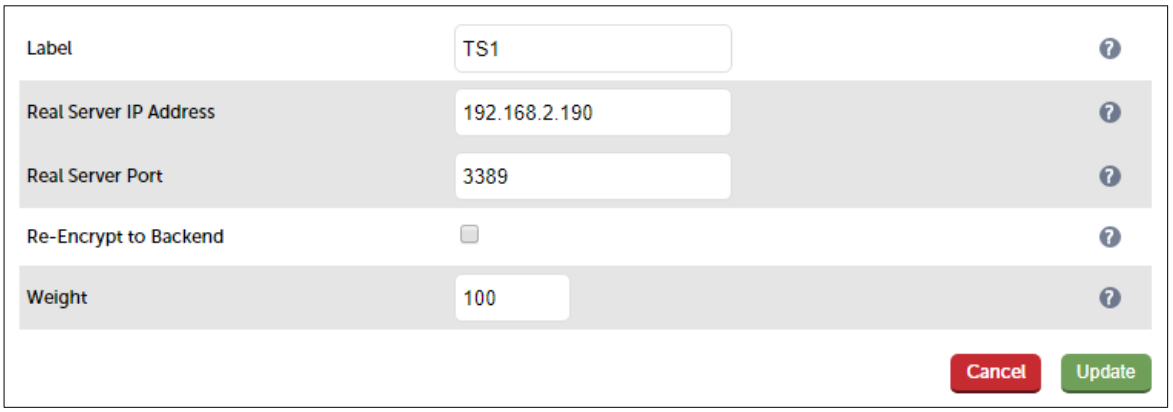
1. Using the WebUI, navigate to: *Cluster Configuration > Layer 7 – Virtual Services* and click **Add a New Virtual Service**
2. Enter the following details:



3. Enter an appropriate name (Label) for the Virtual Service, e.g. **TS-Cluster**
4. Set the *Virtual Service IP address* field to the required IP address, e.g. **192.168.2.180**
5. Set the *Virtual Service Ports* field to **3389**
6. Click **Update**
7. Now click Modify next to the newly created Virtual Service
8. Change *Persistence Mode* to **MS Session Broker**
9. Click **Update**

Configure the Real Servers (RIPs)

1. Using the WebUI, navigate to: *Cluster Configuration > Layer 7 – Real Servers* and click **Add a New Real Server** next to the newly created Virtual Service
2. Enter the following details:



3. Enter an appropriate name (Label) for the first Terminal Server, e.g. **TS1**
4. Change the *Real Server IP Address* field to the required IP address (e.g. **192.168.2.190)**
5. Set the *Real Server Port* field to **3389**
6. Click **Update**
7. Now repeat for your remaining Terminal Server(s)

Terminal Server Configuration

Windows 2003

The Terminal Services Session Directory service should be started on the server designated for this purpose. (Note that Windows 2003 Enterprise Edition is required to support Session Directory)

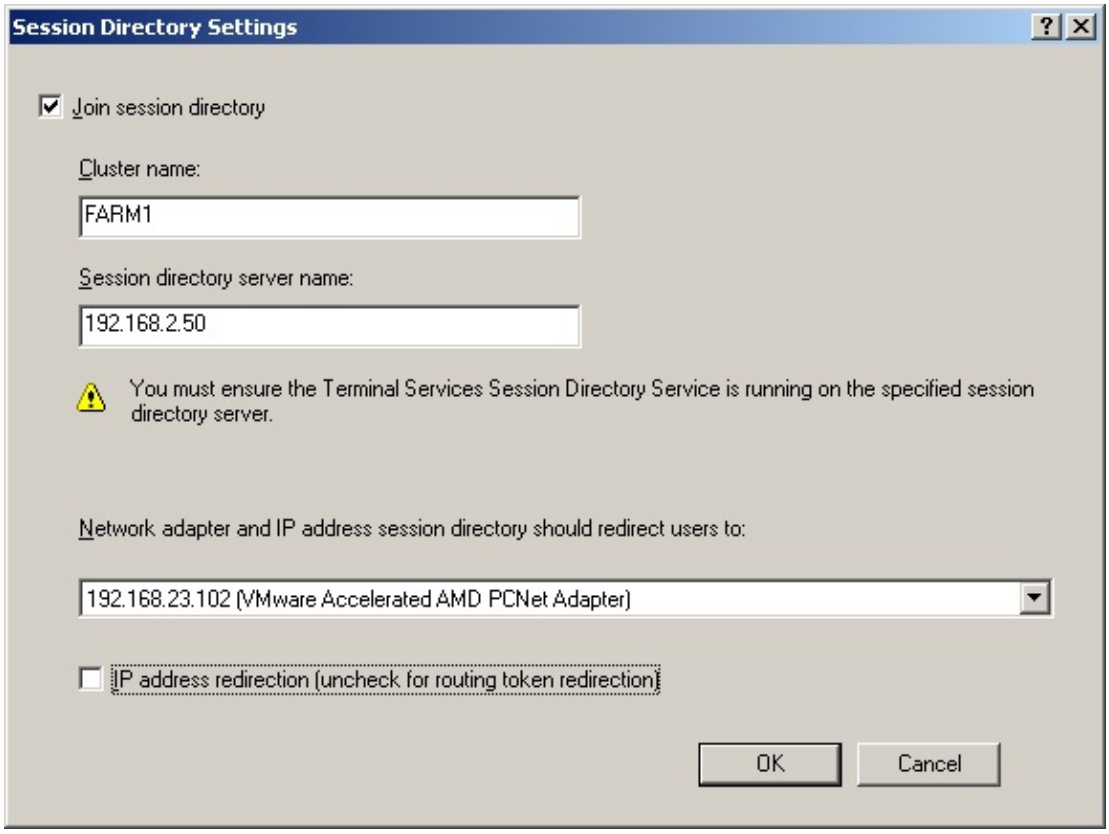
Then on each Terminal Server to be included in the cluster/Farm:

1. Open Terminal Services Configuration
2. Click Server Settings
3. Right-click Session Directory and select Properties
4. Tick the check box to Join Session Directory
5. Enter a name for the Terminal Server Cluster (Farm), e.g. FARM1 (all servers within the same farm require the same name to be specified)
6. Enter the DNS name or IP address for the server running the Session Directory service
7. Un-select the **IP Address Redirection** check-box

> Note: This is a critical step which enables Routing Token Redirection Mode. In this mode the load balancer is able to interact with routing tokens from the client to determine which real server is running a previously disconnected session.

8. Click **OK**

Session Directory Settings:



Windows 2008 R1

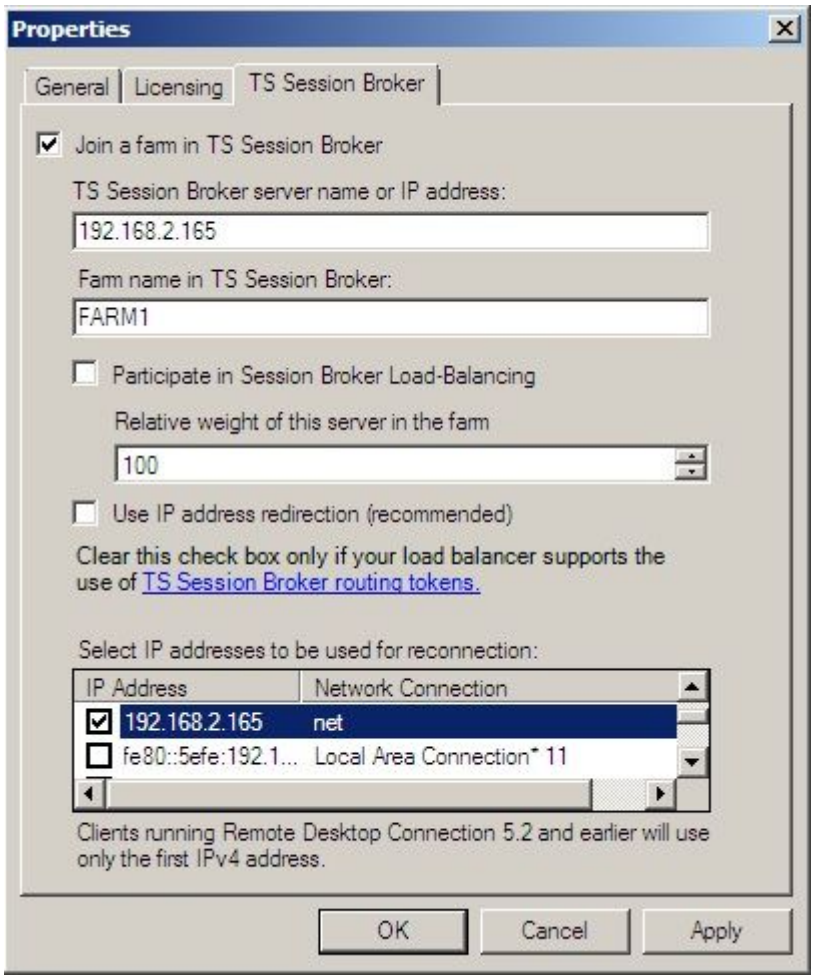
Install Session Broker on the server designated to hold the Session Broker role. Then on each Terminal Server to be included in the cluster/Farm:

1. Open Terminal Services Configuration
2. Right–click 'Member of farm in TS Session Broker' and select Properties
3. Tick the check box to Join a farm in TS Session Broker
4. Enter the DNS name or IP address for the server running the Session Broker Role Service
5. Enter a name for the Terminal Server Cluster (Farm), e.g. FARM1 (all servers within the same farm require the same name to be specified)
6. Un–select the **Use IP Address Redirection** check–box.

> Note: This is a critical step which enables Routing Token Redirection Mode. In this mode the load balancer is able to interact with routing tokens from the client to determine which real server is running a previously disconnected session.

7. Click **OK**

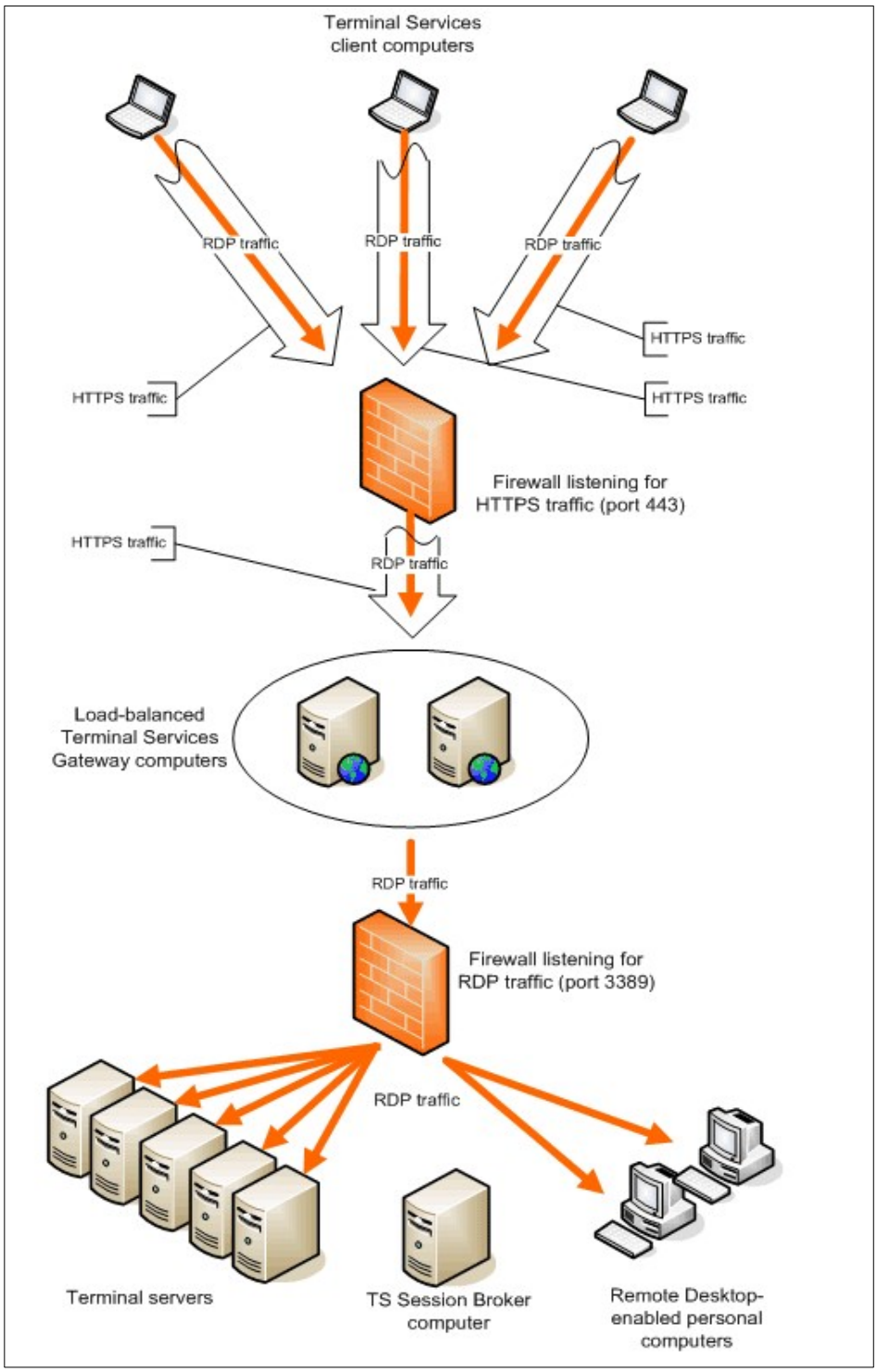
Session Broker Settings:



Load Balancing TS Gateway Servers

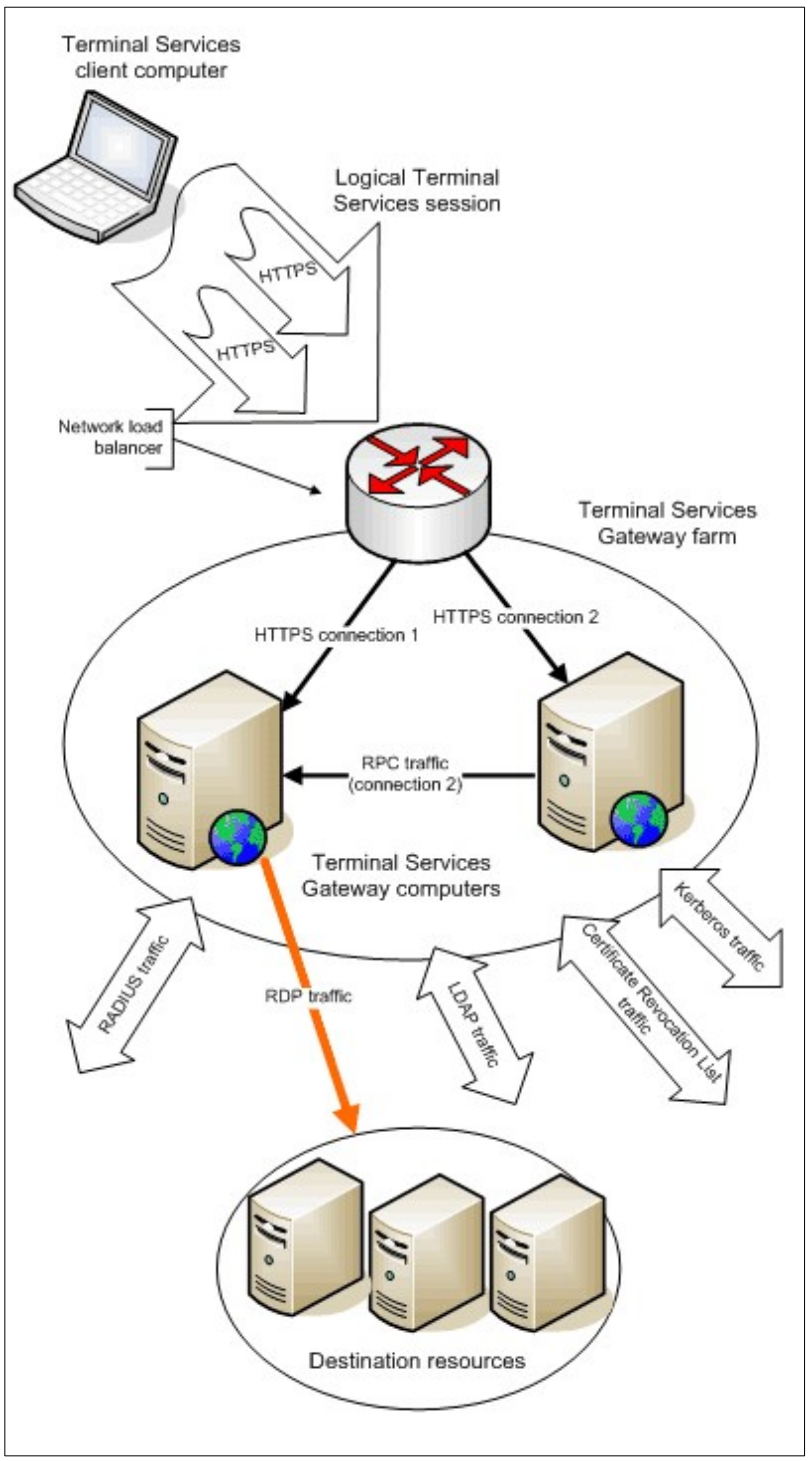
Terminal Services Gateway (TS Gateway) enables authorized remote users to connect to resources on an internal corporate or private network, from any Internet-connected device. The network resources can be Terminal Servers, Terminal Servers running RemoteApp programs, or computers with Remote Desktop enabled.

TS Gateway uses Remote Desktop Protocol (RDP) over HTTPS to establish a secure, encrypted connection between remote users on the Internet and the internal network resources on which their productivity applications run.



To load balance multiple Gateway servers, simply create a VIP that listens on port 443 (HTTPS) with no persistence. Then define the TS gateway servers as related real servers. A layer 4 DR mode VIP is recommended for optimum performance although a layer 4 NAT mode or layer 7 SNAT mode VIP can also be used.

For each TS client connection, two SSL connection are made. If the second of these 2 SSL connections gets load balanced to a different server it will get automatically redirected to the server that received the 1st connection provided that all the Gateway servers are all correctly configured as farm/collection members.



A second load balancer can then be used with Connection Broker/Session Broker to load balance the Terminal Servers/ Remote Data Servers as described in the previous section of this guide.

For further information on deploying TS Gateway, please refer to the following archived Microsoft Technet article: https://web.archive.org/web/20150121152832/https://technet.microsoft.com/en-us/library/cc304366.aspx

9. Technical Support

If you have any questions regarding the appliance or would like assistance designing your deployment, please don't hesitate to contact our support team: support@loadbalancer.org.

10. Further Documentation

The Administration Manual contains much more information about configuring and deploying the appliance. It's available here: http://pdfs.loadbalancer.org/loadbalanceradministrationv8.pdf

11. Conclusion

Loadbalancer.org appliances provide a very cost effective and flexible solution for highly available load balanced Terminal Server environments.

12. Appendix

1 – Clustered Pair Configuration – Adding a Slave Unit

If you initially configured just the master unit and now need to add a slave - our recommended procedure, please refer to the relevant section below for more details:

Note: A number of settings are not replicated as part of the master/slave pairing process and therefore must be manually configured on the slave appliance. These are listed below:

- Hostname & DNS settings
- Network settings including IP addresses, bonding configuration and VLANs
- Routing configuration including default gateways and static routes
- Date & time settings
- Physical – Advanced Configuration settings including Internet Proxy IP address & port, Firewall table size, SMTP relay and Syslog server
- SNMP settings
- Graphing settings
- Firewall Script & Firewall Lockdown Script settings
- Software updates

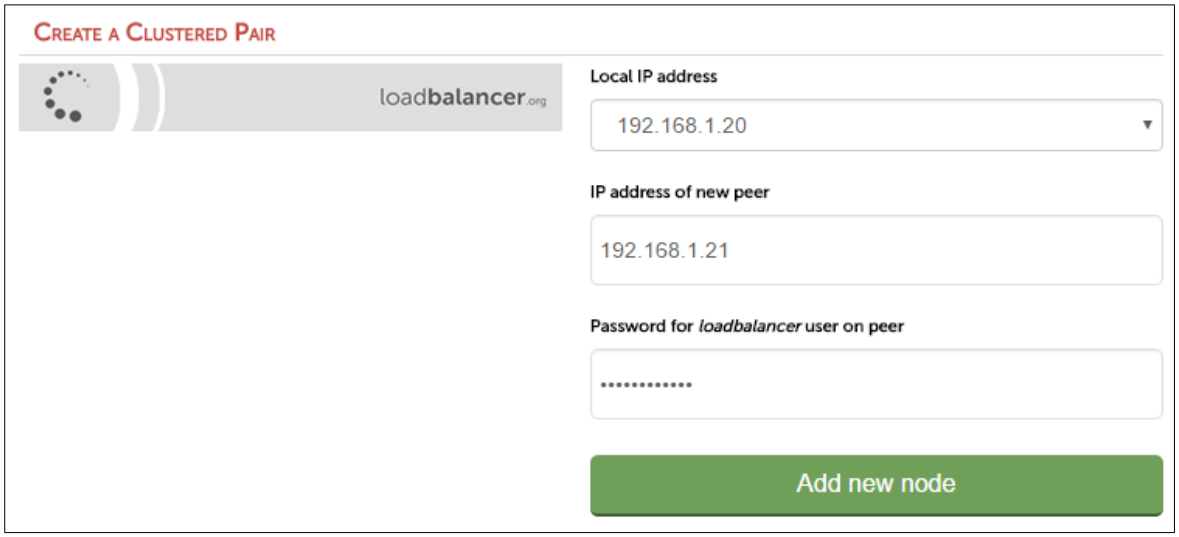
Version 7:

Please refer to **Chapter 8 – Appliance Clustering for HA** in the v7 Administration Manual.

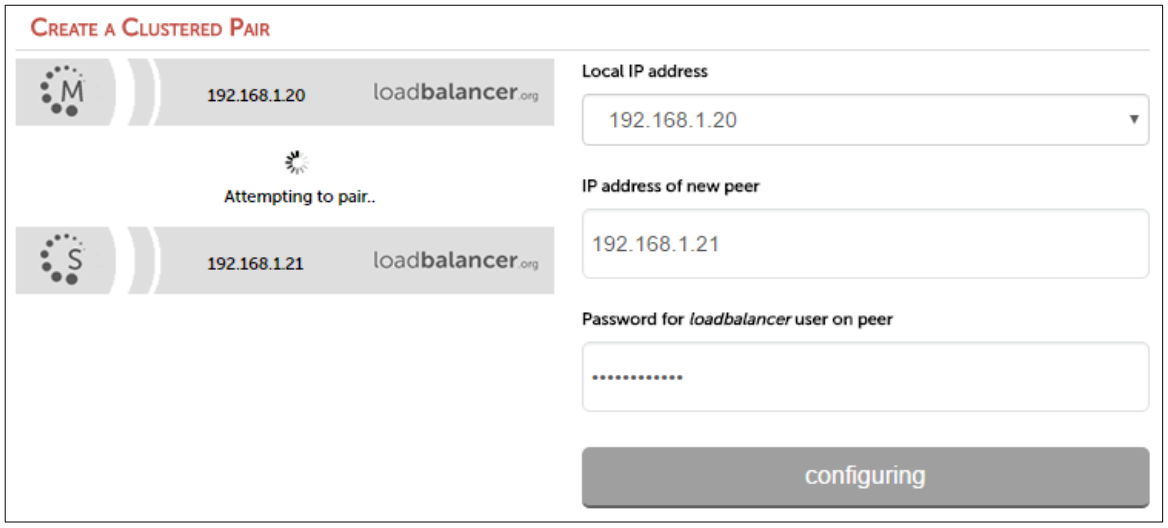
Version 8:

To add a slave node – i.e. create a highly available clustered pair:

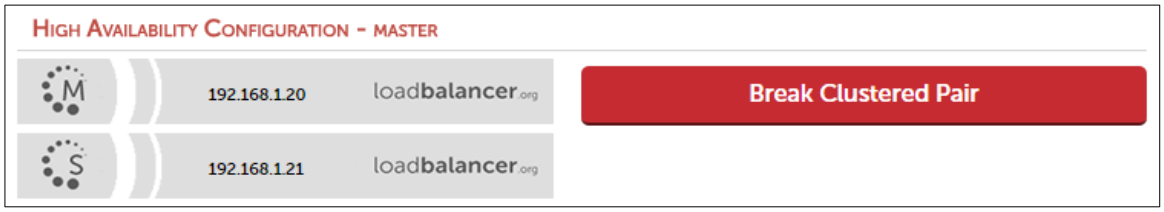
- Deploy a second appliance that will be the slave and configure initial network settings
- Using the WebUI, navigate to: *Cluster Configuration > High-Availability Configuration*



- Specify the IP address and the *loadbalancer* users password (the default is 'loadbalancer') for the slave (peer) appliance as shown above
- Click **Add new node**
- The pairing process now commences as shown below:



- Once complete, the following will be displayed:



- To finalize the configuration, restart heartbeat and any other services as prompted in the blue message box at the top of the screen

Note: Clicking the **Restart Heartbeat** button on the master appliance will also automatically restart heartbeat on the slave appliance.

Note: Please refer to chapter 9 – Appliance Clustering for HA in the Administration Manual for more detailed information on configuring HA with 2 appliances.

2 - Server Feedback Agent

The load balancer can modify the weight (amount of traffic) of each server by gathering data from either a custom agent or an HTTP server. For layer 4 VIPs the feedback method can be set to either agent or HTTP, for Layer 7 VIPs, only the agent method is supported.

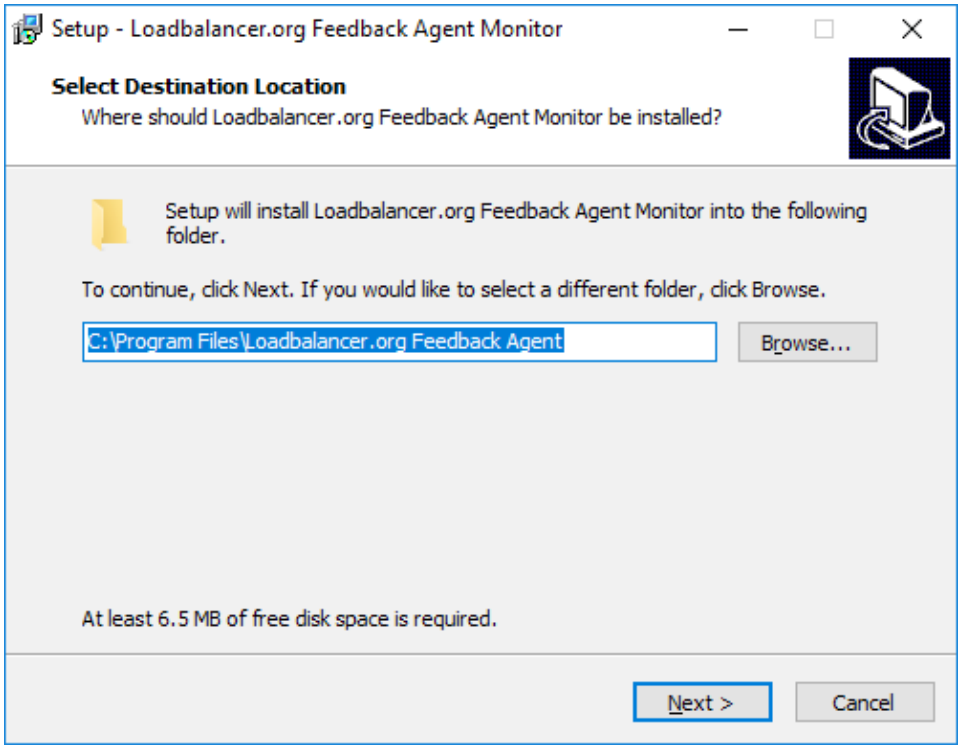
A telnet to port 3333 on a Real Server with the agent installed will return the current idle stats as an integer value in the range 0 – 100. The figure returned can be related to CPU utilization, RAM usage or a combination of both. This can be configured using the XML configuration file located in the agents installation folder (by default C:\ProgramData\LoadBalancer.org\LoadBalancer).

The load balancer typically expects a 0-99 integer response from the agent which by default relates to the current CPU idle state, e.g. a response of 92 would imply that the Real Servers CPU is 92% idle. The load balancer will then use the formula (92/100*requested_weight) to find the new optimized weight.

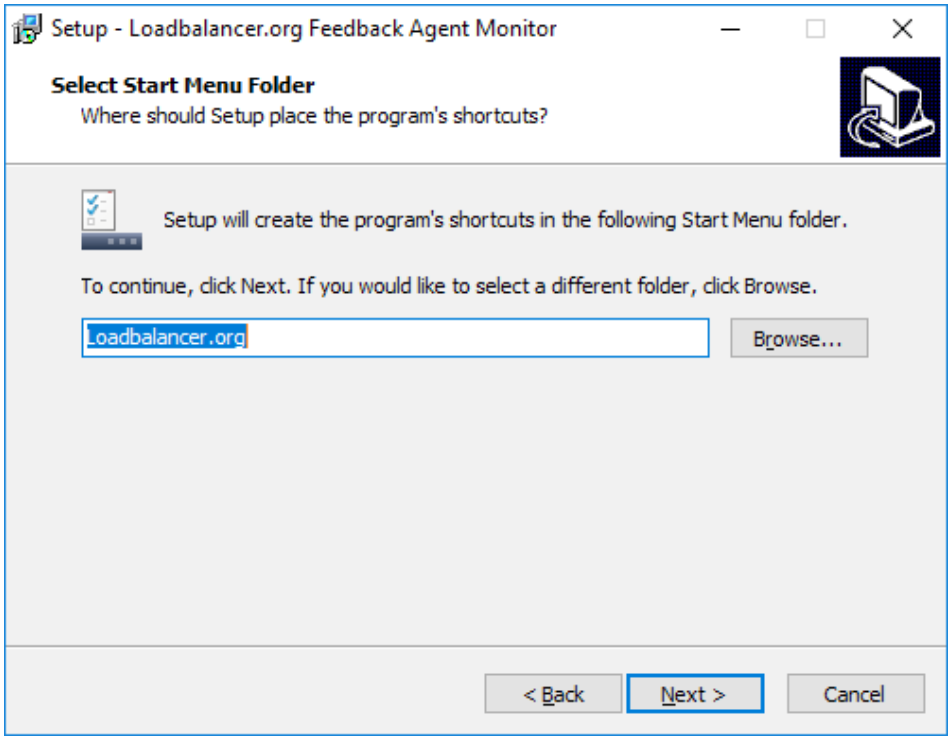
> Note: The 'Requested Weight' is the weight set in the WebUI for each Real Server. For more information please also refer to this blog.

Agent Download

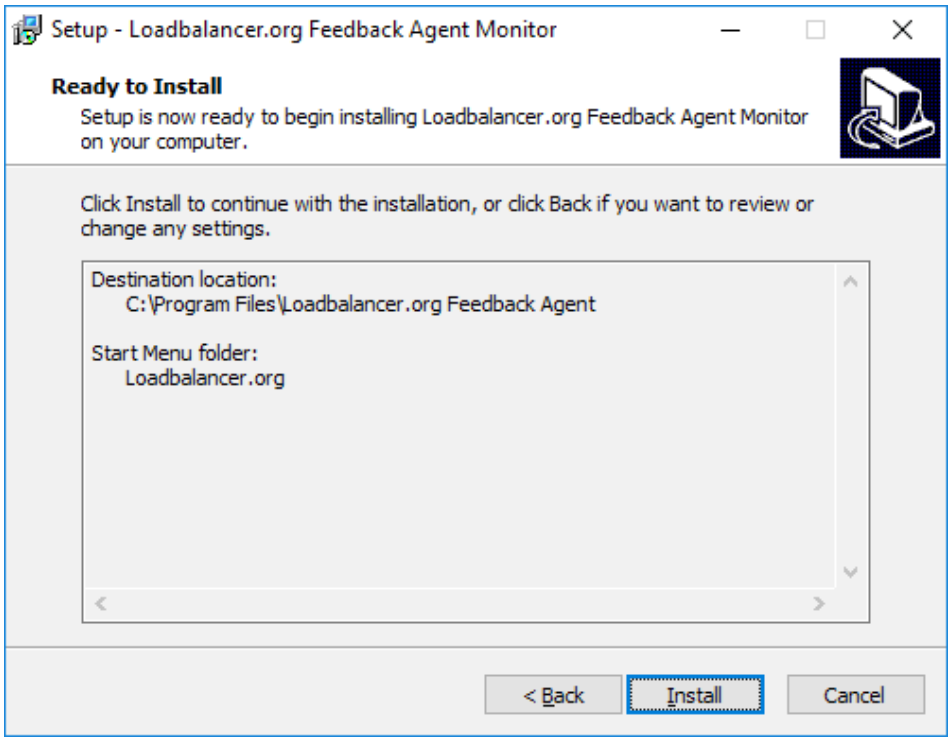
The latest Windows feedback agent (v4.5.5) can be downloaded from here. To install the agent, run loadbalanceragent-4.5.5.msi on each Terminal Server:



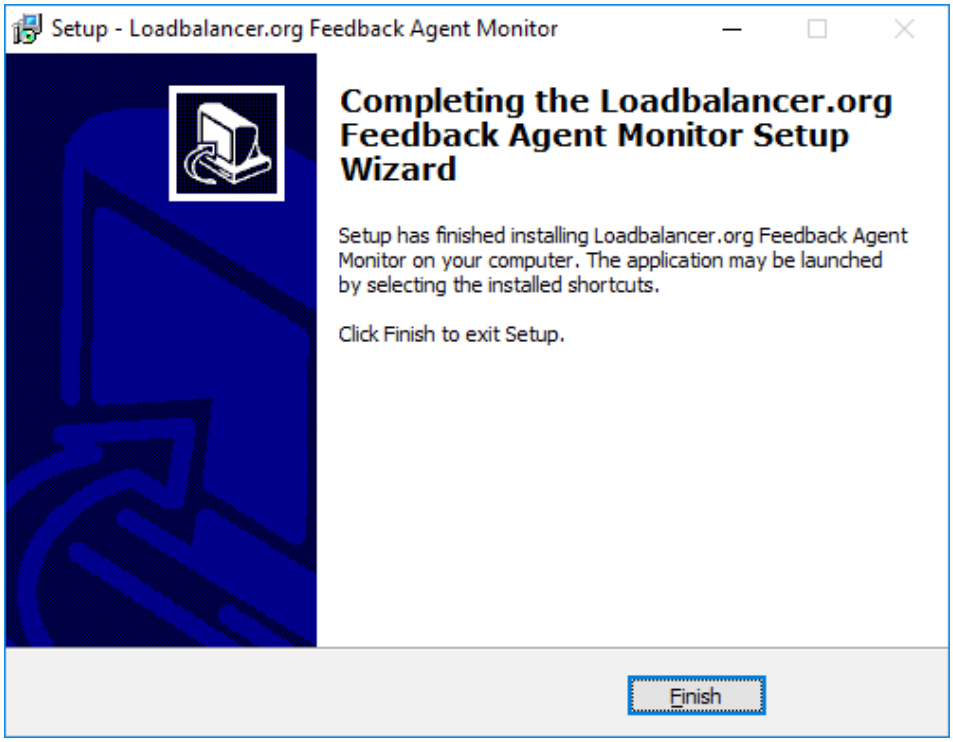
Leave the default location or change according to your requirements, click **Next**



Leave the default location or change according to your requirements, click **Next**



Click **Install** to start the installation process

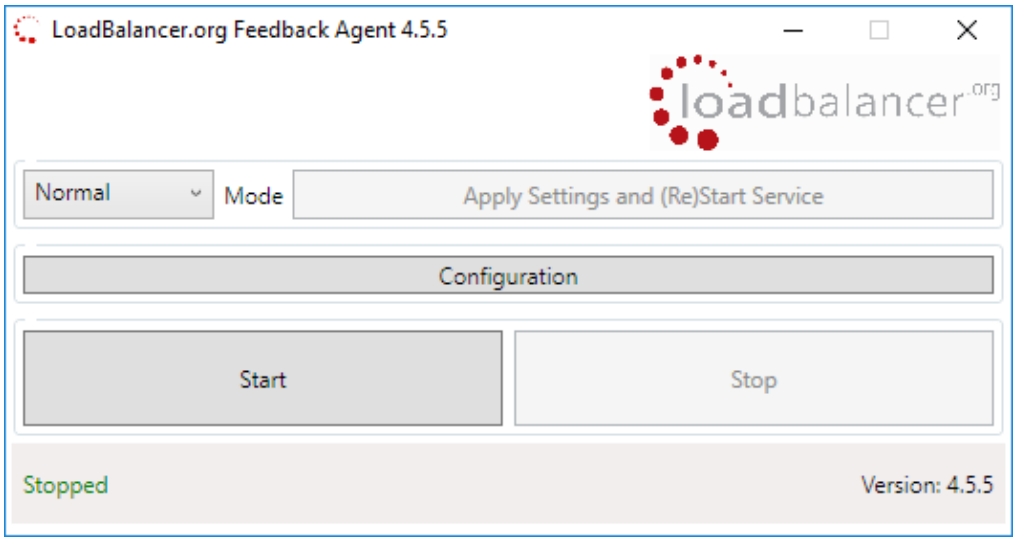


Click **Finish**

> Note: The agent should be installed on all Terminal Servers in the cluster.

Starting the Agent

Once the installation has completed, you'll need to start the service on the Real Servers. The service is controlled by the Feedback Agent monitor & control program that is also installed along with the Agent. This can be accessed on the Windows server from: *Start> Loadbalancer.org > Loadbalancer.org Feedback Agent*. It's also possible to start the service using the services snap-in – the service is called 'LBCPUMon'.



- To start the service, click the **Start** button
- To stop the service, click the **Stop** button

Configuration

To Configure Virtual Services to use the feedback agent, follow the steps below:

1. Using the WebUI, navigate to:

   Cluster Configuration > Layer 4 Virtual Services

   or

   Cluster Configuration > Layer 7 Virtual Services

2. Click **Modify** next to the Virtual Service

3. Change the Feedback Method to **Agent**
4. Click **Update**
5. Reload/Restart services as prompted

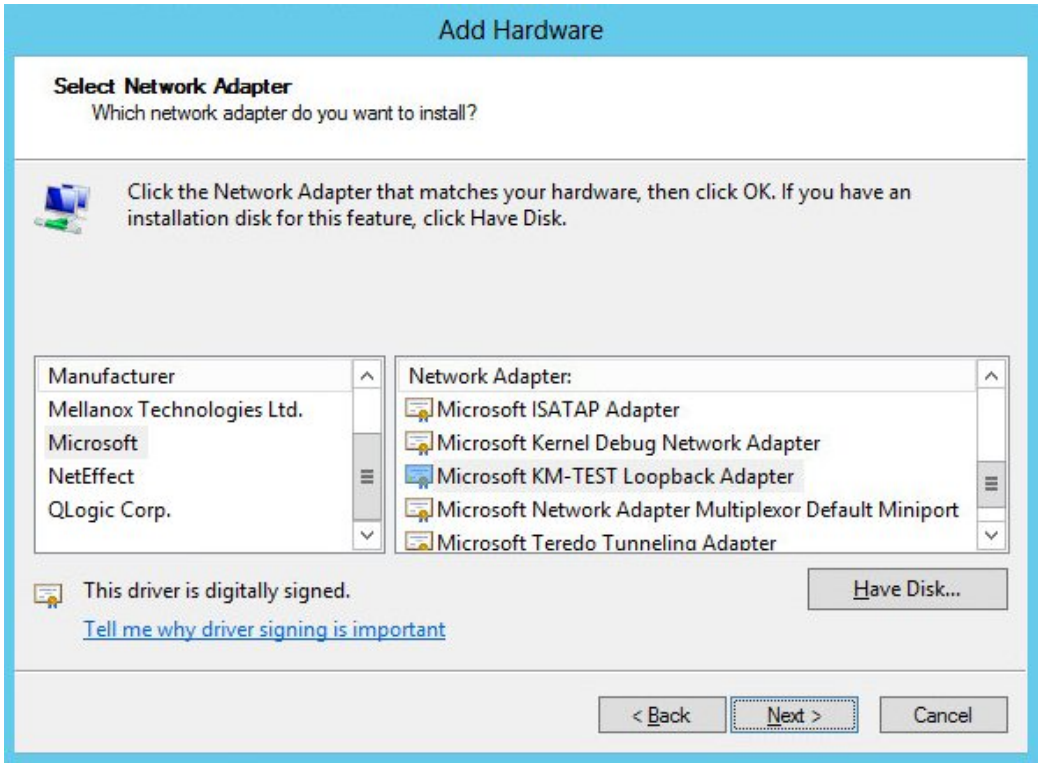
3 – Solving the ARP Problem

When using Layer 4 DR mode, the ARP problem must be solved. This involves configuring each Real Server to be able to receive traffic destined for the VIP, and ensuring that each Real Server does not respond to ARP requests for the VIP address – only the load balancer should do this.

The steps below are for Windows 2012 / 2016, for other versions of Windows please refer to chapter 6 in the Administration Manual.

Step 1: Install the Microsoft Loopback Adapter

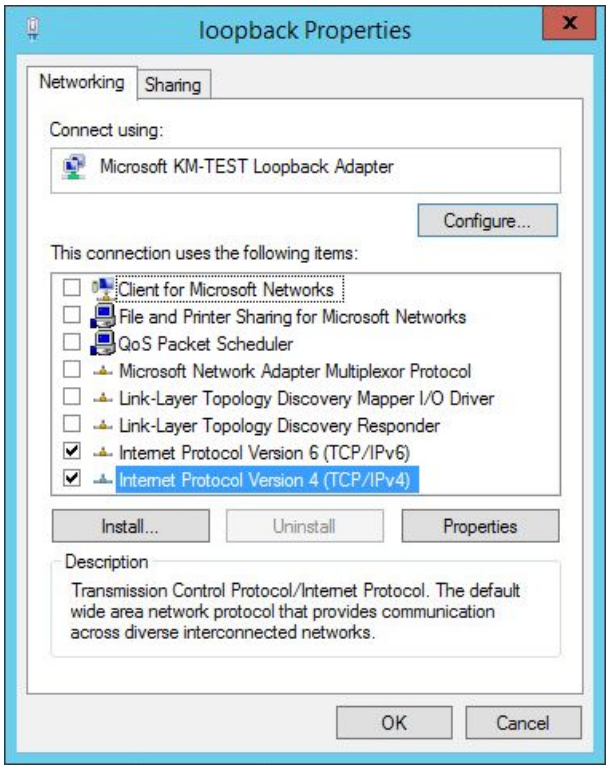
1. Click **Start**, then run **hdwwiz** to start the Hardware Installation Wizard
2. When the Wizard has started, click **Next**
3. Select **Install the hardware that I manually select from a list (Advanced)**, click **Next**
4. Select **Network adapters**, click **Next**
5. Select **Microsoft** & **Microsoft KM-Test Loopback Adapter**, click **Next**



6. Click **Next** to start the installation, when complete click **Finish**

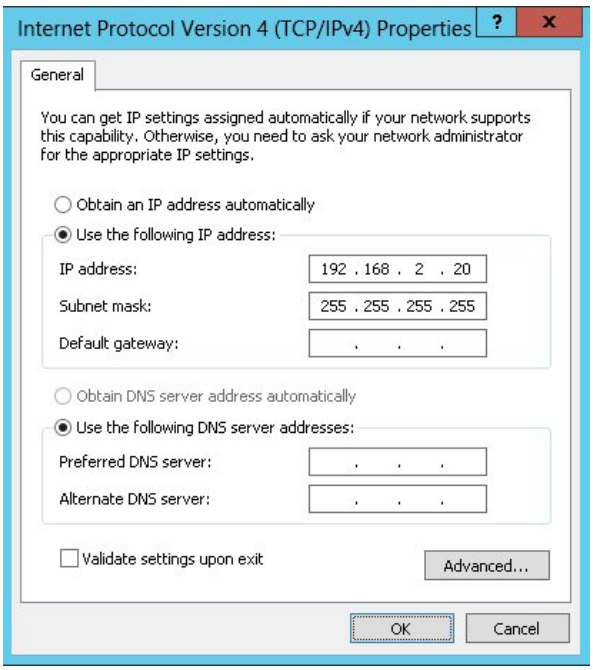
Step 2: Configure the Loopback Adapter

1. Open Control Panel and click **Network and Sharing Center**
2. Click **Change adapter settings**
3. Right-click the new Loopback Adapter and select **Properties**
4. Un-check all items except **Internet Protocol Version 4 (TCP/IPv4)** and **Internet Protocol Version 6 (TCP/IPv6)** as shown below:

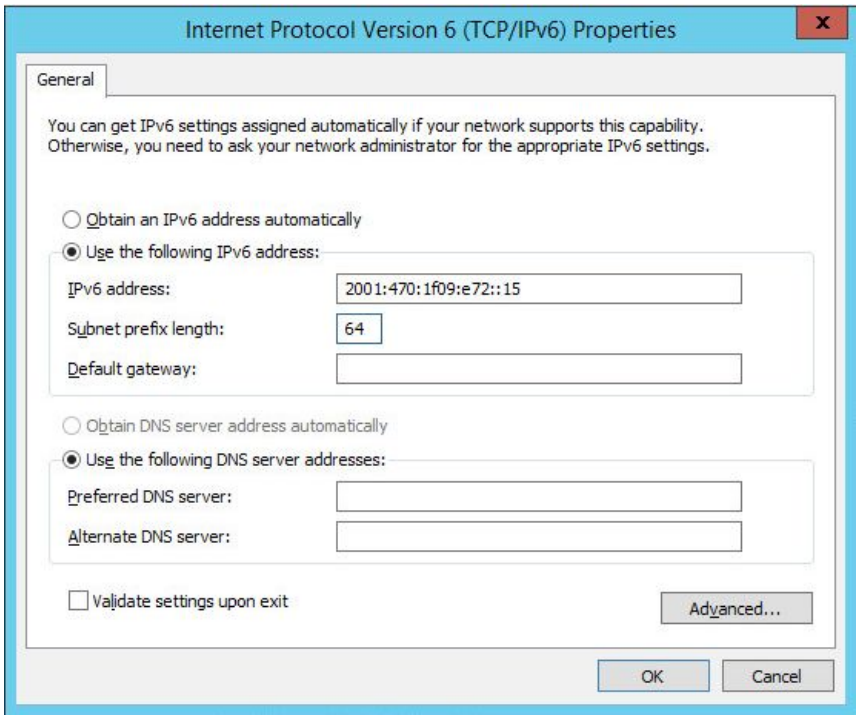


Note: Leaving both checked ensures that both IPv4 and IPv6 are supported. Select one If preferred.

5. If configuring IPv4 addresses select **Internet Protocol Version (TCP/IPv4)**, click **Properties** and configure the IP address to be the same as the Virtual Service (VIP) with a subnet mask of 255.255.255.255 , e.g. 192.168.2.20/255.255.255.255 as shown below:



6. If configuring IPv6 addresses select **Internet Protocol Version (TCP/IPv6)**, click **Properties** and configure the IP address to be the same as the Virtual Service (VIP) and set the *Subnet Prefix Length* to be the same as your network setting , e.g. 2001:470:1f09:e72::15/64 as shown below:



7. Click **OK** on TCP/IP Properties, then click **Close** on Ethernet Properties to save and apply the new settings
8. Now repeat the above process on the other Windows 2012/2016 Real Servers

Step 3: Configure the strong/weak host behavior

Windows Server 2000 and Windows Server 2003 use the weak host model for sending and receiving for all IPv4 interfaces and the strong host model for sending and receiving for all IPv6 interfaces. You cannot configure this behavior. The Next Generation TCP/IP stack in Windows 2008 and later supports strong host sends and receives for both IPv4 and IPv6 by default. To ensure that Windows 2012/2016 is running in the correct mode to be able to respond to the VIP, the following commands must be run on each Real Server:

For IPv4 addresses:

```
netsh interface ipv4 set interface "net" weakhostreceive=enabled
netsh interface ipv4 set interface "loopback" weakhostreceive=enabled
netsh interface ipv4 set interface "loopback" weakhostsend=enabled
```

For these commands to work, the LAN connection NIC must be named "net" and the loopback NIC must be named "loopback" as shown below. If you prefer to leave your current NIC names, then the commands above must be modified accordingly. For example, if your network adapters are named "LAN" and "LOOPBACK", the commands required would be:

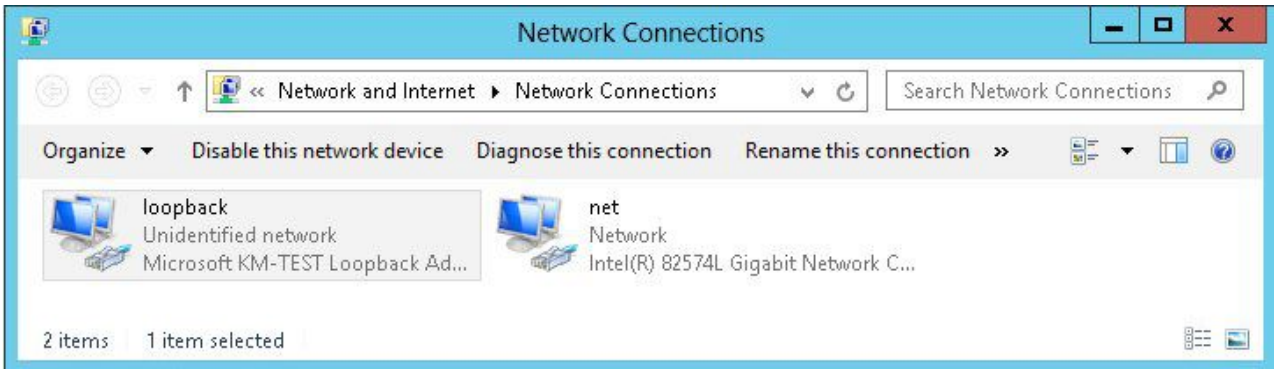
```
netsh interface ipv4 set interface "LAN" weakhostreceive=enabled
netsh interface ipv4 set interface "LOOPBACK" weakhostreceive=enabled
netsh interface ipv4 set interface "LOOPBACK" weakhostsend=enabled
```

For IPv6 addresses:

```
netsh interface ipv6 set interface "net" weakhostreceive=enabled
netsh interface ipv6 set interface "loopback" weakhostreceive=enabled
netsh interface ipv6 set interface "loopback" weakhostsend=enabled
netsh interface ipv6 set interface "loopback" dadtransmits=0
```

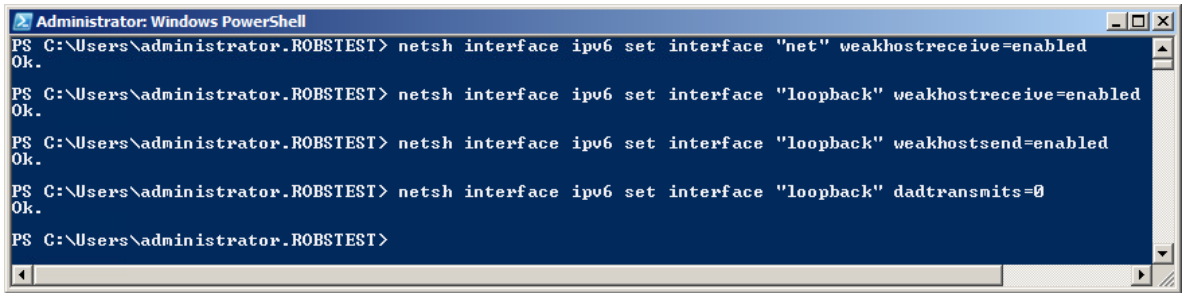
For these commands to work, the LAN connection NIC must be named "net" and the loopback NIC must be named "loopback" as shown below. If you prefer to leave your current NIC names, then the commands above must be modified accordingly. For example, if your network adapters are named "LAN" and "LOOPBACK", the commands required would be:

```
netsh interface ipv6 set interface "LAN" weakhostreceive=enabled
netsh interface ipv6 set interface "LOOPBACK" weakhostreceive=enabled
netsh interface ipv6 set interface "LOOPBACK" weakhostsend=enabled
netsh interface ipv6 set interface "LOOPBACK" dadtransmits=0
```



Note: The names for the NICs are case sensitive, so make sure that the name used for the interface and the name used in the commands match exactly.

1. Start Powershell or use a command window to run the appropriate netsh commands as shown in the example below:



Note: This shows an IPv6 example, use the IPv4 commands if you're using IPv4 addresses.

2. Now repeat these 4 commands on the other Windows 2012 Real Servers

13. Document Revision History

Version	Date	Change	Reason for Change	Changed By
1.2.0	10 September 2019	Styling and layout	General styling updates	RJC
1.2.1	21 July 2020	New title page Updated Canadian contact details Added additional instructions for configuring persistence settings Updated broken and invalid hyperlinks	Branding update Change to Canadian contact details Changes to the appliance WebUI External content linked to from this document had been moved or retired	AH

About Loadbalancer.org

Loadbalancer.org's mission is to ensure that its clients' businesses are never interrupted. The load balancer experts ask the right questions to get to the heart of what matters, bringing a depth of understanding to each deployment. Experience enables Loadbalancer.org engineers to design less complex, unbreakable solutions - and to provide exceptional personalized support.

United Kingdom

Loadbalancer.org Ltd.
Compass House, North Harbour
Business Park, Portsmouth, PO6 4PS
UK:+44 (0) 330 380 1064
sales@loadbalancer.org
support@loadbalancer.org

Canada

Loadbalancer.org Appliances Ltd.
300-422 Richards Street, Vancouver,
BC, V6B 2Z4, Canada
TEL:+1 866 998 0508
sales@loadbalancer.org
support@loadbalancer.org

United States

Loadbalancer.org, Inc.
4550 Linden Hill Road, Suite 201
Wilmington, DE 19808, USA
TEL: +1 833.274.2566
sales@loadbalancer.org
support@loadbalancer.org

Germany

Loadbalancer.org GmbH
Tengstraße 2780798,
München, Germany
TEL: +49 (0)89 2000 2179
sales@loadbalancer.org
support@loadbalancer.org