SonicWall® Management Services High Availability Setup

Administration

SONICWALL®

Contents

1

Firewall High Availability

High Availability allows an administrator to specify a primary and secondary SonicWall appliance for the Management Service. In the event that the connection to the primary device fails, connectivity transfers to the backup device.

In addition, the Management Service can utilize the same device pairing technology to implement different forms of load balancing. Load balancing helps regulate the flow of network traffic by splitting that traffic between primary and secondary SonicWall devices.

NOTE: High Availability is supported on TZ series and above firewalls. Active/Active Clustering is supported on NSA 3600 and above firewalls.

NOTE: High Availability is available at the appliance level; it cannot be configured at the group level.

Topics:

About High Availability

Topics:

What Is High Availability?

High Availability (HA) is a redundancy design that allows two identical SonicWall firewalls running the Management Service to be configured to provide a reliable, continuous connection to the public Internet. One SonicWall firewall is configured as the Primary unit, and an identical SonicWall firewall is configured as the Secondary unit. If the Primary firewall fails, the Secondary firewall takes over to secure a reliable connection between the protected network and the Internet. Two firewalls configured in this way are also known as a High Availability Pair (HA Pair).

High Availability provides a way to share SonicWall licenses between two SonicWall firewalls when one is acting as a high-availability system for the other. Both firewalls must be the same SonicWall model.

To use this feature, you must register the SonicWall firewalls on MySonicWall as Associated Products.

High Availability Terminology

Active	The operative condition of a hardware unit. The Active identifier is a logical role that can be assumed by either a Primary or Secondary hardware unit.
Failover	The actual process in which the Standby unit assumes the Active role following a qualified failure of the Active unit. Qualification of failure is achieved by various configurable physical and logical monitoring facilities described in Configuring High Availability.
HA	High Availability: non-state, hardware failover capability.
IDV	Interface Disambiguation via VLAN.
PoE	Power over Ethernet is a technology that lets network cables carry electrical power.
PPP	Point-to-point protocol that provides a standard method for transporting multi-protocol diagrams over point-to-point links.
PPPoE	A method for transmitting PPP over ethernet.
PPPoE HA	HA PPPoE support function without State.
Preempt	Applies to a post-failover condition in which the Primary unit has failed, and the Secondary unit has assumed the Active role. Enabling Preempt causes the Primary unit to seize the Active role from the Secondary after the Primary has been restored to a verified operational state.
Primary	The principal hardware unit itself. The Primary identifier is a manual designation and is not subject to conditional changes. Under normal operating conditions, the Primary hardware unit operates in an Active role.
Secondary (Backup)	The subordinate hardware unit itself. The Secondary identifier is a relational designation and is assumed by a unit when paired with a Primary unit. Under normal operating conditions, the Secondary unit operates in a Standby mode. Upon failure of the Primary unit, the Secondary unit assumes the Active role.
SHF	State Hardware Failover, a Management Service feature that allows existing network flows to remain active when the primary firewall fails and the backup firewall takes over.
Standby (Idle)	The passive condition of a hardware unit. The Standby identifier is a logical role that can be assumed by either a Primary or Secondary hardware unit. The Standby unit assumes the Active role upon a determinable failure of the Active unit.
STP	Spanning Tree Protocol.

High Availability Modes

High Availability has several operation modes, which can be selected on the **High Availability > Settings** page:

- **None**—Selecting **None** activates a standard high availability configuration and hardware failover functionality, with the option of enabling Stateful HA and Active/Active DPI.
- **Active/Standby**—Active/Standby mode provides basic high availability with the configuration of two identical firewalls as a High Availability Pair. The Active unit handles all traffic, while the Standby unit shares its configuration settings and can take over at any time to provide continuous network connectivity if the Active unit stops working.

  By default, Active/Standby mode is stateless, meaning that network connections and VPN tunnels must be re-established after a failover. To avoid this, Stateful Synchronization can be licensed and enabled with Active/Standby mode. In this Stateful HA mode, the dynamic state is continuously synchronized between the Active and Standby units. When the Active unit encounters a fault condition, stateful failover occurs as the Standby firewall takes over the Active role with no interruptions to the existing network connections.

  NOTE: Stateful HA is:
  - Included on NSA 4600 and higher NSA platforms and SuperMassive Series platforms.
  - Supported on the NSA 2600 and NSA 3600 platforms with a SonicOS Expanded License or a High Availability License.
  - Supported on the TZ500 and higher TZ platforms with a Management Service Expanded License or a High Availability (Stateful) License.

  For licensing information, see Registering and Associating Firewalls on MySonicWall and Licensing High Availability Features.

- **Active/Active DPI**—The Active/Active Deep Packet Inspection (DPI) mode can be used along with the Active/Standby mode. When Active/Active DPI mode is enabled, the processor intensive DPI services, such as Intrusion Prevention (IPS), Gateway Anti-Virus (GAV), and Anti-Spyware are processed on the standby firewall, while other services, such as firewall, NAT, and other types of traffic are processed on the Active firewall concurrently.

  NOTE: Active/Active DPI is:
  - Included on the SM 9000 series platforms.
  - Supported on the NSA 5600 and NSA 6600 platforms with a Management Service Expanded License or a High Availability (Stateful) License.

  For licensing information, see Registering and Associating Firewalls on MySonicWall and Licensing High Availability Features.

- **Active/Active Clustering**—In this mode, multiple firewalls are grouped together as cluster nodes, with multiple Active units processing traffic (as multiple gateways), doing DPI and sharing the network load. Each cluster node consists of two units acting as a Stateful HA pair. Active/Active Clustering provides Stateful Failover support in addition to load-sharing. Optionally, each cluster node can also consist of a single unit, in which case Stateful Failover and Active/Active DPI are not available.

  NOTE: Active/Active Clustering and Active/Active DPI Clustering are:
  - Included on the SM 9000 series platforms
  - Supported on NSA 5600 and NSA 6600 platforms only with the purchase of a Management Service Expanded License.

  For licensing information, see Registering and Associating Firewalls on MySonicWall and Licensing High Availability Features.

- **Active/Active DPI Clustering**—This mode allows for the configuration of up to four HA cluster nodes for failover and load sharing, where the nodes load balance the application of DPI security services to network traffic. This mode can be enabled for additional performance gain, utilizing the standby units in each cluster node.

  NOTE: Active/Active DPI Clustering is:
  - Included on the SM 9000 series platforms
  - Supported on NSA 3600 and above platforms only with the purchase of a Management Service Expanded License.

  For licensing information, see Registering and Associating Firewalls on MySonicWall and Licensing High Availability Features.

Crash Detection

The HA feature has a thorough self-diagnostic mechanism for both the Active and Standby firewalls. The failover to the standby unit occurs when critical services are affected, physical (or logical) link failure is detected on monitored interfaces, or when the firewall loses power.

The self-checking mechanism is managed by software diagnostics, which check the complete system integrity of the firewall. The diagnostics check internal system status, system process status, and network connectivity. There is a weighting mechanism on both sides to decide which side has better connectivity to avoid potential failover looping.

Critical internal system processes such as NAT, VPN, and DHCP (among others) are checked in real time. The failing service is isolated as early as possible, and the failover mechanism repairs it automatically.

Virtual MAC Address

The Virtual MAC address allows the High Availability pair to share the same MAC address, which dramatically reduces convergence time following a failover. Convergence time is the amount of time it takes for the devices in a network to adapt their routing tables to the changes introduced by high availability.

Without Virtual MAC enabled, the Active and Standby firewalls each have their own MAC addresses. Because the firewalls are using the same IP address, when a failover occurs, it breaks the mapping between the IP address and MAC address in the ARP cache of all clients and network resources. The Secondary firewall must issue an ARP request, announcing the new MAC address/IP address pair. Until this ARP request propagates through the network, traffic intended for the Primary firewall's MAC address can be lost.

The Virtual MAC address greatly simplifies this process by using the same MAC address for both the Primary and Secondary firewalls. When a failover occurs, all routes to and from the Primary firewall are still valid for the Secondary firewall. All clients and remote sites continue to use the same Virtual MAC address and IP address without interruption.

By default, this Virtual MAC address is provided by the SonicWall firmware and is different from the physical MAC address of either the Primary or Secondary firewalls. This eliminates the possibility of configuration errors and ensures the uniqueness of the Virtual MAC address, which prevents possible conflicts. Optionally, you can manually configure the Virtual MAC address on the **High Availability > Monitoring** page.

The Virtual MAC setting is available even if Stateful High Availability is not licensed. When Virtual MAC is enabled, it is always used even if Stateful Synchronization is not enabled.

Dynamic WAN Interfaces with PPPoE HA

NOTE: Dynamic WAN interfaces with PPPoE HA is not supported on the SuperMassive 9800. Only the DHCP Server dynamic WAN mode is supported.

With the Management Service, PPPoE can be enabled on interfaces in non-stateful mode, HA Active/Standby mode. PPPoE HA provides HA where a Secondary firewall assumes connection to the PPPoE server when the Active firewall fails.

NOTE: One WAN interface must be configured as PPPoE Unnumbered.

After the Active unit connects to the PPPoE server, the firewall synchronizes the PPPoE session ID and server name to the Secondary unit.

When the Active firewall fails, it terminates the PPPoE HA connection on the client side by timing out. The Secondary firewall connects to the PPPoE server, terminates the original connection on the server side, and starts a new PPPoE connection. All pre-existing network connections are rebuilt, the PPPoE sessions are re-established, and the PPP process is renegotiated.

Stateful Synchronization with DHCP

With the Management Service, DHCP can now be enabled on interfaces in both Active/Standby (non-stateful) and Stateful Synchronization modes.

Only the Active firewall can get a DHCP lease. The Active firewall synchronizes the DHCP IP address along with the DNS and gateway addresses to the Secondary firewall. The DHCP client ID is also synchronized, allowing this feature to work even without enabling Virtual MAC.

During a failover, the Active firewall releases the DHCP lease and, as it becomes the Active unit, the Secondary firewall renews the DHCP lease using the existing DHCP IP address and client ID. The IP address does not change, and network traffic, including VPN tunnel traffic, continues to pass.

If the Active firewall does not have an IP address when failover occurs, the Secondary firewall starts a new DHCP discovery.

Stateful Synchronization with DNS Proxy

DNS Proxy supports stateful synchronization of DNS cache. When the DNS cache is added, deleted, or updated dynamically, it synchronizes to the idle firewall.

About HA Monitoring

On the **High Availability > Monitoring** page, you can configure both physical and logical interface monitoring:

- By enabling physical interface monitoring, you enable link detection for the designated HA interfaces. The link is sensed at the physical layer to determine link viability.
- Logical monitoring involves configuring the SonicWall to monitor a reliable device on one or more of the connected networks.

Failure to periodically communicate with the device by the Active unit in the HA Pair triggers a failover to the Standby unit. If neither unit in the HA Pair can connect to the device, no action is taken.

The Primary and Secondary IP addresses configured on the **High Availability > Monitoring** page can be configured on LAN or WAN interfaces, and are used for multiple purposes:

- As independent management addresses for each unit (supported on all physical interfaces)
- To allow synchronization of licenses between the Standby unit and the SonicWall licensing server
- As the source IP addresses for the probe pings sent out during logical monitoring

Configuring unique management IP addresses for both units in the HA Pair allows you to log in to each unit independently for management purposes. Note that non-management traffic is ignored if it is sent to one of these IP addresses. The Primary and Secondary firewalls' unique LAN IP addresses cannot act as an active gateway; all systems connected to the internal LAN will need to use the virtual LAN IP address as their gateway.

If WAN monitoring IP addresses are configured, then X0 monitoring IP addresses are not required. If WAN monitoring IP addresses are not configured, then X0 monitoring IP addresses are required, since in such a scenario the Standby unit uses the X0 monitoring IP address to connect to the licensing server with all traffic routed via the Active unit.

The management IP address of the Secondary/Standby unit is used to allow license synchronization with the SonicWall licensing server, which handles licensing on a per-firewall basis (not per-HA Pair). Even if the Secondary unit was already registered on MySonicWall before creating the HA association, you must use the link on the **CONSOLE | Licenses** page to connect to the SonicWall server while accessing the Secondary firewall through its management IP address.

When using logical monitoring, the HA Pair pings the specified Logical Probe IP address target from the Primary as well as from the Secondary unit. The IP address set in the Primary IP Address or Secondary IP Address field is used as the source IP address for the ping. If both units can successfully ping the target, no failover occurs. If both cannot successfully ping the target, no failover occurs, as the Management Service assumes that the problem is with the target, and not the firewalls. If one firewall can ping the target but the other cannot, however, the HA Pair failovers to the unit that can ping the target.

The configuration tasks on the **High Availability > Monitoring** page are performed on the Primary unit and then are automatically synchronized to the Secondary.

About Active/Standby HA

HA allows two identical firewalls running the Management Service to be configured to provide a reliable, continuous connection to the public Internet. One firewall is configured as the Primary unit, and an identical firewall is configured as the Secondary unit. In the event of the failure of the Primary firewall, the Secondary firewall takes over to secure a reliable connection between the protected network and the Internet. Two firewalls configured in this way are also known as a High Availability Pair (HA Pair).

Active/Standby HA provides standard, high availability, and hardware failover functionality with the option of enabling stateful HA and Active/Active DPI.

HA provides a way to share licenses between two firewalls when one is acting as a high availability system for the other. To use this feature, you must register the firewalls on MySonicWall as Associated Products. Both firewalls must be the same SonicWall model.

Topics:

Benefits of Active/Standby HA

- **Increased network reliability** – In a High Availability configuration, the Secondary firewall assumes all network responsibilities when the Primary unit fails, ensuring a reliable connection between the protected network and the Internet.
- **Cost-effectiveness** – High Availability is a cost-effective option for deployments that provide high availability by using redundant firewalls. You do not need to purchase a second set of licenses for the Secondary unit in a High Availability Pair.
- **Virtual MAC for reduced convergence time after failover** – The Virtual MAC address setting allows the HA Pair to share the same MAC address, which dramatically reduces convergence time following a failover. Convergence time is the amount of time it takes for the devices in a network to adapt their routing tables to the changes introduced by high availability. By default, the Virtual MAC address is provided by the SonicWall firmware and is different from the physical MAC address of either the Primary or Secondary firewalls.

How Active/Standby HA Works

NOTE: The TZ300 series and TZ400 series firewalls can be operated in Active/Standby HA mode without Stateful Synchronization. The SOHO W does not support High Availability with or without Stateful Synchronization.

HA requires one SonicWall firewall configured as the Primary SonicWall, and an identical SonicWall firewall configured as the Secondary SonicWall. During normal operation, the Primary SonicWall is in an Active state and the Secondary SonicWall in an Standby state. If the Primary device loses connectivity, the Secondary SonicWall transitions to Active mode and assumes the configuration and role of Primary, including the interface IP addresses of the configured interfaces.

Basic Active/Standby HA provides stateless high availability. After a failover to the Secondary firewall, all the pre-existing network connections must be re-established, including the VPN tunnels that must be re-negotiated. Stateful Synchronization can be licensed and enabled separately. For more information, see About Stateful Synchronization.

The failover applies to loss of functionality or network-layer connectivity on the Primary SonicWall. The failover to the Secondary SonicWall occurs when critical services are affected, physical (or logical) link failure is detected on monitored interfaces, or when the Primary SonicWall loses power. The Primary and Secondary SonicWall devices are currently only capable of performing Active/Standby High Availability or Active/Active DPI – complete Active/Active high availability is not supported at present.

There are two types of synchronization for all configuration settings:

- **Incremental** – If the timestamps are in sync and a change is made on the Active unit, an incremental synchronization is pushed to the Standby unit.
- **Complete** –If the timestamps are out of sync and the Standby unit is available, a complete synchronization is pushed to the Standby unit. When incremental synchronization fails, a complete synchronization is automatically attempted.

About Stateful Synchronization

Stateful Synchronization provides dramatically improved failover performance. When enabled, the network connections and VPN tunnel information is continuously synchronized between the two units so that the Secondary can seamlessly assume all network responsibilities if the Primary firewall fails, with no interruptions to existing network connections.

NOTE: Stateful HA is included on NSA 4600 and higher NSA platforms and on all SuperMassive platforms. Stateful HA is supported on the TZ500 and higher TZ platforms and NSA 2600 and NSA 3600 platforms with an Extended or Stateful HA upgrade license. For licensing information, see Registering and Associating Firewalls on MySonicWall and Licensing High Availability Features.

Topics:

Benefits of Stateful Synchronization

- **Improved reliability** - By synchronizing most critical network connection information, Stateful Synchronization prevents down time and dropped connections in case of firewall failure.
- **Faster failover performance** - By maintaining continuous synchronization between the Primary and Secondary firewalls, Stateful Synchronization enables the Secondary firewall to take over in case of a failure with virtually no down time or loss of network connections.
- **Minimal impact on CPU performance** - Typically less than 1% usage.
- **Minimal impact on bandwidth** - Transmission of synchronization data is throttled so as not interfere with other data.

How Does Stateful Synchronization Work?

Stateful Synchronization is not load-balancing. It is an active-standby configuration where the Primary firewall handles all traffic. When Stateful Synchronization is enabled, the Primary firewall actively communicates with the Secondary to update most network connection information. As the Primary firewall creates and updates network connection information (such as VPN tunnels, active users, connection cache entries), it immediately informs the Secondary firewall. This ensures that the Secondary firewall is always ready to transition to the Active state without dropping any connections.

The synchronization traffic is throttled to ensure that it does not interfere with regular network traffic. All configuration changes are performed on the Primary firewall and automatically propagated to the Secondary firewall. The High Availability pair uses the same LAN and WAN IP addresses—regardless of which firewall is currently Active.

When using SonicWall Global Management System (GMS) to manage the firewalls, GMS logs into the shared WAN IP address. In case of a failover, GMS administration continues seamlessly, and GMS administrators currently logged into the firewall are not logged out; however, **Get** and **Post** commands may result in a time out with no reply returned.

See Synchronized and Non-synchronized Information that follows for a list of information that is synchronized and information that is not currently synchronized by Stateful Synchronization.

Synchronized and Non-synchronized Information

Information that is Synchronized	Information that is not Synchronized
VPN information	Dynamic WAN clients (L2TP, PPPoE, and PPTP)
Basic connection cache	Deep Packet Inspection (GAV, IPS, and Anti Spyware)
FTP	IPHelper bindings (such as NetBIOS and DHCP)
Oracle SQL*NET	SYNFlood protection information
Real Audio	Content Filtering Service information
RTSP	VoIP protocols
GVC information	Dynamic ARP entries and ARP cache time outs
Dynamic Address Objects	Active wireless client information
DHCP server information	Wireless client packet statistics
Multicast and IGMP	Rogue AP list
Active users	
ARP	
SonicPoint status	
Wireless guest status	
License information	
Weighted Load Balancing information	
RIP and OSPF information	

Stateful Synchronization Example

In case of a failover, the following sequence of events occurs:

1. A PC user connects to the network, and the Primary firewall creates a session for the user.
2. The Primary firewall synchronizes with the Secondary firewall. The Secondary now has all of the user's session information.
3. The administrator restarts the Primary unit.
4. The Secondary unit detects the restart of the Primary unit and switches from Standby to Active.
5. The Secondary firewall begins to send gratuitous ARP messages to the LAN and WAN switches using the same Virtual MAC address and IP address as the Primary firewall. No routing updates are necessary for downstream or upstream network devices.
6. When the PC user attempts to access a Web page, the Secondary firewall has all of the user's session information and is able to continue the user's session without interruption.

About Active/Active DPI HA

IMPORTANT: Capture functionality is not supported in Active/Active DPI mode.

With Active/Active DPI enabled on a Stateful HA pair, the Deep Packet Inspection services are processed on the standby firewall of an HA pair concurrently with the processing of firewall, NAT, and other modules on the active firewall. The following DPI services are affected:

- Intrusion Prevention Service (IPS)
- Gateway Anti-Virus (GAV)
- Gateway Anti-Spyware
- Application Control

To use the Active/Active DPI feature, the administrator must configure an additional interface as the **Active/Active DPI Interface**. For example, if you choose to make X5 the Active/Active DPI Interface, you must physically connect X5 on the active unit to X5 on the standby unit in the HA pair. Certain packet flows on the active unit are selected and offloaded to the standby unit on the Active/Active DPI Interface. DPI is performed on the standby unit and then the results are returned to the active unit over the same interface. The remaining processing is performed on the active unit.

NOTE: Active/Active DPI is included on SuperMassive 9200, 9400, and 9600 platforms and is supported on the NSA 3600 and NSA 6600 only with extended licenses. For licensing information, see Registering and Associating Firewalls on MySonicWall and Licensing High Availability Features.

Benefits of Active/Active DPI HA

Active/Active DPI taps into the unused CPU cycles available in the standby unit, but the traffic still arrives and leaves through the active unit. The standby unit only sees the network traffic offloaded by the active unit, and processing of all modules other than DPI services is restricted to the active unit.

Active/Standby and Active/Active DPI Prerequisites

This section lists the supported platforms, provides recommendations and requirements for physically connecting the units, and describes how to register, associate, and license the units for High Availability.

Topics:

Supported Platforms for HA

Active/Active DPI is supported only on these SonicWall models:

SuperMassive 9600	NSA 6600
SuperMassive 9400	NSA 5600
SuperMassive 9200	

NOTE: Active/Active DPI is supported on the NSA 5600 and NSA 6600 with the purchase of an expanded license.

Active/Active DPI is not supported on these SonicWall models:

NSA 4600	TZ series
NSA 3600	SOHO Wireless
NSA 2600	

Physically Connecting your Firewalls

NOTE: For complete procedures for connecting your firewalls, see the *Getting Started Guide* for your firewall. For procedures for connecting Active/Active Cluster firewalls, see Connecting the HA Ports for Active/Active Clustering and Connecting Redundant Port Interfaces.

If you are connecting the Primary and Secondary firewalls to an Ethernet switch that uses the spanning tree protocol, be aware that it may be necessary to adjust the link activation time on the switch port to which the SonicWall interfaces connect. For example, on a Cisco Catalyst-series switch, it is necessary to activate **spanning tree port fast** for each port connecting to the SonicWall firewall's interfaces.

High Availability requires additional physical connections among the affected SonicWall firewalls.For all modes, you need connections for HA Control and HA Data. Active/Active DPI requires an additional connection.

In any High Availability deployment, you must physically connect the LAN and WAN ports of all units to the appropriate switches.

It is important that the X0 interfaces from all units be connected to the same broadcast domain. Otherwise, traffic failover will not work. Also, X0 is the default redundant HA port; if the normal HA Control link fails, X0 is used to communicate heartbeats between units. Without X0 in the same broadcast domain, both units would become active if the HA Control link fails.

A WAN connection to the Internet is useful for registering your firewalls on MySonicWall and for synchronizing licensing information. Unless live communication with SonicWall's licensing server is not permitted due to network policy, the WAN (X1) interface should be connected before registration and licensing are performed.

Connecting the Active/Active DPI Interfaces for Active/Active DPI

For Active/Active DPI, you must physically connect at least one additional interface, called the **Active/Active DPI Interface**, between the two firewalls in each HA pair, or Cluster Node. The connected interfaces must be the same number on both firewalls, and must initially appear as unused, unassigned interfaces in the **Network > Interfaces** page. For example, you could connect X5 on the Primary unit to X5 on the Secondary if X5 is an unassigned interface. After enabling Active/Active DPI, the connected interface will have a Zone assignment of **HA Data-Link**.

Certain packet flows on the active unit are selected and offloaded to the standby unit on the Active/Active DPI Interface. DPI is performed on the standby unit and then the results are returned to the active unit over the same interface.

Optionally, for port redundancy with Active/Active DPI, you can physically connect a second Active/Active DPI Interface between the two firewalls in each HA pair. This interface takes over transferring data between the two units during Active/Active DPI processing if the first Active/Active DPI Interface has a fault.

To connect the Active/Active DPI Interfaces for Active/Active DPI:

1. Decide which interface to use for the additional connection between the firewalls in the HA pair. The same interface must be selected on each firewall.
2. In the Management Service management interface, navigate to the **Network > Interfaces** page and ensure that the **Zone** is **Unassigned** for the intended Active/Active DPI Interface.
3. Using a standard Ethernet cable, connect the two interfaces directly to each other.
4. Optionally, for port redundancy with Active/Active DPI, physically connect a second Active/Active DPI Interface between the two firewalls in each HA pair.

Registering and Associating Firewalls on MySonicWall

To use High Availability, you must register both firewalls and associate them for HA on MySonicWall. When you click the link for a registered firewall in your MySonicWall page, the Service Management page displays for that firewall. At the bottom of the Service Management page, you can click the HA Secondary link under Associated Products. Then follow the instructions to select and associate the other unit for your HA Pair. For further information about registering your firewalls, see the *Getting Started Guide* for your firewalls.

After the firewalls are associated as an HA pair, they can share licenses. In addition to High Availability licenses, this includes the Management Service license, the Support subscription, and the security services licenses. The only licenses that are not shareable are for consulting services, such as the SonicWall GMS Preventive Maintenance Service.

It is not required that the Primary and Secondary firewalls have the same security services enabled. The security services settings will be automatically updated as part of the initial synchronization of settings. License synchronization is used so that the Secondary firewall can maintain the same level of network protection provided before the failover.

MySonicWall provides several methods of associating the two firewalls. You can start by registering a new firewall, and then choosing an already-registered unit to associate it with. Or, you can associate two units that are both already registered. You can also start the process by selecting a registered unit and adding a new firewall with which to associate it.

NOTE: Even if you first register your firewalls on MySonicWall, you must individually register both the Primary and the Secondary firewalls from the Management Service management interface while logged into the individual management IP address of each firewall. This allows the Secondary unit to synchronize with the SonicWall license server and share licenses with the associated Primary firewall. When Internet access is restricted, you can manually apply the shared licenses to both firewall.

For information about configuring and using the individual management IP address of each firewall, see About High Availability Monitoring with Active/Clustering and Monitoring High Availability.

Licensing High Availability Features

The HA licenses included with the purchase of the SonicWall network firewall is shown in HA Licenses Available with SonicWall Network Security Firewalls. Some platforms require additional licensing to use the Stateful Synchronization or Active/Active DPI features. Management Service Expanded licenses or High Availability licenses can be purchased on MySonicWall or from a SonicWall reseller.

NOTE: Stateful High Availability licenses must be activated on each firewall, either by registering the unit on MySonicWall from the Management Service management interface, or by applying the license keyset to each unit if Internet access is not available.

HA Licenses Available with SonicWall Network Security Firewalls

Platform	Active/Standby HA [1]	Stateful HA	A/A Clustering	A/A DPI
SOHO W	N/A	N/A	N/A	N/A
TZ300/TZ300 W	Included	N/A	N/A	N/A
TZ400/TZ400 W	Included	N/A	N/A	N/A
TZ500/TZ500 W	Included	Expanded License Stateful HA Upgrade License	N/A	N/A
TZ600	Included	Expanded License Stateful HA Upgrade License	N/A	N/A
NSA 2600	Included	Expanded license HA license	N/A	N/A
NSA 3600	Included	Expanded license HA license	N/A	N/A
NSA 4600	Included	Included	N/A	N/A
NSA 5600	Included	Included	Expanded license	Expanded license
NSA 6600	Included	Included	Expanded license	Expanded license
SM 9200	Included	Included	Included	Included
SM 9400	Included	Included	Included	Included
SM 9600	Included	Included	Included	Included

1. NA = Feature not available

You can view system licenses on the **System > Licenses** page. This page also provides a way to log into MySonicWall and to apply licenses to a firewall.

There is also a way to synchronize licenses for an HA pair whose firewalls do not have Internet access. When live communication with SonicWall's licensing server is not permitted due to network policy, you can use license keysets to manually apply security services licenses to your firewalls. When you register a firewall on MySonicWall, a license keyset is generated for the firewall. If you add a new security service license, the keyset is updated. However, until you apply the licenses to the firewall, it cannot perform the licensed services.

IMPORTANT: In a High Availability deployment without Internet connectivity, you must apply the license keyset to both of the firewalls in the HA pair.

Active/Active Clustering

NOTE: Active/Active Clustering is supported on NSA 2600 and above firewalls. As with NSA 5600 and 6600, Active/Active Clustering is supported on NSA 3600 and NSA 4600 platforms only with the purchase of a SonicOS Expanded License.

An Active/Active Cluster is formed by a collection of up to four Cluster Nodes. A Cluster Node can consist of a Stateful HA pair, a Stateless HA pair with standard failover, or a single standalone unit. Dynamic state synchronization is only available in a Cluster Node if it is a Stateful HA pair. The traditional SonicWall High Availability protocol or Stateful HA protocol is used for communication within the Cluster Node, between the units in the HA pair.

When a Cluster Node is a Stateful HA pair, Active/Active DPI can be enabled within the Cluster Node for higher performance.

With Active/Active Clustering, you can assign certain traffic flows to each node in the cluster, providing load sharing in addition to redundancy, and supporting a much higher throughput without a single point of failure.

Topics:

About Active/Active Clustering

This section provides an introduction to the Active/Active Clustering feature. With Active/Active Clustering, you can assign certain traffic flows to each node in the cluster, providing load sharing in addition to redundancy, and supporting a much higher throughput without a single point of failure.

A typical recommended setup includes four firewalls of the same SonicWall model configured as two Cluster Nodes, where each node consists of one Stateful HA pair. For larger deployments, the cluster can include eight firewalls, configured as four Cluster Nodes (or HA pairs). Within each Cluster Node, Stateful HA keeps the dynamic state synchronized for seamless failover with zero loss of data on a single point of failure. Stateful HA is not required, but is highly recommended for best performance during failover.

Load sharing is accomplished by configuring different Cluster Nodes as different gateways in your network. Typically this is handled by another device downstream (closer to the LAN devices) from the Active/Active Cluster, such as a DHCP server or a router.

A Cluster Node can also be a single firewall, allowing an Active/Active cluster setup to be built using two firewalls. In case of a fault condition on one of the firewalls in this deployment, the failover is not stateful since neither firewall in the Cluster Node has an HA Secondary.

Redundancy is achieved at several levels with Active/Active Clustering:

- The cluster provides redundant Cluster Nodes, each of which can handle the traffic flows of any other Cluster Node, if a failure occurs.
- The Cluster Node consists of a Stateful HA pair, in which the Secondary firewall can assume the duties of the Primary unit in case of failure.
- Port redundancy, in which an unused port is assigned as a secondary to another port, provides protection at the interface level without requiring failover to another firewall or node.
- Active/Active DPI can be enabled, providing increased throughput within each Cluster Node.

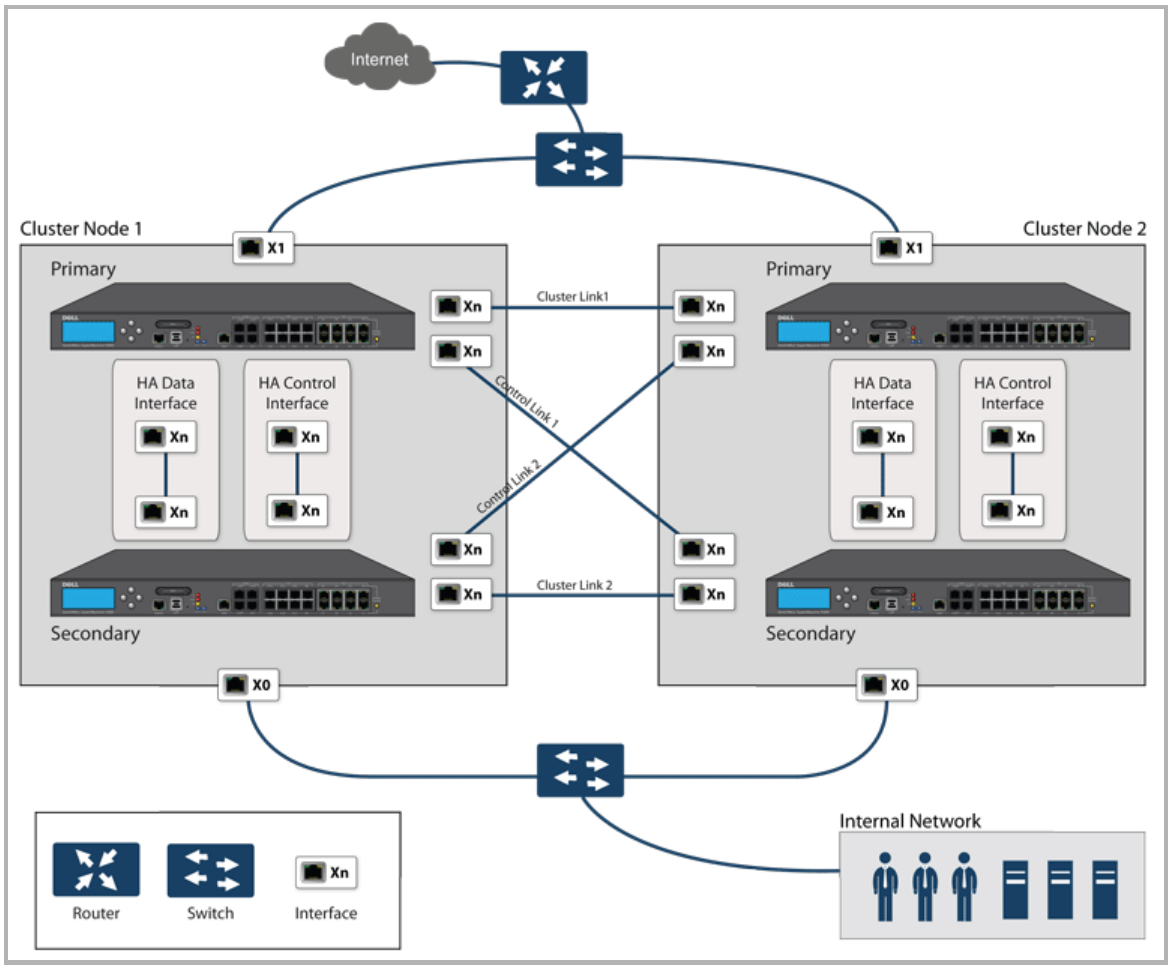
Topics:

- Example: Active/Active Clustering – Four-unit Deployment
- Example: Active/Active Clustering – Two-unit Deployment
- Benefits of Active/Active Clustering
- How Does Active/Active Clustering Work?

Example: Active/Active Clustering – Four-unit Deployment

Active/Active Four-unit Cluster shows a four-unit cluster. Each Cluster Node contains one HA pair. The designated HA ports of all four firewalls are connected to a Layer 2 switch. These ports are used for Cluster Node management and monitoring state messages sent over SVRRP, and for configuration synchronization. The two units in each HA pair are also connected to each other using another interface (shown as the Xn interface). This is the Active/Active DPI Interface necessary for Active/Active DPI. With Active/Active DPI enabled, certain packets are offloaded to the standby unit of the HA pair for DPI processing.

Active/Active Four-unit Cluster

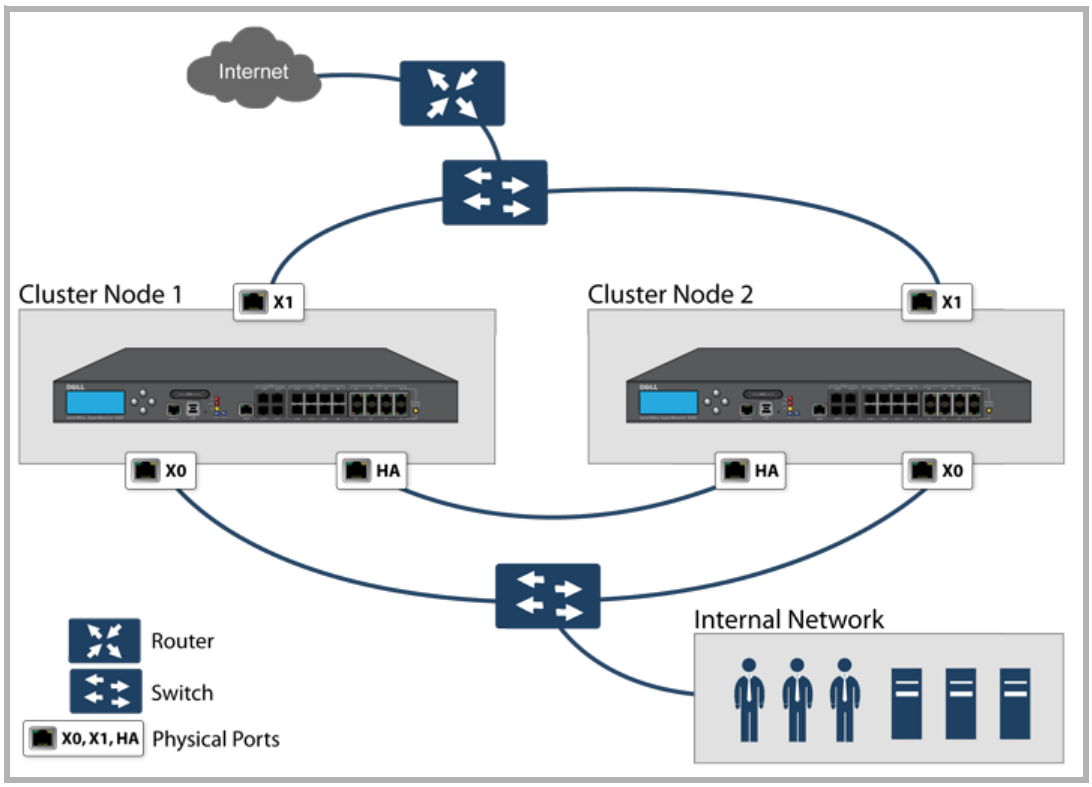


For more information about physically connecting redundant ports and redundant switches, see the *Active/Active Clustering Full Mesh Deployment Technote*.

Example: Active/Active Clustering – Two-unit Deployment

Active/Active Two-unit Cluster shows a two-unit cluster. In a two-unit cluster, HA pairs are not used. Instead, each Cluster Node contains a single firewall. The designated HA ports on the two firewalls are connected directly to each other using a cross-over cable. The SonicWall Virtual Router Redundancy Protocol (SVRRP) uses this HA port connection to send Cluster Node management and monitoring state messages. SVRRP management messages are initiated on the Master Node, and monitoring information is communicated from every firewall in the cluster. The HA port connection is also used for configuration synchronization between Cluster Nodes.

Active/Active Two-unit Cluster



Benefits of Active/Active Clustering

The benefits of Active/Active Clustering include the following:

- All the firewalls in the cluster are utilized to derive maximum throughput
- Can run in conjunction with Active/Active DPI to perform concurrent processing of IPS, GAV, Anti-Spyware, and App Rules services, which are the most processor intensive, on the standby firewall in each HA pair while the active firewall performs other processing
- Load sharing is supported by allowing the assignment of particular traffic flows to each node in the cluster
- All nodes in the cluster provide redundancy for the other nodes, handling traffic as needed if other nodes go down
- Interface redundancy provides secondary for traffic flow without requiring failover
- Both Full Mesh and non-Full Mesh deployments are supported

How Does Active/Active Clustering Work?

There are several important concepts that are introduced for Active/Active Clustering.

Topics:

About Cluster Nodes

An Active/Active Cluster is formed by a collection of Cluster Nodes. A Cluster Node can consist of a Stateful HA pair, a Stateless HA pair or a single standalone unit. Dynamic state synchronization is only available in a Cluster Node if it is a Stateful HA pair. The traditional SonicWall High Availability protocol or Stateful HA protocol is used for communication within the Cluster Node, between the units in the HA pair.

When a Cluster Node is a Stateful HA pair, Active/Active DPI can be enabled within the Cluster Node for higher performance.

About the Cluster

All firewalls in the Cluster must be of same product model and be running the same firmware version.

Within the cluster, all firewalls are connected and communicating with each other; see Active/Active Two-node Cluster. For communication between Cluster Nodes, a new protocol, called SonicWall Virtual Router Redundancy Protocol (SVRRP), is used. Cluster Node management and monitoring state messages are sent using SVRRP.

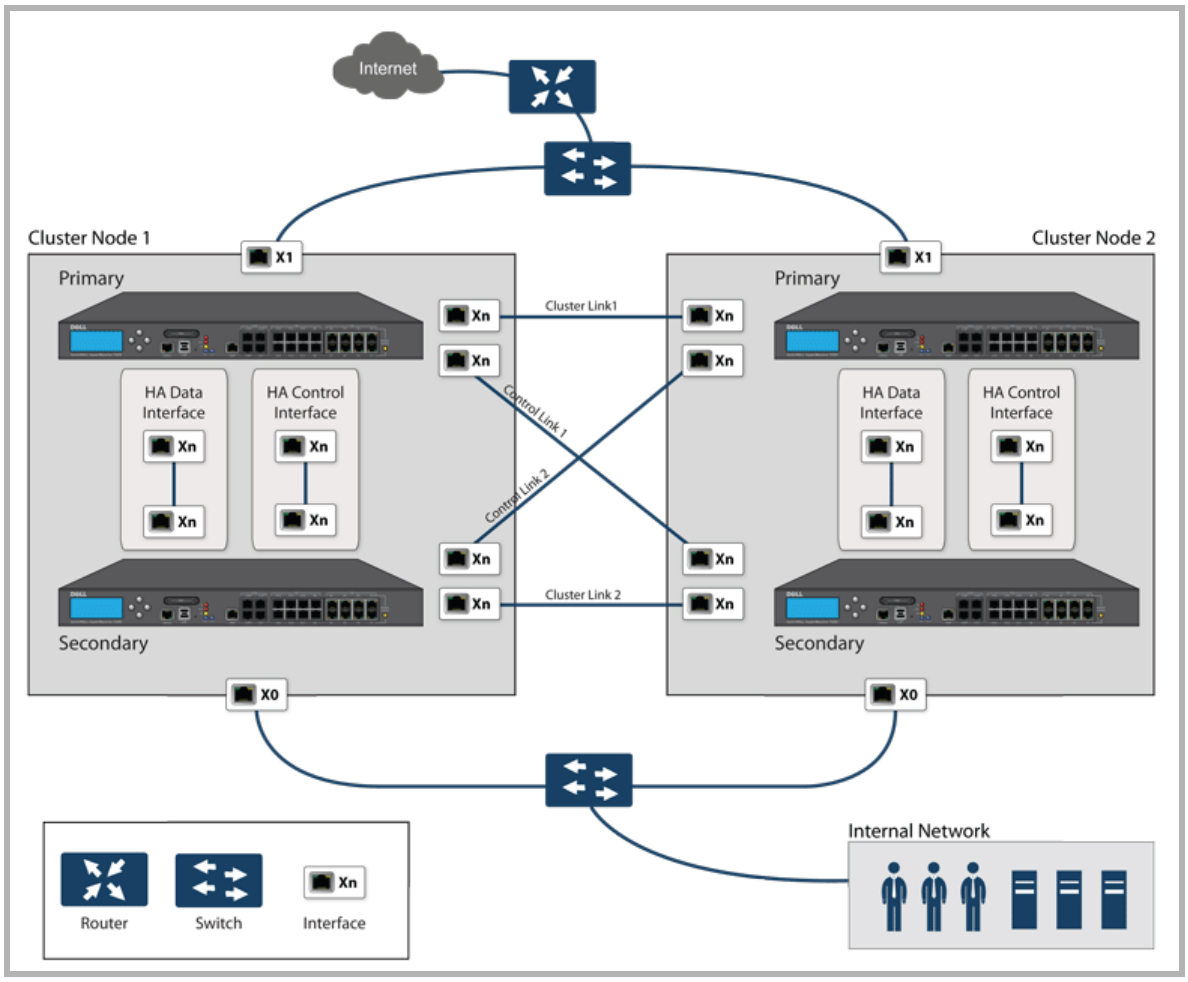
All Cluster Nodes share the same configuration, which is synchronized by the Master Node. The Master Node is also responsible for synchronizing firmware to the other nodes in the cluster. The HA port connection is used to synchronize configuration and firmware updates.

Dynamic state is not synchronized across Cluster Nodes, but only within a Cluster Node. When a Cluster Node contains an HA pair, Stateful HA can be enabled within that Cluster Node, with the advantages of dynamic state synchronization and stateful failover as needed. In the event of the failure of an entire Cluster Node, the failover will be stateless. This means that pre-existing network connections must be rebuilt. For example, Telnet and FTP sessions must be re-established and VPN tunnels must be renegotiated.

About Failover provides more information about how failover works.

The maximum number of Cluster Nodes in a cluster is currently limited to four. If each Cluster Node is an HA pair, the cluster includes eight firewalls.

Active/Active Two-node Cluster



Actions Allowed within the Cluster

The types of administrative actions that are allowed differ based on the state of the firewall in the cluster. All actions are allowed for admin users with appropriate privileges on the active firewall of the Master Node, including all configuration actions. A subset of actions are allowed on the active firewall of Non-Master nodes, and even fewer actions are allowed on firewalls in the standby state. See Administrative Actions Allowed for a list of the allowed actions for active firewalls of Non-Master nodes and standby firewalls in the cluster.

Administrative Actions Allowed

Administrative Action	Active Non-master	Standby
Read-only actions	Allowed	Allowed
Registration on MySonicWall	Allowed	Allowed
License Synchronization with SonicWall License Manager	Allowed	Allowed
Diagnostic tools in **System > Diagnostics**	Allowed	Allowed
Packet capture	Allowed	Allowed
HA Synchronize Settings (syncs settings to the HA peer within the node)	Allowed	Not allowed
HA Synchronize Firmware (syncs firmware to the HA peer within the node)	Allowed	Not allowed

Administrative Actions Allowed (Continued)

Administrative Action	Active Non-master	Standby
Administrative logout of users	Allowed	Not allowed
Authentication tests (such as test LDAP, test RADIUS, test Authentication Agent)	Allowed	Not allowed

About Virtual Groups

Active/Active Clustering also supports the concept of Virtual Groups. Currently, a maximum of four Virtual Groups are supported.

A Virtual Group is a collection of virtual IP addresses for all the configured interfaces in the cluster configuration (unused/unassigned interfaces do not have virtual IP addresses). When Active/Active Clustering is enabled for the first time, the configured IP addresses for the interfaces on that firewall are converted to virtual IP addresses for Virtual Group 1. Thus, Virtual Group 1 includes virtual IP addresses for X0, X1, and any other interfaces that are configured and assigned to a zone.

A Virtual Group can also be thought of as a logical group of traffic flows within a failover context, in that the logical group of traffic flows can failover from one node to another depending upon the fault conditions encountered. Each Virtual Group has one Cluster Node acting as the owner and one or more Cluster Nodes acting as standby. A Virtual Group is only owned by one Cluster Node at a time, and that node becomes the owner of all the virtual IP addresses associated with that Virtual Group. The owner of Virtual Group 1 is designated as the Master Node, and is responsible for synchronizing configuration and firmware to the other nodes in the cluster. If the owner node for a Virtual Group encounters a fault condition, one of the standby nodes will become the owner.

As part of the configuration for Active/Active Clustering, the serial numbers of other firewalls in the cluster are entered into the Management Service management interface, and a ranking number for the standby order is assigned to each. When the Active/Active Clustering configuration is applied, up to three additional Virtual Groups are created, corresponding to the additional Cluster Nodes added, but virtual IP addresses are not created for these Virtual Groups. You need to configure these virtual IP addresses on the **Network > Interfaces** page.

There are two factors in determining Virtual Group ownership (which Cluster Node owns which Virtual Group):

- **Rank of the Cluster Node** – The rank is configured in the Management Service management interface to specify the priority of each node for taking over the ownership of a Virtual Group.
- **Virtual Group Link Weight of the Cluster Nodes** – This is the number of interfaces in the Virtual Group that are up and have a configured virtual IP address.

When more than two Cluster Nodes are configured in a cluster, these factors determine the Cluster Node that is best able to take ownership of the Virtual Group. In a cluster with two Cluster Nodes, one of which has a fault, naturally the other will take ownership.

SVRRP is used to communicate Virtual Group link status and ownership status to all Cluster Nodes in the cluster.

The owner of Virtual Group 1 is designated as the Master Node. Configuration changes and firmware updates are only allowed on the Master Node, which uses SVRRP to synchronize the configuration and firmware to all the nodes in the cluster. On a particular interface, virtual IP addresses for Virtual Group 1 must be configured before other Virtual Groups can be configured.

Load Sharing and Multiple Gateway Support

The traffic for the Virtual Group is processed only by the owner node. A packet arriving on a Virtual Group leaves the firewall on the same Virtual Group. In a typical configuration, each Cluster Node owns a Virtual Group, and therefore processes traffic corresponding to one Virtual Group.

This Virtual Group functionality supports a multiple gateway model with redundancy. In a deployment with two Cluster Nodes, the X0 Virtual Group 1 IP address can be one gateway and the X0 Virtual Group 2 IP address can be another gateway. It is up to the network administrator to determine how the traffic is allocated to each gateway. For example, you could use a smart DHCP server which distributes the gateway allocation to the PCs on the directly connected client network, or you could use policy based routes on a downstream router.

When Active/Active Clustering is enabled, the Management Service internal DHCP server is turned off and cannot be enabled. Networks needing a DHCP server can use an external DHCP server which is aware of the multiple gateways, so that the gateway allocation can be distributed.

NOTE: When Active/Active Clustering is enabled, the Management Service internal DHCP server is turned off.

Effect on Related Configuration Pages

When Active/Active Clustering is initially enabled, the existing IP addresses for all configured interfaces are automatically converted to virtual IP addresses for Virtual Group 1. When Virtual Group 1 or any Virtual Group is created, default interface objects are created for virtual IP addresses with appropriate names, such as “Virtual Group 1” or “Virtual Group 2”. The same interface can have multiple virtual IP addresses, one for each Virtual Group that is configured. You can view these virtual IP addresses in the **Network > Interfaces** page.

NOTE: All Cluster Nodes in the Active/Active cluster share the same configuration

A virtual MAC address is associated with each virtual IP address on an interface and is generated automatically by the Management Service. The virtual MAC address is created in the format 00-17-c5-6a-XX-YY, where XX is the interface number such as “03” for port X3, and YY is the internal group number such as “00” for Virtual Group 1, or “01” for Virtual Group 2.

NOTE: The Active/Active virtual MAC address is different from the High Availability virtual MAC address. The High Availability virtual MAC address functionality is not supported when Active/Active Clustering is enabled.

NAT policies are automatically created for the affected interface objects of each Virtual Group. These NAT policies extend existing NAT policies for particular interfaces to the corresponding virtual interfaces. You can view these NAT policies in the **Network > NAT Policies** page. Additional NAT policies can be configured as needed and can be made specific to a Virtual Group if desired.

After Active/Active Clustering is enabled, you must select the Virtual Group number during configuration when adding a VPN policy.

About SVRRP

For communication between Cluster Nodes in an Active/Active cluster, a new protocol called SonicWall Virtual Router Redundancy Protocol (SVRRP) is used. Cluster Node management and monitoring state messages are sent using SVRRP over the Active/Active Cluster links.

SVRRP is also used to synchronize configuration changes, firmware updates, and signature updates from the Master Node to all nodes in the cluster. In each Cluster Node, only the active unit processes the SVRRP messages.

In the case of failure of the Active/Active Cluster links, SVRRP heartbeat messages are sent on the X0 interface. However, while the Active/Active Cluster links are down, configuration is not synchronized. Firmware or signature updates, changes to policies, and other configuration changes cannot be synchronized to other Cluster Nodes until the Active/Active Cluster links are fixed.

About Failover

There are two types of failover that can occur when Active/Active Clustering is enabled:

- **High Availability failover** – Within an HA pair, the Secondary unit takes over for the Primary. If Stateful HA is enabled for the pair, the failover occurs without interruption to network connections.
- **Active/Active failover** – If all the units in the owner node for a Virtual Group encounter a fault condition, then the standby node for the Virtual Group takes over the Virtual Group ownership. Active/Active failover transfers ownership of a Virtual Group from one Cluster Node to another. The Cluster Node that becomes the Virtual Group owner also becomes the owner of all the virtual IP addresses associated with the Virtual Group and starts using the corresponding virtual MAC addresses.

  Active/Active failover is stateless, meaning that network connections are reset and VPN tunnels must be renegotiated. Layer 2 broadcasts inform the network devices of the change in topology as the Cluster Node which is the new owner of a Virtual Group generates ARP requests with the virtual MACs for the newly owned virtual IP addresses. This greatly simplifies the failover process as only the connected switches need to update their learning tables. All other network devices continue to use the same virtual MAC addresses and do not need to update their ARP tables, because the mapping between the virtual IP addresses and virtual MAC addresses is not broken.

When both High Availability failover and Active/Active failover are possible, HA failover is given precedence over Active/Active failover for the following reasons:

- HA failover can be stateful, whereas Active/Active failover is stateless.
- The standby firewall in an HA pair is lightly loaded and has resources available for taking over the necessary processing, although it may already be handling DPI traffic if Active/Active DPI is enabled. The alternative Cluster Node might already be processing traffic comparable in amount to the failed unit, and could become overloaded after failover.

Active/Active failover always operates in Active/Active preempt mode. Preempt mode means that, after failover between two Cluster Nodes, the original owner node for the Virtual Group will seize the active role from the standby node after the owner node has been restored to a verified operational state. The original owner has a higher priority for a Virtual Group due to its higher ranking if all virtual IP interfaces are up and the link weight is the same between the two Cluster Nodes.

About DPI with Active/Active Clustering

Active/Active DPI can be used along with Active/Active Clustering. When Active/Active DPI is enabled, it utilizes the standby firewall in the HA pair for DPI processing.

For increased performance in an Active/Active cluster, enabling Active/Active DPI is recommended, as it utilizes the standby firewall in the HA pair for Deep Packet Inspection (DPI) processing.

About High Availability Monitoring with Active/Clustering

When Active/Active Clustering is enabled, HA monitoring configuration is supported for the HA pair in each Cluster Node. The HA monitoring features are consistent with previous versions. HA monitoring can be configured for both physical/link monitoring and logical/probe monitoring. After logging into the Master Node, monitoring configuration needs to be added on a per Node basis from the **High Availability > Monitoring** page.

NOTE: The **High Availability > Monitoring** page applies only to the HA pair that you are logged into, not to the entire cluster.

Physical interface monitoring enables link detection for the monitored interfaces. The link is sensed at the physical layer to determine link viability.

When physical interface monitoring is enabled, with or without logical monitoring enabled, HA failover takes precedence over Active/Active failover. If a link fails or a port is disconnected on the active unit, the standby unit in the HA pair will become active.

NOTE: For interfaces with configured virtual IP addresses, Active/Active physical monitoring is implicit and is used to calculate the Virtual Group Link Weight. Physical monitoring cannot be disabled for these interfaces. This is different from HA monitoring.

Logical monitoring involves configuring the Management Service to monitor a reliable device on one or more of the connected networks. Failure to periodically communicate with the device by the active unit in the HA pair will trigger a failover to the standby unit. If neither unit in the HA pair can connect to the device, the problem is assumed to be with the device and no failover will occur.

If both physical monitoring and logical monitoring are disabled, Active/Active failover will occur on link failure or port disconnect.

The Primary and Secondary IP addresses configured on the **High Availability > Monitoring** page can be configured on LAN or WAN interfaces, and are used for multiple purposes:

- As independent management addresses for each unit, regardless of the Active or Standby status of the unit (supported on all physical interfaces)
- To allow synchronization of licenses between the standby unit and the SonicWall licensing server
- As the source IP addresses for the probe pings sent out during logical monitoring

Configuring monitoring IP addresses for both units in the HA pair allows you to log in to each unit independently for management purposes. Note that non-management traffic is ignored if it is sent to one of the monitoring IP addresses. The Primary and Secondary firewall's unique LAN IP addresses cannot act as an active gateway; all systems connected to the internal LAN need to use a virtual LAN IP address as their gateway.

NOTE: When HA Monitoring/Management IP addresses are configured only on WAN interfaces, they need to be configured on all the WAN interfaces for which a Virtual IP address has been configured.

The management IP address of the Secondary unit is used to allow license synchronization with the SonicWall licensing server, which handles licensing on a per-firewall basis (not per-HA pair). Even if the standby unit was already registered on MySonicWall before creating the HA association, you must use the link on the **System > Licenses** page to connect to the SonicWall server while accessing the Secondary firewall through its management IP address. This allows synchronization of licenses (such as the Active/Active Clustering or the Stateful HA license) between the standby unit and the SonicWall licensing server.

When using logical monitoring, the HA pair will ping the specified Logical Probe IP address target from the Primary as well as from the Secondary SonicWall. The IP address set in the Primary IP Address or Secondary IP Address field is used as the source IP address for the ping. If both units can successfully ping the target, no failover occurs. If both cannot successfully ping the target, no failover occurs, as the SonicWalls will assume that the problem is with the target, and not the SonicWalls. But, if one SonicWall can ping the target but the other SonicWall cannot, the HA pair will failover to the SonicWall that can ping the target.

The configuration tasks on the **High Availability > Monitoring** page are performed on the Primary unit and then are automatically synchronized to the Secondary.

Active/Active Clustering Prerequisites

NOTE: In addition to the requirements described in this section, ensure that you have completed the prerequisites described in Active/Standby and Active/Active DPI Prerequisites.

For Active/Active Clustering, additional physical connections are required:

- **Active/Active Cluster Link**—Each Active/Active cluster link must be a 1GB interface

Active/Active Clustering configuration can include configuring Virtual Group IDs and redundant ports. Procedures are provided in this section for both of these tasks within the Configuring High Availability.

Topics:

Licensing Requirements for Active/Active Clustering

Active/Active Clustering licenses included with the purchase of the SonicWall network firewall are shown in Licensing Requirements for A/A Clustering. Some platforms require additional licensing to use the Active/Active Clustering features. Management Service Expanded licenses can be purchased on MySonicWall or from a SonicWall reseller.

NOTE: Active/Active Clustering licenses must be activated on each firewall, either by registering the unit on MySonicWall from the Management Service management interface, or by applying the license keyset to each unit if Internet access is not available.

Licensing Requirements for A/A Clustering

Platform	A/A Clustering [1]
SOHO W	N/A
TZ300/TZ300 W	N/A
TZ400/TZ400 W	N/A
TZ500/TZ500 W	N/A
TZ600	N/A
NSA 2600	N/A
NSA 3600	N/A
NSA 4600	N/A
NSA 5600	Expanded license: • 01-SSC-4480
NSA 6600	Expanded license: • 01-SSC-4481
SM 9200	Included
SM 9400	Included
SM 9600	Included

1. N/A = A/A Clustering not available

You can view system licenses on the **System > Licenses** page. This page also provides a way to log into MySonicWall. For information about licensing, see Registering and Associating Firewalls on MySonicWall.

When the firewalls in the Active/Active cluster have Internet access, each firewall in the cluster must be individually registered from the Management Service management interface while you are logged into the individual management IP address of each firewall. This allows the Secondary units to synchronize with the SonicWall licensing server and share licenses with the associated Primary firewalls in each HA pair.

Connecting the HA Ports for Active/Active Clustering

For Active/Active Clustering, you must physically connect the designated HA ports of all units in the Active/Active cluster to the same Layer 2 network.

SonicWall recommends connecting all designated HA ports to the same Layer 2 switch. You can use a dedicated switch or simply use some ports on an existing switch in your internal network. All of these switch ports must be configured to allow Layer 2 traffic to flow freely amongst them.

In the case of a two-unit Active/Active cluster deployment, where the two Cluster Nodes each have only a single firewall, you can connect the HA ports directly to each other using a cross-over cable. No switch is necessary in this case.

The SonicWall Virtual Router Redundancy Protocol (SVRRP) uses this HA port connection to send Cluster Node management and monitoring state messages. SVRRP management messages are initiated on the Master Node, and monitoring information is communicated from every firewall in the cluster.

The HA port connection is also used to synchronize configuration from the Master Node to the other Cluster Nodes in the deployment. This includes firmware or signature upgrades, policies for VPN and NAT, and other configuration.

Connecting Redundant Port Interfaces

You can assign an unused physical interface as a redundant port to a configured physical interface called the “primary interface”. On each Cluster Node, each primary and redundant port pair must be physically connected to the same switch, or preferably, to redundant switches in the network.

NOTE: Because all Cluster Nodes share the same configuration, each node must have the same redundant ports configured and connected to the same switch(es).

To use Active/Active Clustering, you must register all SonicWall firewalls in the cluster on MySonicWall. The two firewalls in *each* HA pair must also be associated as HA Primary and HA Secondary on MySonicWall. That is, associate the two firewalls in the HA pair for Cluster Node 1, then associate the firewalls in the HA pair for Cluster Node 2, and so on for any other Cluster Nodes.

2

Configuring High Availability

IMPORTANT: High Availability cannot be used along with PortShield except with the SonicWall X-Series Solution. Before configuring HA, remove any existing PortShield configuration from the **Network > PortShield Groups** page.

The High Availability feature configures a pair of SonicWall appliances as a primary and backup. The backup monitors the primary through a series of heartbeats. If the backup detects that the primary is unavailable or has failed, it replaces the primary.

The High Availability feature is available on the following SonicWall appliances:

- SonicWall NSA Series
- SonicWall NSA E-Class Series
- SonicWall PRO 2040/3060/4060/4100/5060

For more information on High Availability, see Firewall High Availability and Active/Standby and Active/Active DPI Prerequisites. If your Active/Active Clustering environment will use VPN or NAT, see Configuring VPN and NAT with Active/Active Clustering after you have finished the Active/Active configuration.

Configuring Active/Standby High Availability Settings

The configuration tasks on the High Availability > Settings page are performed on the Primary firewall and then are automatically synchronized to the Secondary firewall.

To configure Active/Standby:

1. Select a SonicWall appliance and click the **Policies** tab.
2. Navigate to the **High Availability > Settings** page. The High Availability page **General** tab displays.
3. **In the Mode drop-down menu, select Active/Standby.**

   When a SonicWall appliance becomes active after startup, it looks for an active SonicWall appliance that is configured for High Availability. If the other appliance is active, it transitions to **Standby** mode. Sometimes, because of network latency and other issues, it might take a while to find the other SonicWall appliance.
4. Select **Enable Stateful Synchronization**. This option is not selected by default.

   When Stateful High Availability is not enabled, session state is not synchronized between the Primary and Secondary firewalls. If a failover occurs, any session that had been active at the time of failover needs to be renegotiated.

   A confirmation message displays.
5. Click **OK**.

6 Select **Generate/Overwrite Backup Firmware and Settings When Upgrading Firmware** to overwrite the current firmware backup settings when upgrading. With this option, the current settings at the time of upgrade are saved as backup settings.

7 Select **Enable Preempt Mode** to configure the primary SonicWall appliance to take over from the backup SonicWall appliance when it becomes available. Otherwise, the backup SonicWall appliance remains active.

Preempt mode is recommended to be disabled when enabling Stateful High Availability, because preempt mode can be over-aggressive about failing over to the Secondary firewall.

8 Select **Enable Virtual MAC**. When the Stateful High Availability Upgrade is licensed, Virtual MAC capability is also licensed. Virtual MAC allows the backup unit in an HF pair to use the MAC address of the primary unit when a failover occurs. Alternatively, you can manually set a virtual MAC address for both units to use. Virtual MAC addressing contributes to network continuity and efficiency during a failover in the same way as the use of virtual IP addresses. During a failover, the backup unit uses the same virtual IP address that was used by the primary unit. The Virtual MAC feature avoids the need to update the whole network to associate the virtual IP address with the actual physical MAC address of the backup unit.

Only the switch to which the two firewalls are connected needs to be notified. All outside devices continue to route to the single shared MAC address.

9 To encrypt HA control communication between the active and standby firewalls, select **Enable Encryption for Control Communication**. This option is not selected by default.

IMPORTANT: Firewall performance may be affected if you choose encryption.

After the confirmation message appears, click **OK**.

10 Click the **HA Devices** tab to configure the Secondary firewall serial number. The Serial Number for the Primary Device is displayed, and the field is dimmed and cannot be edited.

11 Under the **HA Devices** tab, enter the serial number of the **Secondary Device**.

12 When you are finished, click **Update**. The settings are changed for each selected SonicWall appliance.

13 Click the **HA Interfaces** tab.

14 Select the interface for the **HA Control Interface**. This option is dimmed and the interface displayed if the firewall detects that the interface is already configured.

15 Select the interface for the **Active/Active DPI Interface**. This option is dimmed and the interface displayed out if the firewall detects that the interface is already configured.

16 When finished with all High Availability configuration, click **Apply**. All settings are synchronized to the Standby unit, and the Standby unit reboots.

17 Go to the **High Availability > Advanced** page and follow the steps in Configuring Advanced High Availability Settings.

18 Go to the **High Availability > Monitoring** page and follow the steps in Monitoring High Availability.

19 Go to the **Network > Interfaces** page to verify that you have successfully configured the interfaces that you want.

20 Go to the **High Availability > Status** page to verify your settings for clustering.

Configuring HA with Dynamic WAN Interfaces

The configuration tasks on the **High Availability > Settings** page are performed on the Primary firewall and then are automatically synchronized to the Secondary.

To configure HA with a dynamic WAN interface:

1. Navigate to the **Network > Interfaces** page.
2. Configure a WAN interface as PPPoE Unnumbered.
3. Navigate to the **High Availability > Settings** page.
4. Ensure **Enable Stateful Synchronization** is not selected. This option is not selected by default.
5. Ensure **Enable Preempt Mode** is not selected. This option is not selected by default.
6. Select **Enable Virtual MAC**. This option is not selected by default.
7. Configure the **HA Devices** and **HA Interfaces** tabs as described in Configuring Active/Standby High Availability Settings.
8. Click **Apply**.
9. Navigate to **High Availability > Monitoring**.
10. Click the **Configure** icon for the PPPoE Unnumbered interface. The **Edit HA Monitoring** dialog displays.
11. Select **Enable Physical/Link Monitoring**. This option is not selected by default.
12. Ensure the **Primary Address** and **Secondary Address** fields are set to 0.0.0.0.
13. Ensure none of the other checkboxes are selected.
14. Click **OK**.

Configuring Active/Active DPI High Availability Settings

The configuration tasks on the High Availability > Settings page are performed on the Primary firewall and then are automatically synchronized to the Secondary.

To configure Active/Active DPI:

1. Navigate to the **High Availability > Settings** page.
2. In the **Mode** drop-down menu, select **Active/Active DPI**.
3. The **Enable Stateful Synchronization** option is automatically enabled for Active/Active DPI, and the option is dimmed.
4. To back up the settings when you upgrade the firmware version, select **Generate/Overwrite Backup Firmware and Settings When Upgrading Firmware**. This option is not selected by default.
5. Under normal conditions, the **Enable Preempt Mode** option should be disabled for Active/Active DPI.

   This option is not selected by default.

   NOTE: This option instructs the Primary firewall to take back the Primary role when it restarts after a failure; thus, this option only applies to Active/Standby configurations.

6 Select **Enable Virtual MAC** to allow both firewalls in the HA pair to share a single MAC address. This greatly simplifies the process of updating network ARP tables and caches when a failover occurs. Only the switch to which the two firewalls are connected needs to be notified. All outside devices continue to route to the single shared MAC address. This option is not selected by default.

7 Click the **HA Devices** tab. The Serial Number for the Primary Device is displayed, and the field is dimmed and cannot be edited.

8 Enter the **Serial Number** of the **Secondary Device**.

9 Click the **HA Interfaces** tab.

10 Select the interface for the **HA Control Interface**. This option is dimmed and the interface displayed if the firewall detects that the interface is already configured.

11 Select the interface number for the **HA Data Interface**. This option is dimmed and the interface displayed if the firewall detects that the interface is already configured.

12 Select the interface number for the **Active/Active DPI Interface**. This option is dimmed and the interface displayed if the firewall detects that the interface is already configured.

   This interface is used for transferring data between the two firewalls during Active/Active DPI processing. Only unassigned, available interfaces appear in the drop-down menu. The connected interfaces must be the same number on both appliances, and must initially appear as unused, unassigned interfaces in the **Network > Interfaces** page. For example, you could connect X5 on the Primary unit to X5 on the Secondary if X5 is an unassigned interface. After enabling Active/Active DPI, the connected interface will have a **Zone** assignment of **HA Data-Link**.

13 When finished with all High Availability configuration, click **Apply**. All settings are synchronized to the Standby firewall, and the Standby firewall reboots.

3

Configuring Advanced High Availability Settings

The **High Availability > Advanced** page is used to configure the stateful synchronization and Active/Active UTM features. The Advanced page also provides the ability to fine tune a number of High Availability options that manage the settings that trigger the High Availability pair to fail over from the primary to the backup appliance.

To configure advanced High Availability settings:

1 Select a SonicWall appliance and click the **Policies** tab. Expand the **High Availability** tree and click **Advanced**.

2 Select **Enable Stateful Synchronization** to configure stateful High Availability. With Stateful High Availability, the primary unit actively communicates with the backup on a per connection and VPN level. As the primary creates and updates connection cache entries or VPN tunnels, the backup unit is informed of such changes. The backup unit remains in a continuously synchronized state so that it can seamlessly assume the network responsibilities upon failure of the primary unit with no interruption to existing network connections.

NOTE: Stateful High Availability requires an additional license for the primary SonicWall appliance. The license is shared between the primary and backup appliances.

3 To configure Active/Active UTM select **Enable Active/Active UTM**.

In an active/active model, both SonicWall firewall appliances share the processing of Deep Packet Inspection (DPI) UTM services. When Active/Active UTM is enabled on a Stateful HA pair, these DPI UTM services can be processed concurrently with firewall, NAT, and other modules on both the active and idle SonicWall firewall appliances. Processing of all modules other than DPI UTM services is restricted to the active unit.

4 If enabling Active/Active UTM, select an interface in the **HA Data Interface** pull-down list. This interface is used for transferring data between the two units during Active/Active UTM processing. Only unassigned, available interfaces appear in the pull-down list.

5 Optionally, you can fine tune the following options:

- Enter the heartbeat interval (in seconds) in the **Heartbeat Interval** field.
- Specify how long the backup waits before replacing the primary (in seconds) in the **Failover Trigger Level** field.
- To specify how long the SonicWall appliance searches, enter the number of seconds in the **Election Delay Time** field. You can enter a value between 0 and 300 seconds, but the default value of 0 seconds is sufficient in most cases.
- Optionally, change the value in the **Dynamic Route Hold-Down Time** field. This setting is used when a failover occurs on a High Availability pair that is using either RIP or OSPF dynamic routing. When a failover occurs, Dynamic Route Hold-Down Time is the number of seconds the newly-active appliance keeps the dynamic routes it had previously learned in its route table. During this time, the newly-active appliance relearns the dynamic routes in the network. When the Dynamic Route Hold-Down Time duration expires, it deletes the old routes and implements the new routes it has learned from RIP or OSPF. The default value is 45 seconds. In large or complex networks, a larger value might improve network stability during a failover.

6 When changes are made to the Primary or Secondary SonicWall firewall appliance, the changes are automatically synchronized between the two SonicWall firewall appliances. To cause the synchronization to occur now, click **Synchronize Settings**. Additionally, selecting **Include Certificates/Keys** synchronizes certificates and keys between devices.

7 To force the backup device to load and reboot to current firmware from the primary device, click **Synchronize Firmware**.

8 When you are finished, click **Update**. The settings are changed for each selected SonicWall appliance. To clear all screen settings and start over, click **Reset**.

4

Monitoring High Availability

On the **High Availability > Monitoring** page, you can specify IP addresses that the SonicWall security appliance uses to complete an ICMP ping on to determine link viability. When using logical monitors, the SonicWall pings the defined Probe IP Address target from the Primary as well as the Backup SonicWall. If both can successfully ping the target, no failover occurs. If both cannot successfully ping the target, no failover occurs, as the SonicWalls assume that the problem is with the target, and not the SonicWalls. But, if one SonicWall can ping the target but the other SonicWall cannot, it will failover to the SonicWall that can ping the target.

Pv6 High Availability (HA) Monitoring is implemented as an extension of HA Monitoring in IPv4. After configuring HA Monitoring for IPv6, both the primary and backup appliances can be managed from the IPv6 monitoring address, and IPv6 Probing is capable of detecting the network status of HA pairs. The IPv6 HA Monitoring configuration page is inherited from IPv4, so the configuration procedures are almost identical.

Consider the following when configuring IPv6 HA Monitoring:

- The Physical/Link Monitoring and Virtual MAC check boxes are greyed out because they are layer two properties. That is, the properties are used by both IPv4 and IPv6, so user has to configure them in the IPv4 monitoring page.
- The primary/backup IPv6 address must be in the same subnet of the interface, and it cannot be same as the global IP and Link-Local-IP of the primary/backup appliance.
- If the primary/backup monitoring IP is set to (not ::), then they cannot be the same.
- If **Management** is enabled, then the primary/backup monitoring IP cannot be unspecified (such as ::).
- If the probe check box is enabled, then the probe IP cannot be unspecified.

Topics:

Configuring High Availability Monitoring

To configure interface monitoring between the primary and backup appliances:

1. Expand the **High Availability** tree and click **Monitoring**. The Monitoring Settings page displays.
2. Click the configure icon for the X0 interface. The **Interface X0 Monitoring Settings** window displays.
3. Enter the LAN management IP address for the primary appliance in the **Primary IP Address** field.
4. Enter the LAN management IP address for the backup appliance in the **Secondary IP Address** field.
5. Select **Allow Management on Primary/Secondary IP Address**. When this option is enabled for an interface, a green icon appears in the interface's **Management** column in the **Monitoring Settings** table on the **High Availability > Monitoring** page. Management is only allowed on an interface when this option is enabled.

6 In the **Logical/Probe IP Address** field, enter the IP address of a downstream device on the LAN network that should be monitored for connectivity. Typically, this should be a downstream router or server. (If probing is desired on the WAN side, an upstream device should be used.) The Primary and Secondary firewalls regularly ping this probe IP address. If both can successfully ping the target, no failover occurs. If neither can successfully ping the target, no failover occurs, because it is assumed that the problem is with the target and not the firewalls. But, if one firewall can ping the target and the other firewall cannot, failover occurs to the firewall that can ping the target.

   The **Primary IP Address** and **Secondary IP Address** fields must be configured with independent IP addresses on a LAN interface, such as X0, (or a WAN interface, such as X1, for probing on the WAN) to allow logical probing to function correctly.

7 (Optional) To manually override the virtual MAC address, check **Override Virtual MAC** and enter a MAC address. SonicWall recommends that you manually configure the virtual MAC address only if the appliances do not have Internet access (for example, in secure network environments). Allowing the appliances to retrieve the virtual MAC address from the SonicWall back end eliminates the possibility of configuration errors and ensures the uniqueness of the virtual MAC address, which prevents possible conflicts.

8 To configure monitoring on any of the other interfaces, repeat the above steps.

9 When finished with all High Availability monitoring configuration for the selected Cluster Node, click **Update**.

10 Optionally, select a different Cluster Node, repeat the configuration steps, and then click **Apply**.

11 Click the configure icon for the X1 interface and repeat steps 3 through 7 for the WAN IP addresses on the primary and backup appliances.

Verifying High Availability Status

Under the unit view, the Management Service displays whether an appliance is the primary or secondary unit on the **System > Status** page under the **Management** heading.

Another method to determine which SonicWall is active is to check the High Availability Settings Status indicator on the **High Availability > Settings** page. If the primary SonicWall is active, the first line in the page indicates that the primary SonicWall is currently Active. It is also possible to check the status of the backup SonicWall by logging into the LAN IP Address of the backup SonicWall. If the primary SonicWall is operating normally, the status indicates that the backup SonicWall is currently Idle. If the backup has taken over for the primary, the status indicates that the backup is currently Active.

Using the GEM framework, you can also configure the Management Service to send email alerts when there is a change in the status of the High Availability pair. You can configure an alert using the **Unit HF Status** alert type.

You can also view details on High Availability events in the Management Service log that is available on the **Console** tab under the **Log** tree.

5

SonicWall Support

Technical support is available to customers who have purchased SonicWall products with a valid maintenance contract and to customers who have trial versions.

The Support Portal provides self-help tools you can use to solve problems quickly and independently, 24 hours a day, 365 days a year. To access the Support Portal, go to https://www.sonicwall.com/support.

The Support Portal enables you to:

- View knowledge base articles and technical documentation
- View video tutorials
- Access MySonicWall
- Learn about SonicWall professional services
- Review SonicWall Support services and warranty information
- Register for training and certification
- Request technical support or customer service

To contact SonicWall Support, visit https://www.sonicwall.com/support/contact-support.

About This Document

Legend

WARNING: A WARNING icon indicates a potential for property damage, personal injury, or death.

CAUTION: A CAUTION icon indicates potential damage to hardware or loss of data if instructions are not followed.

IMPORTANT, NOTE, TIP, MOBILE, or VIDEO: An information icon indicates supporting information.

Management Services High Availability Setup Administration
Updated - February 2019
232-004734-00 Rev A