# SonicWall® Global Management System High Availability

Administration

SONICWALL®

1

# Firewall High Availability

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

In addition, the GMS 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 or global 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 GMS to be configured to provide a reliable, continuous connection to the public Internet. One SonicWall firewall is configured as the primary firewall, and an identical SonicWall firewall is configured as the secondary firewall. 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 firewall. The Active identifier is a logical role that can be assumed by either a primary or secondary hardware firewall. |
| **Failover** | The actual process in which the Standby firewall assumes the Active role following a qualified failure of the Active firewall. 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 through 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** | High Availability PPPoE support function without State. |
| **Preempt** | Applies to a post-failover condition in which the primary firewall has failed, and the secondary firewall has assumed the Active role. Enabling Preempt causes the primary firewall to seize the Active role from the secondary after the primary firewall has been restored to a verified operational state. |
| **Primary** | The principal hardware firewall itself. The primary identifier is a manual designation and is not subject to conditional changes. Under normal operating conditions, the primary hardware firewall operates in an Active role. |
| **Secondary (Backup)** | The subordinate hardware firewall itself. The secondary identifier is a relational designation and is assumed by a firewall when paired with a primary firewall. Under normal operating conditions, the secondary firewall operates in a standby mode. Upon failure of the primary firewall, the secondary firewall assumes the Active role. |
| **SHF** | State Hardware Failover, a GMS 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 firewall. The standby identifier is a logical role that can be assumed by either a primary or secondary hardware firewall. The Standby firewall assumes the Active role upon a determinable failure of the Active firewall. |
| **STP** | Spanning Tree Protocol. |

# High Availability Modes

High Availability has several operation modes that can be selected on the **High Availability > Settings | General** page:

- **None**—Selecting **None** activates a standard high availability configuration and hardware failover functionality, with the option of enabling stateful High Availability 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 firewall handles all traffic, while the Standby firewall shares its configuration settings and can take over at any time to provide continuous network connectivity if the Active firewall 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 High Availability mode, the dynamic state is continuously synchronized between the Active and Standby firewalls. When the Active firewall 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 High Availability 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 GMS 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 GMS 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 firewalls processing traffic (as multiple gateways), doing DPI and sharing the network load. Each cluster node consists of two firewalls acting as a stateful High Availability pair. Active/Active Clustering provides stateful failover support in addition to load-sharing. Optionally, each cluster node can also consist of a single firewall, 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 GMS 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 High Availability 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 firewalls 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 GMS Expanded License.

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

# Crash Detection

The High Availability feature has a thorough self-diagnostic mechanism for both the Active and Standby firewalls. The failover to the Standby firewall 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 that 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 realtime. 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 Address 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 the 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 that prevents possible conflicts. Optionally, you can manually configure the Virtual MAC Address on the **High Availability > Monitoring** page.

The Virtual MAC Address setting is available even when stateful High Availability is not licensed. When Virtual MAC Address is enabled, it is always used even when stateful synchronization is not enabled.

# Dynamic WAN Interfaces with PPPoE High Availability

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

With GMS, PPPoE can be enabled on interfaces in non-stateful mode or in High Availability Active/Standby mode. PPPoE High Availability provides High Availability 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 firewall connects to the PPPoE server, the firewall synchronizes the PPPoE session ID and server name to the secondary firewall.

When the Active firewall fails, it terminates the PPPoE High Availability 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 preexisting network connections are rebuilt, the PPPoE sessions are reestablished, and the PPP process is renegotiated.

# Stateful Synchronization with DHCP

With the GMS, 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 Address.

During a failover, the Active firewall releases the DHCP lease and, as it becomes the Active firewall, 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 the DNS cache. When the DNS cache is added, deleted, or updated dynamically, it synchronizes to the idle firewall.

# About High Availability 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 High Availability interfaces. The link is sensed at the physical layer to determine link viability.
- Logical monitoring involves configuring the firewall to monitor a reliable device on one or more of the connected networks.

Failure to periodically communicate with the device by the Active firewall in the High Availability pair triggers a failover to the Standby firewall. If neither firewall in the High Availability 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 firewall (supported on all physical interfaces)
- To allow synchronization of licenses between the Standby firewall 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 firewalls in the High Availability pair allows you to log in to each firewall independently for management purposes. Note that non-management traffic is ignored when it is sent to one of these IP addresses. The primary and secondary firewalls' unique LAN IP addresses cannot act as active gateways; all systems connected to the internal LAN 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 is required, because in such a scenario the Standby firewall uses the X0 monitoring IP address to connect to the licensing server with all traffic routed through the Active firewall.

The management IP address of the secondary/Standby firewall is used to allow license synchronization with the SonicWall licensing server, which handles licensing on a per-firewall basis (not per-High Availability pair). Even if the secondary firewall is already registered on MySonicWall before creating the High Availability 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 High Availability pair pings the specified logical probe IP address target from the primary as well as from the secondary firewall. 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 firewalls can successfully ping the target, no failover occurs. If both cannot successfully ping the target, no failover occurs, as GMS assumes the problem is with the target, and not the firewalls. If one firewall can ping the target but the other cannot, however, the High Availability pair fails over to the firewall that can ping the target.

The configuration tasks on the **High Availability > Monitoring** page are performed on the primary firewall and then are automatically synchronized to the secondary.

# About Active/Standby High Availability

High Availability allows two identical firewalls running GMS to be configured to provide a reliable, continuous connection to the public Internet. One firewall is configured as the primary firewall, and an identical firewall is configured as the secondary firewall. 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 High Availability provides standard, High Availability, and hardware failover functionality with the option of enabling stateful High Availability and Active/Active DPI.

High Availability 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 models.

**Topics:**

# Benefits of Active/Standby High Availability

- **Increased network reliability** – In a High Availability configuration, the secondary firewall assumes all network responsibilities when the primary firewall 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 firewall in a High Availability pair.
- **Virtual MAC for reduced convergence time after failover** – The Virtual MAC Address setting 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. 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 High Availability Works

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

High Availability requires one SonicWall firewall configured as the primary firewall, and an identical firewall configured as the secondary firewall. During normal operation, the primary firewall is in an Active state and the secondary SonicWall in an Standby state. If the primary firewall 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 High Availability provides stateless high availability. After a failover to the secondary firewall, all the preexisting network connections must be reestablished, 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 firewall. 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 firewall 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.

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 firewall, an incremental synchronization is pushed to the Standby firewall.
- **Complete** –If the timestamps are out of sync and the Standby firewall is available, a complete synchronization is pushed to the Standby firewall. 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 firewalls so that the secondary can seamlessly assume all network responsibilities if the primary firewall fails, with no interruptions to existing network connections.

**NOTE:** Stateful High Availability is included on NSA 4600 and higher NSA platforms and on all SuperMassive platforms. Stateful High Availability is supported on the TZ500 and higher TZ platforms and NSA 2600 and NSA 3600 platforms with an Extended or Stateful High Availability 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 one percent 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, and 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 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 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 might result in a timeout 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 firewall.
4. The secondary firewall detects the restart of the primary firewall 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 High Availability

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

With Active/Active DPI enabled on a stateful High Availability pair, the Deep Packet Inspection services are processed on the Standby firewall of an High Availability 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, you 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 firewall to X5 on the Standby firewall of the High Availability pair. Certain packet flows on the Active firewall are selected and offloaded to the Standby firewall using the Active/Active DPI Interface. DPI is performed on the Standby firewall and then the results are returned to the Active firewall over the same interface. The remaining processing is performed on the Active firewall.

**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 High Availability

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

# Active/Standby and Active/Active DPI Prerequisites

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

**Topics:**

# Supported Platforms for High Availability

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 High Availability 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 might 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 High Availability Control and High Availability Data. Active/Active DPI requires an additional connection.

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

It is important that the X0 interfaces from all firewalls be connected to the same broadcast domain. Otherwise, traffic failover does not work. Also, X0 is the default redundant High Availability port; if the normal High Availability Control link fails, X0 is used to communicate heartbeats between firewalls. Without X0 in the same broadcast domain, both firewalls would become active if the High Availability 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 because of network policy, the WAN (X1) interface should be connected before registration and licensing are performed.

SonicWall network security firewalls require the following interface link speeds for each designated High Availability interface:

- **High Availability Control Interface**—Can be a 1GB or 10GB interface. 1GB is recommended.

  **NOTE:** Link Aggregation and Port Redundancy are not supported for the High Availability Control Interface.

- **High Availability Data Interface**—Can be a 1GB or 10GB interface. 10GB is recommended.

  The High Availability Control Interface and the High Availability Data Interface can share the same single interface.

  If they share a single interface, 10GB is recommended.

- **Active/Active DPI Interface**—Can be a 1GB or 10GB interface.

# 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 High Availability pair, or cluster node. The connected interfaces must be the same number on both firewalls, and must initially appear as unused, unassigned interfaces on the **Network > Interfaces** page. For example, you could connect X5 on the primary firewall to X5 on the secondary if X5 is an unassigned interface. After enabling Active/Active DPI, the connected interface has a zone assignment of **High Availability Data-Link**.

Certain packet flows on the Active firewall are selected and offloaded to the Standby firewall on the Active/Active DPI Interface. DPI is completed on the Standby firewall and then the results are returned to the Active firewall 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 High Availability pair. This interface takes over transferring data between the two firewalls 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 High Availability pair. The same interface must be selected on each firewall.
2. In the GMS 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 High Availability pair.

# Registering and Associating Firewalls on MySonicWall

To use High Availability, you must register both firewalls and associate them for High Availability 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 High Availability secondary link under Associated Products. Then follow the instructions to select and associate the other firewall for your High Availability pair. For more information about registering your firewalls, see the *Getting Started Guide* for your firewalls.

After the firewalls are associated as an High Availability pair, they can share licenses. In addition to High Availability licenses, this includes the GMS 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 are 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 firewall to associate it with. Or, you can associate two firewalls

that are both already registered. You can also start the process by selecting a registered firewall 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 GMS management interface while logged into the individual management IP address of each firewall. This allows the secondary firewall 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 firewalls.

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 High Availability licenses included with the purchase of the SonicWall network firewall is shown in High Availability Licenses Available with SonicWall Network Security Firewalls. Some platforms require additional licensing to use the stateful synchronization or Active/Active DPI features. GMS 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 firewall on MySonicWall from the GMS management interface, or by applying the license keyset to each firewall when Internet access is not available.

**High Availability Licenses Available with SonicWall Network Security Firewalls**

| Platform | Active/Standby High Availability [1] | Stateful High Availability | 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<br>Stateful High Availability Upgrade License | N/A | N/A |
| TZ600 | Included | Expanded License<br>Stateful High Availability Upgrade License | N/A | N/A |
| NSA 2600 | Included | Expanded license<br>High Availability license | N/A | N/A |
| NSA 3600 | Included | Expanded license<br>High Availability 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

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 High Availability pair whose firewalls do not have Internet access. When live communication with SonicWall's licensing server is not permitted because of a network policy, you can use license keysets to manually apply security service 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 High Availability 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 High Availability pair, a stateless High Availability pair with standard failover, or a single standalone firewall. Dynamic state synchronization is only available in a cluster node if it is a stateful High Availability pair. The traditional SonicWall High Availability protocol or stateful High Availability protocol is used for communication within the cluster node, between the firewalls in the High Availability pair.

When a cluster node is a stateful High Availability 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 High Availability pair. For larger deployments, the cluster can include eight firewalls, configured as four cluster nodes (or High Availability pairs). Within each cluster node, stateful High Availability keeps the dynamic state synchronized for seamless failover with zero loss of data on a single point of failure. Stateful High Availability 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 because neither firewall in the cluster node has an High Availability 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 High Availability pair, in which the secondary firewall can assume the duties of the primary firewall 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.

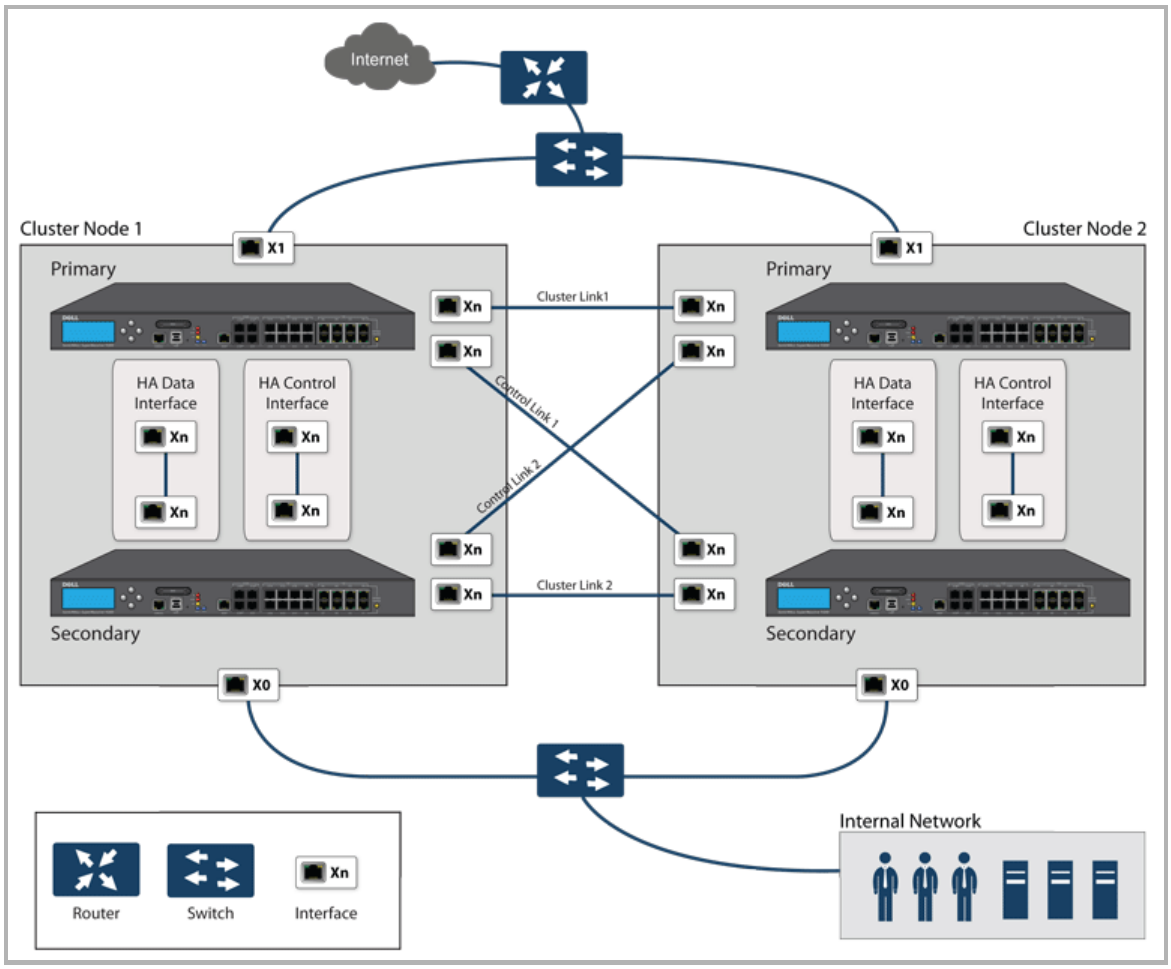
**Topics:**



# Example: Active/Active Clustering – Four-unit Deployment

Active/Active Four-unit Cluster shows a four-unit cluster. Each cluster node contains one High Availability pair. The designated High Availability 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 firewalls in each High Availability 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 firewall of the High Availability pair for DPI processing.

### Active/Active Four-unit Cluster

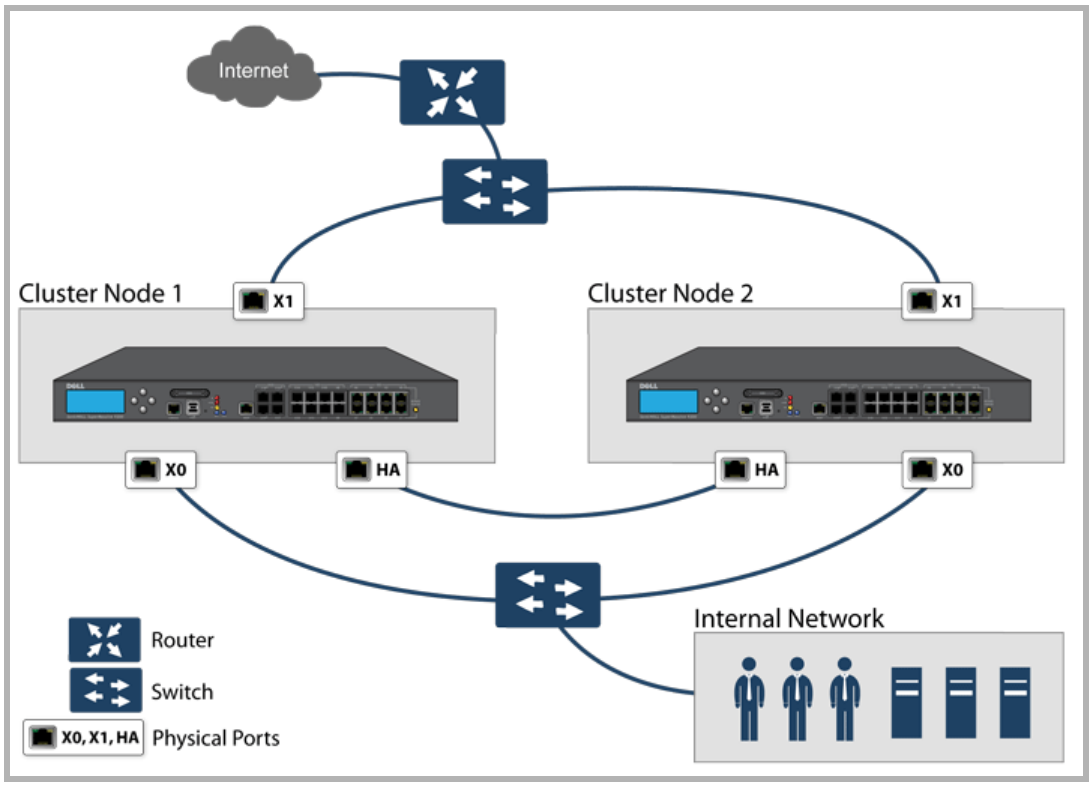



For more information about physically connecting redundant ports and redundant switches, see the *Active/Active Clustering Full Mesh Deployment Technote*.

## Example: Active/Active Clustering – Two-unit Deployment

Active/Active Two-unit Cluster shows a two-unit cluster. In a two-unit cluster, High Availability pairs are not used. Instead, each cluster node contains a single firewall. The designated High Availability 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 High Availability 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 High Availability 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 that are the most processor intensive, on the Standby firewall in each High Availability 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
- About the Cluster

- About Virtual Groups
- About SVRRP
- About Failover
- About DPI with Active/Active Clustering
- About High Availability Monitoring with Active/Clustering

## About Cluster Nodes

An Active/Active Cluster is formed by a collection of cluster nodes. A cluster node can consist of a stateful High Availability pair, a Stateless High Availability pair or a single standalone firewall. Dynamic state synchronization is only available in a cluster node if it is a stateful High Availability pair. The traditional SonicWall High Availability protocol or stateful High Availability protocol is used for communication within the cluster node, between the firewalls in the High Availability pair.

When a cluster node is a stateful High Availability 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 the same product model and 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.

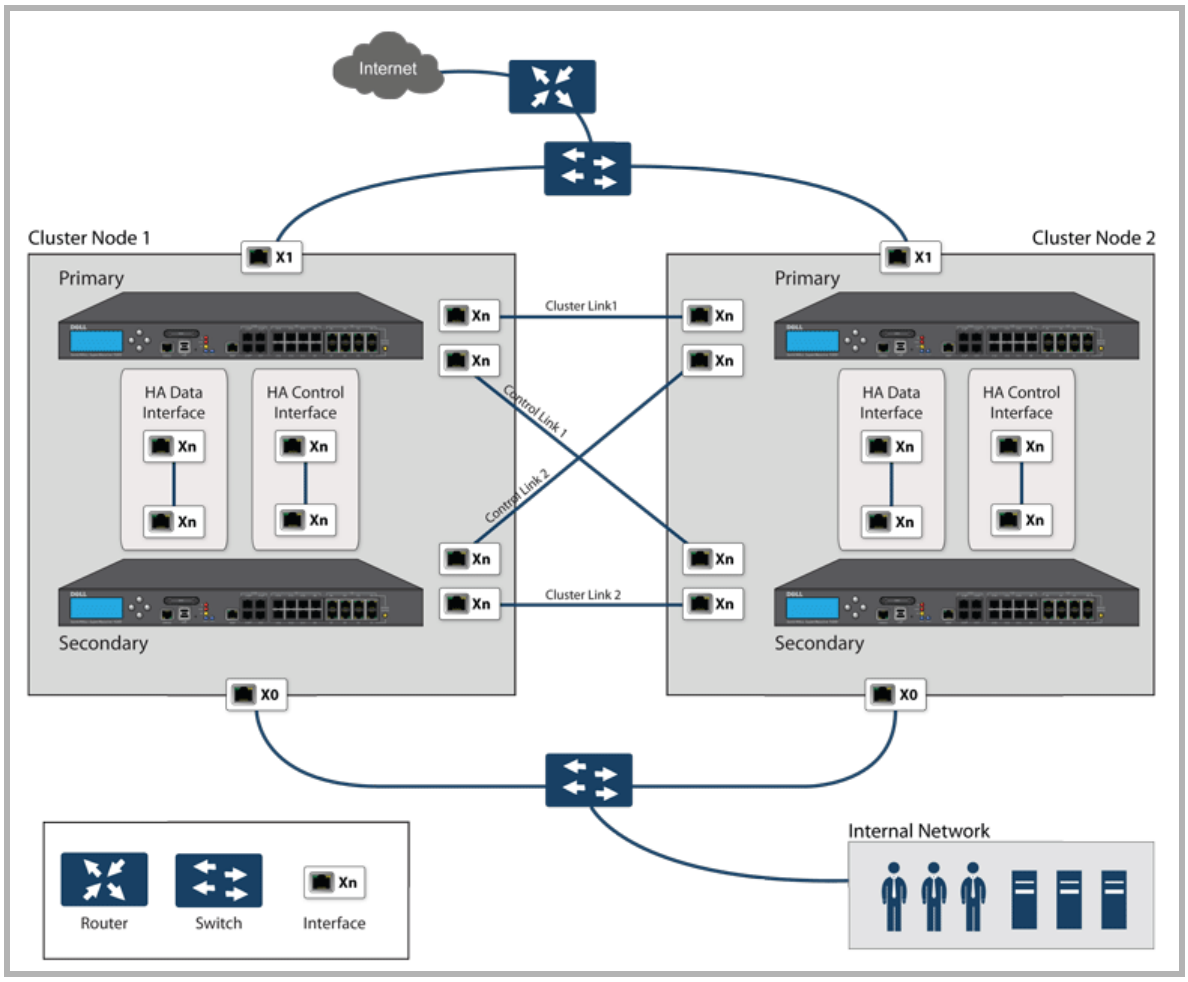
All cluster nodes share the same configuration that is synchronized by the master node. The master node is also responsible for synchronizing firmware to the other nodes in the cluster. The High Availability port connection is used to synchronize configuration and firmware updates.

Dynamic state is not synchronized across cluster nodes, but only within a cluster node. When a cluster node contains an High Availability pair, stateful High Availability 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 is stateless. This means that preexisting network connections must be rebuilt. For example, Telnet and FTP sessions must be reestablished 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 High Availability 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 |
| High Availability synchronize settings (syncs settings to the High Availability peer within the node) | Allowed | Not allowed |

**Administrative Actions Allowed (Continued)**

| Administrative Action | Active Non-master | Standby |
|---|---|---|
| High Availability Synchronize Firmware (syncs firmware to the High Availability peer within the node) | Allowed | Not allowed |
| 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. As a result, 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 becomes the owner.

As part of the configuration for Active/Active Clustering, the serial numbers of other firewalls in the cluster are entered into the GMS 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 GMS 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 takes 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 that 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 that 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 GMS internal DHCP server is turned off and cannot be enabled. Networks needing a DHCP server can use an external DHCP server that is aware of the multiple gateways, so that the gateway allocation can be distributed.

**NOTE:** When Active/Active Clustering is enabled, the GMS 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 GMS. 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 firewall 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 High Availability pair, the secondary firewall takes over for the primary. If stateful High Availability is enabled for the pair, the failover occurs without interruption to network connections.
- **Active/Active failover** – If all the firewalls 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 that 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, High Availability failover is given precedence over Active/Active failover for the following reasons:

- High Availability failover can be stateful, whereas Active/Active failover is stateless.
- The Standby firewall in an High Availability pair is lightly loaded and has resources available for taking over the necessary processing, although it might 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 firewall, 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 seizes 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 because of 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 High Availability 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 High Availability pair for Deep Packet Inspection (DPI) processing.

# About High Availability Monitoring with Active/Clustering

When Active/Active Clustering is enabled, High Availability monitoring configuration is supported for the High Availability pair in each cluster node. The High Availability monitoring features are consistent with previous versions. High Availability 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 High Availability 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, High Availability failover takes precedence over Active/Active failover. If a link fails or a port is disconnected on the Active firewall, the Standby firewall in the High Availability pair becomes 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 High Availability monitoring.

Logical monitoring involves configuring GMS to monitor a reliable device on one or more of the connected networks. Failure to periodically communicate with the device by the Active firewall in the High Availability pair triggers a failover to the Standby firewall. If neither firewall in the High Availability pair can connect to the device, the problem is assumed to be with the device and no failover occurs.

If both physical monitoring and logical monitoring are disabled, Active/Active failover occurs 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 firewall, regardless of the Active or Standby status of the firewall (supported on all physical interfaces)
- To allow synchronization of licenses between the Standby firewall 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 firewalls in the High Availability pair allows you to log in to each firewall 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 High Availability 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 firewall is used to allow license synchronization with the SonicWall licensing server that handles licensing on a per-firewall basis (not per-High Availability pair). Even if the Standby firewall was already registered on MySonicWall before creating the High Availability 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 High Availability license) between the Standby firewall and the SonicWall licensing server.

When using logical monitoring, the High Availability pair pings 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 firewalls 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, the High Availability pair failsover to the SonicWall that can ping the target.

The configuration tasks on the **High Availability > Monitoring** page are performed on the primary firewall 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
- Connecting the High Availability Ports for Active/Active Clustering
- Connecting Redundant Port Interfaces

## 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. GMS 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 firewall on MySonicWall from the GMS interface, or by applying the license keyset to each firewall when 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:<br>• 01-SSC-4480 |
| NSA 6600 | Expanded license:<br>• 01-SSC-4481 |
| SM 9200 | Included |

**Licensing Requirements for A/A Clustering (Continued)**

| Platform | A/A Clustering [1] |
|---|---|
| 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 GMS interface while you are logged into the individual management IP address of each firewall. This allows the secondary firewalls to synchronize with the SonicWall licensing server and share licenses with the associated primary firewalls in each High Availability pair.

## Connecting the High Availability Ports for Active/Active Clustering

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

SonicWall recommends connecting all designated High Availability 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-firewall Active/Active Cluster deployment, where the two cluster nodes each have only a single firewall, you can connect the High Availability ports directly to each other using a crossover cable. No switch is necessary in this case.

The SonicWall Virtual Router Redundancy Protocol (SVRRP) uses this High Availability 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 High Availability 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* High Availability pair must also be associated as High Availability primary and High Availability secondary on MySonicWall. That is, associate the two firewalls in the High Availability pair for cluster node 1, then associate the firewalls in the High Availability 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 High Availability, 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 uses 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 **Manage** view.

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 might already be 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 the failover needs to be renegotiated.

5 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 an upgrade are saved as backup settings.

6 Select **Enable Preempt Mode** to configure the primary SonicWall appliance to take over for 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 overly-aggressive about failing over to the secondary firewall.

7 Select **Enable Virtual MAC**. When the stateful High Availability Upgrade is licensed, Virtual MAC capabilities are also licensed. Virtual MAC allows the backup firewall in a High Availability pair to use the MAC address of the primary firewall when a failover occurs. Alternatively, you can manually set a virtual MAC address for both firewalls 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 firewall uses the same virtual IP address that was used by the primary firewall. 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 firewall.

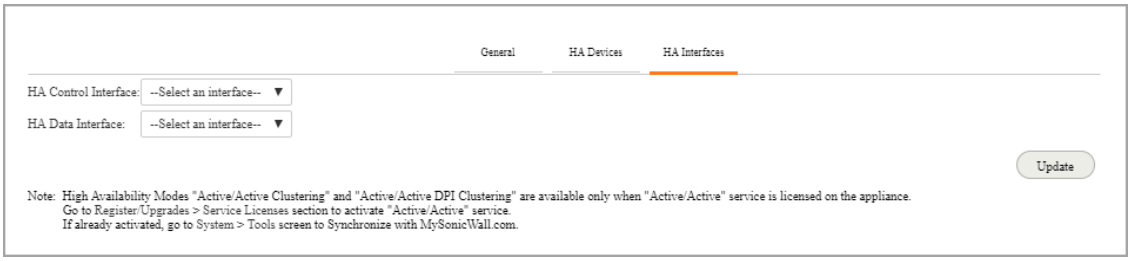
Only the switch to which the two firewalls are connected needs to be notified. All outside devices continue to route to a single shared MAC address.

8 Click **HA Devices** 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.


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

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

11 Click **HA Interfaces**.




12 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.

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

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

15 Go to the **High Availability > Advanced** page and follow the steps in Advanced High Availability Configuration.

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

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

# Configuring High Availability 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 High Availability 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 Choose **Active/Standby** from the **Mode** drop-down menu.

5 Ensure **Enable Stateful Synchronization** is not selected. This option is selected by default.

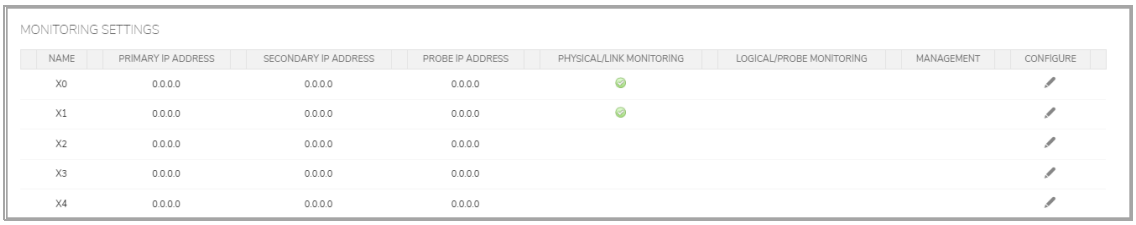
6 Ensure **Enable Preempt Mode** is not selected. This option is not selected by default.

7 Select **Enable Virtual MAC**. This option is not selected by default.

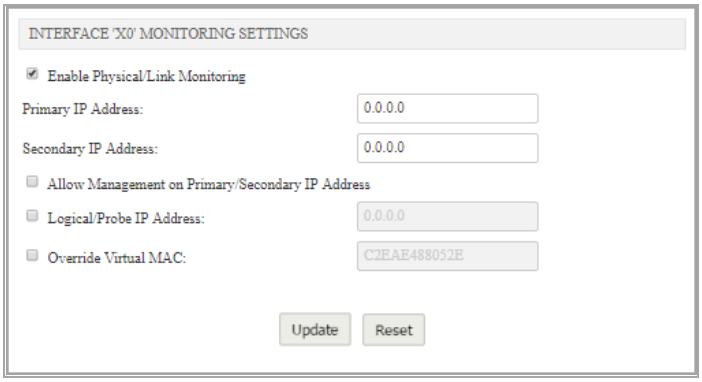
8 Configure the **HA Devices** and **HA Interfaces** tabs as described in Configuring Active/Standby High Availability Settings.

9 Click **Update**.

10 Navigate to **High Availability > Monitoring**.




11 Click the **Configure** icon for the PPPoE Unnumbered interface. The **Edit HA Monitoring** dialog displays.




12 Select **Enable Physical/Link Monitoring**. This option is not selected by default.

13 Ensure the **Primary Address** and **Secondary Address** fields are set to 0.0.0.0.

14 Ensure none of the other checkboxes are selected.

15 Click **Update**.

# 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 backup 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; for that reason, this option only applies to Active/Standby configurations.

6 Select **Enable Virtual MAC** to allow both firewalls in the High Availability 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 **HA Devices**. 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 **HA Interfaces**.


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 firewall to X5 on the secondary if X5 is an unassigned interface. After enabling Active/Active DPI, the connected interface has 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

# Advanced High Availability Configuration

The **High Availability > Advanced** page provides the ability to fine-tune the High Availability configuration as well as synchronize settings and firmware among the High Availability Security Appliances. **High Availability > Advanced** settings are identical for both Active/Standby and Active/Active configurations.

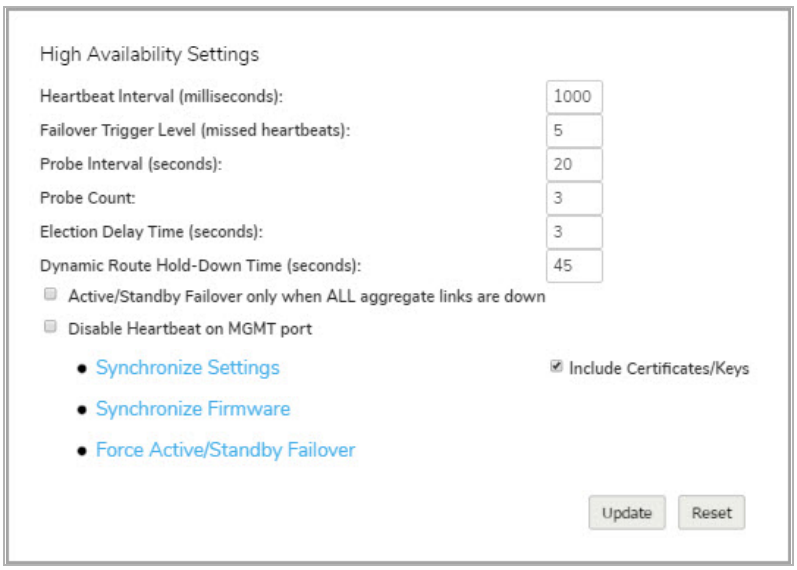
The **Heartbeat Interval** and **Failover Trigger Level (missed heartbeats)** settings apply to both the SVRRP heartbeats (Active/Active Clustering heartbeat) and HA heartbeats. Other settings on **High Availability > Advanced** apply only to the HA pairs within the cluster nodes.

## Configuring Advanced Settings

***To configure advanced settings:***

1 Log in as an administrator to the GMS master node, that is, on the Virtual Group1 IP address (on X0 or another interface with HTTP management enabled).

2 Navigate to **High Availability** | **Advanced**.




3 Optionally adjust the **Heartbeat Interval** to control how often the security appliances in the Active/Active cluster communicate. This setting applies to all units in the Active/Active cluster. The default is **1,000** milliseconds (1 second), the minimum value is 1,000 milliseconds, and the maximum is 300000.

**NOTE:** SonicWall recommends that you set the Heartbeat Interval to at least 1000.

You can use higher values if your deployment handles a lot of network traffic. Lower values might cause unnecessary failovers, especially when the security appliances are under a heavy load.

This timer is linked to the **Failover Trigger Level (missed heartbeats)** timer.

4 Set the **Failover Trigger Level** to the number of heartbeats that can be missed before failing over. This setting applies to all units in the Active/Active cluster. The default is **5**, the minimum is 4, and the maximum is 99.

This timer is linked to the Heartbeat Interval timer. If the **Failover Trigger Level** is set to **5** and the **Heartbeat Interval** is set to 10000 milliseconds (10 seconds), it takes 50 seconds without a heartbeat before a failover is triggered.

5 Set the **Probe Interval** to the interval, in seconds, between probes sent to specified IP addresses to monitor that the network critical path is still reachable. This interval is used in logical monitoring for the local HA pair. The default is **20 seconds**, and the allowed range is 5 to 255 seconds.

**TIP:** SonicWall recommends that you set the interval for at least **5** seconds.

6 Set the **Probe Count** to the number of consecutive probes before GMS concludes that the network critical path is unavailable or the probe target is unreachable. This count is used in logical monitoring for the local High Availability pair. The default is **3**, and the allowed range is 3 to 10.

7 Set the **Election Delay Time** to the number of seconds the Primary Security Appliance waits to consider an interface up and stable. The default is **3** seconds, the minimum is 3 seconds, and the maximum is 255 seconds.

**TIP:** This timer is useful with switch ports that have a spanning-tree delay set.

8 Set the **Dynamic Route Hold-Down Time** to the number of seconds the newly-active Security Appliance keeps the dynamic routes it had previously learned in its route table. The default value is **45** seconds, the minimum is 0 seconds, and the maximum is 1200 seconds (20 minutes).

**NOTE:** The **Dynamic Route Hold-Down Time** setting is displayed only when the **Advanced Routing** option is selected on **Network > Routing**.

**TIP:** In large or complex networks, a larger value might improve network stability during a failover.

This setting is used when a failover occurs on a High Availability pair that is using either RIP or OSPF dynamic routing. During this time, the newly-active appliance relearns the dynamic routes in the network. When the **Dynamic Route Hold-Down Time** duration expires, GMS deletes the old routes and implements the new routes it has learned from RIP or OSPF.

9 If you want failover to occur only when ALL aggregate links are down, select **Active/Standby Failover only when ALL aggregate links are down**. This option is not selected by default.

10 To have the firewalls synchronize all certificates and keys within the High Availability pair, select **Include Certificates/Keys**. This option is selected by default.

11 (Optional) To synchronize the GMS preference settings between your primary and secondary High Availability firewalls, click **Synchronize Settings**.

12 (Optional) To synchronize the firmware version between your primary and secondary High Availability firewalls, click **Synchronize Firmware**.

13 (Optional) To test that the High Availability failover functionality is working properly by attempting an Active/Standby High Availability failover to the secondary firewall, click **Force Active/Standby Failover**.

14 When you are finished with all High Availability configuration, click **Update**. All settings are synchronized to the secondary firewall or to other units in the cluster.

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 fails over to the SonicWall that can ping the target.

IPv6 High Availability (HA) Monitoring is implemented as an extension of High Availability Monitoring in IPv4. After configuring High Availability 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 High Availability pairs. The IPv6 High Availability Monitoring configuration page is inherited from IPv4, so the configuration procedures are almost identical.

Consider the following when configuring IPv6 High Availability Monitoring:

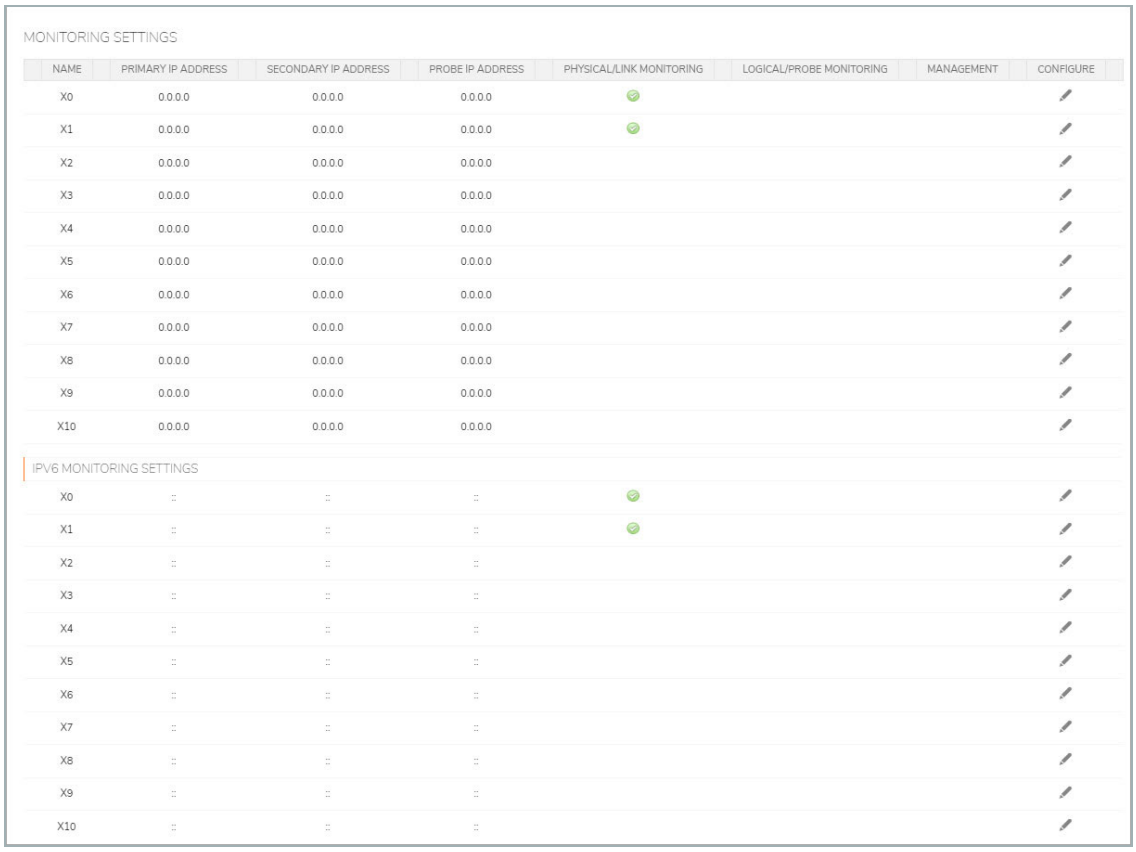
- The Physical/Link Monitoring and Virtual MAC checkboxes are greyed out because they are layer two properties. That is, the properties are used by both IPv4 and IPv6, so user has to configure them in the IPv4 monitoring page.
- The primary/backup IPv6 address must be in the same subnet of the interface, 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 checkbox is enabled, then the probe IP cannot be unspecified.

**Topics:**

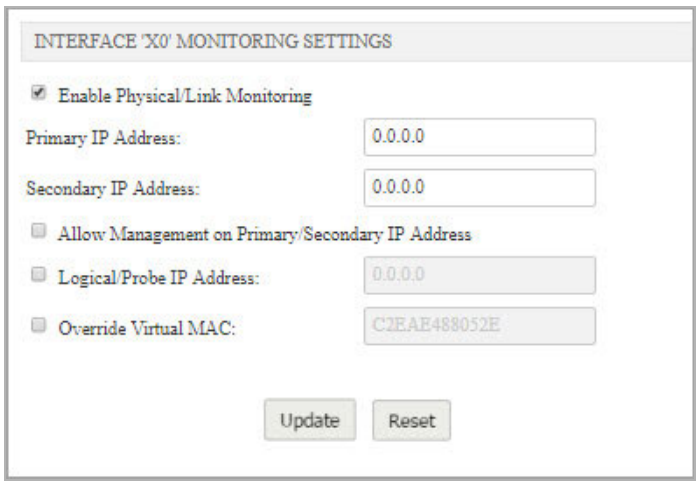
# Configuring High Availability Monitoring

***To configure interface monitoring between the primary and backup appliances:***

1 Navigate to **High Availability** | **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 previous 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 **Update**.

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

# Verifying High Availability Status

Under the firewall view, GMS displays whether an appliance is the primary or secondary firewall 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 GMS 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 GMS 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.

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 and participate in the Community forum discussions at https://community.sonicwall.com/technology-and-support.
- 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

**NOTE:** A NOTE icon indicates supporting information.

**IMPORTANT:** An IMPORTANT icon indicates supporting information that may need a little extra attention.

**TIP:** A TIP indicates helpful information.

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

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

Global Management System High Availability Administration
Updated - November 2020
Software Version - 9.3
232-005130-00 Rev B