

ASA Cluster

Clustering lets you group multiple ASAs together as a single logical device. A cluster provides all the convenience of a single device (management, integration into a network) while achieving the increased throughput and redundancy of multiple devices.

Note Some features are not supported when using clustering. See Unsupported Features with Clustering, on page 71.

- About ASA Clustering, on page 1
- Licenses for ASA Clustering, on page 5
- Requirements and Prerequisites for ASA Clustering, on page 5
- Guidelines for ASA Clustering, on page 7
- Configure ASA Clustering, on page 12
- Manage Cluster Members, on page 44
- Monitoring the ASA Cluster, on page 50
- Examples for ASA Clustering, on page 51
- Reference for Clustering, on page 71
- History for ASA Clustering, on page 86

About ASA Clustering

This section describes the clustering architecture and how it works.

How the ASA Cluster Fits into Your Network

The cluster consists of multiple ASAs acting as a single unit. To act as a cluster, the ASAs need the following infrastructure:

- Isolated, high-speed backplane network for intra-cluster communication, known as the *cluster control link*.
- Management access to each ASA for configuration and monitoring.

When you place the cluster in your network, the upstream and downstream routers need to be able to load-balance the data coming to and from the cluster using one of the following methods:

- Spanned EtherChannel (Recommended)—Interfaces on multiple members of the cluster are grouped into a single EtherChannel; the EtherChannel performs load balancing between units.
- Policy-Based Routing (Routed firewall mode only)—The upstream and downstream routers perform load balancing between units using route maps and ACLs.
- Equal-Cost Multi-Path Routing (Routed firewall mode only)—The upstream and downstream routers perform load balancing between units using equal cost static or dynamic routes.

Cluster Members

Cluster members work together to accomplish the sharing of the security policy and traffic flows. This section describes the nature of each member role.

Bootstrap Configuration

On each device, you configure a minimal bootstrap configuration including the cluster name, cluster control link interface, and other cluster settings. The first unit on which you enable clustering typically becomes the *control* unit. When you enable clustering on subsequent units, they join the cluster as *data* units.

Control and Data Unit Roles

One member of the cluster is the control unit. The control unit is determined by the priority setting in the bootstrap configuration; the priority is set between 1 and 100, where 1 is the highest priority. All other members are data units. Typically, when you first create a cluster, the first unit you add becomes the control unit simply because it is the only unit in the cluster so far.

You must perform all configuration (aside from the bootstrap configuration) on the control unit only; the configuration is then replicated to the data units. In the case of physical assets, such as interfaces, the configuration of the control unit is mirrored on all data units. For example, if you configure GigabitEthernet 0/1 as the inside interface and GigabitEthernet 0/0 as the outside interface, then these interfaces are also used on the data units as inside and outside interfaces.

Some features do not scale in a cluster, and the control unit handles all traffic for those features.

Cluster Interfaces

You can configure data interfaces as either Spanned EtherChannels or as Individual interfaces. All data interfaces in the cluster must be one type only. See About Cluster Interfaces, on page 13 for more information.

Cluster Control Link

Each unit must dedicate at least one hardware interface as the cluster control link. See About the Cluster Control Link, on page 13 for more information.

Configuration Replication

All units in the cluster share a single configuration. You can only make configuration changes on the control unit, and changes are automatically synced to all other units in the cluster.

ASA Cluster Management

One of the benefits of using ASA clustering is the ease of management. This section describes how to manage the cluster.

Management Network

We recommend connecting all units to a single management network. This network is separate from the cluster control link.

Management Interface

For the management interface, we recommend using one of the dedicated management interfaces. You can configure the management interfaces as Individual interfaces (for both routed and transparent modes) or as a Spanned EtherChannel interface.

We recommend using Individual interfaces for management, even if you use Spanned EtherChannels for your data interfaces. Individual interfaces let you connect directly to each unit if necessary, while a Spanned EtherChannel interface only allows remote connection to the current control unit.



Note If you use Spanned EtherChannel interface mode, and configure the management interface as an Individual interface, you cannot enable dynamic routing for the management interface. You must use a static route.

For an Individual interface, the Main cluster IP address is a fixed address for the cluster that always belongs to the current control unit. For each interface, you also configure a range of addresses so that each unit, including the current control unit, can use a Local address from the range. The Main cluster IP address provides consistent management access to an address; when a control unit changes, the Main cluster IP address moves to the new control unit, so management of the cluster continues seamlessly. The Local IP address is used for routing, and is also useful for troubleshooting.

For example, you can manage the cluster by connecting to the Main cluster IP address, which is always attached to the current control unit. To manage an individual member, you can connect to the Local IP address.

For outbound management traffic such as TFTP or syslog, each unit, including the control unit, uses the Local IP address to connect to the server.

For a Spanned EtherChannel interface, you can only configure one IP address, and that IP address is always attached to the control unit. You cannot connect directly to a data unit using the EtherChannel interface; we recommend configuring the management interface as an Individual interface so that you can connect to each unit. Note that you can use a device-local EtherChannel for management.

Control Unit Management Vs. Data Unit Management

All management and monitoring can take place on the control unit. From the control unit, you can check runtime statistics, resource usage, or other monitoring information of all units. You can also issue a command to all units in the cluster, and replicate the console messages from data units to the control unit.

You can monitor data units directly if desired. Although also available from the control unit, you can perform file management on data units (including backing up the configuration and updating images). The following functions are not available from the control unit:

- Monitoring per-unit cluster-specific statistics.
- Syslog monitoring per unit (except for syslogs sent to the console when console replication is enabled).
- SNMP
- NetFlow

RSA Key Replication

When you create an RSA key on the control unit, the key is replicated to all data units. If you have an SSH session to the Main cluster IP address, you will be disconnected if the control unit fails. The new control unit uses the same key for SSH connections, so that you do not need to update the cached SSH host key when you reconnect to the new control unit.

ASDM Connection Certificate IP Address Mismatch

By default, a self-signed certificate is used for the ASDM connection based on the Local IP address. If you connect to the Main cluster IP address using ASDM, then a warning message about a mismatched IP address might appear because the certificate uses the Local IP address, and not the Main cluster IP address. You can ignore the message and establish the ASDM connection. However, to avoid this type of warning, you can enroll a certificate that contains the Main cluster IP address and all the Local IP addresses from the IP address pool. You can then use this certificate for each cluster member. See https://www.cisco.com/c/en/us/td/docs/security/asdm/identity-cert/cert-install.html for more information.

Inter-Site Clustering

For inter-site installations, you can take advantage of ASA clustering as long as you follow the recommended guidelines.

You can configure each cluster chassis to belong to a separate site ID.

Site IDs work with site-specific MAC addresses and IP addresses. Packets egressing the cluster use a site-specific MAC address and IP address, while packets received by the cluster use a global MAC address and IP address. This feature prevents the switches from learning the same global MAC address from both sites on two different ports, which causes MAC flapping; instead, they only learn the site MAC address. Site-specific MAC addresses and IP address are supported for routed mode using Spanned EtherChannels only.

Site IDs are also used to enable flow mobility using LISP inspection, director localization to improve performance and reduce round-trip time latency for inter-site clustering for data centers, and site redundancy for connections where a backup owner of a traffic flow is always at a different site from the owner.

See the following sections for more information about inter-site clustering:

- Sizing the Data Center Interconnect—Requirements and Prerequisites for ASA Clustering, on page 5
- Inter-Site Guidelines—Guidelines for ASA Clustering, on page 7
- Configure Cluster Flow Mobility—Configure Cluster Flow Mobility, on page 41
- Enable Director Localization—Configure Basic ASA Cluster Parameters, on page 36

- Enable Site Redundancy—Configure Basic ASA Cluster Parameters, on page 36
- Inter-Site Examples—Examples for Inter-Site Clustering, on page 67

Licenses for ASA Clustering

Cluster units do not require the same license on each unit. Typically, you buy a license only for the control unit; data units inherit the control unit license. If you have licenses on multiple units, they combine into a single running ASA cluster license.

There are exceptions to this rule. See the following table for precise licensing requirements for clustering.

Model	License Requirement
ASA 5585-X	Cluster License, supports up to 16 units. **Note** Each unit must have the same encryption license; each unit must have the same 10 GE I/O/Security Plus license (ASA 5585-X with SSP-10 and -20).
ASA 5516-X	Base license, supports 2 units. **Note** Each unit must have the same encryption license.
ASA 5512-X	Security Plus license, supports 2 units. **Note** Each unit must have the same encryption license.
ASA 5515-X, ASA 5525-X, ASA 5545-X, ASA 5555-X	Base License, supports 2 units. **Note** Each unit must have the same encryption license.
Firepower 4100/9300 Chassis	See ASA Cluster Licenses for the ASA on the Firepower 4100/9300 Chassis.
All other models	No support.

Requirements and Prerequisites for ASA Clustering

Model Requirements

- ASA 5516-X—Maximum 2 units
- ASA 5512-X, 5515-X, 5525-X, 5545-X, and 5555-X—Maximum 2 units
- ASA 5585-X—Maximum 16 units

  For the ASA 5585-X with SSP-10 and SSP-20, which include two Ten Gigabit Ethernet interfaces, we recommend using one interface for the cluster control link, and the other for data (you can use subinterfaces

for data). Although this setup does not accommodate redundancy for the cluster control link, it does satisfy the need to size the cluster control link to match the size of the data interfaces.

- ASA FirePOWER module—The ASA FirePOWER module does not support clustering directly, but you can use these modules in a cluster. You are responsible for maintaining consistent policies on the ASA FirePOWER modules in the cluster.



Note Create the cluster before you configure the ASA FirePOWER modules. If the modules are already configured on the data units, clear the interface configuration on the devices before adding them to the cluster. From the CLI, enter the **clear configure interface** command.

ASA Hardware and Software Requirements

All units in a cluster:

- Must be the same model with the same DRAM. You do not have to have the same amount of flash memory.
- Must run the identical software except at the time of an image upgrade. Hitless upgrade is supported.
- Must be in the same security context mode, single or multiple.
- (Single context mode) Must be in the same firewall mode, routed or transparent.
- New cluster members must use the same SSL encryption setting (the **ssl encryption** command) as the control unit for initial cluster control link communication before configuration replication.
- Must have the same cluster, encryption and, for the ASA 5585-X, 10 GE I/O licenses.

Switch Requirements

- Be sure to complete the switch configuration before you configure clustering on the ASAs.
- For a list of supported switches, see Cisco ASA Compatibility.

ASA Requirements

- Provide each unit with a unique IP address before you join them to the management network.
  - See the Getting Started chapter for more information about connecting to the ASA and setting the management IP address.
  - Except for the IP address used by the control unit (typically the first unit you add to the cluster), these management IP addresses are for temporary use only.
  - After a data unit joins the cluster, its management interface configuration is replaced by the one replicated from the control unit.
- To use jumbo frames on the cluster control link (recommended), you must enable Jumbo Frame Reservation before you enable clustering.

Sizing the Data Center Interconnect for Inter-Site Clustering

You should reserve bandwidth on the data center interconnect (DCI) for cluster control link traffic equivalent to the following calculation:

$$\frac{\text{\# of cluster members per site}}{2} \times \text{cluster control link size per member}$$

If the number of members differs at each site, use the larger number for your calculation. The minimum bandwidth for the DCI should not be less than the size of the cluster control link for one member.

For example:

- For 4 members at 2 sites:
  - 4 cluster members total
  - 2 members at each site
  - 5 Gbps cluster control link per member

  Reserved DCI bandwidth = 5 Gbps (2/2 x 5 Gbps).
- For 6 members at 3 sites, the size increases:
  - 6 cluster members total
  - 3 members at site 1, 2 members at site 2, and 1 member at site 3
  - 10 Gbps cluster control link per member

  Reserved DCI bandwidth = 15 Gbps (3/2 x 10 Gbps).
- For 2 members at 2 sites:
  - 2 cluster members total
  - 1 member at each site
  - 10 Gbps cluster control link per member

  Reserved DCI bandwidth = 10 Gbps (1/2 x 10 Gbps = 5 Gbps; but the minimum bandwidth should not be less than the size of the cluster control link (10 Gbps)).

Other Requirements

We recommend using a terminal server to access all cluster member unit console ports. For initial setup, and ongoing management (for example, when a unit goes down), a terminal server is useful for remote management.

Guidelines for ASA Clustering

Context Mode

The mode must match on each member unit.

Firewall Mode

For single mode, the firewall mode must match on all units.

Failover

Failover is not supported with clustering.

IPv6

The cluster control link is only supported using IPv4.

Switches

- Make sure connected switches match the MTU for both cluster data interfaces and the cluster control link interface. You should configure the cluster control link interface MTU to be at least 100 bytes higher than the data interface MTU, so make sure to configure the cluster control link connecting switch appropriately. Because the cluster control link traffic includes data packet forwarding, the cluster control link needs to accommodate the entire size of a data packet plus cluster traffic overhead.
- For Cisco IOS XR systems, if you want to set a non-default MTU, set the IOS interface MTU to be 14 bytes higher than the cluster device MTU. Otherwise, OSPF adjacency peering attempts may fail unless the **mtu-ignore** option is used. Note that the cluster device MTU should match the IOS *IPv4* MTU. This adjustment is not required for Cisco Catalyst and Cisco Nexus switches.
- On the switch(es) for the cluster control link interfaces, you can optionally enable Spanning Tree PortFast on the switch ports connected to the cluster unit to speed up the join process for new units.
- On the switch, we recommend that you use one of the following EtherChannel load-balancing algorithms: **source-dest-ip** or **source-dest-ip-port** (see the Cisco Nexus OS and Cisco IOS **port-channel load-balance** command). Do not use a **vlan** keyword in the load-balance algorithm because it can cause unevenly distributed traffic to the devices in a cluster. *Do not* change the load-balancing algorithm from the default on the cluster device.
- If you change the load-balancing algorithm of the EtherChannel on the switch, the EtherChannel interface on the switch temporarily stops forwarding traffic, and the Spanning Tree Protocol restarts. There will be a delay before traffic starts flowing again.
- Some switches do not support dynamic port priority with LACP (active and standby links). You can disable dynamic port priority to provide better compatibility with Spanned EtherChannels.
- Switches on the cluster control link path should not verify the L4 checksum. Redirected traffic over the cluster control link does not have a correct L4 checksum. Switches that verify the L4 checksum could cause traffic to be dropped.
- Port-channel bundling downtime should not exceed the configured keepalive interval.
- On Supervisor 2T EtherChannels, the default hash distribution algorithm is adaptive. To avoid asymmetric traffic in a VSS design, change the hash algorithm on the port-channel connected to the cluster device to fixed:

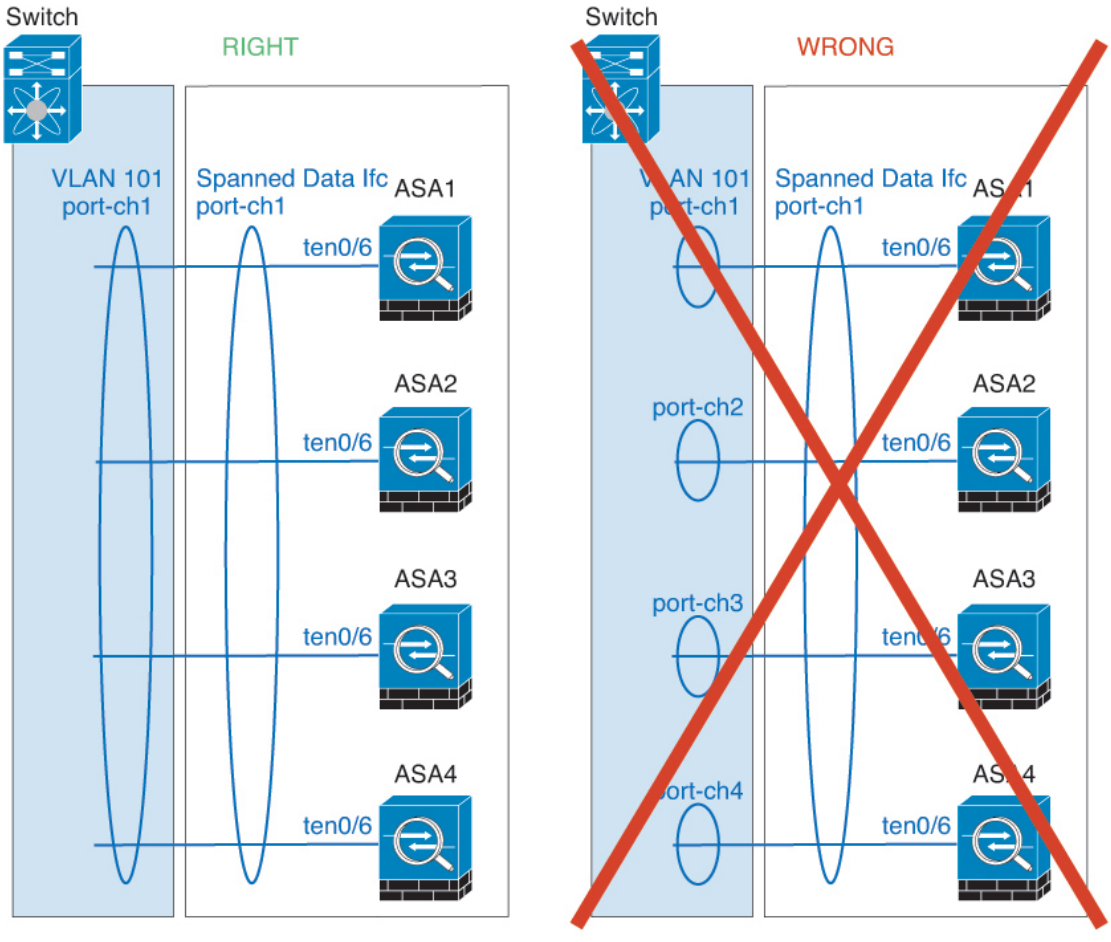
  router(config)# **port-channel** *id* **hash-distribution fixed**

  Do not change the algorithm globally; you may want to take advantage of the adaptive algorithm for the VSS peer link.

- You should disable the LACP Graceful Convergence feature on all cluster-facing EtherChannel interfaces for Cisco Nexus switches.

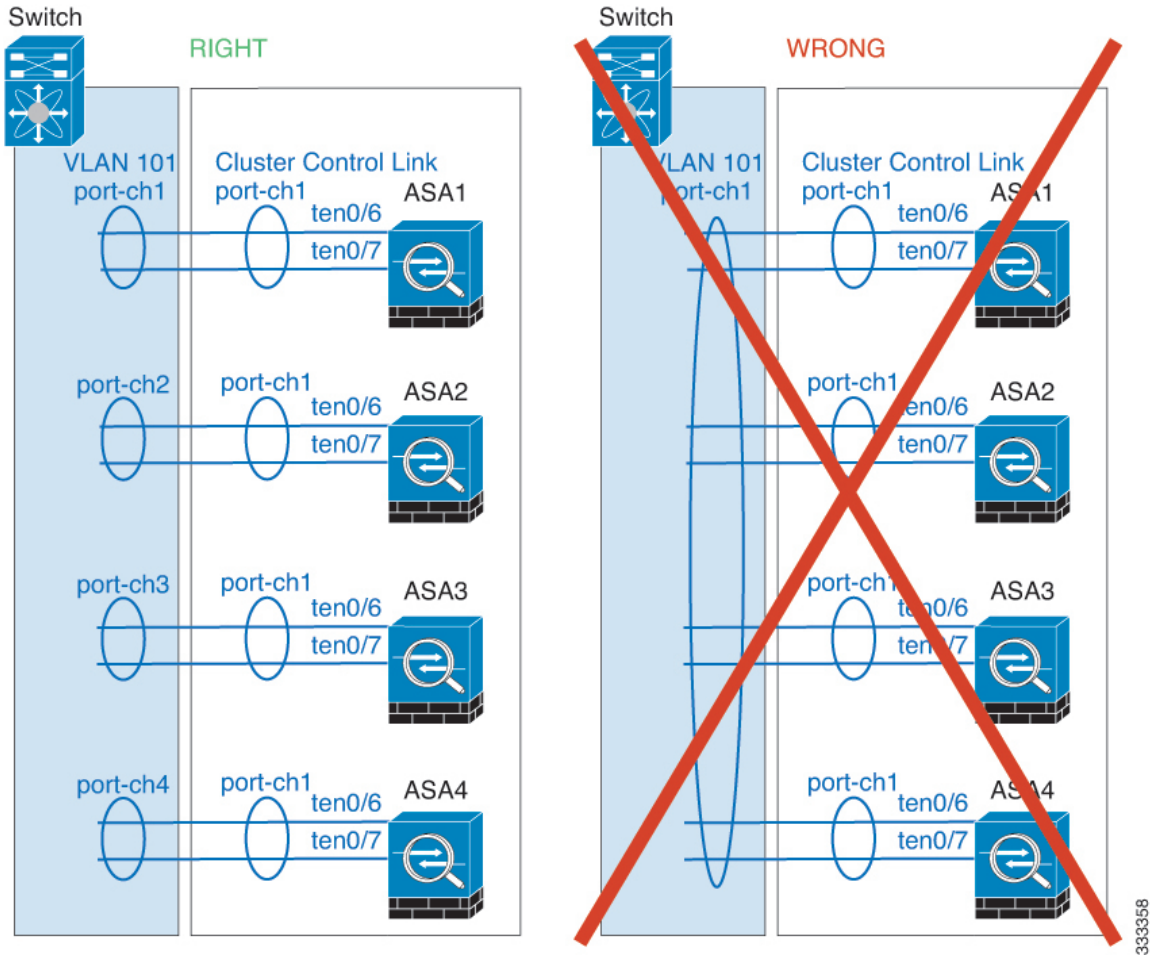
EtherChannels

- In Catalyst 3750-X Cisco IOS software versions earlier than 15.1(1)S2, the cluster unit did not support connecting an EtherChannel to a switch stack. With default switch settings, if the cluster unit EtherChannel is connected cross stack, and if the control unit switch is powered down, then the EtherChannel connected to the remaining switch will not come up. To improve compatibility, set the **stack-mac persistent timer** command to a large enough value to account for reload time; for example, 8 minutes or 0 for indefinite. Or, you can upgrade to more a more stable switch software version, such as 15.1(1)S2.
- Spanned vs. Device-Local EtherChannel Configuration—Be sure to configure the switch appropriately for Spanned EtherChannels vs. Device-local EtherChannels.
  - Spanned EtherChannels—For cluster unit *Spanned* EtherChannels, which span across all members of the cluster, the interfaces are combined into a single EtherChannel on the switch. Make sure each interface is in the same channel group on the switch.



  - Device-local EtherChannels—For cluster unit *Device-local* EtherChannels including any EtherChannels configured for the cluster control link, be sure to configure discrete EtherChannels

on the switch; do not combine multiple cluster unit EtherChannels into one EtherChannel on the switch.



Inter-Site Guidelines

See the following guidelines for inter-site clustering:

- Supports inter-site clustering in the following interface and firewall modes:

Interface Mode	Firewall Mode	
	Routed	**Transparent**
Individual Interface	**Yes**	N/A
Spanned EtherChannel	**Yes**	**Yes**

- For individual interface mode, when using ECMP towards a multicast Rendezvous Point (RP), we recommend that you use a static route for the RP IP address using the Main cluster IP address as the next hop. This static route prevents sending unicast PIM register packets to data units. If a data unit receives a PIM register packet, then the packet is dropped, and the multicast stream cannot be registered.

- The cluster control link latency must be less than 20 ms round-trip time (RTT).
- The cluster control link must be reliable, with no out-of-order or dropped packets; for example, you should use a dedicated link.
- Do not configure connection rebalancing; you do not want connections rebalanced to cluster members at a different site.
- The ASA does not encrypt forwarded data traffic on the cluster control link because it is a dedicated link, even when used on a Data Center Interconnect (DCI). If you use Overlay Transport Virtualization (OTV), or are otherwise extending the cluster control link outside of the local administrative domain, you can configure encryption on your border routers such as 802.1AE MacSec over OTV.
- The cluster implementation does not differentiate between members at multiple sites for incoming connections; therefore, connection roles for a given connection may span across sites. This is expected behavior. However, if you enable director localization, the local director role is always chosen from the same site as the connection owner (according to site ID). Also, the local director chooses a new owner at the same site if the original owner fails (Note: if the traffic is asymmetric across sites, and there is continuous traffic from the remote site after the original owner fails, then a unit from the remote site might become the new owner if it receives a data packet within the re-hosting window.).
- For director localization, the following traffic types do not support localization: NAT or PAT traffic; SCTP-inspected traffic; Fragmentation owner query.
- For transparent mode, if the cluster is placed between a pair of inside and outside routers (AKA North-South insertion), you must ensure that both inside routers share a MAC address, and also that both outside routers share a MAC address. When a cluster member at site 1 forwards a connection to a member at site 2, the destination MAC address is preserved. The packet will only reach the router at site 2 if the MAC address is the same as the router at site 1.
- For transparent mode, if the cluster is placed between data networks and the gateway router at each site for firewalling between internal networks (AKA East-West insertion), then each gateway router should use a First Hop Redundancy Protocol (FHRP) such as HSRP to provide identical virtual IP and MAC address destinations at each site. The data VLANs are extended across the sites using Overlay Transport Virtualization (OTV), or something similar. You need to create filters to prevent traffic that is destined to the local gateway router from being sent over the DCI to the other site. If the gateway router becomes unreachable at one site, you need to remove any filters so traffic can successfully reach the other site's gateway.
- For transparent mode, if the cluster is connected to an HSRP router, you must add the router HSRP MAC address as a static MAC address table entry on the ASA (see Add a Static MAC Address for Bridge Groups). When adjacent routers use HSRP, traffic destined to the HSRP IP address will be sent to the HSRP MAC Address, but return traffic will be sourced from the MAC address of a particular router's interface in the HSRP pair. Therefore, the ASA MAC address table is typically only updated when the ASA ARP table entry for the HSRP IP address expires, and the ASA sends an ARP request and receives a reply. Because the ASA's ARP table entries expire after 14400 seconds by default, but the MAC address table entry expires after 300 seconds by default, a static MAC address entry is required to avoid MAC address table expiration traffic drops.
- For routed mode using Spanned EtherChannel, configure site-specific MAC addresses. Extend the data VLANs across the sites using OTV, or something similar. You need to create filters to prevent traffic that is destined to the global MAC address from being sent over the DCI to the other site. If the cluster becomes unreachable at one site, you need to remove any filters so traffic can successfully reach the other site's cluster units. Dynamic routing is not supported when an inter-site cluster acts as the first hop router for an extended segment.

Additional Guidelines

- When significant topology changes occur (such as adding or removing an EtherChannel interface, enabling or disabling an interface on the ASA or the switch, adding an additional switch to form a VSS or vPC) you should disable the health check feature and also disable interface monitoring for the disabled interfaces. When the topology change is complete, and the configuration change is synced to all units, you can re-enable the interface health check feature.
- When adding a unit to an existing cluster, or when reloading a unit, there will be a temporary, limited packet/connection drop; this is expected behavior. In some cases, the dropped packets can hang your connection; for example, dropping a FIN/ACK packet for an FTP connection will make the FTP client hang. In this case, you need to reestablish the FTP connection.
- If you use a Windows 2003 server connected to a Spanned EtherChannel, when the syslog server port is down and the server does not throttle ICMP error messages, then large numbers of ICMP messages are sent back to the ASA cluster. These messages can result in some units of the ASA cluster experiencing high CPU, which can affect performance. We recommend that you throttle ICMP error messages.
- We do not support VXLAN in Individual Interface mode. Only Spanned EtherChannel mode supports VXLAN.
- We do not support IS-IS in Spanned EtherChannel mode. Only Individual Interface mode supports IS-IS.
- It takes time to replicate changes to all the units in a cluster. If you make a large change, for example, adding an access control rule that uses object groups (which, when deployed, are broken out into multiple rules), the time needed to complete the change can exceed the timeout for the cluster units to respond with a success message. If this happens, you might see a "failed to replicate command" message. You can ignore the message.

Defaults for ASA Clustering

- When using Spanned EtherChannels, the cLACP system ID is auto-generated and the system priority is 1 by default.
- The cluster health check feature is enabled by default with the holdtime of 3 seconds. Interface health monitoring is enabled on all interfaces by default.
- The cluster auto-rejoin feature for a failed cluster control link is unlimited attempts every 5 minutes.
- The cluster auto-rejoin feature for a failed data interface is 3 attempts every 5 minutes, with the increasing interval set to 2.
- Connection rebalancing is disabled by default. If you enable connection rebalancing, the default time between load information exchanges is 5 seconds.
- Connection replication delay of 5 seconds is enabled by default for HTTP traffic.

Configure ASA Clustering

To configure clustering, perform the following tasks.

Note To enable or disable clustering, you must use a console connection (for CLI) or an ASDM connection.

Back Up Your Configurations (Recommended)

When you enable clustering on a secondary unit, the current configuration is replaced with one synced from the primary unit. If you ever want to leave the cluster entirely, it may be useful to have a backup configuration with a usable management interface configuration.

Before you begin

Perform a backup on each unit.

Procedure

Step 1 Choose **Tools** > **Backup Configurations**.

Step 2 Back up at least the running configuration. See Back Up and Restore Configurations or Other Files for a detailed procedure.

Cable the Units and Configure Interfaces

Before configuring clustering, cable the cluster control link network, management network, and data networks. Then configure your interfaces.

About Cluster Interfaces

You can configure data interfaces as either Spanned EtherChannels or as Individual interfaces. All data interfaces in the cluster must be one type only. Each unit must also dedicate at least one hardware interface as the cluster control link.

About the Cluster Control Link

Each unit must dedicate at least one hardware interface as the cluster control link.

Cluster Control Link Traffic Overview

Cluster control link traffic includes both control and data traffic.

Control traffic includes:

- Control unit election.
- Configuration replication.
- Health monitoring.

Data traffic includes:

- State replication.

- Connection ownership queries and data packet forwarding.

Cluster Control Link Interfaces and Network

You can use any data interface(s) for the cluster control link, with the following exceptions:

- You cannot use a VLAN subinterface as the cluster control link.
- You cannot use a Management *x*/*x* interface as the cluster control link, either alone or as an EtherChannel.
- For the ASA 5585-X with an ASA FirePOWER module, Cisco recommends that you use ASA interfaces for the cluster control link, and not interfaces on the ASA FirePOWER module. Module interfaces can drop traffic for up to 30 seconds during a module reload, including reloads that occur during a software upgrade. However, if needed, you can use module interfaces and ASA interfaces in the same cluster control link EtherChannel. When the module interfaces drop, the remaining interfaces in the EtherChannel are still up. The ASA 5585-X Network Module does not run a separate operating system, so it is not affected by this issue.

  Be aware that data interfaces on the module are also affected by reload drops. Cisco recommends always using ASA interfaces redundantly with module interfaces in an EtherChannel.

  For the ASA 5585-X with SSP-10 and SSP-20, which include two Ten Gigabit Ethernet interfaces, we recommend using one interface for the cluster control link, and the other for data (you can use subinterfaces for data). Although this setup does not accommodate redundancy for the cluster control link, it does satisfy the need to size the cluster control link to match the size of the data interfaces.

You can use an EtherChannel or redundant interface.

Each cluster control link has an IP address on the same subnet. This subnet should be isolated from all other traffic, and should include only the ASA cluster control link interfaces.

For a 2-member cluster, do not directly-connect the cluster control link from one ASA to the other ASA. If you directly connect the interfaces, then when one unit fails, the cluster control link fails, and thus the remaining healthy unit fails. If you connect the cluster control link through a switch, then the cluster control link remains up for the healthy unit.

Size the Cluster Control Link

If possible, you should size the cluster control link to match the expected throughput of each chassis so the cluster-control link can handle the worst-case scenarios. For example, if you have the ASA 5585-X with SSP-60, which can pass 14 Gbps per unit maximum in a cluster, then you should also assign interfaces to the cluster control link that can pass at least 14 Gbps. In this case, you could use 2 Ten Gigabit Ethernet interfaces in an EtherChannel for the cluster control link, and use the rest of the interfaces as desired for data links.

Cluster control link traffic is comprised mainly of state update and forwarded packets. The amount of traffic at any given time on the cluster control link varies. The amount of forwarded traffic depends on the load-balancing efficacy or whether there is a lot of traffic for centralized features. For example:

- NAT results in poor load balancing of connections, and the need to rebalance all returning traffic to the correct units.
- AAA for network access is a centralized feature, so all traffic is forwarded to the control unit.
- When membership changes, the cluster needs to rebalance a large number of connections, thus temporarily using a large amount of cluster control link bandwidth.

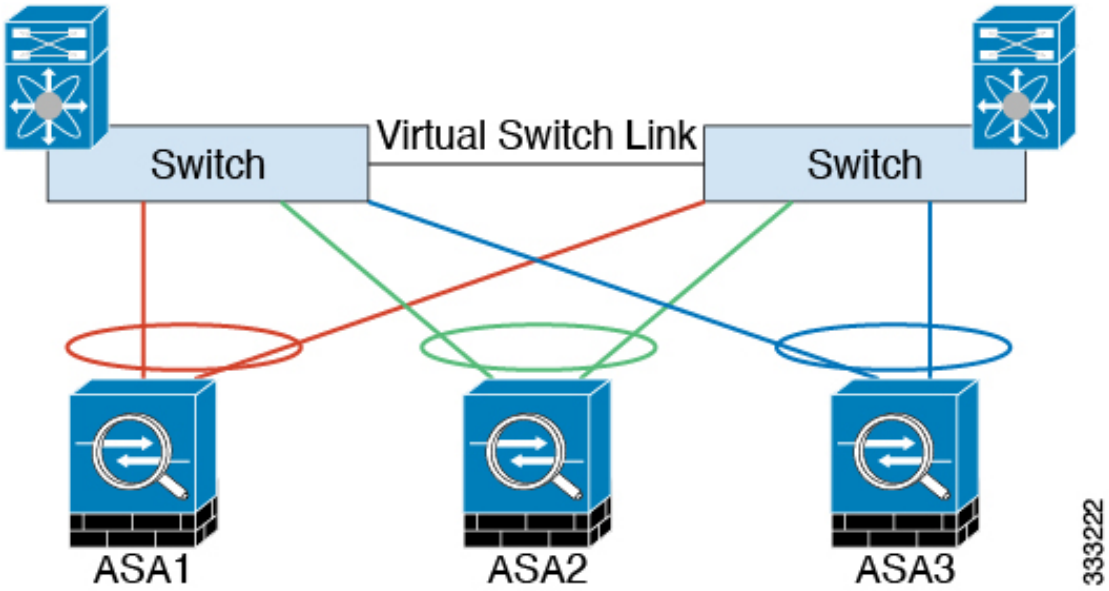
A higher-bandwidth cluster control link helps the cluster to converge faster when there are membership changes and prevents throughput bottlenecks.

Note If your cluster has large amounts of asymmetric (rebalanced) traffic, then you should increase the cluster control link size.

Cluster Control Link Redundancy

We recommend using an EtherChannel for the cluster control link, so that you can pass traffic on multiple links in the EtherChannel while still achieving redundancy.

The following diagram shows how to use an EtherChannel as a cluster control link in a Virtual Switching System (VSS) or Virtual Port Channel (vPC) environment. All links in the EtherChannel are active. When the switch is part of a VSS or vPC, then you can connect ASA interfaces within the same EtherChannel to separate switches in the VSS or vPC. The switch interfaces are members of the same EtherChannel port-channel interface, because the separate switches act like a single switch. Note that this EtherChannel is device-local, not a Spanned EtherChannel.



Cluster Control Link Reliability

To ensure cluster control link functionality, be sure the round-trip time (RTT) between units is less than 20 ms. This maximum latency enhances compatibility with cluster members installed at different geographical sites. To check your latency, perform a ping on the cluster control link between units.

The cluster control link must be reliable, with no out-of-order or dropped packets; for example, for inter-site deployment, you should use a dedicated link.

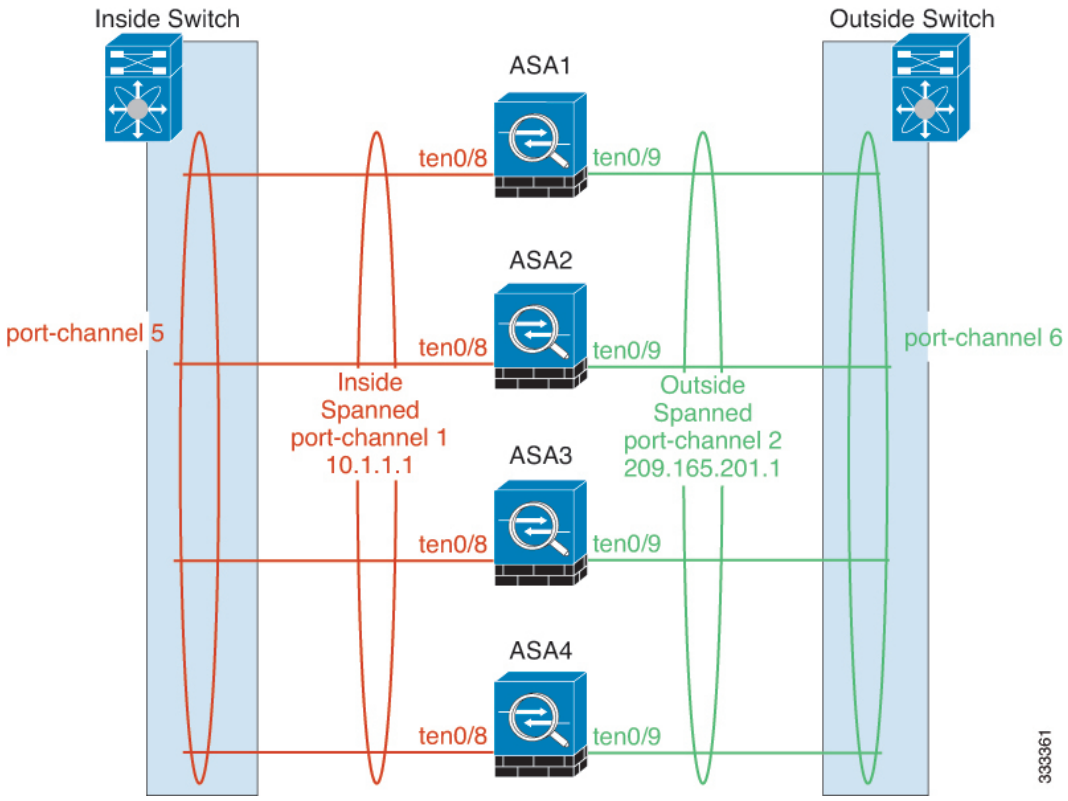
Cluster Control Link Failure

If the cluster control link line protocol goes down for a unit, then clustering is disabled; data interfaces are shut down. After you fix the cluster control link, you must manually rejoin the cluster by re-enabling clustering.

Note When the ASA becomes inactive, all data interfaces are shut down; only the management-only interface can send and receive traffic. The management interface remains up using the IP address the unit received from the cluster IP pool. However if you reload, and the unit is still inactive in the cluster, the management interface is not accessible (because it then uses the Main IP address, which is the same as the control unit). You must use the console port for any further configuration.

Spanned EtherChannels (Recommended)

You can group one or more interfaces per chassis into an EtherChannel that spans all chassis in the cluster. The EtherChannel aggregates the traffic across all the available active interfaces in the channel. A Spanned EtherChannel can be configured in both routed and transparent firewall modes. In routed mode, the EtherChannel is configured as a routed interface with a single IP address. In transparent mode, the IP address is assigned to the BVI, not to the bridge group member interface. The EtherChannel inherently provides load balancing as part of basic operation.



Spanned EtherChannel Benefits

The EtherChannel method of load-balancing is recommended over other methods for the following benefits:

- Faster failure discovery.
- Faster convergence time. Individual interfaces rely on routing protocols to load-balance traffic, and routing protocols often have slow convergence during a link failure.
- Ease of configuration.

Guidelines for Maximum Throughput

To achieve maximum throughput, we recommend the following:

- Use a load balancing hash algorithm that is "symmetric," meaning that packets from both directions will have the same hash, and will be sent to the same ASA in the Spanned EtherChannel. We recommend using the source and destination IP address (the default) or the source and destination port as the hashing algorithm.
- Use the same type of line cards when connecting the ASAs to the switch so that hashing algorithms applied to all packets are the same.

Load Balancing

The EtherChannel link is selected using a proprietary hash algorithm, based on source or destination IP addresses and TCP and UDP port numbers.



Note On the ASA, do not change the load-balancing algorithm from the default. On the switch, we recommend that you use one of the following algorithms: **source-dest-ip** or **source-dest-ip-port** (see the Cisco Nexus OS or Cisco IOS **port-channel load-balance** command). Do not use a **vlan** keyword in the load-balance algorithm because it can cause unevenly distributed traffic to the ASAs in a cluster.

The number of links in the EtherChannel affects load balancing.

Symmetric load balancing is not always possible. If you configure NAT, then forward and return packets will have different IP addresses and/or ports. Return traffic will be sent to a different unit based on the hash, and the cluster will have to redirect most returning traffic to the correct unit.

EtherChannel Redundancy

The EtherChannel has built-in redundancy. It monitors the line protocol status of all links. If one link fails, traffic is re-balanced between remaining links. If all links in the EtherChannel fail on a particular unit, but other units are still active, then the unit is removed from the cluster.

Connecting to a VSS or vPC

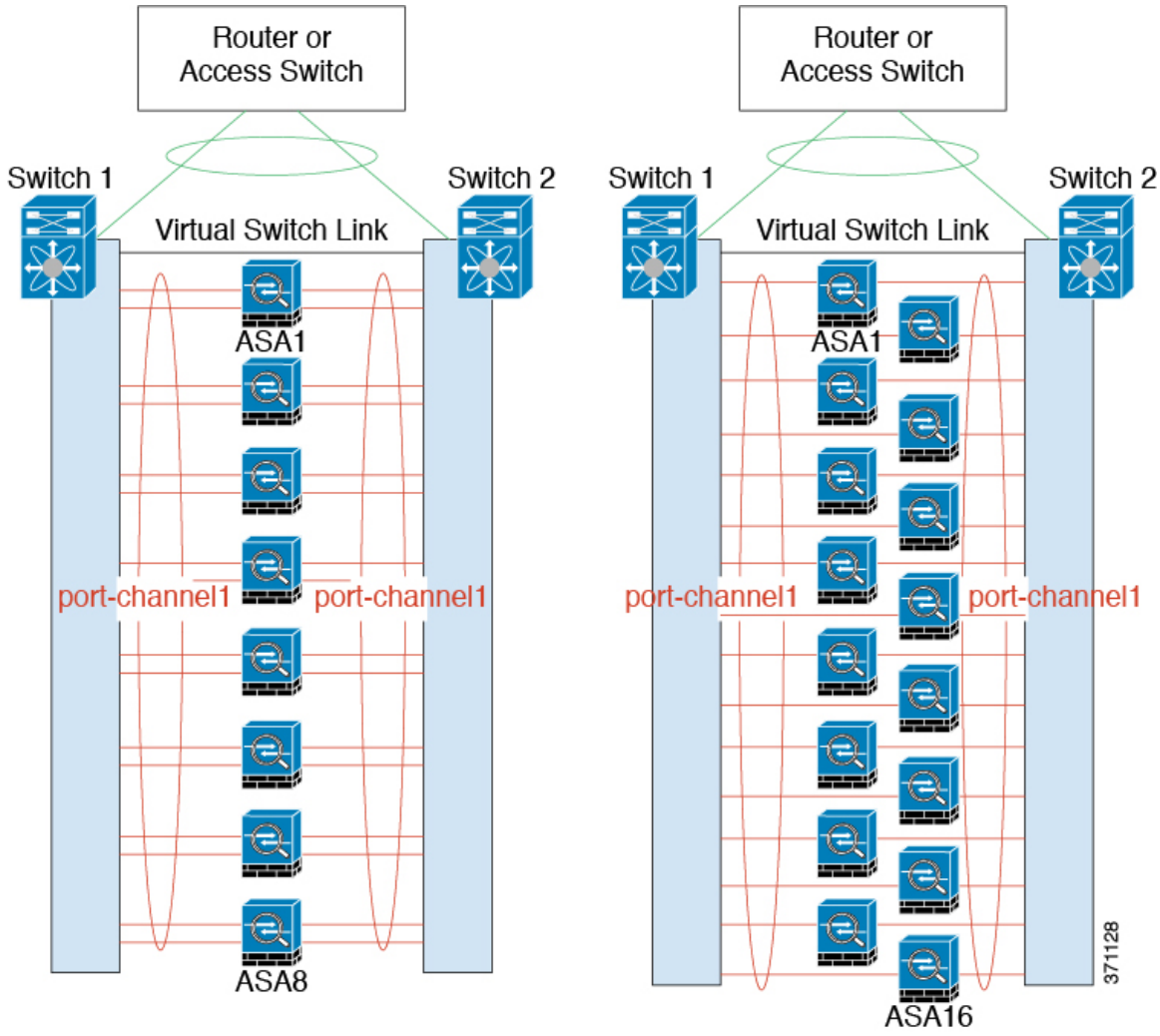
You can include multiple interfaces per ASA in the Spanned EtherChannel. Multiple interfaces per ASA are especially useful for connecting to both switches in a VSS or vPC.

Depending on your switches, you can configure up to 32 active links in the spanned EtherChannel. This feature requires both switches in the vPC to support EtherChannels with 16 active links each (for example the Cisco Nexus 7000 with F2-Series 10 Gigabit Ethernet Module).

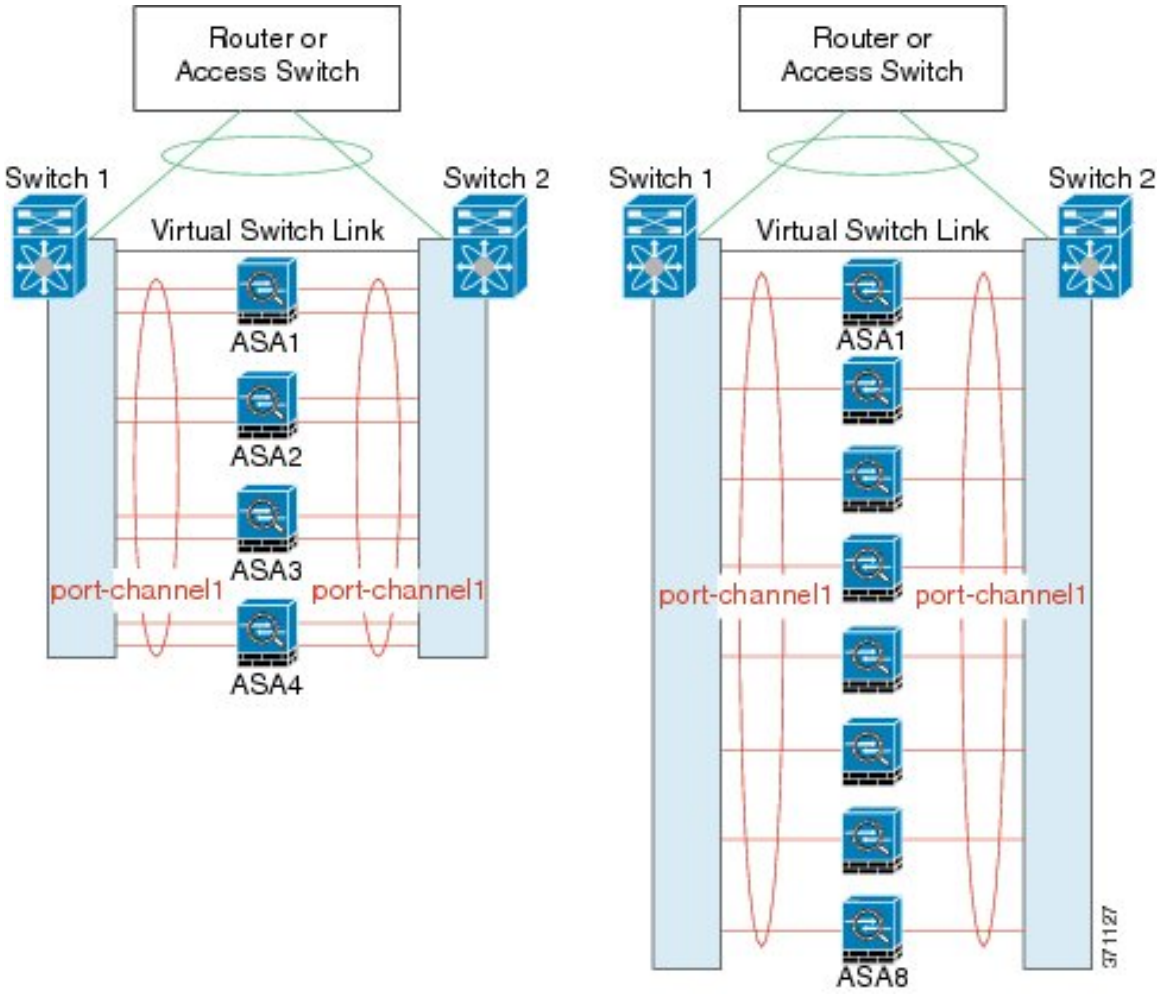
For switches that support 8 active links in the EtherChannel, you can configure up to 16 active links in the spanned EtherChannel when connecting to two switches in a VSS/vPC.

If you want to use more than 8 active links in a spanned EtherChannel, you cannot also have standby links; the support for 9 to 32 active links requires you to disable cLACP dynamic port priority that allows the use of standby links. You can still use 8 active links and 8 standby links if desired, for example, when connecting to a single switch.

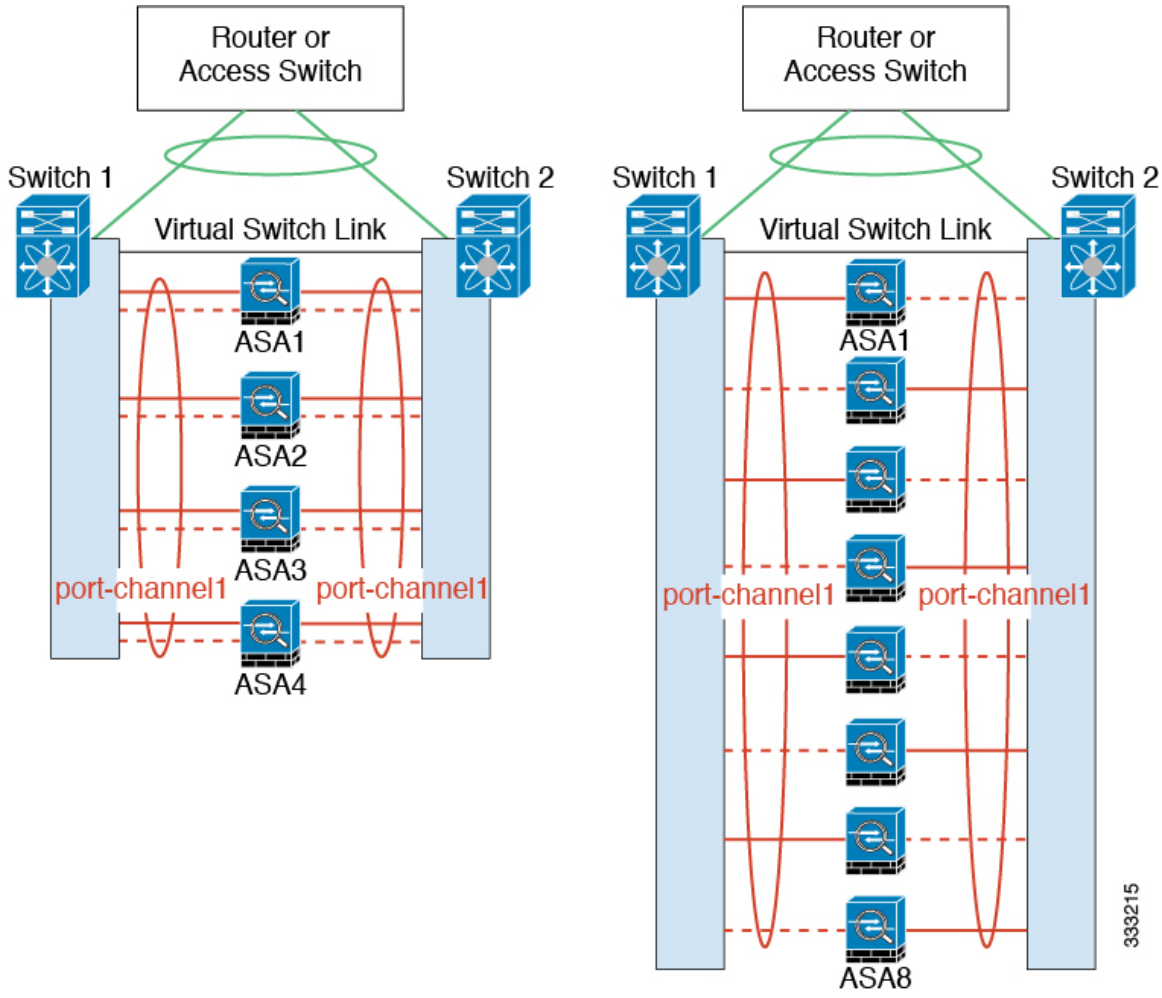
The following figure shows a 32 active link spanned EtherChannel in an 8-ASA cluster and a 16-ASA cluster.



The following figure shows a 16 active link spanned EtherChannel in a 4-ASA cluster and an 8-ASA cluster.



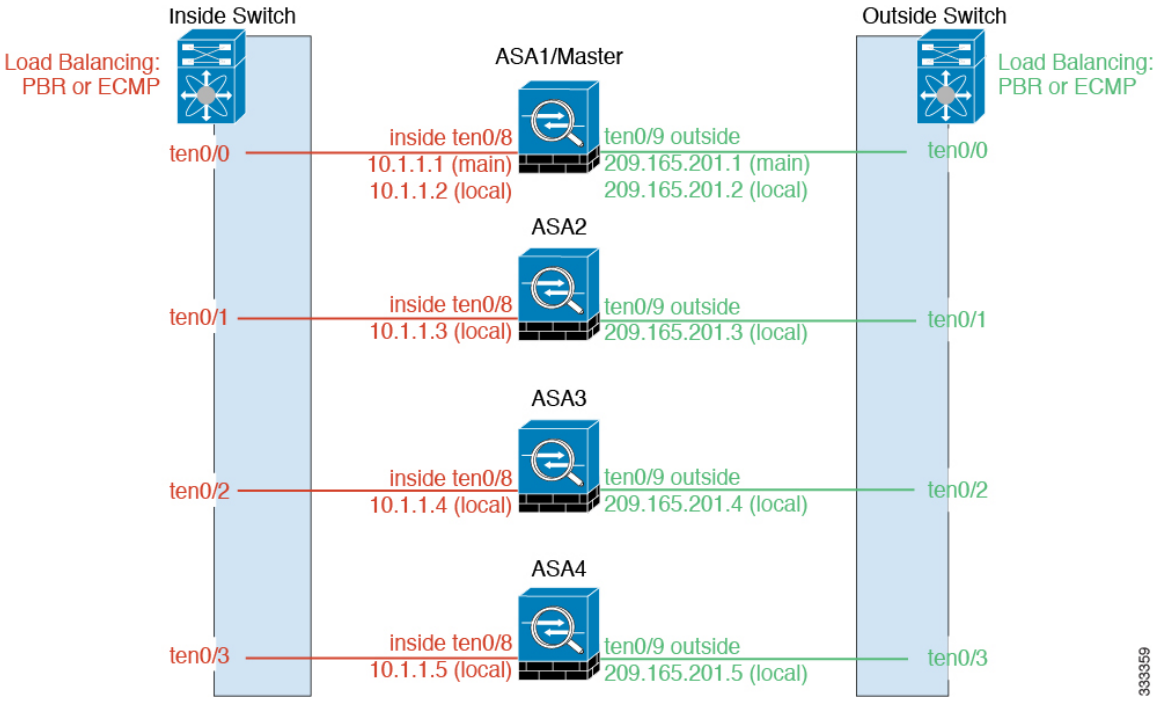
The following figure shows a traditional 8 active/8 standby link spanned EtherChannel in a 4-ASA cluster and an 8-ASA cluster. The active links are shown as solid lines, while the inactive links are dotted. cLACP load-balancing can automatically choose the best 8 links to be active in the EtherChannel. As shown, cLACP helps achieve load balancing at the link level.



Individual Interfaces (Routed Firewall Mode Only)

Individual interfaces are normal routed interfaces, each with their own *Local IP address*. Because interface configuration must be configured only on the control unit, the interface configuration lets you set a pool of IP addresses to be used for a given interface on the cluster members, including one for the control unit. The *Main cluster IP address* is a fixed address for the cluster that always belongs to the current control unit. The Main cluster IP address is a data unit IP address for the control unit; the Local IP address is always the control unit address for routing. The Main cluster IP address provides consistent management access to an address; when a control unit changes, the Main cluster IP address moves to the new control unit, so management of the cluster continues seamlessly. Load balancing, however, must be configured separately on the upstream switch in this case.

Note We recommend Spanned EtherChannels instead of Individual interfaces because Individual interfaces rely on routing protocols to load-balance traffic, and routing protocols often have slow convergence during a link failure.



Policy-Based Routing (Routed Firewall Mode Only)

When using Individual interfaces, each ASA interface maintains its own IP address and MAC address. One method of load balancing is Policy-Based Routing (PBR).

We recommend this method if you are already using PBR, and want to take advantage of your existing infrastructure. This method might offer additional tuning options vs. Spanned EtherChannel as well.

PBR makes routing decisions based on a route map and ACL. You must manually divide traffic between all ASAs in a cluster. Because PBR is static, it may not achieve the optimum load balancing result at all times. To achieve the best performance, we recommend that you configure the PBR policy so that forward and return packets of a connection are directed to the same physical ASA. For example, if you have a Cisco router, redundancy can be achieved by using Cisco IOS PBR with Object Tracking. Cisco IOS Object Tracking monitors each ASA using ICMP ping. PBR can then enable or disable route maps based on reachability of a particular ASA. See the following URLs for more details:

http://www.cisco.com/c/en/us/solutions/data-center-virtualization/intelligent-traffic-director/index.html

http://www.cisco.com/en/US/products/ps6599/products_white_paper09186a00800a4409.shtml

Note If you use this method of load-balancing, you can use a device-local EtherChannel as an Individual interface.

Equal-Cost Multi-Path Routing (Routed Firewall Mode Only)

When using Individual interfaces, each ASA interface maintains its own IP address and MAC address. One method of load balancing is Equal-Cost Multi-Path (ECMP) routing.

We recommend this method if you are already using ECMP, and want to take advantage of your existing infrastructure. This method might offer additional tuning options vs. Spanned EtherChannel as well.

ECMP routing can forward packets over multiple "best paths" that tie for top place in the routing metric. Like EtherChannel, a hash of source and destination IP addresses and/or source and destination ports can be used to send a packet to one of the next hops. If you use static routes for ECMP routing, then an ASA failure can cause problems; the route continues to be used, and traffic to the failed ASA will be lost. If you use static routes, be sure to use a static route monitoring feature such as Object Tracking. We recommend using dynamic routing protocols to add and remove routes, in which case, you must configure each ASA to participate in dynamic routing.

Note If you use this method of load-balancing, you can use a device-local EtherChannel as an Individual interface.

Nexus Intelligent Traffic Director (Routed Firewall Mode Only)

When using Individual interfaces, each ASA interface maintains its own IP address and MAC address. Intelligent Traffic Director (ITD) is a high-speed hardware load-balancing solution for Nexus 5000, 6000, 7000, and 9000 switch series. In addition to fully covering the functional capabilities of traditional PBR, it offers a simplified configuration workflow and multiple additional features for a more granular load distribution.

ITD supports IP stickiness, consistent hashing for bi-directional flow symmetry, virtual IP addressing, health monitoring, sophisticated failure handling policies with N+M redundancy, weighted load-balancing, and application IP SLA probes including DNS. Due to the dynamic nature of load-balancing, it achieves a more even traffic distribution across all cluster members as compared to PBR. In order to achieve bi-directional flow symmetry, we recommend configuring ITD such that forward and return packets of a connection are directed to the same physical ASA. See the following URL for more details:

http://www.cisco.com/c/en/us/solutions/data-center-virtualization/intelligent-traffic-director/index.html

Cable the Cluster Units and Configure Upstream and Downstream Equipment

Before configuring clustering, cable the cluster control link network, management network, and data networks.

Procedure

Cable the cluster control link network, management network, and data networks.

Note At a minimum, an active cluster control link network is required before you configure the units to join the cluster.

You should also configure the upstream and downstream equipment. For example, if you use EtherChannels, then you should configure the upstream and downstream equipment for the EtherChannels.

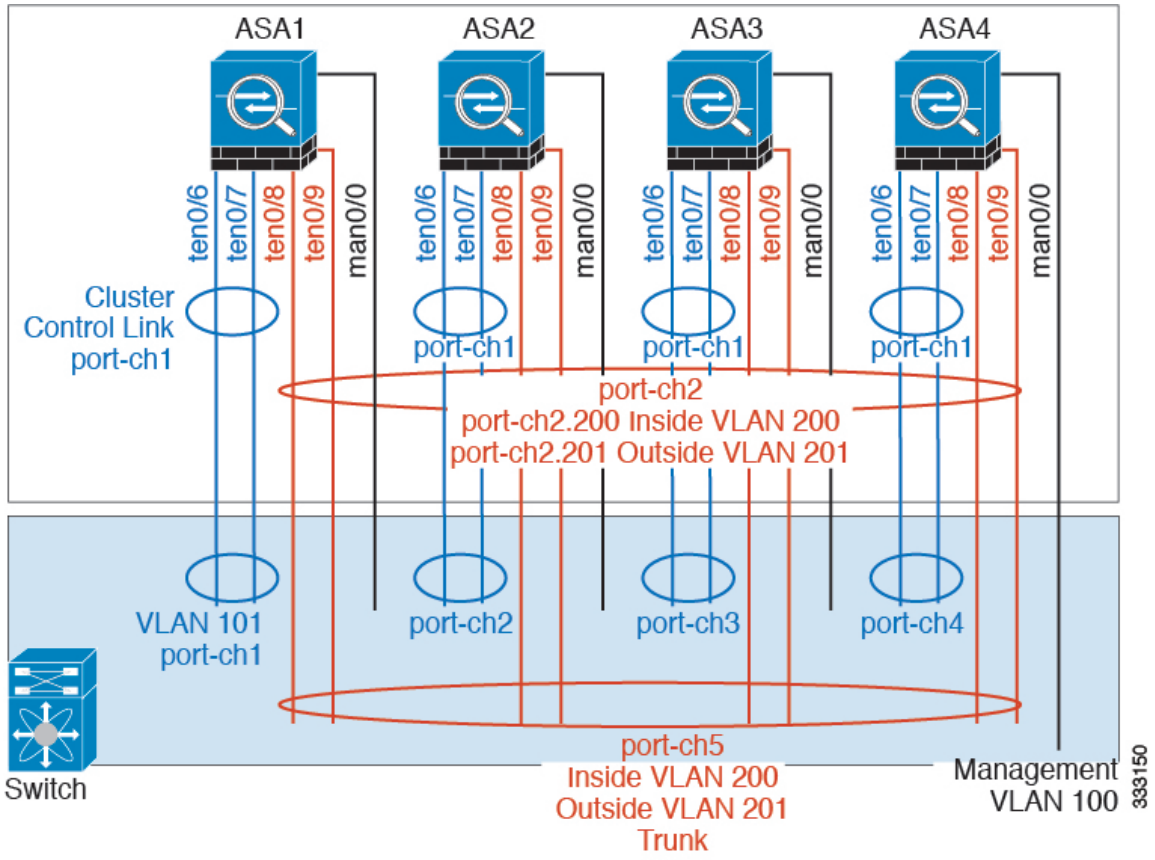
Examples

Note This example uses EtherChannels for load-balancing. If you are using PBR or ECMP, your switch configuration will differ.

For example on each of 4 ASA 5585-Xs, you want to use:

- 2 Ten Gigabit Ethernet interfaces in a device-local EtherChannel for the cluster control link.
- 2 Ten Gigabit Ethernet interfaces in a Spanned EtherChannel for the inside and outside network; each interface is a VLAN subinterface of the EtherChannel. Using subinterfaces lets both inside and outside interfaces take advantage of the benefits of an EtherChannel.
- 1 Management interface.

You have one switch for both the inside and outside networks.



Purpose	Connect Interfaces on each of 4 ASAs	To Switch Ports
Cluster control link	TenGigabitEthernet 0/6 and TenGigabitEthernet 0/7	8 ports total For each TenGigabitEthernet 0/6 and TenGigabitEthernet 0/7 pair, configure 4 EtherChannels (1 EC for each ASA). These EtherChannels must all be on the same isolated cluster control VLAN, for example VLAN 101.

Purpose	Connect Interfaces on each of 4 ASAs	To Switch Ports
Inside and outside interfaces	TenGigabitEthernet 0/8 and TenGigabitEthernet 0/9	8 ports total Configure a single EtherChannel (across all ASAs). On the switch, configure these VLANs and networks now; for example, a trunk including VLAN 200 for the inside and VLAN 201 for the outside.
Management interface	Management 0/0	4 ports total Place all interfaces on the same isolated management VLAN, for example VLAN 100.

Configure the Cluster Interface Mode on the Control Unit

You can only configure one type of interface for clustering: Spanned EtherChannels or Individual interfaces; you cannot mix interface types in a cluster.

Note If you do not add data units from the control unit, you must set the interface mode manually on all units according to this section, not just the control unit; if you add secondaries from the control unit, ASDM sets the interface mode automatically on the data unit.

Before you begin

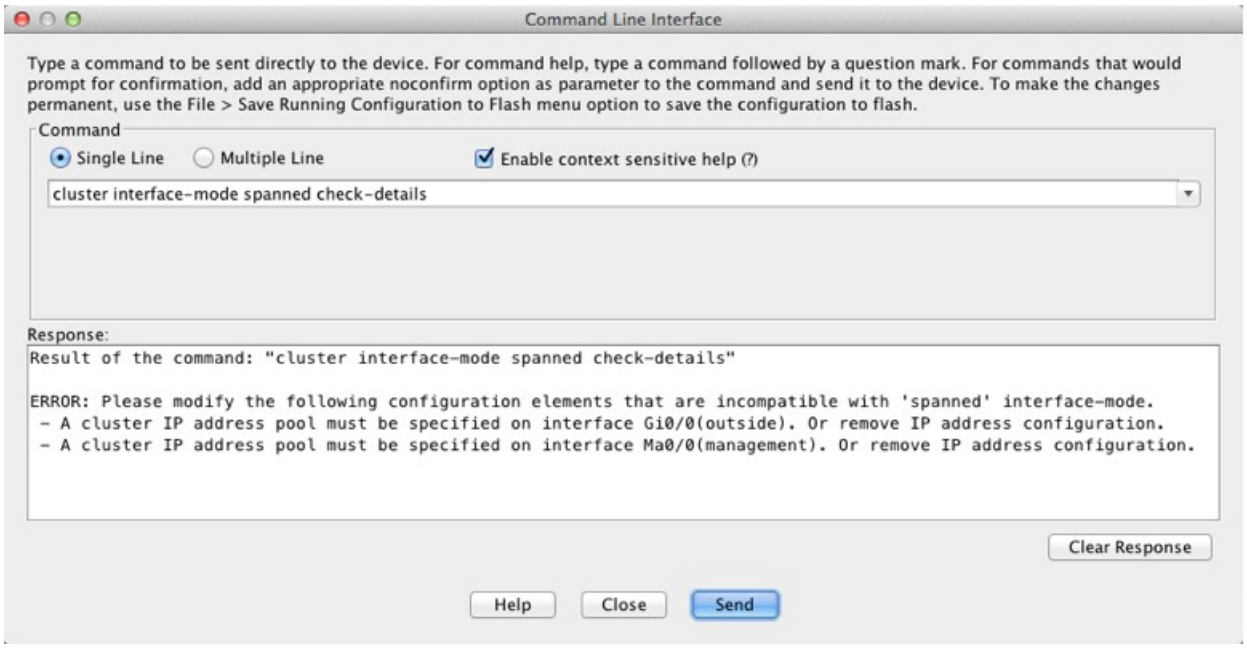
- You can always configure the management-only interface as an Individual interface (recommended), even in Spanned EtherChannel mode. The management interface can be an Individual interface even in transparent firewall mode.
- In Spanned EtherChannel mode, if you configure the management interface as an Individual interface, you cannot enable dynamic routing for the management interface. You must use a static route.
- In multiple context mode, you must choose one interface type for all contexts. For example, if you have a mix of transparent and routed mode contexts, you must use Spanned EtherChannel mode for all contexts because that is the only interface type allowed for transparent mode.

Procedure

Step 1 In ASDM on the control unit, choose **Tools** > **Command Line Interface**.Show any incompatible configuration so that you can force the interface mode and fix your configuration later; the mode is not changed with this command:

cluster interface-mode {**individual** | **spanned**} **check-details**

Example:

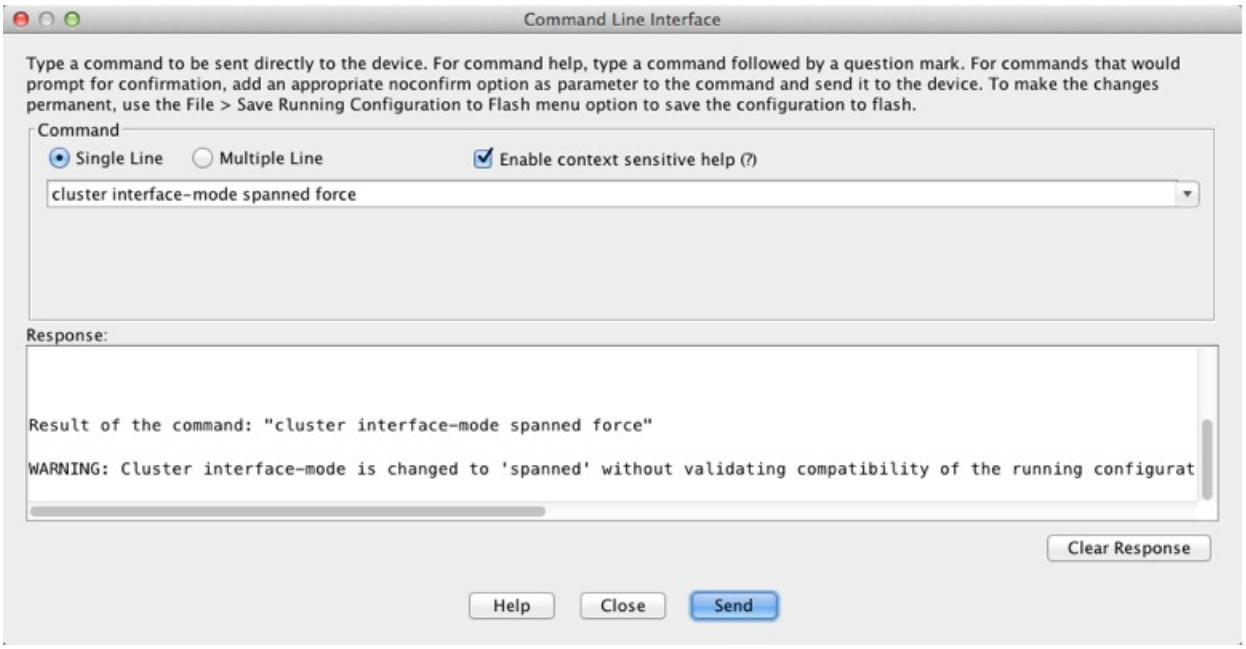


Caution After you set the interface mode, you can continue to connect to the interface; however, if you reload the ASA before you configure your management interface to comply with clustering requirements (for example, adding a cluster IP pool), you will not be able to reconnect because cluster-incompatible interface configuration is removed. In that case, you will have to connect to the console port to fix the interface configuration.

Step 2 Set the interface mode for clustering:

cluster interface-mode {**individual** | **spanned**} **force**

Example:



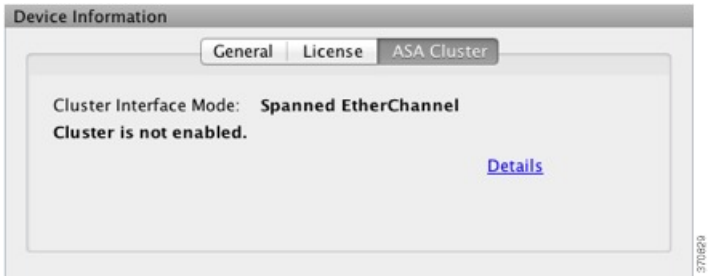
There is no default setting; you must explicitly choose the mode. If you have not set the mode, you cannot enable clustering.

The **force** option changes the mode without checking your configuration for incompatible settings. You need to manually fix any configuration issues after you change the mode. Because any interface configuration can only be fixed after you set the mode, we recommend using the **force** option so that you can at least start from the existing configuration. You can re-run the **check-details** option after you set the mode for more guidance.

Without the **force** option, if there is any incompatible configuration, you are prompted to clear your configuration and reload, thus requiring you to connect to the console port to reconfigure your management access. If your configuration is compatible (rare), the mode is changed and the configuration is preserved. If you do not want to clear your configuration, you can exit the command by typing **n**.

To remove the interface mode, enter the **no cluster interface-mode** command.

Step 3 Quit ASDM and reload. ASDM needs to be restarted to correctly account for the cluster interface mode. After you reload, you see the ASA Cluster tab on the home page:



(Recommended; Required in Multiple Context Mode) Configure Interfaces on the Control Unit

You must modify any interface that is currently configured with an IP address to be cluster-ready before you enable clustering. At a minimum, you must modify the management interface to which ASDM is currently connected. For other interfaces, you can configure them before or after you enable clustering; we recommend pre-configuring all of your interfaces so that the complete configuration is synced to new cluster members. In multiple context mode, you must use the procedures in this section to fix existing interfaces or to configure new interfaces. However, in single mode, you can skip this section and configure common interface parameters within the High Availability and Scalability wizard (see Run the High Availability Wizard, on page 32). Note that advanced interface settings such as creating EtherChannels for Individual interfaces are not available in the wizard.

This section describes how to configure interfaces to be compatible with clustering. You can configure data interfaces as either Spanned EtherChannels or as Individual interfaces. Each method uses a different load-balancing mechanism. You cannot configure both types in the same configuration, with the exception of the management interface, which can be an Individual interface even in Spanned EtherChannel mode.

Configure Individual Interfaces (Recommended for the Management Interface)

Individual interfaces are normal routed interfaces, each with their own IP address taken from a pool of IP addresses. The Main cluster IP address is a fixed address for the cluster that always belongs to the current primary unit.

In Spanned EtherChannel mode, we recommend configuring the management interface as an Individual interface. Individual management interfaces let you connect directly to each unit if necessary, while a Spanned EtherChannel interface only allows connection to the current primary unit.

Before you begin

- Except for the management-only interface, you must be in Individual interface mode.
- For multiple context mode, perform this procedure in each context. If you are not already in the context configuration mode in the Configuration > Device List pane, double-click the context name under the active device IP address.
- Individual interfaces require you to configure load balancing on neighbor devices. External load balancing is not required for the management interface.
- (Optional) Configure the interface as a device-local EtherChannel interface, a redundant interface, and/or configure subinterfaces.
  - For an EtherChannel, this EtherChannel is local to the unit, and is not a Spanned EtherChannel.
  - Management-only interfaces cannot be redundant interfaces.
- If you are connecting remotely to the management interface using ASDM, the current IP address of prospective secondary units are for temporary use.
  - Each member will be assigned an IP address from the cluster IP pool defined on the primary unit.
  - The cluster IP pool cannot include addresses already in use on the network, including prospective secondary IP addresses.

    For example:

    1. You configure the primary unit to use 10.1.1.1.

2. Other units use 10.1.1.2, 10.1.1.3, and 10.1.1.4.
3. When you configure the cluster IP pool on the primary unit, you cannot include the .2, .3, or .4 addresses in the pool, because they are in use.
4. Instead, you need to use other IP addresses on the network, such as .5, .6, .7, and .8.

Note The pool needs as many addresses as there are members of the cluster, including the primary unit; the original .1 address is the main cluster IP address that belongs to the current primary unit.

5. After you join the cluster, the old, temporary addresses are relinquished and can be used elsewhere.

Procedure

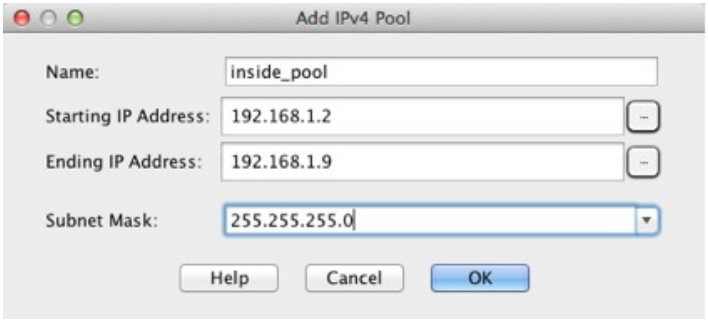
Step 1 Choose the **Configuration** > **Device Setup** > **Interface Settings** > **Interfaces** pane.

Step 2 Choose the interface row, and click **Edit**. Set the interface parameters. See the following guidelines:

- (Required for a management interface in Spanned EtherChannel mode) **Dedicate this interface to management only**—Sets an interface to management-only mode so that it does not pass through traffic. By default, Management type interfaces are configured as management-only. In transparent mode, this command is always enabled for a Management type interface.
- **Use Static IP**—DHCP and PPPoE are not supported.

Step 3 To add the IPv4 cluster IP pool, MAC address pool, and site-specific MAC addresses, click the **Advanced** tab and set **ASA Cluster** area parameters.

a) Create a cluster IP pool by clicking the **...** button next to the **IP Address Pool** field. The valid range shown is determined by the Main IP address you set on the General tab.
b) Click **Add**.
c) Configure a range of addresses that does not include the Main cluster IP address, and that does not include any addresses currently in-use on your network. You should make the range large enough for the size of the cluster, for example, 8 addresses.



d) Click **OK** to create the new pool.
e) Select the new pool you created, and click **Assign**, and then click **OK**.

The pool name appears in the **IP Address Pool** field.

f) (Optional) (Optional) Configure a **MAC Address Pool** if you want to manually configure MAC addresses.

Step 4 To configure an IPv6 address, click the **IPv6** tab.

a) Check the **Enable IPv6** check box.

b) In the **Interface IPv6 Addresses** area, click **Add**.

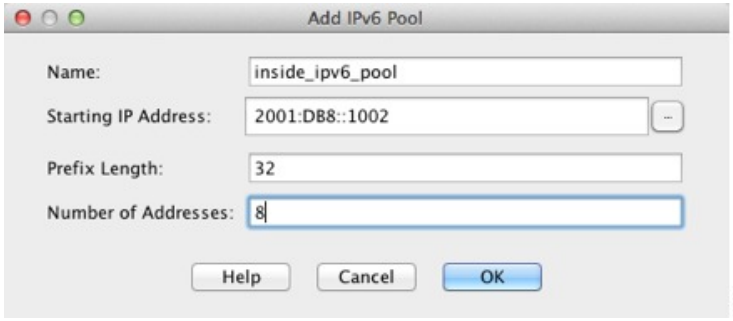
The **Enable address autoconfiguration** option is not supported.

The **Add IPv6 Address for Interface** dialog box appears.

c) In the **Address/Prefix Length** field, enter the global IPv6 address and the IPv6 prefix length. For example, 2001:0DB8::BA98:0:3210/48.

d) Click the **...** button to configure the cluster IP pool.

e) Click **Add**.



f) Configure the starting IP address (network prefix), prefix length, and number of addresses in the pool.

g) Click **OK** to create the new pool.

h) Select the new pool you created, and click **Assign**, and then click **OK**.

The pool appears in the **ASA Cluster IP Pool** field.

i) Click **OK**.

Step 5 Click **OK** to return to the Interfaces pane.

Step 6 Click **Apply**.

Configure Spanned EtherChannels

A Spanned EtherChannel spans all ASAs in the cluster, and provides load balancing as part of the EtherChannel operation.

Before you begin

- You must be in Spanned EtherChannel interface mode.
- For multiple context mode, start this procedure in the system execution space. If you are not already in the System configuration mode in the Configuration > Device List pane, double-click **System** under the active device IP address.
- For transparent mode, configure the bridge group. See Configure the Bridge Virtual Interface (BVI).

- *Do not* specify the maximum and minimum links in the EtherChannel—We recommend that you do not specify the maximum and minimum links in the EtherChannel on either the ASA or the switch. If you need to use them, note the following:
  - The maximum links set on the ASA is the total number of active ports for the whole cluster. Be sure the maximum links value configured on the switch is not larger than the ASA value.
  - The minimum links set on the ASA is the minimum active ports to bring up a port-channel interface *per unit*. On the switch, the minimum links is the minimum links across the cluster, so this value will not match the ASA value.
- *Do not* change the load-balancing algorithm from the default. On the switch, we recommend that you use one of the following algorithms: **source-dest-ip** or **source-dest-ip-port** (see the Cisco Nexus OS and Cisco IOS **port-channel load-balance** command). Do not use a **vlan** keyword in the load-balance algorithm because it can cause unevenly distributed traffic to the ASAs in a cluster.
- When using Spanned EtherChannels, the port-channel interface will not come up until clustering is fully enabled. This requirement prevents traffic from being forwarded to a unit that is not an active unit in the cluster.

Procedure

Step 1 Depending on your context mode:

- For single mode, choose the **Configuration > Device Setup > Interface Settings > Interfaces** pane.
- For multiple mode in the System execution space, choose the **Configuration > Context Management > Interfaces** pane.

Step 2 Choose **Add** > **EtherChannel Interface**.

The **Add EtherChannel Interface** dialog box appears.

Step 3 Enable the following:

- **Port Channel ID**
- **Span EtherChannel across the ASA cluster**
- **Enable Interface** (checked by default)
- **Members in Group**—In the **Members in Group** list, you need to add at least one interface. Multiple interfaces in the EtherChannel per unit are useful for connecting to switches in a VSS or vPC. Keep in mind that by default, a spanned EtherChannel can have only 8 active interfaces out of 16 maximum across all members in the cluster; the remaining 8 interfaces are on standby in case of link failure. To use more than 8 active interfaces (but no standby interfaces), disable dynamic port priority. When you disable dynamic port priority, you can use up to 32 active links across the cluster. For example, for a cluster of 16 ASAs, you can use a maximum of 2 interfaces on each ASA, for a total of 32 interfaces in the spanned EtherChannel.

  Make sure all interfaces are the same type and speed. The first interface you add determines the type and speed of the EtherChannel. Any non-matching interfaces you add will be put into a suspended state. ASDM does not prevent you from adding non-matching interfaces.

The rest of the fields on this screen are described later in this procedure.

Step 4 (Optional) To override the media type, duplex, speed, and pause frames for flow control for all member interfaces, click **Configure Hardware Properties**. This method provides a shortcut to set these parameters because these parameters must match for all interfaces in the channel group.

Click **OK** to accept the **Hardware Properties** changes.

Step 5 To configure the MAC address and optional parameters, click the **Advanced** tab.

- In the **MAC Address Cloning** area, set a manual global MAC address for the EtherChannel. Do not set the Standby MAC Address; it is ignored. You must configure a MAC address for a Spanned EtherChannel to avoid potential network connectivity problems. With a manually-configured MAC address, the MAC address stays with the current control unit. If you do not configure a MAC address, then if the control unit changes, the new control unit uses a new MAC address for the interface, which can cause a temporary network outage.

  In multiple context mode, if you share an interface between contexts, you should instead enable auto-generation of MAC addresses so you do not need to set the MAC address manually. Note that you must manually configure the MAC address using this command for *non-shared* interfaces.

- (Routed mode) In the **ASA Cluster** area, for inter-site clustering set **Site specific MAC Addresses** and IP addresses for a site by clicking **Add** and specifying a MAC address and IP address for the site ID (1 through 8). Repeat for up to 8 sites. The site-specific IP addresses must be on the same subnet as the global IP address. The site-specific MAC address and IP address used by a unit depends on the site ID you specify in each unit's bootstrap configuration.

- (Optional) If you are connecting the ASA to two switches in a VSS or vPC, then you should enable VSS load balancing by checking the **Enable load balancing between switch pairs in VSS or vPC** mode check box. This feature ensures that the physical link connections between the ASAs to the VSS (or vPC) pair are balanced.

  In the **Member Interface Configuration** area, you must then identify to which switch a given interface is connected, **1** or **2**.

Note We recommend that you do *not* set the **Minimum Active Members** and the **Maximum Active Members**.

Step 6 (Optional) Configure VLAN subinterfaces on this EtherChannel. The rest of this procedure applies to the subinterfaces.

Step 7 (Multiple context mode) Before you complete this procedure, you need to allocate interfaces to contexts.

a) Click **OK** to accept your changes.
b) Allocate interfaces.
c) Change to the context that you want to configure: in the **Device List** pane, double-click the context name under the active device IP address.
d) Choose the **Configuration > Device Setup > Interface Settings > Interfaces** pane, select the port-channel interface that you want to customize, and click **Edit**.

   The **Edit Interface** dialog box appears.

Step 8 Click the **General** tab.

Step 9 (Transparent Mode) From the **Bridge Group** drop-down list, choose the bridge group to which you want to assign this interface.

Step 10 In the **Interface Name** field, enter a name up to 48 characters in length.

Step 11 In the **Security level** field, enter a level between 0 (lowest) and 100 (highest).

Step 12 (Routed Mode) For an IPv4 address, click the **Use Static IP** radio button and enter the IP address and mask. DHCP and PPPoE are not supported. For point-to-point connections, you can specify a 31-bit subnet mask (255.255.255.254). In this case, no IP addresses are reserved for the network or broadcast addresses. For transparent mode, you configure the IP address for the bridge group interface, not the EtherChannel interface.

Step 13 (Routed Mode) To configure an IPv6 address, click the **IPv6** tab.

For transparent mode, you configure the IP address for the bridge group interface, not the EtherChannel interface.

a) Check the **Enable IPv6** check box.
b) In the **Interface IPv6 Addresses** area, click **Add**.

The **Add IPv6 Address for Interface** dialog box appears.

Note The **Enable address autoconfiguration** option is not supported.

c) In the **Address/Prefix Length** field, enter the global IPv6 address and the IPv6 prefix length. For example, 2001:DB8::BA98:0:3210/64.
d) (Optional) To use the Modified EUI-64 interface ID as the host address, check the **EUI-64** check box. In this case, just enter the prefix in the **Address/Prefix Length** field.
e) Click **OK**.

Step 14 Click **OK** to return to the **Interfaces** screen.

Step 15 Click **Apply**.

Create or Join an ASA Cluster

Each unit in the cluster requires a bootstrap configuration to join the cluster.

Run the High Availability Wizard

Each unit in the cluster requires a bootstrap configuration to join the cluster. Run the High Availability and Scalability wizard on one unit (that will become the control unit) to create the cluster, and then add data units to it.



Note For the control unit, if you want to change the default of the cLACP system ID and priority, you cannot use the wizard; you must configure the cluster manually.

Before you begin

- For multiple context mode, complete this procedure in the system execution space. If you are not already in the System configuration mode, in the **Configuration > Device List** pane, double-click **System** under the active device IP address.

- We suggest setting the cluster control link MTU to the maximum 9198 which requires you to enable jumbo frame reservation *on each unit* before continuing with this procedure. Jumbo frame reservation requires a reload of the ASA.
- The interfaces you intend to use for the cluster control link interface must be in an up state on the connected switch.
- When you add a unit to a running cluster, you may see temporary, limited packet/connection drops; this is expected behavior.

Procedure

Step 1 Choose **Wizards** > **High Availability and Scalability Wizard**. See select wizard guidelines in the following steps.

Step 2 On the **Interfaces** screen, you cannot create new EtherChannels from this screen (except for the cluster control link).

Step 3 On the ASA Cluster Configuration screen, configure bootstrap settings including:

- **Member Priority**—Sets the priority of this unit for control unit elections, between 1 and 100, where 1 is the highest priority.
- **(Routed mode; Spanned EtherChannel mode) Site Index**—If you use inter-site clustering, set the site ID for this unit so it uses a site-specific MAC address, between 1 and 8.
- (Optional) **Shared Key**—Sets an encryption key for control traffic on the cluster control link. The shared secret is an ASCII string from 1 to 63 characters. The shared secret is used to generate the encryption key. This parameter does not affect datapath traffic, including connection state update and forwarded packets, which are always sent in the clear. You must configure this parameter if you also enable the password encryption service.
- (Optional) **Enable connection rebalancing for TCP traffic across all the ASAs in the cluster**—Enables connection rebalancing. This parameter is disabled by default. If enabled, ASAs in a cluster exchange load information periodically, and offload new connections from more loaded devices to less loaded devices. The frequency, between 1 and 360 seconds, specifies how often the load information is exchanged. This parameter is not part of the bootstrap configuration, and is replicated from the control unit to the data units.

  Note Do not configure connection rebalancing for inter-site topologies; you do not want connections rebalanced to cluster members at a different site.

- (Optional) **Enable health monitoring of this device within the cluster**—Enables the cluster unit health check feature. To determine unit health, the ASA cluster units send heartbeat messages on the cluster control link to other units. If a unit does not receive any heartbeat messages from a peer unit within the holdtime period, the peer unit is considered unresponsive or dead.

  Note When any topology changes occur (such as adding or removing a data interface, enabling or disabling an interface on the ASA or the switch, adding an additional switch to form a VSS or vPC) you must disable the health check and also disable interface monitoring for the disabled interfaces. When the topology change is complete, and the configuration change is synced to all units, you can re-enable the health check.

- **Time to Wait Before Device Considered Failed**—This value determines the amount of time between unit keepalive status messages, between .3 and 45 seconds; The default is 3 seconds.
- (Optional) **Broadcast keepalive messages to all EtherChannel cluster control link ports for VSS/vPC support**—If you configure the cluster control link as an EtherChannel (recommended), and it is connected to a VSS or vPC pair, then you might need to enable this option. For some switches, when one unit in the VSS/vPC is shutting down or booting up, EtherChannel member interfaces connected to that switch may appear to be Up to the ASA, but they are not passing traffic on the switch side. The ASA can be erroneously removed from the cluster if you set the ASA holdtime timeout to a low value (such as .8 seconds), and the ASA sends heartbeat messages on one of these EtherChannel interfaces. When you enable this option, the ASA floods the heartbeat messages on all EtherChannel interfaces in the cluster control link to ensure that at least one of the switches can receive them.

- (Optional) **Replicate console output**—Enables console replication from data units to the control unit. This feature is disabled by default. The ASA may print out some messages directly to the console for certain critical events. If you enable console replication, data units send the console messages to the control unit so that you only need to monitor one console port for the cluster. This parameter is not part of the bootstrap configuration, and is replicated from the control unit to the data units.
- **Cluster Control Link**—Specifies the cluster control link interface.
  - **MTU**—Specifies the maximum transmission unit for the cluster control link interface to be at least 100 bytes higher than the highest MTU of the data interfaces, between 1400 and 9198 bytes. The default MTU is 1500 bytes. If you already enabled jumbo frame reservation, we suggest setting the MTU to the maximum. Because the cluster control link traffic includes data packet forwarding, the cluster control link needs to accommodate the entire size of a data packet plus cluster traffic overhead. For example, because the maximum MTU is 9198 bytes, then the highest data interface MTU can be 9098, while the cluster control link can be set to 9198. **Note:** If you have not pre-enabled jumbo frame reservation, you should quit the wizard, enable jumbo frames, and then restart this procedure.

Step 4 On the **Interfaces for Health Monitoring** screen, you can exempt some interfaces from monitoring for failure. You might want to disable health monitoring of non-essential interfaces, for example, the management interface. To exempt a hardware module such as the ASA Firepower module from monitoring, check the **Exempt Service Module from Cluster health monitoring** check box.

Note When any topology changes occur (such as adding or removing a data interface, enabling or disabling an interface on the ASA or the switch, adding an additional switch to form a VSS or vPC) you must disable the health check and also disable interface monitoring for the disabled interfaces. When the topology change is complete, and the configuration change is synced to all units, you can re-enable the health check.

Step 5 On the **Interface Auto Rejoin settings** screen, customize the auto-rejoin settings in case of an interface or cluster control link failure. For each type, you can set the following:

- **Maximum Rejoin Attempts**—Define the number of attempts at rejoining the cluster by setting **Unlimited** or a value between 0 and 65535. **0** disables auto-rejoining. The default value is **Unlimited** for the cluster-interface and **3** for the data-interface.
- **Rejoin Interval**—Define the interval duration in minutes between rejoin attempts by setting the interval between 2 and 60. The default value is **5** minutes. The maximum total time that the unit attempts to rejoin the cluster is limited to 14400 minutes (10 days) from the time of last failure.

- **Interval Variation**—Define if the interval duration increases by setting the interval variation between 1 and 3: **1** (no change); **2** (2 x the previous duration), or **3** (3 x the previous duration). For example, if you set the interval duration to 5 minutes, and set the variation to 2, then the first attempt is after 5 minutes; the 2nd attempt is 10 minutes (2 x 5); the 3rd attempt 20 minutes (2 x 10), and so on. The default value is **1** for the cluster-interface and **2** for the data-interface.

Step 6 Click **Finish**.

Step 7 The ASA scans the running configuration for incompatible commands for features that are not supported with clustering, including commands that may be present in the default configuration. Click **OK** to delete the incompatible commands. If you click **Cancel**, then clustering is not enabled.

After a period of time while ASDM enables clustering and reconnects to the ASA, the Information screen appears confirming that the ASA was added to the cluster.

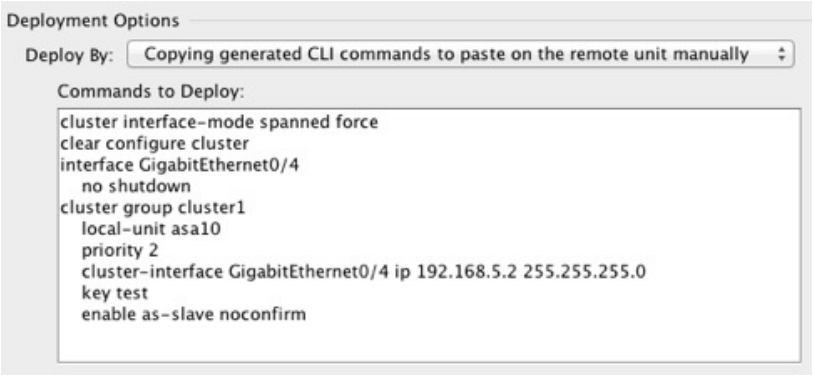
Note In some cases, there might be an error when joining the cluster after you finish the wizard. If ASDM was disconnected, ASDM will not receive any subsequent errors from the ASA. If clustering remains disabled after you reconnect ASDM, you should connect to the ASA console port to determine the exact error condition that disabled clustering; for example, the cluster control link might be down.

Step 8 To add a data unit, click **Yes**.

If you are re-running the wizard from the control unit, you can add data units by choosing the **Add another member to the cluster** option when you first start the wizard.

Step 9 In the **Deployment Options** area, choose one of the following **Deploy By** options:

- **Sending CLI commands to the remote unit now—**Send the bootstrap configuration to the data unit (temporary) management IP address. Enter the data unit management IP address, username, and password.
- **Copying generated CLI commands to paste on the remote unit manually**—Generates the commands so that you can cut and paste them at the data unit CLI or using the CLI tool in ASDM. In the Commands to Deploy box, select and copy the generated commands for later use.



Customize the Clustering Operation

You can customize clustering health monitoring, TCP connection replication delay, flow mobility and other optimizations.

Perform these procedures on the control unit.

Configure Basic ASA Cluster Parameters

You can customize cluster settings on the control unit. If you do not use the wizard to add a unit to the cluster, you can configure the cluster parameters manually. If you already enabled clustering, you can edit some cluster parameters; others that cannot be edited while clustering is enabled are grayed out. This procedure also includes advanced parameters that are not included in the wizard.

Before you begin

- Pre-configure the cluster control link interfaces on each unit before joining the cluster. For a single interface, you must enable it; do not configure any other settings. For an EtherChannel interface, enable it and set the EtherChannel mode to On.
- For multiple context mode, complete this procedure in the system execution space on the control unit. If you are not already in the System configuration mode, in the **Configuration** > **Device List** pane, double-click **System** under the active device IP address.

Procedure

Step 1 Choose **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster**.

If your device is already in the cluster, and is the control unit, then this pane is on the Cluster Configuration tab.

Step 2 Check the **Configure ASA cluster settings** check box.

If you uncheck the check box, the settings are erased. Do not check **Participate in ASA cluster** until after you have set all your parameters.

Note After you enable clustering, do not uncheck the **Configure ASA cluster settings** check box without understanding the consequences. This action clears all cluster configuration, and also shuts down all interfaces including the management interface to which ASDM is connected. To restore connectivity in this case, you need to access the CLI at the console port.

Step 3 Configure the following bootstrap parameters:

- **Cluster Name**—Names the cluster. The name must be an ASCII string from 1 to 38 characters. You can only configure one cluster per unit. All members of the cluster must use the same name.
- **Member Name**—Names this member of the cluster with a unique ASCII string from 1 to 38 characters.
- **Member Priority**—Sets the priority of this unit for control unit elections, between 1 and 100, where 1 is the highest priority.
- **Site Index**—If you use inter-site clustering, set the site ID for this unit so it uses a site-specific MAC address, between 1 and 8.
- (Optional) **Site Periodic GARP**—The ASA generates gratuitous ARP (GARP) packets to keep the switching infrastructure up to date: the highest priority member at each site periodically generates GARP traffic for the global MAC/IP addresses. GARP is enabled by default when you set the site ID for each unit and the site MAC and IP address for each Spanned EtherChannel. Set the GARP interval between 1 and 1000000 seconds. The default is 290 seconds.

When using per-site MAC and IP addresses, packets sourced from the cluster use a site-specific MAC address and IP address, while packets received by the cluster use a global MAC address and IP address. If traffic is not generated from the global MAC address periodically, you could experience a MAC address timeout on your switches for the global MAC address. After a timeout, traffic destined for the global MAC address will be flooded across the entire switching infrastructure, which can cause performance and security concerns.

- (Optional) **Shared Key**—Sets an encryption key for control traffic on the cluster control link. The shared secret is an ASCII string from 1 to 63 characters. The shared secret is used to generate the encryption key. This parameter does not affect datapath traffic, including connection state update and forwarded packets, which are always sent in the clear. You must configure this parameter if you also enable the password encryption service.
- (Optional) **Enable connection rebalancing for TCP traffic across all the ASAs in the cluster**—Enables connection rebalancing. This parameter is disabled by default. If enabled, ASAs in a cluster exchange load information periodically, and offload new connections from more loaded devices to less loaded devices. The frequency, between 1 and 360 seconds, specifies how often the load information is exchanged. This parameter is not part of the bootstrap configuration, and is replicated from the control unit to the data units.
- (Optional) **Enable health monitoring of this device within the cluster**—Enables the cluster unit health check feature, and determines the amount of time between unit heartbeat status messages, between .3 and 45 seconds; The default is 3 seconds. **Note**: When you are adding new units to the cluster, and making topology changes on the ASA or the switch, you should disable this feature temporarily until the cluster is complete, and also disable interface monitoring for the disabled interfaces (**Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster** > **Cluster Interface Health Monitoring**). You can re-enable this feature after cluster and topology changes are complete. To determine unit health, the ASA cluster units send heartbeat messages on the cluster control link to other units. If a unit does not receive any heartbeat messages from a peer unit within the holdtime period, the peer unit is considered unresponsive or dead.
  - (Optional) **Broadcast keepalive messages to all EtherChannel cluster control link ports for VSS/vPC support**—If you configure the cluster control link as an EtherChannel (recommended), and it is connected to a VSS or vPC pair, then you might need to enable this option. For some switches, when one unit in the VSS/vPC is shutting down or booting up, EtherChannel member interfaces connected to that switch may appear to be Up to the ASA, but they are not passing traffic on the switch side. The ASA can be erroneously removed from the cluster if you set the ASA holdtime timeout to a low value (such as .8 seconds), and the ASA sends heartbeat messages on one of these EtherChannel interfaces. When you enable this option, the ASA floods the heartbeat messages on all EtherChannel interfaces in the cluster control link to ensure that at least one of the switches can receive them.
- (Optional) **Debounce Time**—Configures the debounce time before the ASA considers an interface to be failed and the unit is removed from the cluster. This feature allows for faster detection of interface failures. Note that configuring a lower debounce time increases the chances of false-positives. When an interface status update occurs, the ASA waits the number of milliseconds specified before marking the interface as failed and the unit is removed from the cluster. In the case of an EtherChannel that transitions from a down state to an up state (for example, the switch reloaded, or the switch enabled an EtherChannel), a longer debounce time can prevent the interface from appearing to be failed on a cluster unit just because another cluster unit was faster at bundling the ports. The default debounce time is 500 ms, with a range of 300 ms to 9 seconds.

- (Optional) **Replicate console output**—Enables console replication from data units to the control unit. This feature is disabled by default. The ASA may print out some messages directly to the console for certain critical events. If you enable console replication, data units send the console messages to the control unit so that you only need to monitor one console port for the cluster. This parameter is not part of the bootstrap configuration, and is replicated from the control unit to the data units.
- (Optional) **Enable Clustering Flow Mobility**. See Configure LISP Inspection, on page 42.
- (Optional) **Enable Director Localization for inter-DC cluster**—To improve performance and reduce round-trip time latency for inter-site clustering for data centers, you can enable director localization. New connections are typically load-balanced and owned by cluster members within a given site. However, the ASA assigns the Director role to a member at *any* site. Director localization enables additional Director roles: a Local Director at the same site as the Owner, and a Global Director that can be at any site. Keeping the Owner and Director at the same site improves performance. Also, if the original Owner fails, the Local Director will choose a new connection Owner at the same site. The Global Director is used if a cluster member receives packets for a connection that is owned on a different site.
- (Optional) **Site Redundancy**—To protect flows from a site failure, you can enable site redundancy. If the connection backup owner is at the same site as the owner, then an additional backup owner will be chosen from another site to protect flows from a site failure. Director localization and site redundancy are separate features; you can configure one or the other, or configure both.
- **Cluster Control Link**—Specifies the cluster control link interface. This interface cannot have a name configured; available interfaces are shown in the drop-down list.
  - **Interface**—Specifies the interface ID, preferably an EtherChannel. Subinterfaces and Management type interfaces are not allowed.
  - **IP Address**—Specifies an IPv4 address for the IP address; IPv6 is not supported for this interface.
  - **Subnet Mask**—Specifies the subnet mask.
  - **MTU**—Specifies the maximum transmission unit for the cluster control link interface to be at least 100 bytes higher than the highest MTU of the data interfaces, between 1400 and 9198 bytes. The default MTU is 1500 bytes. Because the cluster control link traffic includes data packet forwarding, the cluster control link needs to accommodate the entire size of a data packet plus cluster traffic overhead. We suggest setting the cluster control link MTU to the maximum, which requires you to enable jumbo frame reservation. Jumbo frame reservation requires a reload of the ASA. For example, because the maximum MTU is 9198 bytes, then the highest data interface MTU can be 9098, while the cluster control link can be set to 9198.
- (Optional) **Cluster LACP**—When using Spanned EtherChannels, the ASA uses cLACP to negotiate the EtherChannel with the neighbor switch. ASAs in a cluster collaborate in cLACP negotiation so that they appear as a single (virtual) device to the switch.
  - **Enable static port priority**—Disables dynamic port priority in LACP. Some switches do not support dynamic port priority, so this parameter improves switch compatibility. Moreover, it enables support of more than 8 active spanned EtherChannel members, up to 32 members. Without this parameter, only 8 active members and 8 standby members are supported. If you enable this parameter, then you cannot use any standby members; all members are active. This parameter is not part of the bootstrap configuration, and is replicated from the control unit to the data units.
  - **Virtual System MAC Address**—Sets the cLACP system ID, which is in the format of a MAC address. All ASAs use the same system ID: auto-generated by the control unit (the default) and replicated to all secondaries; or manually specified in the form *H.H.H*, where H is a 16-bit

hexadecimal digit. For example, the MAC address 00-0C-F1-42-4C-DE is entered as 000C.F142.4CDE. This parameter is not part of the bootstrap configuration, and is replicated from the control unit to the data units. However, you can only change this value if you disable clustering.

- **System Priority**—Sets the system priority, between 1 and 65535. The priority is used to decide which unit is in charge of making a bundling decision. By default, the ASA uses priority 1, which is the highest priority. The priority needs to be higher than the priority on the switch. This parameter is not part of the bootstrap configuration, and is replicated from the control unit to the data units. However, you can only change this value if you disable clustering.

Step 4 Check the **Participate in ASA cluster** check box to join the cluster.

Step 5 Click **Apply**.

Configure Interface Health Monitoring and Auto-Rejoin Settings

You might want to disable health monitoring of non-essential interfaces, for example, the management interface. You can monitor any port-channel ID, redundant ID, or single physical interface ID, or the software or hardware module, such as the ASA Firepower module. Health monitoring is not performed on VLAN subinterfaces or virtual interfaces such as VNIs or BVIs. You cannot configure monitoring for the cluster control link; it is always monitored.

Procedure

Step 1 Choose **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster** > **Cluster Interface Health Monitoring**.

Step 2 In the **Monitored Interfaces** box, select an interface, and click **Add** to move it to the **Unmonitored Interfaces** box.

Interface status messages detect link failure. If all physical ports for a given logical interface fail on a particular unit, but there are active ports under the same logical interface on other units, then the unit is removed from the cluster. If a unit does not receive interface status messages within the holdtime, then the amount of time before the ASA removes a member from the cluster depends on the type of interface and whether the unit is an established member or is joining the cluster. Health check is enabled by default for all interfaces.

You might want to disable health monitoring of non-essential interfaces, for example, the management interface. You can specify any port-channel ID, redundant ID, or single physical interface ID. Health monitoring is not performed on VLAN subinterfaces or virtual interfaces such as VNIs or BVIs. You cannot configure monitoring for the cluster control link; it is always monitored.

When any topology changes occur (such as adding or removing a data interface, enabling or disabling an interface on the ASA or the switch, or adding an additional switch to form a VSS or vPC) you should disable the health check feature (**Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster**) and also disable interface monitoring for the disabled interfaces. When the topology change is complete, and the configuration change is synced to all units, you can re-enable the health check feature.

Step 3 (Optional) Check the **Exempt Service Module from Cluster Health Monitoring** check box to exempt a hardware or software module such as the ASA FirePOWER module.

For the ASA 5585-X, if you disable monitoring of the service module, you may also want to disable monitoring of the interfaces on the module, which are monitored separately.

Step 4 Click the **Auto Rejoin** tab to customize the auto-rejoin settings in case of an interface, system, or cluster control link failure. For each type, click **Edit** to set the following:

- **Maximum Rejoin Attempts**—Define the number of attempts at rejoining the cluster by setting **Unlimited** or a value between 0 and 65535. **0** disables auto-rejoining. The default value is **Unlimited** for the cluster-interface and **3** for the data-interface and system.
- **Rejoin Interval**—Define the interval duration in minutes between rejoin attempts by setting the interval between 2 and 60. The default value is **5** minutes. The maximum total time that the unit attempts to rejoin the cluster is limited to 14400 minutes (10 days) from the time of last failure.
- **Interval Variation**—Define if the interval duration increases by setting the interval variation between 1 and 3: **1** (no change); **2** (2 x the previous duration), or **3** (3 x the previous duration). For example, if you set the interval duration to 5 minutes, and set the variation to 2, then the first attempt is after 5 minutes; the 2nd attempt is 10 minutes (2 x 5); the 3rd attempt 20 minutes (2 x 10), and so on. The default value is **1** for the cluster-interface and **2** for the data-interface and system.

Click **Restore Defaults** to restore the default settings.

Step 5 Click **Apply**.

Configure the Cluster TCP Replication Delay

Enable the cluster replication delay for TCP connections to help eliminate the “unnecessary work” related to short-lived flows by delaying the director/backup flow creation. Note that if a unit fails before the director/backup flow is created, then those flows cannot be recovered. Similarly, if traffic is rebalanced to a different unit before the flow is created, then the flow cannot be recovered. You should not enable the TCP replication delay for traffic on which you disable TCP randomization.

Procedure

Step 1 Choose **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster Replication**.

Step 2 Click **Add** and set the following values:

- **Replication delay**—Set the seconds between 1 and 15.
- **HTTP**—Set the delay for all HTTP traffic. This setting is enabled by default for 5 seconds for the Firepower 4100/9300 chassis only.
- **Source Criteria**
  - **Source**—Set the source IP address.
  - **Service**—(Optional) Set the source port. Typically you set either the source or the destination port, but not both.
- **Destination Criteria**
  - **Source**—Set the destination IP address.
  - **Service**—(Optional) Set the destination port. Typically you set either the source or the destination port, but not both.

Step 3 Click **OK**.

Step 4 Click **Apply**.

Configure Inter-Site Features

For inter-site clustering, you can customize your configuration to enhance redundancy and stability.

Configure Cluster Flow Mobility

You can inspect LISP traffic to enable flow mobility when a server moves between sites.

About LISP Inspection

You can inspect LISP traffic to enable flow mobility between sites.

About LISP

Data center virtual machine mobility such as VMware VMotion enables servers to migrate between data centers while maintaining connections to clients. To support such data center server mobility, routers need to be able to update the ingress route towards the server when it moves. Cisco Locator/ID Separation Protocol (LISP) architecture separates the device identity, or endpoint identifier (EID), from its location, or routing locator (RLOC), into two different numbering spaces, making server migration transparent to clients. For example, when a server moves to a new site and a client sends traffic to the server, the router redirects traffic to the new location.

LISP requires routers and servers in certain roles, such as the LISP egress tunnel router (ETR), ingress tunnel router (ITR), first hop routers, map resolver (MR), and map server (MS). When the first hop router for the server senses that the server is connected to a different router, it updates all of the other routers and databases so that the ITR connected to the client can intercept, encapsulate, and send traffic to the new server location.

ASA LISP Support

The ASA does not run LISP itself; it can, however, inspect LISP traffic for location changes and then use this information for seamless clustering operation. Without LISP integration, when a server moves to a new site, traffic comes to an ASA cluster member at the new site instead of to the original flow owner. The new ASA forwards traffic to the ASA at the old site, and then the old ASA has to send traffic back to the new site to reach the server. This traffic flow is sub-optimal and is known as “tromboning” or “hair-pinning.”

With LISP integration, the ASA cluster members can inspect LISP traffic passing between the first hop router and the ETR or ITR, and can then change the flow owner to be at the new site.

LISP Guidelines

- The ASA cluster members must reside between the first hop router and the ITR or ETR for the site. The ASA cluster itself cannot be the first hop router for an extended segment.
- Only fully-distributed flows are supported; centralized flows, semi-distributed flows, or flows belonging to individual units are not moved to new owners. Semi-distributed flows include applications, such as SIP, where all child flows are owned by the same ASA that owns the parent flow.
- The cluster only moves Layer 3 and 4 flow states; some application data might be lost.
- For short-lived flows or non-business-critical flows, moving the owner may not be worthwhile. You can control the types of traffic that are supported with this feature when you configure the inspection policy, and should limit flow mobility to essential traffic.

ASA LISP Implementation

This feature includes several inter-related configurations (all of which are described in this chapter):

1. (Optional) Limit inspected EIDs based on the host or server IP address—The first hop router might send EID-notify messages for hosts or networks the ASA cluster is not involved with, so you can limit the EIDs to only those servers or networks relevant to your cluster. For example, if the cluster is only involved with 2 sites, but LISP is running on 3 sites, you should only include EIDs for the 2 sites involved with the cluster.
2. LISP traffic inspection—The ASA inspects LISP traffic on UDP port 4342 for the EID-notify message sent between the first hop router and the ITR or ETR. The ASA maintains an EID table that correlates the EID and the site ID. For example, you should inspect LISP traffic with a source IP address of the first hop router and a destination address of the ITR or ETR. Note that LISP traffic is not assigned a director, and LISP traffic itself does not participate in cluster state sharing.
3. Service Policy to enable flow mobility on specified traffic—You should enable flow mobility on business-critical traffic. For example, you can limit flow mobility to only HTTPS traffic, and/or to traffic to specific servers.
4. Site IDs—The ASA uses the site ID for each cluster unit to determine the new owner.
5. Cluster-level configuration to enable flow mobility—You must also enable flow mobility at the cluster level. This on/off toggle lets you easily enable or disable flow mobility for a particular class of traffic or applications.

Configure LISP Inspection

You can inspect LISP traffic to enable flow mobility when a server moves between sites.

Before you begin

- Assign each cluster unit to a site ID according to Configure Basic ASA Cluster Parameters, on page 36.
- LISP traffic is not included in the default-inspection-traffic class, so you must configure a separate class for LISP traffic as part of this procedure.

Procedure

Step 1 (Optional) Configure a LISP inspection map to limit inspected EIDs based on IP address, and to configure the LISP pre-shared key:

a) Choose **Configuration** > **Firewall** > **Objects** > **Inspect Maps** > **LISP**.
b) Click **Add** to add a new map.
c) Enter a name (up to 40 characters) and description.
d) For the **Allowed-EID access-list**, click **Manage**.

The **ACL Manager** opens.

The first hop router or ITR/ETR might send EID-notify messages for hosts or networks the ASA cluster is not involved with, so you can limit the EIDs to only those servers or networks relevant to your cluster. For example, if the cluster is only involved with 2 sites, but LISP is running on 3 sites, you should only include EIDs for the 2 sites involved with the cluster.

e) Add an ACL with at least one ACE according to the firewall configuration guide.
f) If necessary, enter the **Validation Key**.

   If you copied an encrypted key, click the **Encrypted** radio button.

g) Click **OK**.

Step 2 Add a service policy rule to configure LISP inspection:

a) Choose **Configuration** > **Firewall** > **Service Policy Rules**.
b) Click **Add**.
c) On the **Service Policy** page, apply the rule to an interface or globally.

   If you have an existing service policy you want to use, add a rule to that policy. By default, the ASA includes a global policy called **global_policy**. You can also create one service policy per interface if you do not want to apply the policy globally. LISP inspection is applied to traffic bidirectionally so you do not need to apply the service policy on both the source and destination interfaces; all traffic that enters or exits the interface to which you apply the rule is affected if the traffic matches the class for both directions.

d) On the **Traffic Classification Criteria** page, click **Create a new traffic class**, and under **Traffic Match Criteria**, check **Source and Destination IP Address (uses ACL)**.
e) Click **Next**.
f) Specify the traffic you want to inspect. You should specify traffic between the first hop router and the ITR or ETR on UDP port 4342. Both IPv4 and IPv6 ACLs are accepted.
g) Click **Next**.
h) On the **Rule Actions** wizard page or tab, select the **Protocol Inspection** tab.
i) Check the **LISP** check box,.
j) (Optional) Click **Configure** to choose the inspection map you created.
k) Click **Finish** to save the service policy rule.

Step 3 Add a service policy rule to enable Flow Mobility for critical traffic:

a) Choose **Configuration** > **Firewall** > **Service Policy Rules**.
b) Click **Add**.
c) On the **Service Policy** page, choose the same service policy you used for LISP inspection.
d) On the **Traffic Classification Criteria** page, click **Create a new traffic class**, and under **Traffic Match Criteria**, check **Source and Destination IP Address (uses ACL)**.
e) Click **Next**.
f) Specify the business critical traffic that you want to re-assign to the most optimal site when servers change sites. For example, you can limit flow mobility to only HTTPS traffic, and/or to traffic to specific servers. Both IPv4 and IPv6 ACLs are accepted.
g) Click **Next**.
h) On the **Rule Actions** wizard page or tab, select the **Cluster** tab.
i) Check the **Enable Cluster flow-mobility triggered by LISP EID messages** check box.
j) Click **Finish** to save the service policy rule.

Step 4 Choose **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster** > **Cluster Configuration**, and check the **Enable Clustering flow mobility** check box.

Step 5 Click **Apply**.

Manage Cluster Members

After you deploy the cluster, you can change the configuration and manage cluster members.

Add a New Data Unit from the Control Unit

You can add additional secondaries to the cluster from the control unit. You can also add secondaries using the High Availability and Scalability wizard. Adding a data unit from the control unit has the benefit of configuring the cluster control link and setting the cluster interface mode on each data unit you add.

You can alternatively log into the data unit and configure clustering directly on the unit. However, after you enable clustering, your ASDM session will be disconnected, and you will have to reconnect.

Before you begin

- For multiple context mode, complete this procedure in the system execution space. If you are not already in the System configuration mode, in the Configuration > Device List pane, double-click **System** under the active device IP address.
- If you want to send the bootstrap configuration over the management network, be sure the data unit has an accessible IP address.

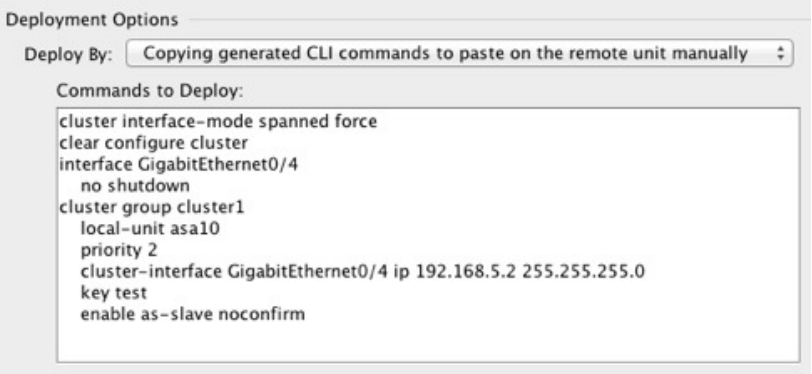
Procedure

Step 1 Choose **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster** > **Cluster Members**.

Step 2 Click **Add**.

Step 3 Configure the following parameters:

- **Member Name**—Names this member of the cluster with a unique ASCII string from 1 to 38 characters.
- **Member Priority**—Sets the priority of this unit for control unit elections, between 1 and 100, where 1 is the highest priority.
- **Cluster Control Link > IP Address**—Specifies a unique IP address for this member for the cluster control link, on the same network as the control unit cluster control link.
- In the **Deployment Options** area, choose one of the following **Deploy By** options:
  - **Sending CLI commands to the remote unit now**—Send the bootstrap configuration to the data unit (temporary) management IP address. Enter the data unit management IP address, username, and password.
  - **Copying generated CLI commands to paste on the remote unit manually**—Generates the commands so that you can cut and paste them at the data unit CLI or using the CLI tool in ASDM. In the Commands to Deploy box, select and copy the generated commands for later use.



Step 4 Click **OK**, then **Apply**.

Become an Inactive Member

To become an inactive member of the cluster, disable clustering on the unit while leaving the clustering configuration intact.

Note When an ASA becomes inactive (either manually or through a health check failure), all data interfaces are shut down; only the management-only interface can send and receive traffic. To resume traffic flow, re-enable clustering; or you can remove the unit altogether from the cluster. The management interface remains up using the IP address the unit received from the cluster IP pool. However if you reload, and the unit is still inactive in the cluster (for example, you saved the configuration with clustering disabled), then the management interface is disabled. You must use the console port for any further configuration.

Before you begin

- For multiple context mode, perform this procedure in the system execution space. If you are not already in the System configuration mode in the Configuration > Device List pane, double-click **System** under the active device IP address.

Procedure

Step 1 Choose **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster** > **Cluster Configuration**.

Step 2 Uncheck the **Participate in ASA cluster** check box.

Note Do not uncheck the **Configure ASA cluster settings** check box; this action clears all cluster configuration, and also shuts down all interfaces including the management interface to which ASDM is connected. To restore connectivity in this case, you need to access the CLI at the console port.

Step 3 Click **Apply**.

Deactivate a Data Unit from the Control Unit

To deactivate a data unit, perform the following steps.



Note When an ASA becomes inactive, all data interfaces are shut down; only the management-only interface can send and receive traffic. To resume traffic flow, re-enable clustering. The management interface remains up using the IP address the unit received from the cluster IP pool. However if you reload, and the unit is still inactive in the cluster (for example, if you saved the configuration with clustering disabled), the management interface is disabled. You must use the console port for any further configuration.

Before you begin

For multiple context mode, perform this procedure in the system execution space. If you are not already in the System configuration mode in the **Configuration > Device List** pane, double-click **System** under the active device IP address.

Procedure

Step 1 Choose **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster**.

Step 2 Select the data unit that you want to remove, and click **Delete**.

The data unit bootstrap configuration remains intact, so that you can later re-add the data unit without losing your configuration.

Step 3 Click **Apply**.

Rejoin the Cluster

If a unit was removed from the cluster, for example for a failed interface or if you manually deactivated a member, you must manually rejoin the cluster.

Before you begin

- You must use the console port to reenable clustering. Other interfaces are shut down. The exception is if you manually disabled clustering in ASDM, then you can reenable clustering in ASDM if you did not save the configuration and reload. After reloading, the management interface is disabled, so console access is the only method to reenable clustering.
- For multiple context mode, perform this procedure in the system execution space. If you are not already in the System configuration mode in the **Configuration > Device List** pane, double-click **System** under the active device IP address.
- Make sure the failure is resolved before you try to rejoin the cluster.

Procedure

Step 1 If you still have ASDM access, you can reenable clustering in ASDM by connecting ASDM to the unit you want to reenable.

You cannot reenable clustering for a data unit from the control unit unless you add it as a new member.

a) Choose **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster**.
b) Check the **Participate in ASA cluster** check box.
c) Click **Apply**.

Step 2 If you cannot use ASDM: At the console, enter cluster configuration mode:

cluster group *name*

Example:

```
ciscoasa(config)# cluster group pod1
```

Step 3 Enable clustering.

enable

Leave the Cluster

If you want to leave the cluster altogether, you need to remove the entire cluster bootstrap configuration. Because the current configuration on each member is the same (synced from the primary unit), leaving the cluster also means either restoring a pre-clustering configuration from backup, or clearing your configuration and starting over to avoid IP address conflicts.

Before you begin

You must use the console port; when you remove the cluster configuration, all interfaces are shut down, including the management interface and cluster control link.

Procedure

Step 1 For a secondary unit, disable clustering:

cluster group *cluster_name*
no enable

Example:

```
ciscoasa(config)# cluster group cluster1
ciscoasa(cfg-cluster)# no enable
```

You cannot make configuration changes while clustering is enabled on a secondary unit.

Step 2 Clear the cluster configuration:

clear configure cluster

The ASA shuts down all interfaces including the management interface and cluster control link.

Step 3 Disable cluster interface mode:

no cluster interface-mode

The mode is not stored in the configuration and must be reset manually.

Step 4 If you have a backup configuration, copy the backup configuration to the running configuration:

copy *backup_cfg* **running-config**

Example:

```
ciscoasa(config)# copy backup_cluster.cfg running-config

Source filename [backup_cluster.cfg]?

Destination filename [running-config]?
ciscoasa(config)#
```

Step 5 Save the configuration to startup:

write memory

Step 6 If you do not have a backup configuration, reconfigure management access. Be sure to change the interface IP addresses, and restore the correct hostname, for example.

Change the Control Unit



Caution The best method to change the control unit is to disable clustering on the control unit, wait for a new control election, and then re-enable clustering. If you must specify the exact unit you want to become the control unit, use the procedure in this section. Note, however, that for centralized features, if you force a control unit change using this procedure, then all connections are dropped, and you have to re-establish the connections on the new control unit.

To change the control unit, perform the following steps.

Before you begin

For multiple context mode, perform this procedure in the system execution space. If you are not already in the System configuration mode in the Configuration > Device List pane, double-click **System** under the active device IP address.

Procedure

Step 1 Choose **Monitoring** > **ASA Cluster** > **Cluster Summary**.

Step 2 From the drop-down list, choose a data unit to become master, and click the button to make it the control unit.

Step 3 You are prompted to confirm the control unit change. Click **Yes**.

Step 4 Quit ASDM, and reconnect using the Main cluster IP address.

Execute a Command Cluster-Wide

To send a command to all members in the cluster, or to a specific member, perform the following steps. Sending a **show** command to all members collects all output and displays it on the console of the current unit. Other commands, such as **capture** and **copy**, can also take advantage of cluster-wide execution.

Before you begin

Perform this procedure at the Command Line Interface tool: choose **Tools > Command Line Interface**.

Procedure

Send a command to all members, or if you specify the unit name, a specific member:

cluster exec [**unit** *unit_name*] *command*

Example:

```
ciscoasa# cluster exec show xlate
```

To view member names, enter **cluster exec unit ?** (to see all names except the current unit), or enter the **show cluster info** command.

Examples

To copy the same capture file from all units in the cluster at the same time to a TFTP server, enter the following command on the control unit:

```
ciscoasa# cluster exec copy /pcap capture: tftp://10.1.1.56/capture1.pcap
```

Multiple PCAP files, one from each unit, are copied to the TFTP server. The destination capture file name is automatically attached with the unit name, such as capture1_asa1.pcap, capture1_asa2.pcap, and so on. In this example, asa1 and asa2 are cluster unit names.

The following sample output for the **cluster exec show port-channel** summary command shows EtherChannel information for each member in the cluster:

```
ciscoasa# cluster exec show port-channel summary
master(LOCAL):**********************************************************
 Number of channel-groups in use: 2
Group  Port-channel  Protocol  Span-cluster  Ports
------+-------------+-----------+-----------------------------------------------
1         Po1             LACP       Yes  Gi0/0(P)
```

```
2         Po2              LACP       Yes  Gi0/1(P)
 slave:********************************************************************
 Number of channel-groups in use: 2
Group  Port-channel  Protocol  Span-cluster   Ports
------+-------------+-----------+-----------------------------------------------
1         Po1              LACP       Yes   Gi0/0(P)
2         Po2              LACP       Yes   Gi0/1(P)
```

Monitoring the ASA Cluster

You can monitor and troubleshoot cluster status and connections.

Monitoring Cluster Status

See the following screens for monitoring cluster status:

- **Monitoring > ASA Cluster > Cluster Summary**

  This pane shows cluster information about the unit to which you are connected, as well as other units in the cluster. You can also change the primary unit from this pane.

- **Cluster Dashboard**

  On the home page on the primary unit, you can monitor the cluster using the Cluster Dashboard and the Cluster Firewall Dashboard.

Capturing Packets Cluster-Wide

See the following screen for capturing packets in a cluster:

Wizards > Packet Capture Wizard

To support cluster-wide troubleshooting, you can enable capture of cluster-specific traffic on the control unit, which is then automatically enabled on all of the data units in the cluster.

Monitoring Cluster Resources

See the following screens for monitoring cluster resources:

- **Monitoring > ASA Cluster > System Resources Graphs > CPU**

  This pane lets you create graphs or tables showing the CPU utilization across the cluster members.

- **Monitoring > ASA Cluster > System Resources Graphs > Memory**. This pane lets you create graphs or tables showing the Free Memory and Used Memory across the cluster members.

Monitoring Cluster Traffic

See the following screens for monitoring cluster traffic:

- **Monitoring > ASA Cluster > Traffic Graphs > Connections**.

This pane lets you create graphs or tables showing the Connections across the cluster members.

- **Monitoring > ASA Cluster > Traffic Graphs > Throughput**.

  This pane lets you create graphs or tables showing the traffic throughput across the cluster members.

Monitoring the Cluster Control Link

See the following screen for monitoring cluster status:

Monitoring > Properties > System Resources Graphs > Cluster Control Link.

This pane lets you create graphs or tables showing the cluster control link Receival and Transmittal capacity utilization.

Monitoring Cluster Routing

See the following screen for cluster routing:

- **Monitoring** > **Routing** > **LISP-EID Table**

  Shows the ASA EID table showing EIDs and site IDs.

Configuring Logging for Clustering

See the followingscreen for configuring logging for clustering:

Configuration > Device Management > Logging > Syslog Setup

Each unit in the cluster generates syslog messages independently. You can generate syslog messages with identical or different device IDs to make messages appear to come from the same or different units in the cluster.

Examples for ASA Clustering

These examples include all cluster-related ASA configuration for typical deployments.

Sample ASA and Switch Configuration

The following sample configurations connect the following interfaces between the ASA and the switch:

ASA Interface	Switch Interface
GigabitEthernet 0/2	GigabitEthernet 1/0/15
GigabitEthernet 0/3	GigabitEthernet 1/0/16
GigabitEthernet 0/4	GigabitEthernet 1/0/17
GigabitEthernet 0/5	GigabitEthernet 1/0/18

ASA Configuration

Interface Mode on Each Unit

```
cluster interface-mode spanned force
```

ASA1 Control Unit Bootstrap Configuration

```
interface GigabitEthernet0/0
 channel-group 1 mode on
 no shutdown
!
interface GigabitEthernet0/1
 channel-group 1 mode on
 no shutdown
!
interface Port-channel1
 description Clustering Interface
!
cluster group Moya
 local-unit A
 cluster-interface Port-channel1 ip 10.0.0.1 255.255.255.0
 priority 10
 key emphyri0
 enable noconfirm
```

ASA2 Data Unit Bootstrap Configuration

```
interface GigabitEthernet0/0
 channel-group 1 mode on
 no shutdown
!
interface GigabitEthernet0/1
 channel-group 1 mode on
 no shutdown
!
interface Port-channel1
 description Clustering Interface
!
cluster group Moya
 local-unit B
 cluster-interface Port-channel1 ip 10.0.0.2 255.255.255.0
 priority 11
 key emphyri0
 enable as-slave
```

Control Unit Interface Configuration

```
ip local pool mgmt-pool 10.53.195.231-10.53.195.232

interface GigabitEthernet0/2
 channel-group 10 mode active
 no shutdown
!
interface GigabitEthernet0/3
```

```
 channel-group 10 mode active
 no shutdown
!
interface GigabitEthernet0/4
 channel-group 11 mode active
 no shutdown
!
interface GigabitEthernet0/5
 channel-group 11 mode active
 no shutdown
!
interface Management0/0
 management-only
 nameif management
 ip address 10.53.195.230 cluster-pool mgmt-pool
 security-level 100
 no shutdown
!
interface Port-channel10
 port-channel span-cluster
 mac-address aaaa.bbbb.cccc
 nameif inside
 security-level 100
 ip address 209.165.200.225 255.255.255.224
!
interface Port-channel11
 port-channel span-cluster
 mac-address aaaa.dddd.cccc
 nameif outside
 security-level 0
 ip address 209.165.201.1 255.255.255.224
```

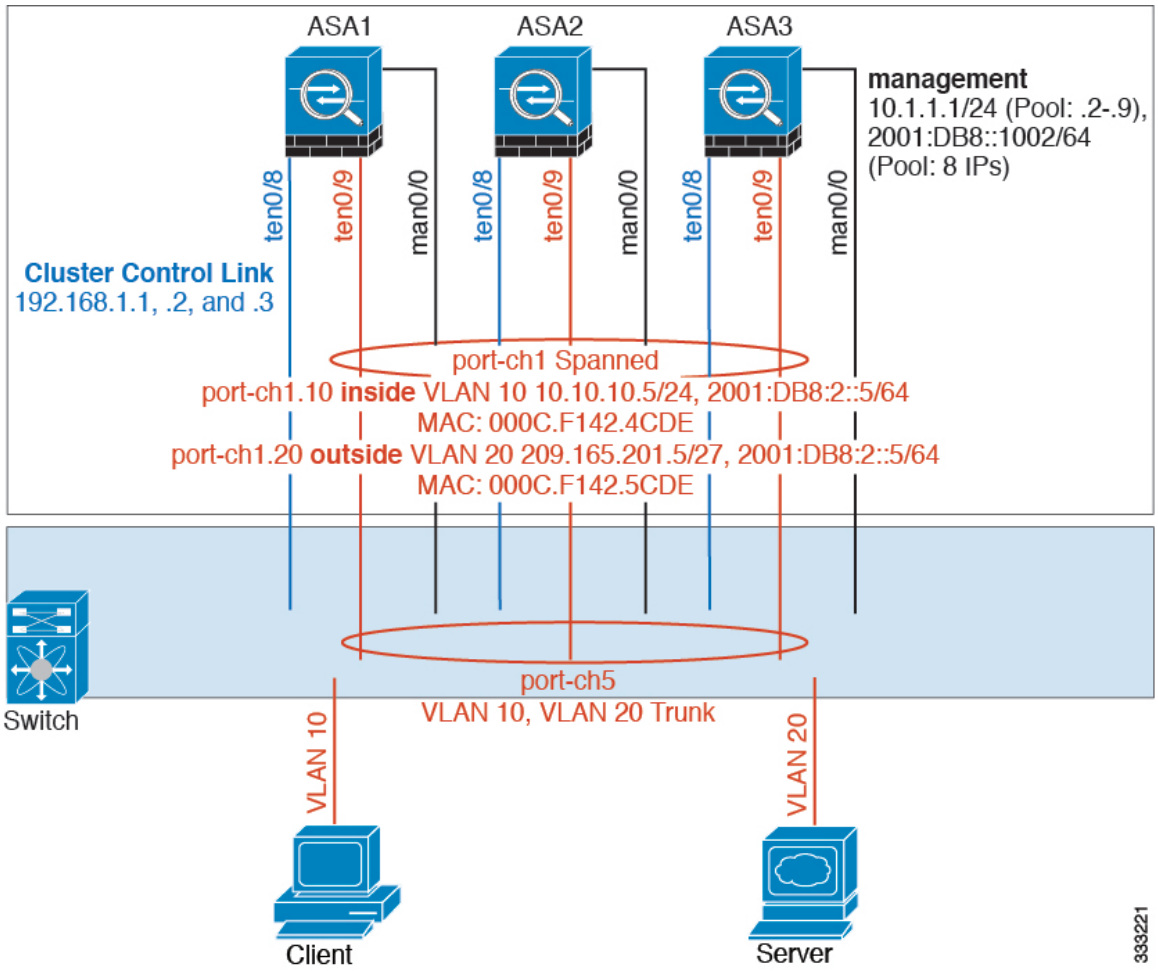
Cisco IOS Switch Configuration

```
interface GigabitEthernet1/0/15
 switchport access vlan 201
 switchport mode access
 spanning-tree portfast
 channel-group 10 mode active
!
interface GigabitEthernet1/0/16
 switchport access vlan 201
 switchport mode access
 spanning-tree portfast
 channel-group 10 mode active
!
interface GigabitEthernet1/0/17
 switchport access vlan 401
 switchport mode access
 spanning-tree portfast
 channel-group 11 mode active
!
interface GigabitEthernet1/0/18
 switchport access vlan 401
 switchport mode access
 spanning-tree portfast
 channel-group 11 mode active

interface Port-channel10
 switchport access vlan 201
 switchport mode access
```

```
interface Port-channel11
 switchport access vlan 401
 switchport mode access
```

Firewall on a Stick



Data traffic from different security domains are associated with different VLANs, for example, VLAN 10 for the inside network and VLAN 20 for the outside network. Each ASA has a single physical port connected to the external switch or router. Trunking is enabled so that all packets on the physical link are 802.1q encapsulated. The ASA is the firewall between VLAN 10 and VLAN 20.

When using Spanned EtherChannels, all data links are grouped into one EtherChannel on the switch side. If an ASA becomes unavailable, the switch will rebalance traffic between the remaining units.

Interface Mode on Each Unit

```
cluster interface-mode spanned force
```

ASA1 Control Unit Bootstrap Configuration

```
interface tengigabitethernet 0/8

no shutdown
description CCL

cluster group cluster1

local-unit asa1
cluster-interface tengigabitethernet0/8 ip 192.168.1.1 255.255.255.0
priority 1
key chuntheunavoidable
enable noconfirm
```

ASA2 Data Unit Bootstrap Configuration

```
interface tengigabitethernet 0/8

no shutdown
description CCL

cluster group cluster1

local-unit asa2
cluster-interface tengigabitethernet0/8 ip 192.168.1.2 255.255.255.0
priority 2
key chuntheunavoidable
enable as-slave
```

ASA3 Data Unit Bootstrap Configuration

```
interface tengigabitethernet 0/8

no shutdown
description CCL

cluster group cluster1

local-unit asa3
cluster-interface tengigabitethernet0/8 ip 192.168.1.3 255.255.255.0
priority 3
key chuntheunavoidable
enable as-slave
```

Control Unit Interface Configuration

```
ip local pool mgmt 10.1.1.2-10.1.1.9
ipv6 local pool mgmtipv6 2001:DB8::1002/64 8
interface management 0/0

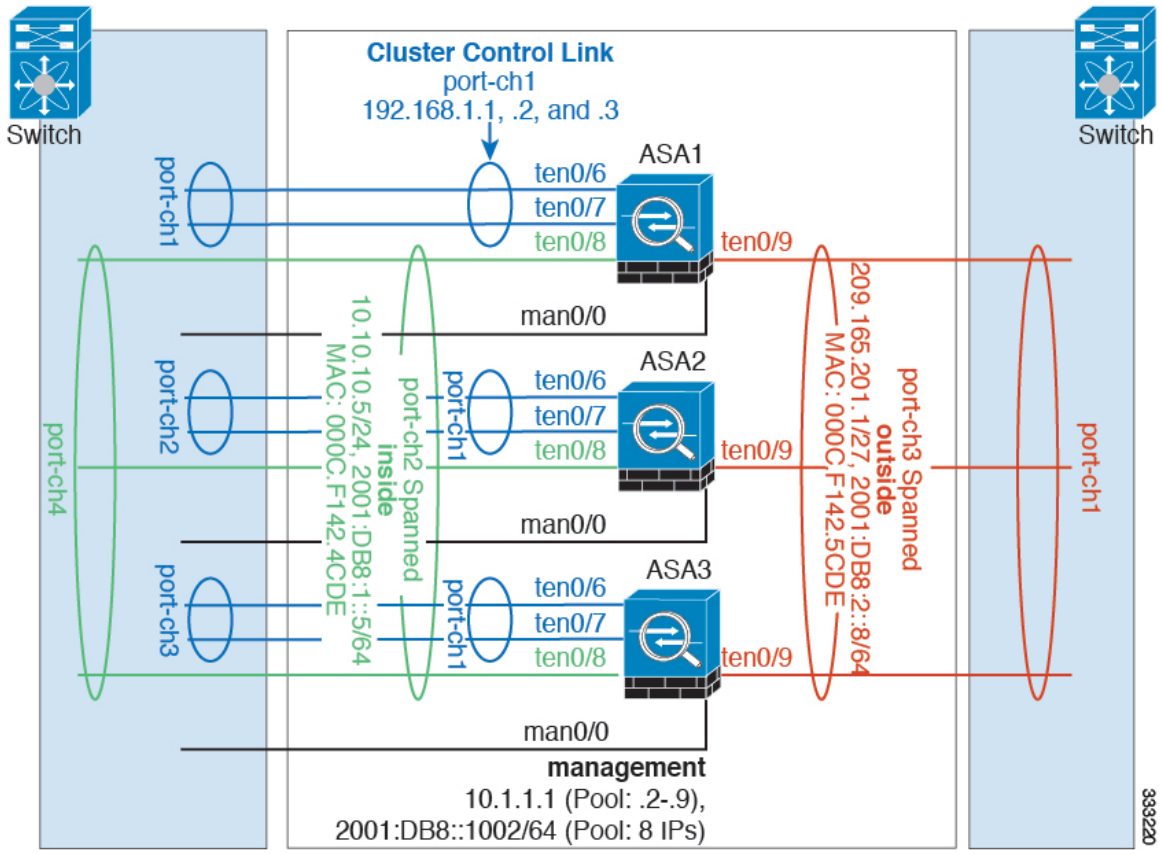
nameif management
ip address 10.1.1.1 255.255.255.0 cluster-pool mgmt
ipv6 address 2001:DB8::1001/32 cluster-pool mgmtipv6
security-level 100
management-only
```

```
no shutdown

interface tengigabitethernet 0/9

channel-group 2 mode active
no shutdown
interface port-channel 2
port-channel span-cluster
interface port-channel 2.10
vlan 10
nameif inside
ip address 10.10.10.5 255.255.255.0
ipv6 address 2001:DB8:1::5/64
mac-address 000C.F142.4CDE
interface port-channel 2.20
vlan 20
nameif outside
ip address 209.165.201.1 255.255.255.224
ipv6 address 2001:DB8:2::8/64
mac-address 000C.F142.5CDE
```

Traffic Segregation



You may prefer physical separation of traffic between the inside and outside network.

As shown in the diagram above, there is one Spanned EtherChannel on the left side that connects to the inside switch, and the other on the right side to outside switch. You can also create VLAN subinterfaces on each EtherChannel if desired.

Interface Mode on Each Unit

```
cluster interface-mode spanned force
```

ASA1 Control Unit Bootstrap Configuration

```
interface tengigabitethernet 0/6

channel-group 1 mode on
no shutdown

interface tengigabitethernet 0/7

channel-group 1 mode on
no shutdown
interface port-channel 1
description CCL

cluster group cluster1

local-unit asa1
cluster-interface port-channel1 ip 192.168.1.1 255.255.255.0
priority 1
key chuntheunavoidable
enable noconfirm
```

ASA2 Data Unit Bootstrap Configuration

```
interface tengigabitethernet 0/6

channel-group 1 mode on
no shutdown

interface tengigabitethernet 0/7

channel-group 1 mode on
no shutdown
interface port-channel 1
description CCL

cluster group cluster1

local-unit asa2
cluster-interface port-channel1 ip 192.168.1.2 255.255.255.0
priority 2
key chuntheunavoidable
enable as-slave
```

ASA3 Data Unit Bootstrap Configuration

```
interface tengigabitethernet 0/6
```

```
channel-group 1 mode on
no shutdown

interface tengigabitethernet 0/7

channel-group 1 mode on
no shutdown
interface port-channel 1
description CCL

cluster group cluster1

local-unit asa3
cluster-interface port-channel1 ip 192.168.1.3 255.255.255.0
priority 3
key chuntheunavoidable
enable as-slave
```

Control Unit Interface Configuration

```
ip local pool mgmt 10.1.1.2-10.1.1.9
ipv6 local pool mgmtipv6 2001:DB8::1002/64 8
interface management 0/0

nameif management
ip address 10.1.1.1 255.255.255.0 cluster-pool mgmt
ipv6 address 2001:DB8::1001/32 cluster-pool mgmtipv6
security-level 100
management-only
no shutdown

interface tengigabitethernet 0/8

channel-group 2 mode active
no shutdown
interface port-channel 2
port-channel span-cluster
nameif inside
ip address 10.10.10.5 255.255.255.0
ipv6 address 2001:DB8:1::5/64
mac-address 000C.F142.4CDE

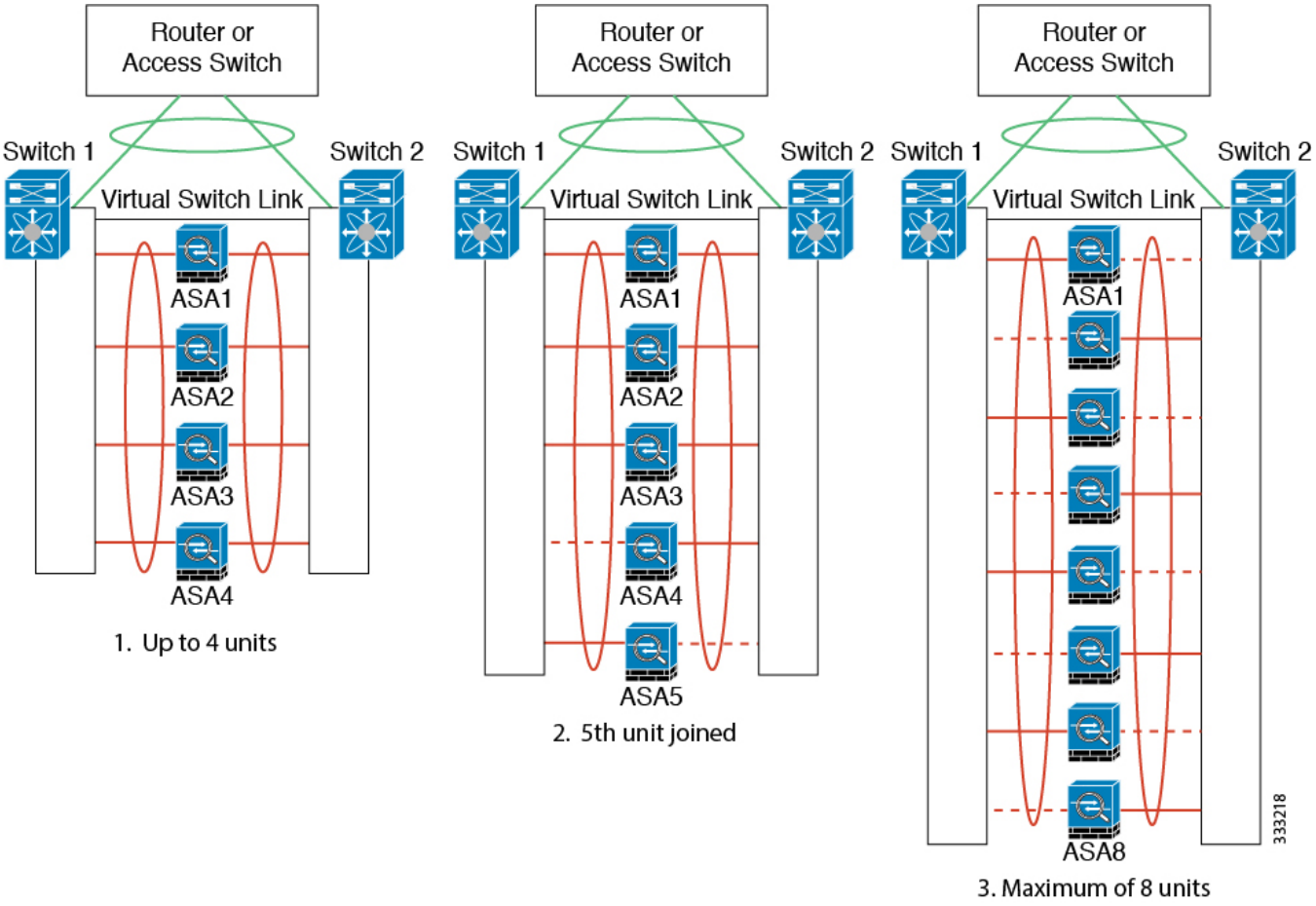
interface tengigabitethernet 0/9

channel-group 3 mode active
no shutdown
interface port-channel 3
port-channel span-cluster
nameif outside
ip address 209.165.201.1 255.255.255.224
ipv6 address 2001:DB8:2::8/64
mac-address 000C.F142.5CDE
```

Spanned EtherChannel with Backup Links (Traditional 8 Active/8 Standby)

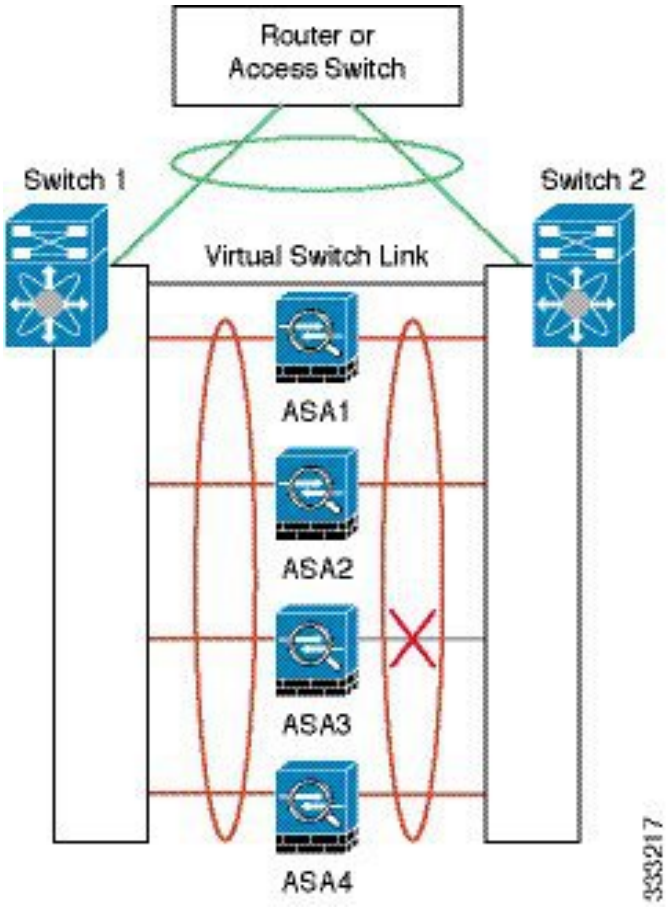
The maximum number of active ports in a traditional EtherChannel is limited to 8 from the switch side. If you have an 8-ASA cluster, and you allocate 2 ports per unit to the EtherChannel, for a total of 16 ports total, then 8 of them have to be in standby mode. The ASA uses LACP to negotiate which links should be active

or standby. If you enable multi-switch EtherChannel using VSS or vPC, you can achieve inter-switch redundancy. On the ASA, all physical ports are ordered first by the slot number then by the port number. In the following figure, the lower ordered port is the "control" port (for example, GigabitEthernet 0/0), and the other one is the "data" port (for example, GigabitEthernet 0/1). You must guarantee symmetry in the hardware connection: all control links must terminate on one switch, and all data links must terminate on another switch if VSS/vPC is used. The following diagram shows what happens when the total number of links grows as more units join the cluster:

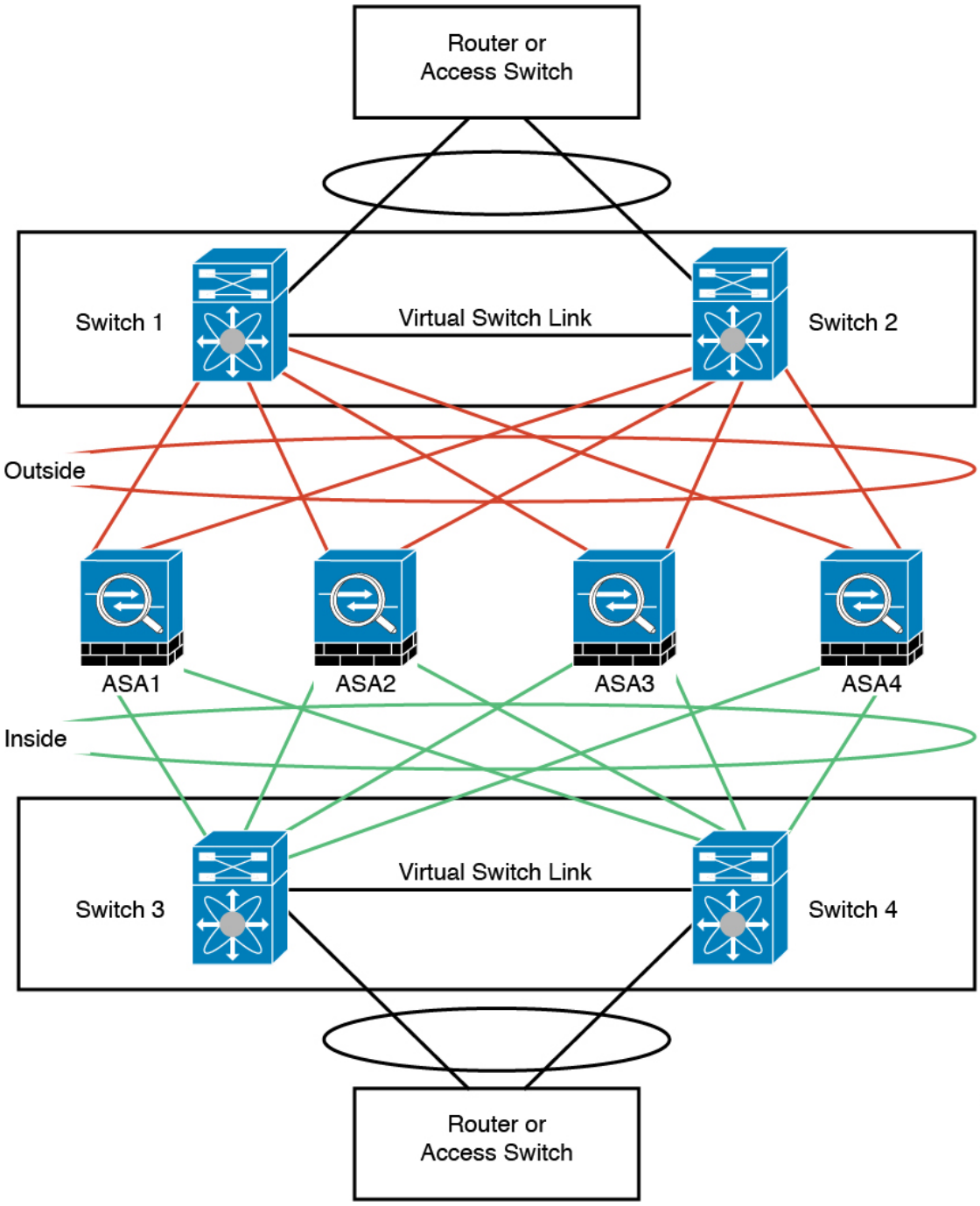


The principle is to first maximize the number of active ports in the channel, and secondly keep the number of active control ports and the number of active data ports in balance. Note that when a 5th unit joins the cluster, traffic is not balanced evenly between all units.

Link or device failure is handled with the same principle. You may end up with a less-than-perfect load balancing situation. The following figure shows a 4-unit cluster with a single link failure on one of the units.



There could be multiple EtherChannels configured in the network. The following diagram shows an EtherChannel on the inside and one on the outside. An ASA is removed from the cluster if both control and data links in one EtherChannel fail. This prevents the ASA from receiving traffic from the outside network when it has already lost connectivity to the inside network.



Interface Mode on Each Unit

```
cluster interface-mode spanned force
```

ASA1 Control Unit Bootstrap Configuration

```
interface tengigabitethernet 0/6

channel-group 1 mode on
no shutdown

interface tengigabitethernet 0/7

channel-group 1 mode on
no shutdown

interface tengigabitethernet 0/8

channel-group 1 mode on
no shutdown

interface tengigabitethernet 0/9

channel-group 1 mode on
no shutdown
interface port-channel 1
description CCL

cluster group cluster1

local-unit asa1
cluster-interface port-channel1 ip 192.168.1.1 255.255.255.0
priority 1
key chuntheunavoidable
enable noconfirm
```

ASA2 Data Unit Bootstrap Configuration

```
interface tengigabitethernet 0/6

channel-group 1 mode on
no shutdown

interface tengigabitethernet 0/7

channel-group 1 mode on
no shutdown

interface tengigabitethernet 0/8

channel-group 1 mode on
no shutdown

interface tengigabitethernet 0/9

channel-group 1 mode on
no shutdown
interface port-channel 1
description CCL

cluster group cluster1

local-unit asa2
cluster-interface port-channel1 ip 192.168.1.2 255.255.255.0
priority 2
```

```
key chuntheunavoidable
enable as-slave
```

ASA3 Data Unit Bootstrap Configuration

```
interface tengigabitethernet 0/6

channel-group 1 mode on
no shutdown

interface tengigabitethernet 0/7

channel-group 1 mode on
no shutdown

interface tengigabitethernet 0/8

channel-group 1 mode on
no shutdown

interface tengigabitethernet 0/9

channel-group 1 mode on
no shutdown
interface port-channel 1
description CCL

cluster group cluster1

local-unit asa3
cluster-interface port-channel1 ip 192.168.1.3 255.255.255.0
priority 3
key chuntheunavoidable
enable as-slave
```

ASA4 Data Unit Bootstrap Configuration

```
interface tengigabitethernet 0/6

channel-group 1 mode on
no shutdown

interface tengigabitethernet 0/7

channel-group 1 mode on
no shutdown

interface tengigabitethernet 0/8

channel-group 1 mode on
no shutdown

interface tengigabitethernet 0/9

channel-group 1 mode on
no shutdown
interface port-channel 1
description CCL
```

```
cluster group cluster1

local-unit asa4
cluster-interface port-channel1 ip 192.168.1.4 255.255.255.0
priority 4
key chuntheunavoidable
enable as-slave
```

Control Unit Interface Configuration

```
ip local pool mgmt 10.1.1.2-10.1.1.9
interface management 0/0

channel-group 2 mode active
no shutdown

interface management 0/1

channel-group 2 mode active
no shutdown
interface port-channel 2
nameif management
ip address 10.1.1.1 255.255.255.0 cluster-pool mgmt
security-level 100
management-only

interface tengigabitethernet 1/6

channel-group 3 mode active vss-id 1
no shutdown

interface tengigabitethernet 1/7

channel-group 3 mode active vss-id 2
no shutdown
interface port-channel 3
port-channel span-cluster vss-load-balance
nameif inside
ip address 10.10.10.5 255.255.255.0
mac-address 000C.F142.4CDE

interface tengigabitethernet 1/8

channel-group 4 mode active vss-id 1
no shutdown

interface tengigabitethernet 1/9

channel-group 4 mode active vss-id 2
no shutdown
interface port-channel 4
port-channel span-cluster vss-load-balance
nameif outside
ip address 209.165.201.1 255.255.255.224
mac-address 000C.F142.5CDE
```

OTV Configuration for Routed Mode Inter-Site Clustering

The success of inter-site clustering for routed mode with Spanned EtherChannels depends on the proper configuration and monitoring of OTV. OTV plays a major role by forwarding the packets across the DCI. OTV forwards unicast packets across the DCI only when it learns the MAC address in its forwarding table. If the MAC address is not learned in the OTV forwarding table, it will drop the unicast packets.

Sample OTV Configuration

```
//Sample OTV config:
//3151 – Inside VLAN, 3152 – Outside VLAN, 202 – CCL VLAN
//aaaa.1111.1234 – ASA inside interface global vMAC
//0050.56A8.3D22 – Server MAC

feature ospf
feature otv

mac access-list ALL_MACs
  10 permit any any
mac access-list HSRP_VMAC
  10 permit aaaa.1111.1234 0000.0000.0000 any
  20 permit aaaa.2222.1234 0000.0000.0000 any
  30 permit any aaaa.1111.1234 0000.0000.0000
  40 permit any aaaa.2222.1234 0000.0000.0000
vlan access-map Local 10
  match mac address HSRP_VMAC
  action drop
vlan access-map Local 20
  match mac address ALL_MACs
  action forward
vlan filter Local vlan-list 3151-3152


//To block global MAC with ARP inspection:
arp access-list HSRP_VMAC_ARP
  10 deny aaaa.1111.1234 0000.0000.0000 any
  20 deny aaaa.2222.1234 0000.0000.0000 any
  30 deny any aaaa.1111.1234 0000.0000.0000
  40 deny any aaaa.2222.1234 0000.0000.0000
  50 permit ip any mac
ip arp inspection filter HSRP_VMAC_ARP 3151-3152

no ip igmp snooping optimise-multicast-flood
vlan 1,202,1111,2222,3151-3152

otv site-vlan 2222
mac-list GMAC_DENY seq 10 deny aaaa.aaaa.aaaa ffff.ffff.ffff
mac-list GMAC_DENY seq 20 deny aaaa.bbbb.bbbb ffff.ffff.ffff
mac-list GMAC_DENY seq 30 permit 0000.0000.0000 0000.0000.0000
route-map stop-GMAC permit 10
  match mac-list GMAC_DENY

interface Overlay1
  otv join-interface Ethernet8/1
  otv control-group 239.1.1.1
  otv data-group 232.1.1.0/28
  otv extend-vlan 202, 3151
  otv arp-nd timeout 60
  no shutdown

interface Ethernet8/1
```

```
  description uplink_to_OTV_cloud
  mtu 9198
  ip address 10.4.0.18/24
  ip igmp version 3
  no shutdown

interface Ethernet8/2

interface Ethernet8/3
  description back_to_default_vdc_e6/39
  switchport
    switchport mode trunk
    switchport trunk allowed vlan 202,2222,3151-3152
  mac packet-classify
  no shutdown

otv-isis default
  vpn Overlay1
    redistribute filter route-map stop-GMAC
otv site-identifier 0x2
//OTV flood not required for ARP inspection:
otv flood mac 0050.56A8.3D22 vlan 3151
```

OTV Filter Modifications Required Because of Site Failure

If a site goes down, the filters need to be removed from OTV because you do not want to block the global MAC address anymore. There are some additional configurations required.

You need to add a static entry for the ASA global MAC address on the OTV switch in the site that is functional. This entry will let the OTV at the other end add these entries on the overlay interface. This step is required because if the server and client already have the ARP entry for the ASA, which is the case for existing connections, then they will not send the ARP again. Therefore, OTV will not get a chance to learn the ASA global MAC address in its forwarding table. Because OTV does not have the global MAC address in its forwarding table, and per OTV design it will not flood unicast packets over the overlay interface, then it will drop the unicast packets to the global MAC address from the server, and the existing connections will break.

```
//OTV filter configs when one of the sites is down

mac-list GMAC_A seq 10 permit 0000.0000.0000 0000.0000.0000
route-map a-GMAC permit 10
  match mac-list GMAC_A

otv-isis default
  vpn Overlay1
    redistribute filter route-map a-GMAC

no vlan filter Local vlan-list 3151

//For ARP inspection, allow global MAC:
arp access-list HSRP_VMAC_ARP_Allow
  50 permit ip any mac
ip arp inspection filter HSRP_VMAC_ARP_Allow 3151-3152

mac address-table static aaaa.1111.1234 vlan 3151 interface Ethernet8/3
//Static entry required only in the OTV in the functioning Site
```

When the other site is restored, you need to add the filters back again and remove this static entry on the OTV. It is very important to clear the dynamic MAC address table on both the OTVs to clear the overlay entry for the global MAC address.

MAC Address Table Clearing

When a site goes down, and a static entry for the global MAC address is added to OTV, you need to let the other OTV learn the global MAC address on the overlay interface. After the other site comes up, these entries should be cleared. Make sure to clear the mac address table to make sure OTV does not have these entries in its forwarding table.

```
cluster-N7k6-OTV# show mac address-table
Legend:
* - primary entry, G - Gateway MAC, (R) - Routed MAC, O - Overlay MAC
age - seconds since last seen,+ - primary entry using vPC Peer-Link,
(T) - True, (F) - False
VLAN MAC Address Type age Secure NTFY Ports/SWID.SSID.LID
---------+-----------------+--------+---------+------+----+------------------
G -    d867.d900.2e42 static  -  F F sup-eth1(R)
O 202  885a.92f6.44a5 dynamic -  F F Overlay1
* 202  885a.92f6.4b8f dynamic 5  F F Eth8/3
O 3151 0050.5660.9412 dynamic -  F F Overlay1
* 3151 aaaa.1111.1234 dynamic 50 F F Eth8/3
```

OTV ARP Cache Monitoring

OTV maintains an ARP cache to proxy ARP for IP addresses that it learned across the OTV interface.

```
cluster-N7k6-OTV# show otv arp-nd-cache
OTV ARP/ND L3->L2 Address Mapping Cache

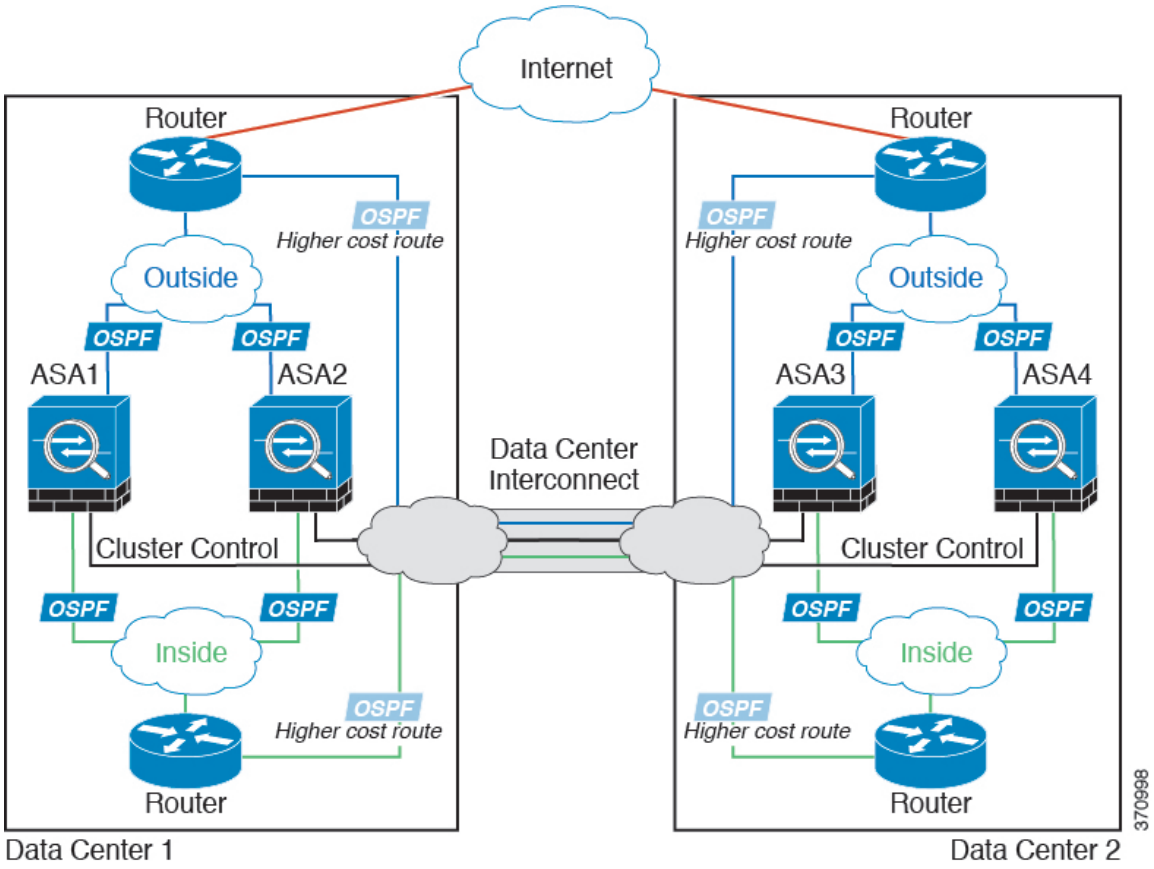
Overlay Interface Overlay1
VLAN MAC Address Layer-3 Address Age Expires In
3151 0050.5660.9412 10.0.0.2 1w0d 00:00:31
cluster-N7k6-OTV#
```

Examples for Inter-Site Clustering

The following examples show supported cluster deployments.

Individual Interface Routed Mode North-South Inter-Site Example

The following example shows 2 ASA cluster members at each of 2 data centers placed between inside and outside routers (North-South insertion). The cluster members are connected by the cluster control link over the DCI. The inside and outside routers at each data center use OSPF and PBR or ECMP to load balance the traffic between cluster members. By assigning a higher cost route across the DCI, traffic stays within each data center unless all ASA cluster members at a given site go down. In the event of a failure of all cluster members at one site, traffic goes from each router over the DCI to the ASA cluster members at the other site.



Spanned EtherChannel Routed Mode Example with Site-Specific MAC and IP Addresses

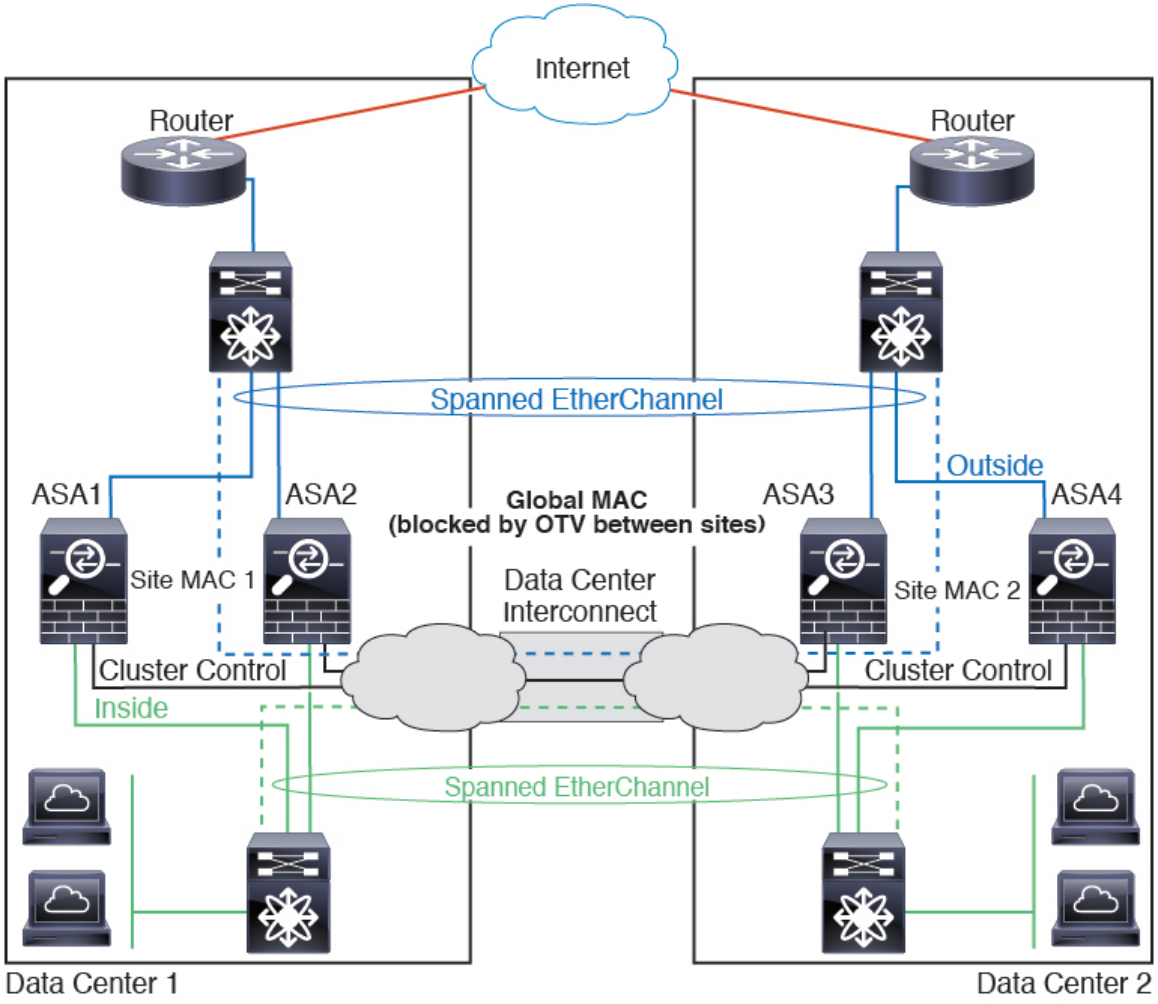
The following example shows 2 cluster members at each of 2 data centers placed between the gateway router and an inside network at each site (East-West insertion). The cluster members are connected by the cluster control link over the DCI. The cluster members at each site connect to the local switches using spanned EtherChannels for both the inside and outside networks. Each EtherChannel is spanned across all chassis in the cluster.

The data VLANs are extended between the sites using Overlay Transport Virtualization (OTV) (or something similar). You must add filters blocking the global MAC address to prevent traffic from traversing the DCI to the other site when the traffic is destined for the cluster. If the cluster units at one site become unreachable, you must remove the filters so traffic can be sent to the other site's cluster units. You should use VACLs to filter the global MAC address.For some switches, such as Nexus with the F3-series line card, you must also use ARP inspection to block ARP packets from the global MAC address. ARP inspection requires you to set both the site MAC address and the site IP address on the ASA. If you only configure the site MAC address be sure to disable ARP inspection. See OTV Configuration for Routed Mode Inter-Site Clustering, on page 65 for more information.

The cluster acts as the gateway for the inside networks. The global virtual MAC, which is shared across all cluster units, is used only to receive packets. Outgoing packets use a site-specific MAC address from each DC cluster. This feature prevents the switches from learning the same global MAC address from both sites on two different ports, which causes MAC flapping; instead, they only learn the site MAC address.

In this scenario:

- All egress packets sent from the cluster use the site MAC address and are localized at the data center.
- All ingress packets to the cluster are sent using the global MAC address, so they can be received by any of the units at both sites; filters at the OTV localize the traffic within the data center.



For a sample OTV configuration and best practices, see OTV Configuration for Routed Mode Inter-Site Clustering, on page 65.

Spanned EtherChannel Transparent Mode North-South Inter-Site Example

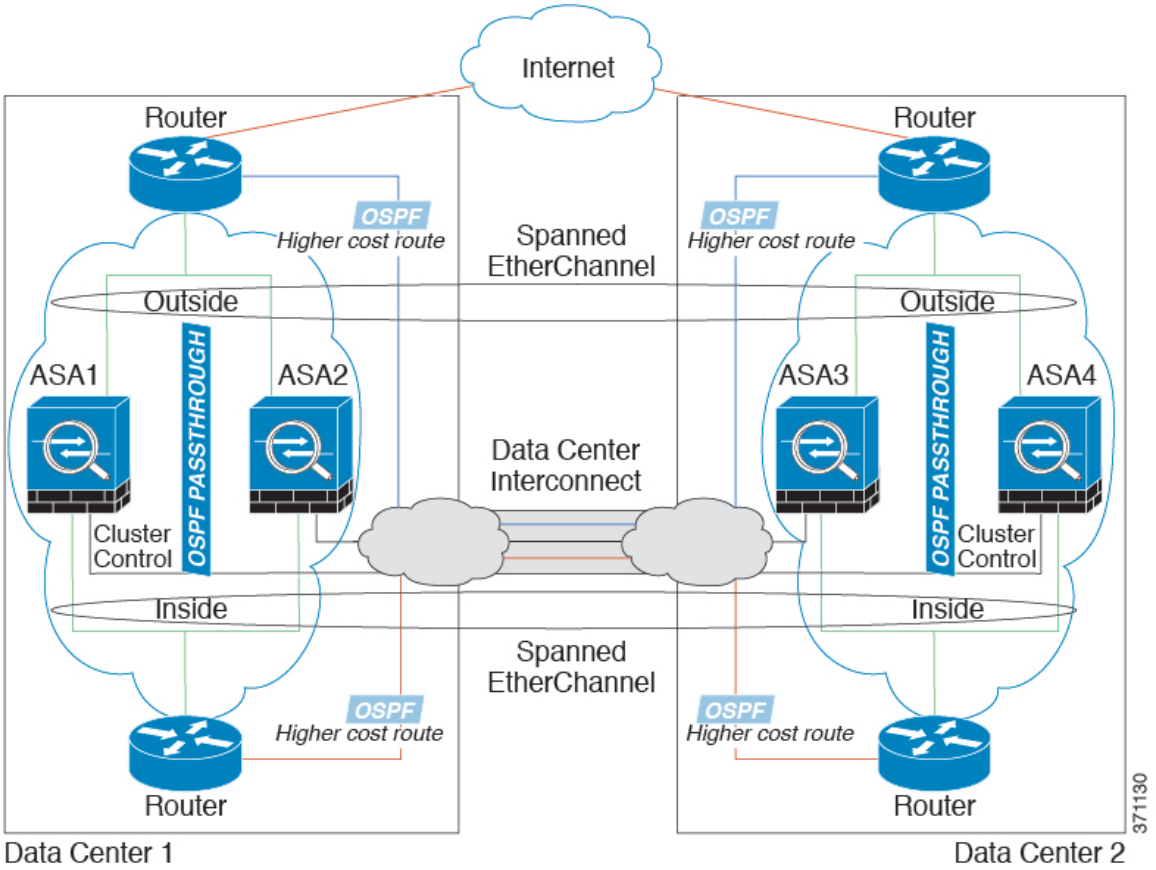
The following example shows 2 cluster members at each of 2 data centers placed between inside and outside routers (North-South insertion). The cluster members are connected by the cluster control link over the DCI. The cluster members at each site connect to the local switches using spanned EtherChannels for the inside and outside. Each EtherChannel is spanned across all chassis in the cluster.

The inside and outside routers at each data center use OSPF, which is passed through the transparent ASAs. Unlike MACs, router IPs are unique on all routers. By assigning a higher cost route across the DCI, traffic stays within each data center unless all cluster members at a given site go down. The lower cost route through the ASAs must traverse the same bridge group at each site for the cluster to maintain asymmetric connections.

In the event of a failure of all cluster members at one site, traffic goes from each router over the DCI to the cluster members at the other site.

The implementation of the switches at each site can include:

- Inter-site VSS/vPC—In this scenario, you install one switch at Data Center 1, and the other at Data Center 2. One option is for the cluster units at each Data Center to only connect to the local switch, while the VSS/vPC traffic goes across the DCI. In this case, connections are for the most part kept local to each datacenter. You can optionally connect each unit to both switches across the DCI if the DCI can handle the extra traffic. In this case, traffic is distributed across the data centers, so it is essential for the DCI to be very robust.
- Local VSS/vPC at each site—For better switch redundancy, you can install 2 separate VSS/vPC pairs at each site. In this case, although the cluster units still have a spanned EtherChannel with Data Center 1 chassis connected only to both local switches, and Data Center 2 chassis connected to those local switches, the spanned EtherChannel is essentially "split." Each local VSS/vPC sees the spanned EtherChannel as a site-local EtherChannel.

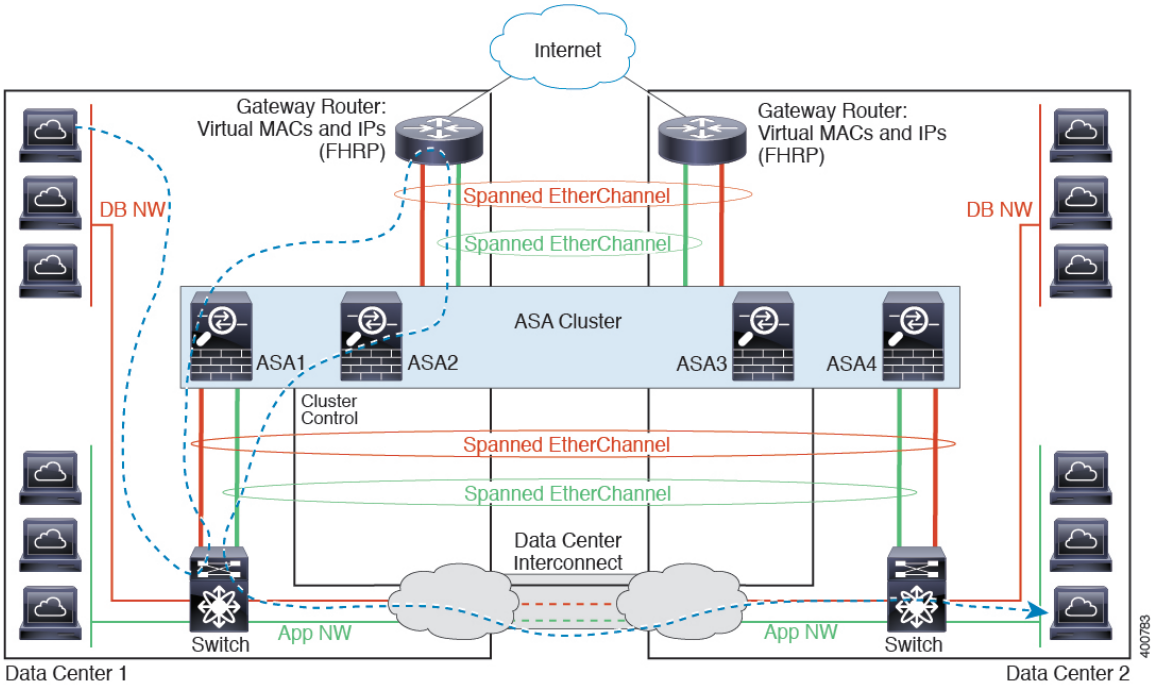


Spanned EtherChannel Transparent Mode East-West Inter-Site Example

The following example shows 2 cluster members at each of 2 data centers placed between the gateway router and two inside networks at each site, the App network and the DB network (East-West insertion). The cluster members are connected by the cluster control link over the DCI. The cluster members at each site connect to

the local switches using spanned EtherChannels for both the App and DB networks on the inside and outside. Each EtherChannel is spanned across all chassis in the cluster.

The gateway router at each site uses an FHRP such as HSRP to provide the same destination virtual MAC and IP addresses at each site. A good practice to avoid unintended MAC address flapping is to statically add the gateway routers real MAC addresses to the ASA MAC address table. Without these entries, if the gateway at site 1 communicates with the gateway at site 2, that traffic might pass through the ASA and attempt to reach site 2 from the inside interface and cause problems. The data VLANs are extended between the sites using Overlay Transport Virtualization (OTV) (or something similar). You must add filters to prevent traffic from traversing the DCI to the other site when the traffic is destined for the gateway router. If the gateway router at one site becomes unreachable, you must remove the filters so traffic can be sent to the other site's gateway router.



See Spanned EtherChannel Transparent Mode North-South Inter-Site Example, on page 69 for information about vPC/VSS options.

Reference for Clustering

This section includes more information about how clustering operates.

ASA Features and Clustering

Some ASA features are not supported with ASA clustering, and some are only supported on the control unit. Other features might have caveats for proper usage.

Unsupported Features with Clustering

These features cannot be configured with clustering enabled, and the commands will be rejected.

- Unified Communication features that rely on TLS Proxy
- Remote access VPN (SSL VPN and IPsec VPN)
- The following application inspections:
  - CTIQBE
  - H323, H225, and RAS
  - IPsec passthrough
  - MGCP
  - MMP
  - RTSP
  - SCCP (Skinny)
  - WAAS
  - WCCP
- Botnet Traffic Filter
- Auto Update Server
- DHCP client, server, and proxy. DHCP relay is supported.
- VPN load balancing
- Failover
- ASA CX module
- Integrated Routing and Bridging
- Dead Connection Detection (DCD)
- FIPS mode

Centralized Features for Clustering

The following features are only supported on the control unit, and are not scaled for the cluster. For example, you have a cluster of eight units (5516-X). The Other VPN license allows a maximum of 300 site-to-site IPsec tunnels for one ASA 5516-X. For the entire cluster of eight units, you can only use 300 tunnels; the feature does not scale.



Note Traffic for centralized features is forwarded from member units to the control unit over the cluster control link.

If you use the rebalancing feature, traffic for centralized features may be rebalanced to non-control units before the traffic is classified as a centralized feature; if this occurs, the traffic is then sent back to the control unit.

For centralized features, if the control unit fails, all connections are dropped, and you have to re-establish the connections on the new control unit.

- Site-to-site VPN
- The following application inspections:
  - DCERPC
  - ESMTP
  - IM
  - NetBIOS
  - PPTP
  - RADIUS
  - RSH
  - SNMP
  - SQLNET
  - SUNRPC
  - TFTP
  - XDMCP
- Dynamic routing (Spanned EtherChannel mode only)
- Multicast routing (Individual interface mode only)
- Static route tracking
- IGMP multicast control plane protocol processing (data plane forwarding is distributed across the cluster)
- PIM multicast control plane protocol processing (data plane forwarding is distributed across the cluster)
- Authentication and Authorization for network access. Accounting is decentralized.
- Filtering Services

Features Applied to Individual Units

These features are applied to each ASA unit, instead of the cluster as a whole or to the control unit.

- QoS—The QoS policy is synced across the cluster as part of configuration replication. However, the policy is enforced on each unit independently. For example, if you configure policing on output, then the conform rate and conform burst values are enforced on traffic exiting a particular ASA. In a cluster with 3 units and with traffic evenly distributed, the conform rate actually becomes 3 times the rate for the cluster.
- Threat detection—Threat detection works on each unit independently; for example, the top statistics is unit-specific. Port scanning detection, for example, does not work because scanning traffic will be load-balanced between all units, and one unit will not see all traffic.
- Resource management—Resource management in multiple context mode is enforced separately on each unit based on local usage.

- LISP traffic—LISP traffic on UDP port 4342 is inspected by each receiving unit, but is not assigned a director. Each unit adds to the EID table that is shared across the cluster, but the LISP traffic itself does not participate in cluster state sharing.
- ASA Firepower module—There is no configuration sync or state sharing between ASA Firepower modules. You are responsible for maintaining consistent policies on the ASA Firepower modules in the cluster using Firepower Management Center. Do not use different ASA-interface-based zone definitions for devices in the cluster.
- ASA IPS module—There is no configuration sync or state sharing between IPS modules. Some IPS signatures require IPS to keep the state across multiple connections. For example, the port scanning signature is used when the IPS module detects that someone is opening many connections to one server but with different ports. In clustering, those connections will be balanced between multiple ASA devices, each of which has its own IPS module. Because these IPS modules do not share state information, the cluster may not be able to detect port scanning as a result.

AAA for Network Access and Clustering

AAA for network access consists of three components: authentication, authorization, and accounting. Authentication and authorization are implemented as centralized features on the clustering control unit with replication of the data structures to the cluster data units. If a control unit is elected, the new control unit will have all the information it needs to continue uninterrupted operation of the established authenticated users and their associated authorizations. Idle and absolute timeouts for user authentications are preserved when a control unit change occurs.

Accounting is implemented as a distributed feature in a cluster. Accounting is done on a per-flow basis, so the cluster unit owning a flow will send accounting start and stop messages to the AAA server when accounting is configured for a flow.

Connection Settings

Connection limits are enforced cluster-wide (see **Configuration** > **Firewall** > **Service Policy** page). Each unit has an estimate of the cluster-wide counter values based on broadcast messages. Due to efficiency considerations, the configured connection limit across the cluster might not be enforced exactly at the limit number. Each unit may overestimate or underestimate the cluster-wide counter value at any given time. However, the information will get updated over time in a load-balanced cluster.

FTP and Clustering

- If FTP data channel and control channel flows are owned by different cluster members, then the data channel owner will periodically send idle timeout updates to the control channel owner and update the idle timeout value. However, if the control flow owner is reloaded, and the control flow is re-hosted, the parent/child flow relationship will not longer be maintained; the control flow idle timeout will not be updated.
- If you use AAA for FTP access, then the control channel flow is centralized on the control unit.

Identity Firewall and Clustering

Only the control unit retrieves the user-group from the AD and the user-ip mapping from the AD agent. The control unit then populates the user information to data units, and data units can make a match decision for user identity based on the security policy.

Multicast Routing and Clustering

Multicast routing behaves differently depending on the interface mode.

Multicast Routing in Spanned EtherChannel Mode

In Spanned EtherChannel mode, the control unit handles all multicast routing packets and data packets until fast-path forwarding is established. After the connection is established, each data unit can forward multicast data packets.

Multicast Routing in Individual Interface Mode

In Individual interface mode, units do not act independently with multicast. All data and routing packets are processed and forwarded by the control unit, thus avoiding packet replication.

NAT and Clustering

NAT can affect the overall throughput of the cluster. Inbound and outbound NAT packets can be sent to different ASAs in the cluster, because the load balancing algorithm relies on IP addresses and ports, and NAT causes inbound and outbound packets to have different IP addresses and/or ports. When a packet arrives at the ASA that is not the NAT owner, it is forwarded over the cluster control link to the owner, causing large amounts of traffic on the cluster control link. Note that the receiving unit does not create a forwarding flow to the owner, because the NAT owner may not end up creating a connection for the packet depending on the results of security and policy checks.

If you still want to use NAT in clustering, then consider the following guidelines:

- No Proxy ARP—For Individual interfaces, a proxy ARP reply is never sent for mapped addresses. This prevents the adjacent router from maintaining a peer relationship with an ASA that may no longer be in the cluster. The upstream router needs a static route or PBR with Object Tracking for the mapped addresses that points to the Main cluster IP address. This is not an issue for a Spanned EtherChannel, because there is only one IP address associated with the cluster interface.
- No interface PAT on an Individual interface—Interface PAT is not supported for Individual interfaces.
- PAT with Port Block Allocation—See the following guidelines for this feature:
  - Maximum-per-host limit is not a cluster-wide limit, and is enforced on each unit individually. Thus, in a 3-node cluster with the maximum-per-host limit configured as 1, if the traffic from a host is load-balanced across all 3 units, then it can get allocated 3 blocks with 1 in each unit.
  - Port blocks created on the backup unit from the backup pools are not accounted for when enforcing the maximum-per-host limit.
  - When a PAT IP address owner goes down, the backup unit will own the PAT IP address, corresponding port blocks, and xlates. If it runs out of ports on its normal PAT address, it can use the address that it took over to service new requests. As the connections eventually time out, the blocks get freed.
  - On-the-fly PAT rule modifications, where the PAT pool is modified with a completely new range of IP addresses, will result in xlate backup creation failures for the xlate backup requests that were still in transit while the new pool became effective. This behavior is not specific to the port block allocation feature, and is a transient PAT pool issue seen only in cluster deployments where the pool is distributed and traffic is load-balanced across the cluster units.

- NAT pool address distribution for dynamic PAT—The control unit evenly pre-distributes addresses across the cluster. If a member receives a connection and they have no addresses assigned, then the connection is forwarded to the control unit for PAT. If a cluster member leaves the cluster (due to failure), a backup member will get the PAT IP address, and if the backup exhausts its normal PAT IP address, it can make use of the new address. Make sure to include at least as many NAT addresses as there are units in the cluster, plus at least one extra address, to ensure that each unit receives an address, and that a failed unit can get a new address if its old address is in use by the member that took over the address. Use the **show nat pool cluster** command to see the address allocations.
- Reusing a PAT pool in multiple rules—To use the same PAT pool in multiple rules, you must be careful about the interface selection in the rules. You must either use specific interfaces in all rules, or "any" in all rules. You cannot mix specific interfaces and "any" across the rules, or the system might not be able to match return traffic to the right node in the cluster. Using unique PAT pools per rule is the most reliable option.
- No round-robin—Round-robin for a PAT pool is not supported with clustering.
- Dynamic NAT xlates managed by the control unit—The control unit maintains and replicates the xlate table to data units. When a data unit receives a connection that requires dynamic NAT, and the xlate is not in the table, it requests the xlate from the control unit. The data unit owns the connection.
- Stale xlates—The xlate idle time on the connection owner does not get updated. Thus, the idle time might exceed the idle timeout. An idle timer value higher than the configured timeout with a refcnt of 0 is an indication of a stale xlate.
- Per-session PAT feature—Although not exclusive to clustering, the per-session PAT feature improves the scalability of PAT and, for clustering, allows each data unit to own PAT connections; by contrast, multi-session PAT connections have to be forwarded to and owned by the control unit. By default, all TCP traffic and UDP DNS traffic use a per-session PAT xlate, whereas ICMP and all other UDP traffic uses multi-session. You can configure per-session NAT rules to change these defaults for TCP and UDP, but you cannot configure per-session PAT for ICMP. For traffic that benefits from multi-session PAT, such as H.323, SIP, or Skinny, you can disable per-session PAT for the associated TCP ports (the UDP ports for those H.323 and SIP are already multi-session by default). For more information about per-session PAT, see the firewall configuration guide.
- No static PAT for the following inspections—
  - FTP
  - PPTP
  - RSH
  - SQLNET
  - TFTP
  - XDMCP
  - SIP
- If you have an extremely large number of NAT rules, over ten thousand, you should enable the transactional commit model using the **asp rule-engine transactional-commit nat** command in the device CLI. Otherwise, the unit might not be able to join the cluster.

Dynamic Routing and Clustering

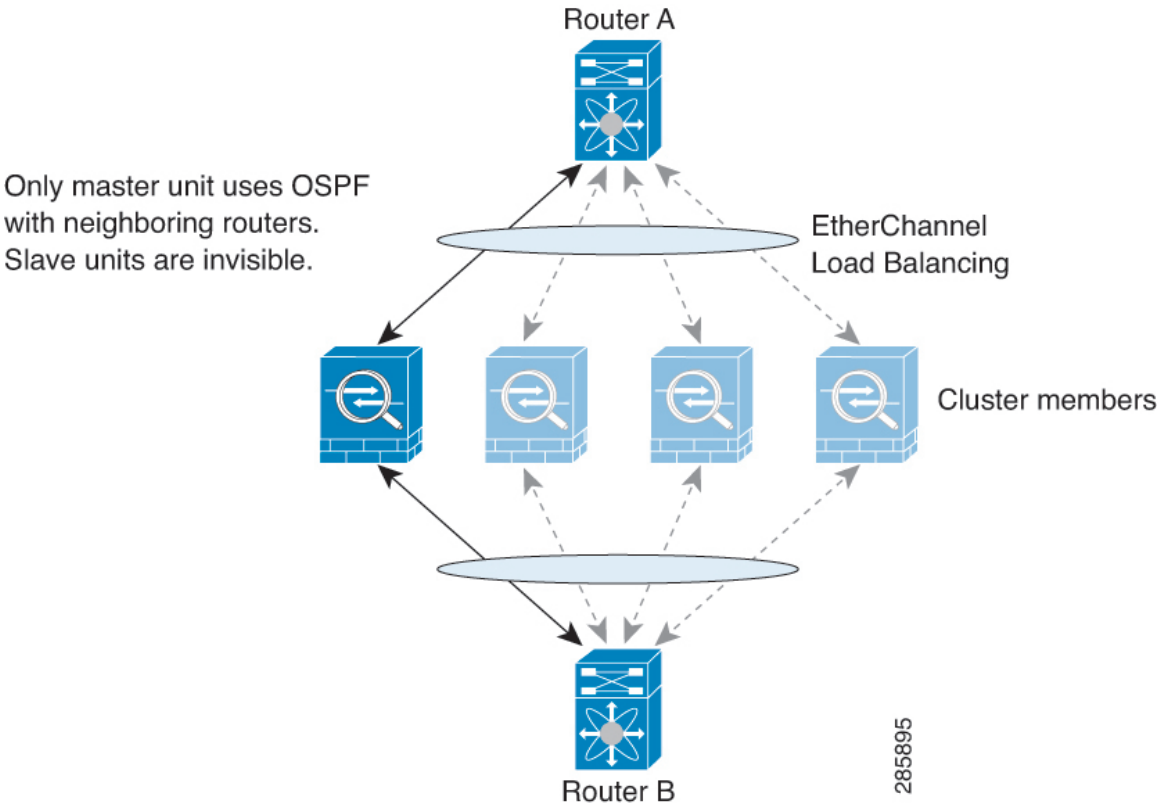
This section describes how to use dynamic routing with clustering.

Dynamic Routing in Spanned EtherChannel Mode

Note IS-IS is not supported in Spanned EtherChannel mode.

In Spanned EtherChannel mode: The routing process only runs on the control unit, and routes are learned through the control unit and replicated to data units. If a routing packet arrives at a data unit, it is redirected to the control unit.

Figure 1: Dynamic Routing in Spanned EtherChannel Mode



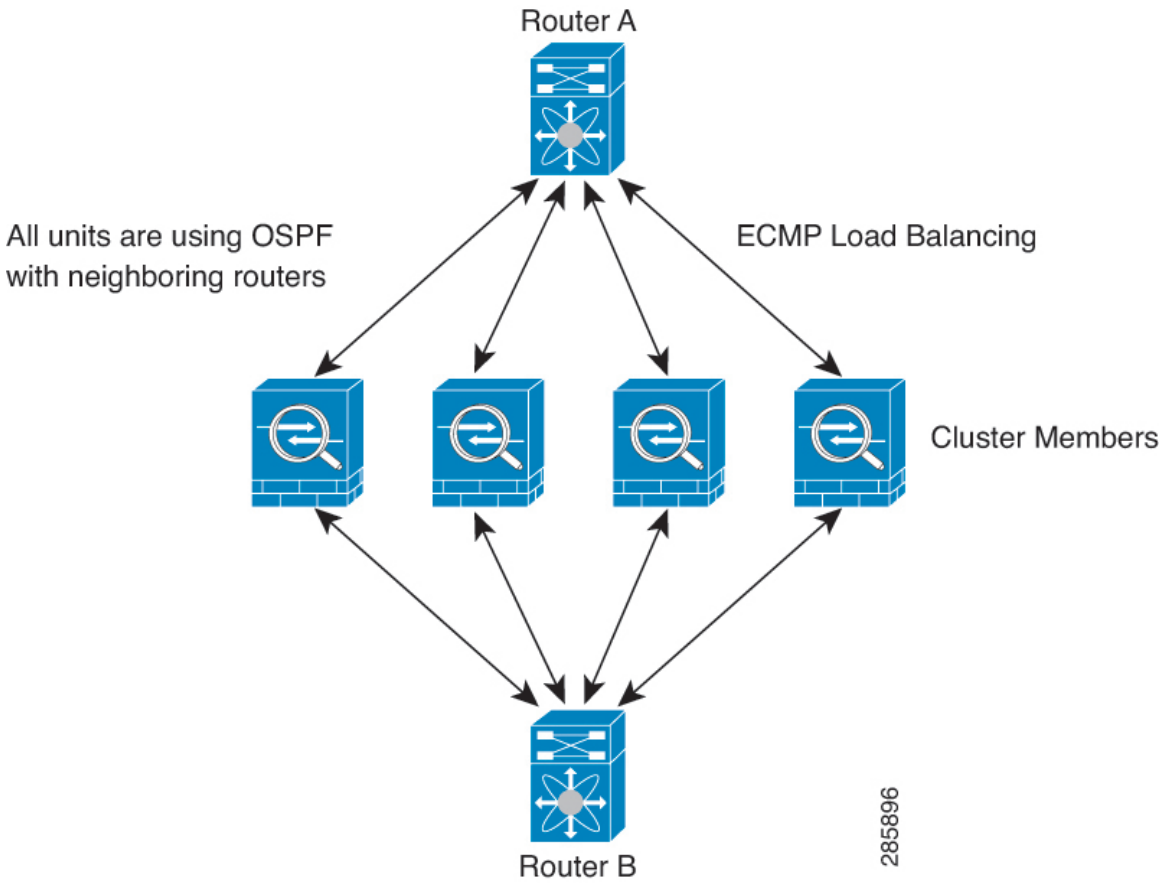
After the data unit learn the routes from the control unit, each unit makes forwarding decisions independently.

The OSPF LSA database is not synchronized from the control unit to data units. If there is a control unit switchover, the neighboring router will detect a restart; the switchover is not transparent. The OSPF process picks an IP address as its router ID. Although not required, you can assign a static router ID to ensure a consistent router ID is used across the cluster. See the OSPF Non-Stop Forwarding feature to address the interruption.

Dynamic Routing in Individual Interface Mode

In Individual interface mode, each unit runs the routing protocol as a standalone router, and routes are learned by each unit independently.

Figure 2: Dynamic Routing in Individual Interface Mode



In the above diagram, Router A learns that there are 4 equal-cost paths to Router B, each through an ASA. ECMP is used to load balance traffic between the 4 paths. Each ASA picks a different router ID when talking to external routers.

You must configure a cluster pool for the router ID so that each unit has a separate router ID.

EIGRP does not form neighbor relationships with cluster peers in individual interface mode.

Note If the cluster has multiple adjacencies to the same router for redundancy purposes, asymmetric routing can lead to unacceptable traffic loss. To avoid asymmetric routing, group all of these ASA interfaces into the same traffic zone. See Configure a Traffic Zone.

SCTP and Clustering

An SCTP association can be created on any unit (due to load balancing); its multi-homing connections must reside on the same unit.

SIP Inspection and Clustering

A control flow can be created on any unit (due to load balancing); its child data flows must reside on the same unit.

TLS Proxy configuration is not supported.

SNMP and Clustering

An SNMP agent polls each individual ASA by its Local IP address. You cannot poll consolidated data for the cluster.

You should always use the Local address, and not the Main cluster IP address for SNMP polling. If the SNMP agent polls the Main cluster IP address, if a new control unit is elected, the poll to the new control unit will fail.

When using SNMPv3 with clustering, if you add a new cluster unit after the initial cluster formation, then SNMPv3 users are not replicated to the new unit.You must re-add them on the control unit to force the users to replicate to the new unit, or directly on the data unit.

STUN and Clustering

STUN inspection is supported in failover and cluster modes, as pinholes are replicated. However, the transaction ID is not replicated among units. In the case where a unit fails after receiving a STUN Request and another unit received the STUN Response, the STUN Response will be dropped.

Syslog and NetFlow and Clustering

- Syslog—Each unit in the cluster generates its own syslog messages. You can configure logging so that each unit uses either the same or a different device ID in the syslog message header field. For example, the hostname configuration is replicated and shared by all units in the cluster. If you configure logging to use the hostname as the device ID, syslog messages generated by all units look as if they come from a single unit. If you configure logging to use the local-unit name that is assigned in the cluster bootstrap configuration as the device ID, syslog messages look as if they come from different units.
- NetFlow—Each unit in the cluster generates its own NetFlow stream. The NetFlow collector can only treat each ASA as a separate NetFlow exporter.

Cisco TrustSec and Clustering

Only the control unit learns security group tag (SGT) information. The control unit then populates the SGT to data units, and data units can make a match decision for SGT based on the security policy.

VPN and Clustering

Site-to-site VPN is a centralized feature; only the control unit supports VPN connections. Distributed site-to-site VPN clustering is supported. Search for High Availability options in this pdf for details.



Note Remote access VPN is not supported with clustering.

VPN functionality is limited to the control unit and does not take advantage of the cluster high availability capabilities. If the control unit fails, all existing VPN connections are lost, and VPN users will see a disruption in service. When a new control unit is elected, you must reestablish the VPN connections.

When you connect a VPN tunnel to a Spanned EtherChannel address, connections are automatically forwarded to the control unit. For connections to an Individual interface when using PBR or ECMP, you must always connect to the Main cluster IP address, not a Local address.

VPN-related keys and certificates are replicated to all units.

Performance Scaling Factor

When you combine multiple units into a cluster, you can expect a performance of approximately:

- 70% of the combined throughput
- 60% of maximum connections
- 50% of connections per second

For example, for throughput, the ASA 5585-X with SSP-40 can handle approximately 10 Gbps of real world firewall traffic when running alone. For a cluster of 8 units, the maximum combined throughput will be approximately 70% of 80 Gbps (8 units x 10 Gbps): 56 Gbps.

Control Unit Election

Members of the cluster communicate over the cluster control link to elect a control unit as follows:

1. When you enable clustering for a unit (or when it first starts up with clustering already enabled), it broadcasts an election request every 3 seconds.
2. Any other units with a higher priority respond to the election request; the priority is set between 1 and 100, where 1 is the highest priority.
3. If after 45 seconds, a unit does not receive a response from another unit with a higher priority, then it becomes the control unit.



Note If multiple units tie for the highest priority, the cluster unit name and then the serial number is used to determine the control unit.

4. If a unit later joins the cluster with a higher priority, it does not automatically become the control unit; the existing control unit always remains as the control unit unless it stops responding, at which point a new control unit is elected.
5. In a "split brain" scenario when there are temporarily multiple control units, then the unit with highest priority retains the role while the other units return to data unit roles.



Note You can manually force a unit to become the control unit. For centralized features, if you force a control unit change, then all connections are dropped, and you have to re-establish the connections on the new control unit.

High Availability Within the ASA Cluster

ASA Clustering provides high availability by monitoring unit and interface health and by replicating connection states between units.

Unit Health Monitoring

Each unit periodically sends a broadcast heartbeat packet over the cluster control link. If the control unit does not receive any heartbeat packets or other packets from a data unit within the configurable timeout period, then the control unit removes the data unit from the cluster. If the data units do not receive packets from the control unit, then a new control unit is elected from the remaining members.

If units cannot reach each other over the cluster control link because of a network failure and not because a unit has actually failed, then the cluster may go into a "split brain" scenario where isolated data units will elect their own control units. For example, if a router fails between two cluster locations, then the original control unit at location 1 will remove the location 2 data units from the cluster. Meanwhile, the units at location 2 will elect their own control unit and form their own cluster. Note that asymmetric traffic may fail in this scenario. After the cluster control link is restored, then the control unit that has the higher priority will keep the control unit's role. See Control Unit Election, on page 80 for more information.

Interface Monitoring

Each unit monitors the link status of all named hardware interfaces in use, and reports status changes to the control unit.

- Spanned EtherChannel—Uses cluster Link Aggregation Control Protocol (cLACP). Each unit monitors the link status and the cLACP protocol messages to determine if the port is still active in the EtherChannel. The status is reported to the control unit.
- Individual interfaces (Routed mode only)—Each unit self-monitors its interfaces and reports interface status to the control unit.

When you enable health monitoring, all physical interfaces (including the main EtherChannel and redundant interface types) are monitored by default; you can optionally disable monitoring per interface. Only named interfaces can be monitored. For example, the named EtherChannel must fail to be considered failed, which means all member ports of an EtherChannel must fail to trigger cluster removal (depending on your minimum port bundling setting).

A unit is removed from the cluster if its monitored interfaces fail. The amount of time before the ASA removes a member from the cluster depends on the type of interface and whether the unit is an established member or is joining the cluster. For EtherChannels (spanned or not), if the interface is down on an established member, then the ASA removes the member after 9 seconds. The ASA does not monitor interfaces for the first 90 seconds that a unit joins the cluster. Interface status changes during this time will not cause the ASA to be removed from the cluster. For non-EtherChannels, the unit is removed after 500 ms, regardless of the member state.

Status After Failure

When a unit in the cluster fails, the connections hosted by that unit are seamlessly transferred to other units; state information for traffic flows is shared over the control unit's cluster control link.

If the control unit fails, then another member of the cluster with the highest priority (lowest number) becomes the control unit.

The ASA automatically tries to rejoin the cluster, depending on the failure event.

Note When the ASA becomes inactive and fails to automatically rejoin the cluster, all data interfaces are shut down; only the management-only interface can send and receive traffic. The management interface remains up using the IP address the unit received from the cluster IP pool. However if you reload, and the unit is still inactive in the cluster, the management interface is disabled. You must use the console port for any further configuration.

Rejoining the Cluster

After a cluster member is removed from the cluster, how it can rejoin the cluster depends on why it was removed:

- Failed cluster control link when initially joining—After you resolve the problem with the cluster control link, you must manually rejoin the cluster by re-enabling clustering.
- Failed cluster control link after joining the cluster—The ASA automatically tries to rejoin every 5 minutes, indefinitely. This behavior is configurable.
- Failed data interface—The ASA automatically tries to rejoin at 5 minutes, then at 10 minutes, and finally at 20 minutes. If the join is not successful after 20 minutes, then the ASA disables clustering. After you resolve the problem with the data interface, you have to manually enable clustering. This behavior is configurable.
- Failed ASA FirePOWER module on the ASA 5585-X—The ASA automatically tries to rejoin at 5 minutes.
- Failed ASA FirePOWER software module—After you resolve the problem with the module, you must manually enable clustering.
- Failed unit—If the unit was removed from the cluster because of a unit health check failure, then rejoining the cluster depends on the source of the failure. For example, a temporary power failure means the unit will rejoin the cluster when it starts up again as long as the cluster control link is up and clustering is still enabled. The ASA attempts to rejoin the cluster every 5 seconds.
- Internal error—Internal failures include: application sync timeout; inconsistent application statuses; and so on. A unit will attempt to rejoin the cluster automatically at the following intervals: 5 minutes, 10 minutes, and then 20 minutes. This behavior is configurable.

See Configure Basic ASA Cluster Parameters, on page 36.

Data Path Connection State Replication

Every connection has one owner and at least one backup owner in the cluster. The backup owner does not take over the connection in the event of a failure; instead, it stores TCP/UDP state information, so that the connection can be seamlessly transferred to a new owner in case of a failure. The backup owner is usually also the director.

Some traffic requires state information above the TCP or UDP layer. See the following table for clustering support or lack of support for this kind of traffic.

Table 1: Features Replicated Across the Cluster

Traffic	State Support	Notes
Up time	Yes	Keeps track of the system up time.
ARP Table	Yes	—
MAC address table	Yes	—
User Identity	Yes	Includes AAA rules (uauth) and identity firewall.
IPv6 Neighbor database	Yes	—
Dynamic routing	Yes	—
SNMP Engine ID	**No**	—
Centralized VPN (Site-to-Site)	**No**	VPN sessions will be disconnected if the control unit fails.
Distributed VPN (Site-to-Site)	Yes	Backup session becomes the active session, then a new backup session is created.

How the ASA Cluster Manages Connections

Connections can be load-balanced to multiple members of the cluster. Connection roles determine how connections are handled in both normal operation and in a high availability situation.

Connection Roles

See the following roles defined for each connection:

- Owner—Usually, the unit that initially receives the connection. The owner maintains the TCP state and processes packets. A connection has only one owner. If the original owner fails, then when new units receive packets from the connection, the director chooses a new owner from those units.
- Backup owner—The unit that stores TCP/UDP state information received from the owner, so that the connection can be seamlessly transferred to a new owner in case of a failure. The backup owner does not take over the connection in the event of a failure. If the owner becomes unavailable, then the first unit to receive packets from the connection (based on load balancing) contacts the backup owner for the relevant state information so it can become the new owner.

  As long as the director (see below) is not the same unit as the owner, then the director is also the backup owner. If the owner chooses itself as the director, then a separate backup owner is chosen.

  For inter-chassis clustering on the Firepower 9300, which can include up to 3 cluster units in one chassis, if the backup owner is on the same chassis as the owner, then an additional backup owner will be chosen from another chassis to protect flows from a chassis failure.

  If you enable director localization for inter-site clustering, then there are two backup owner roles: the local backup and the global backup. The owner always chooses a local backup at the same site as itself

(based on site ID). The global backup can be at any site, and might even be the same unit as the local backup. The owner sends connection state information to both backups.

If you enable site redundancy, and the backup owner is at the same site as the owner, then an additional backup owner will be chosen from another site to protect flows from a site failure. Chassis backup and site backup are independent, so in some cases a flow will have both a chassis backup and a site backup.

- Director—The unit that handles owner lookup requests from forwarders. When the owner receives a new connection, it chooses a director based on a hash of the source/destination IP address and ports, and sends a message to the director to register the new connection. If packets arrive at any unit other than the owner, the unit queries the director about which unit is the owner so it can forward the packets. A connection has only one director. If a director fails, the owner chooses a new director.

  As long as the director is not the same unit as the owner, then the director is also the backup owner (see above). If the owner chooses itself as the director, then a separate backup owner is chosen.

  If you enable director localization for inter-site clustering, then there are two director roles: the local director and the global director. The owner always chooses a local director at the same site as itself (based on site ID). The global director can be at any site, and might even be the same unit as the local director. If the original owner fails, then the local director chooses a new connection owner at the same site.

- Forwarder—A unit that forwards packets to the owner. If a forwarder receives a packet for a connection it does not own, it queries the director for the owner, and then establishes a flow to the owner for any other packets it receives for this connection. The director can also be a forwarder. If you enable director localization, then the forwarder always queries the local director. The forwarder only queries the global director if the local director does not know the owner, for example, if a cluster member receives packets for a connection that is owned on a different site. Note that if a forwarder receives the SYN-ACK packet, it can derive the owner directly from a SYN cookie in the packet, so it does not need to query the director. (If you disable TCP sequence randomization, the SYN cookie is not used; a query to the director is required.) For short-lived flows such as DNS and ICMP, instead of querying, the forwarder immediately sends the packet to the director, which then sends them to the owner. A connection can have multiple forwarders; the most efficient throughput is achieved by a good load-balancing method where there are no forwarders and all packets of a connection are received by the owner.

  Note We do not recommend disabling TCP sequence randomization when using clustering. There is a small chance that some TCP sessions won't be established, because the SYN/ACK packet might be dropped.

- Fragment Owner—For fragmented packets, cluster units that receive a fragment determine a fragment owner using a hash of the fragment source IP address, destination IP address, and the packet ID. All fragments are then forwarded to the fragment owner over the cluster control link. Fragments may be load-balanced to different cluster units, because only the first fragment includes the 5-tuple used in the switch load balance hash. Other fragments do not contain the source and destination ports and may be load-balanced to other cluster units. The fragment owner temporarily reassembles the packet so it can determine the director based on a hash of the source/destination IP address and ports. If it is a new connection, the fragment owner will register to be the connection owner. If it is an existing connection, the fragment owner forwards all fragments to the provided connection owner over the cluster control link. The connection owner will then reassemble all fragments.

When a connection uses Port Address Translation (PAT), then the PAT type (per-session or multi-session) influences which member of the cluster becomes the owner of a new connection:

- Per-session PAT—The owner is the unit that receives the initial packet in the connection.

  By default, TCP and DNS UDP traffic use per-session PAT.

- Multi-session PAT—The owner is always the control unit. If a multi-session PAT connection is initially received by a data unit, then the data unit forwards the connection to the control unit.

  By default, UDP (except for DNS UDP) and ICMP traffic use multi-session PAT, so these connections are always owned by the control unit.

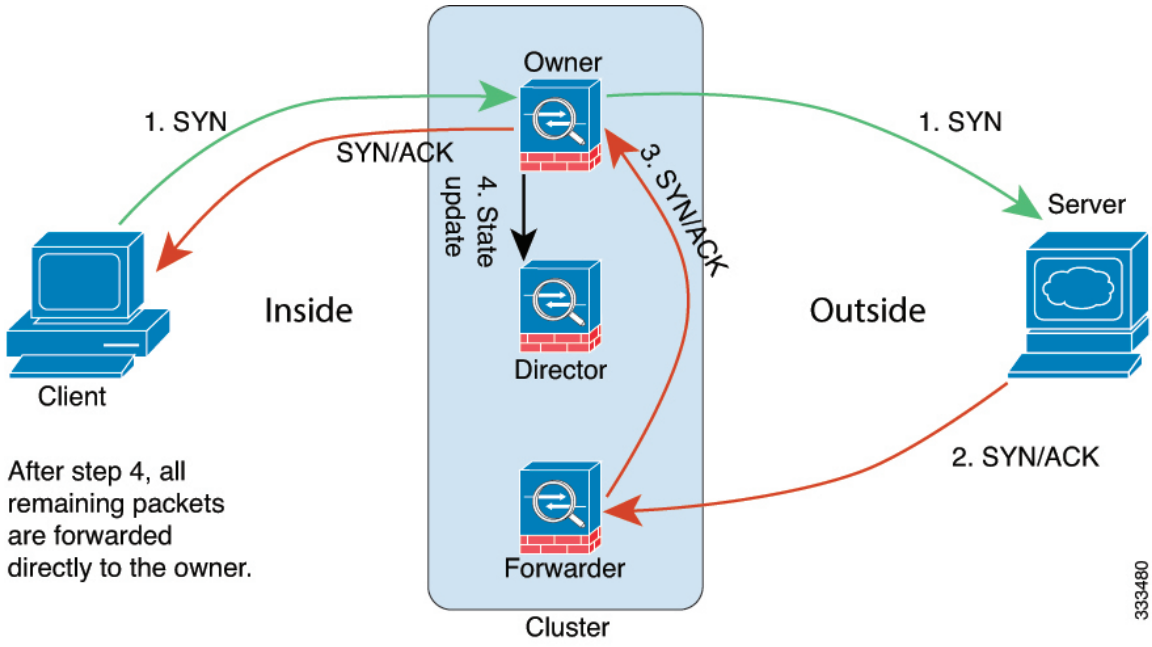
You can change the per-session PAT defaults for TCP and UDP so connections for these protocols are handled per-session or multi-session depending on the configuration. For ICMP, you cannot change from the default multi-session PAT. For more information about per-session PAT, see the firewall configuration guide.

New Connection Ownership

When a new connection is directed to a member of the cluster via load balancing, that unit owns both directions of the connection. If any connection packets arrive at a different unit, they are forwarded to the owner unit over the cluster control link. For best performance, proper external load balancing is required for both directions of a flow to arrive at the same unit, and for flows to be distributed evenly between units. If a reverse flow arrives at a different unit, it is redirected back to the original unit.

Sample Data Flow

The following example shows the establishment of a new connection.



1. The SYN packet originates from the client and is delivered to one ASA (based on the load balancing method), which becomes the owner. The owner creates a flow, encodes owner information into a SYN cookie, and forwards the packet to the server.
2. The SYN-ACK packet originates from the server and is delivered to a different ASA (based on the load balancing method). This ASA is the forwarder.

3. Because the forwarder does not own the connection, it decodes owner information from the SYN cookie, creates a forwarding flow to the owner, and forwards the SYN-ACK to the owner.
4. The owner sends a state update to the director, and forwards the SYN-ACK to the client.
5. The director receives the state update from the owner, creates a flow to the owner, and records the TCP state information as well as the owner. The director acts as the backup owner for the connection.
6. Any subsequent packets delivered to the forwarder will be forwarded to the owner.
7. If packets are delivered to any additional units, it will query the director for the owner and establish a flow.
8. Any state change for the flow results in a state update from the owner to the director.

Rebalancing New TCP Connections Across the Cluster

If the load balancing capabilities of the upstream or downstream routers result in unbalanced flow distribution, you can configure overloaded units to redirect new TCP flows to other units. No existing flows will be moved to other units.

History for ASA Clustering

Feature Name	Version	Feature Information
Per-site gratuitous ARP for clustering	9.12(1)	The ASA now generates gratuitous ARP (GARP) packets to keep the switching infrastructure up to date: the highest priority member at each site periodically generates GARP traffic for the global MAC/IP addresses. When using per-site MAC and IP addresses, packets sourced from the cluster use a site-specific MAC address and IP address, while packets received by the cluster use a global MAC address and IP address. If traffic is not generated from the global MAC address periodically, you could experience a MAC address timeout on your switches for the global MAC address. After a timeout, traffic destined for the global MAC address will be flooded across the entire switching infrastructure, which can cause performance and security concerns. GARP is enabled by default when you set the site ID for each unit and the site MAC address for each Spanned EtherChannel. New/Modified screens: **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster** > **Cluster Configuration** > **Site Periodic GARP** field
Cluster interface debounce time now applies to interfaces changing from a down state to an up state	9.10(1)	When an interface status update occurs, the ASA waits the number of milliseconds specified in the **health-check monitor-interface debounce-time** command or the ASDM **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster** screen before marking the interface as failed and the unit is removed from the cluster. This feature now applies to interfaces changing from a down state to an up state. For example, in the case of an EtherChannel that transitions from a down state to an up state (for example, the switch reloaded, or the switch enabled an EtherChannel), a longer debounce time can prevent the interface from appearing to be failed on a cluster unit just because another cluster unit was faster at bundling the ports. We did not modify any screens.

Feature Name	Version	Feature Information
Automatically rejoin the cluster after an internal failure	9.9(2)	Formerly, many error conditions caused a cluster unit to be removed from the cluster, and you were required to manually rejoin the cluster after resolving the issue. Now, a unit will attempt to rejoin the cluster automatically at the following intervals by default: 5 minutes, 10 minutes, and then 20 minutes. These values are configurable. Internal failures include: application sync timeout; inconsistent application statuses; and so on. New or modified screen: **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster** > **Auto Rejoin**
Show transport related statistics for cluster reliable transport protocol messages	9.9(2)	You can now view per-unit cluster reliable transport buffer usage so you can identify packet drop issues when the buffer is full in the control plane. New or modified command: **show cluster info transport cp detail**
Configurable debounce time to mark an interface as failed for the ASA 5000-X series	9.9(2)	You can now configure the debounce time before the ASA considers an interface to be failed, and the unit is removed from the cluster on the ASA 5500-X series. This feature allows for faster detection of interface failures. Note that configuring a lower debounce time increases the chances of false-positives. When an interface status update occurs, the ASA waits the number of milliseconds specified before marking the interface as failed and the unit is removed from the cluster. The default debounce time is 500 ms, with a range of 300 ms to 9 seconds. This feature was previously available for the Firepower 4100/9300. New or modified screen: **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster**
Inter-site redundancy for clustering	9.9(1)	Inter-site redundancy ensures that a backup owner for a traffic flow will always be at the other site from the owner. This feature guards against site failure. New or modified screen: **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster**
Improved cluster unit health-check failure detection	9.8(1)	You can now configure a lower holdtime for the unit health check: .3 seconds minimum. The previous minimum was .8 seconds. This feature changes the unit health check messaging scheme to *heartbeats* in the data plane from *keepalives* in the control plane. Using heartbeats improves the reliability and the responsiveness of clustering by not being susceptible to control plane CPU hogging and scheduling delays. Note that configuring a lower holdtime increases cluster control link messaging activity. We suggest that you analyze your network before you configure a low holdtime; for example, make sure a ping from one unit to another over the cluster control link returns within the *holdtime*/3, because there will be three heartbeat messages during one holdtime interval. If you downgrade your ASA software after setting the hold time to .3 - .7, this setting will revert to the default of 3 seconds because the new setting is unsupported. We modified the following screen: **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster**

Feature Name	Version	Feature Information
Director localization: inter-site clustering improvement for data centers	9.7(1)	To improve performance and keep traffic within a site for inter-site clustering for data centers, you can enable director localization. New connections are typically load-balanced and owned by cluster members within a given site. However, the ASA assigns the director role to a member at *any* site. Director localization enables additional director roles: a local director at the same site as the owner, and a global director that can be at any site. Keeping the owner and director at the same site improves performance. Also, if the original owner fails, the local director chooses a new connection owner at the same site. The global director is used if a cluster member receives packets for a connection that is owned on a different site. We modified the following screen: **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster** > **Cluster Configuration**
Support for site-specific IP addresses in Routed, Spanned EtherChannel mode	9.6(1)	For inter-site clustering in routed mode with Spanned EtherChannels, you can now configure site-specific IP addresess in addition to site-specific MAC addresses. The addition of site IP addresses allows you to use ARP inspection on the Overlay Transport Virtualization (OTV) devices to prevent ARP responses from the global MAC address from traveling over the Data Center Interconnect (DCI), which can cause routing problems. ARP inspection is required for some switches that cannot use VACLs to filter MAC addresses. We modified the following screen: **Configuration** > **Device Setup** > **Interface Settings** > **Interfaces** > **Add/Edit EtherChannel Interface** > **Advanced**
ASA 5516-X support for clustering	9.5(2)	The ASA 5516-X now supports 2-unit clusters. Clustering for 2 units is enabled by default in the base license. We did not modify any ASDM screens.
LISP Inspection for Inter-Site Flow Mobility	9.5(2)	Cisco Locator/ID Separation Protocol (LISP) architecture separates the device identity from its location into two different numbering spaces, making server migration transparent to clients. The ASA can inspect LISP traffic for location changes and then use this information for seamless clustering operation; the ASA cluster members inspect LISP traffic passing between the first hop router and the egress tunnel router (ETR) or ingress tunnel router (ITR), and then change the flow owner to be at the new site. We introduced or modified the following screens: **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster** > **Cluster Configuration** **Configuration** > **Firewall** > **Objects** > **Inspect Maps** > **LISP** **Configuration** > **Firewall** > **Service Policy Rules** > **Protocol Inspection** **Configuration** > **Firewall** > **Service Policy Rules** > **Cluster** **Monitoring** > **Routing** > **LISP-EID Table**
Carrier Grade NAT enhancements now supported in failover and ASA clustering	9.5(2)	For carrier-grade or large-scale PAT, you can allocate a block of ports for each host, rather than have NAT allocate one port translation at a time (see RFC 6888). This feature is now supported in failover and ASA cluster deployments. We did not modify any screens.

Feature Name	Version	Feature Information
Configurable level for clustering trace entries	9.5(2)	By default, all levels of clustering events are included in the trace buffer, including many low level events. To limit the trace to higher level events, you can set the minimum trace level for the cluster. We did not modify any screens.
Site-specific MAC addresses for inter-site clustering support for Spanned EtherChannel in Routed firewall mode	9.5(1)	You can now use inter-site clustering for Spanned EtherChannels in routed mode. To avoid MAC address flapping, configure a site ID for each cluster member so that a site-specific MAC address for each interface can be shared among a site's units. We modified the following screen: **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster** > **Cluster Configuration**
ASA cluster customization of the auto-rejoin behavior when an interface or the cluster control link fails	9.5(1)	You can now customize the auto-rejoin behavior when an interface or the cluster control link fails. We introduced the following screen: **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster** > **Auto Rejoin**
The ASA cluster supports GTPv1 and GTPv2	9.5(1)	The ASA cluster now supports GTPv1 and GTPv2 inspection. We did not modify any screens.
Disable health monitoring of a hardware module in ASA clustering	9.5(1)	By default when using clustering, the ASA monitors the health of an installed hardware module such as the ASA FirePOWER module. If you do not want a hardware module failure to trigger failover, you can disable module monitoring. We modified the following screen: **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster** > **Cluster Interface Health Monitoring**
Cluster replication delay for TCP connections	9.5(1)	This feature helps eliminate the "unnecessary work" related to short-lived flows by delaying the director/backup flow creation. We introduced the following screen: **Configuration** > **Device Management** > **High Availability and Scalability** > **ASA Cluster Replication**
Enable and disable ASA cluster health monitoring per interface	9.4(1)	You can now enable or disable health monitoring per interface. Health monitoring is enabled by default on all port-channel, redundant, and single physical interfaces. Health monitoring is not performed on VLAN subinterfaces or virtual interfaces such as VNIs or BVIs. You cannot configure monitoring for the cluster control link; it is always monitored. You might want to disable health monitoring of non-essential interfaces, for example, the management interface. We introduced the following screen: **Configuration > Device Management > High Availability and Scalability > ASA Cluster > Cluster Interface Health Monitoring**.
ASA clustering support for DHCP relay	9.4(1)	You can now configure DHCP relay on the ASA cluster. Client DHCP requests are load-balanced to the cluster members using a hash of the client MAC address. DHCP client and server functions are still not supported. We did not modify any screens.

Feature Name	Version	Feature Information
SIP inspection support in ASA clustering	9.4(1)	You can now configure SIP inspection on the ASA cluster. A control flow can be created on any unit (due to load balancing), but its child data flows must reside on the same unit. TLS Proxy configuration is not supported. We did not modify any screens.
Inter-site deployment in transparent mode with the ASA cluster firewalling between inside networks	9.3(2)	You can now deploy a cluster in transparent mode between inside networks and the gateway router at each site (AKA East-West insertion), and extend the inside VLANs between sites. We recommend using Overlay Transport Virtualization (OTV), but you can use any method that ensures that the overlapping MAC Addresses and IP addresses of the gateway router do not leak between sites. Use a First Hop Redundancy Protocol (FHRP) such as HSRP to provide the same virtual MAC and IP addresses to the gateway routers.
BGP support for ASA clustering	9.3(1)	We added support for BGP with ASA clustering. We modified the following screen: **Configuration > Device Setup > Routing > BGP > IPv4 Family > General**.
Support for cluster members at different geographical locations (inter-site) for transparent mode	9.2(1)	You can now place cluster members at different geographical locations when using Spanned EtherChannel mode in transparent firewall mode. Inter-site clustering with spanned EtherChannels in routed firewall mode is not supported. We did not modify any ASDM screens.
Static LACP port priority support for clustering	9.2(1)	Some switches do not support dynamic port priority with LACP (active and standby links). You can now disable dynamic port priority to provide better compatibility with spanned EtherChannels. You should also follow these guidelines: • Network elements on the cluster control link path should not verify the L4 checksum. Redirected traffic over the cluster control link does not have a correct L4 checksum. Switches that verify the L4 checksum could cause traffic to be dropped. • Port-channel bundling downtime should not exceed the configured keepalive interval. We modified the following screen: **Configuration > Device Management > High Availability and Scalability > ASA Cluster**.
Support for 32 active links in a spanned EtherChannel	9.2(1)	ASA EtherChannels now support up to 16 active links. With *spanned* EtherChannels, that functionality is extended to support up to 32 active links across the cluster when used with two switches in a vPC and when you disable dynamic port priority. The switches must support EtherChannels with 16 active links, for example, the Cisco Nexus 7000 with F2-Series 10 Gigabit Ethernet Module. For switches in a VSS or vPC that support 8 active links, you can now configure 16 active links in the spanned EtherChannel (8 connected to each switch). Previously, the spanned EtherChannel only supported 8 active links and 8 standby links, even for use with a VSS/vPC. **Note** If you want to use more than 8 active links in a spanned EtherChannel, you cannot also have standby links; the support for 9 to 32 active links requires you to disable cLACP dynamic port priority that allows the use of standby links. We modified the following screen: **Configuration > Device Management > High Availability and Scalability > ASA Cluster**.

Feature Name	Version	Feature Information
Support for 16 cluster members for the ASA 5585-X	9.2(1)	The ASA 5585-X now supports 16-unit clusters. We did not modify any screens.
ASA 5500-X support for clustering	9.1(4)	The ASA 5512-X, ASA 5515-X, ASA 5525-X, ASA 5545-X, and ASA 5555-X now support 2-unit clusters. Clustering for 2 units is enabled by default in the base license; for the ASA 5512-X, you need the Security Plus license. We did not modify any ASDM screens.
Improved VSS and vPC support for health check monitoring	9.1(4)	If you configure the cluster control link as an EtherChannel (recommended), and it is connected to a VSS or vPC pair, you can now increase stability with health check monitoring. For some switches, such as the Cisco Nexus 5000, when one unit in the VSS/vPC is shutting down or booting up, EtherChannel member interfaces connected to that switch may appear to be Up to the ASA, but they are not passing traffic on the switch side. The ASA can be erroneously removed from the cluster if you set the ASA holdtime timeout to a low value (such as .8 seconds), and the ASA sends keepalive messages on one of these EtherChannel interfaces. When you enable the VSS/vPC health check feature, the ASA floods the keepalive messages on all EtherChannel interfaces in the cluster control link to ensure that at least one of the switches can receive them. We modified the following screen: **Configuration > Device Management > High Availability and Scalability > ASA Cluster**.
Support for cluster members at different geographical locations (inter-site); Individual Interface mode only	9.1(4)	You can now place cluster members at different geographical locations when using Individual Interface mode. We did not modify any ASDM screens.

Feature Name	Version	Feature Information
ASA Clustering for the ASA 5580 and 5585-X	9.0(1)	ASA Clustering lets you group up to 8 ASAs together as a single logical device. A cluster provides all the convenience of a single device (management, integration into a network) while achieving the increased throughput and redundancy of multiple devices. ASA clustering is supported for the ASA 5580 and the ASA 5585-X; all units in a cluster must be the same model with the same hardware specifications. See the configuration guide for a list of unsupported features when clustering is enabled. We introduced or modified the following screens: **Home > Device Dashboard** **Home > Cluster Dashboard** **Home > Cluster Firewall Dashboard** **Configuration > Device Management > Advanced > Address Pools > MAC Address Pools** **Configuration > Device Management > High Availability and Scalability > ASA Cluster** **Configuration > Device Management > Logging > Syslog Setup > Advanced** **Configuration > Device Setup > Interface Settings > Interfaces > Add/Edit Interface > Advanced** **Configuration > Device Setup > Interface Settings > Interfaces > Add/Edit Interface > IPv6** **Configuration > Device Setup > Interface Settings > Interfaces > Add/Edit EtherChannel Interface > Advanced** **Configuration > Firewall > Advanced > Per-Session NAT Rules** **Monitoring > ASA Cluster** **Monitoring > Properties > System Resources Graphs > Cluster Control Link** **Tools > Preferences > General** **Tools > System Reload** **Tools > Upgrade Software from Local Computer** **Wizards > High Availability and Scalability Wizard** **Wizards > Packet Capture Wizard** **Wizards > Startup Wizard**