Deploying a Cluster for ASA on the Firepower 4100/9300 for Scalability and High Availability

Clustering lets you group multiple ASA units 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. Clustering is also supported on ASA hardware models, but because the Firepower 4100/9300 requires separate configuration in FXOS, this document focuses on the entire configuration across FXOS and ASA.

Some features are not supported when using clustering. See Unsupported Features with Clustering, on page 48.

This document covers the latest ASA version features; see History for ASA Clustering on the Firepower 4100/9300 Chassis, on page 60 for details about feature changes. If you are on an old version of software, refer to the procedures in the FXOS configuration guide and ASA configuration guide for your version.

Benefit of this Integration

The FXOS platform lets you run multiple logical devices, including the ASA. Deploying standalone and clustered logical devices is easy for both intra-chassis clusters (for the Firepower 9300) and inter-chassis clusters. When you deploy a cluster from FXOS, you pre-configure the ASA bootstrap configuration so very little customization is required within the ASA application. You can also add additional cluster members by exporting the cluster configuration in FXOS.

Integrated Products

This table lists the products required for this integration.

<table>
<thead>
<tr>
<th>Products</th>
<th>Function</th>
<th>Minimum Version</th>
<th>Required?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firepower 4100 or 9300</td>
<td>Hardware platform to run the ASA</td>
<td>FXOS 1.1.2</td>
<td>Required</td>
</tr>
</tbody>
</table>
### Workflow

This workflow uses Firepower Chassis Manager on FXOS and ASDM on the ASA to complete your clustering deployment.

#### Procedure

**Step 1**  
FXOS prerequisites:  
- Configure Smart Licensing. Smart licensing requires you to configure an NTP server (or at least accurate manual time) and DNS.

**Step 2**  
FXOS tasks:  
- **a)** [FXOS: Configure Interfaces, on page 15](#). Configure one management and all data interfaces that you intend to assign to the ASA. The cluster interface is defined by default as Port-Channel 48, but for inter-chassis clustering, you need to add member interfaces.  
- **b)** [Create an ASA Cluster, on page 18](#).  
- **c)** [Add More Cluster Members, on page 23](#).

**Step 3**  
ASA tasks. Perform these tasks only on the master unit.  
- **a)** (Optional) Configure licenses for the Carrier and context features. See the [ASA general operations configuration guide](#).  
- **b)** (Optional) [ASA: Change the Firewall Mode and Context Mode, on page 24](#). By default, the FXOS chassis deploys a cluster in routed firewall mode, and single context mode.  
- **c)** [ASA: Configure Data Interfaces, on page 25](#). The management interface was pre-configured when you deployed the cluster.  
- **d)** (Optional) [ASA: Customize the Cluster Configuration, on page 27](#). Customize or enable many clustering features including inter-site features and distributed site-to-site VPN.

### About Clustering on the Firepower 4100/9300 Chassis

The cluster consists of multiple devices acting as a single logical unit. When you deploy a cluster on the Firepower 4100/9300 chassis, it does the following:  
- Creates a *cluster-control link* (by default, port-channel 48) for unit-to-unit communication. For intra-chassis clustering (Firepower 9300 only), this link utilizes the Firepower 9300 backplane for cluster...
communications. For inter-chassis clustering, you need to manually assign physical interface(s) to this EtherChannel for communications between chassis.

- Creates the cluster bootstrap configuration within the application.

When you deploy the cluster, the Firepower 4100/9300 chassis supervisor pushes a minimal bootstrap configuration to each unit that includes the cluster name, cluster control link interface, and other cluster settings. Some parts of the bootstrap configuration may be user-configurable within the application if you want to customize your clustering environment.

- Assigns data interfaces to the cluster as *Spanned* interfaces.

For intra-chassis clustering, spanned interfaces are not limited to EtherChannels, like it is for inter-chassis clustering. The Firepower 9300 supervisor uses EtherChannel technology internally to load-balance traffic to multiple modules on a shared interface, so any data interface type works for Spanned mode. For inter-chassis clustering, you must use Spanned EtherChannels for all data interfaces.

<table>
<thead>
<tr>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual interfaces are not supported, with the exception of a management interface.</td>
</tr>
</tbody>
</table>

- Assigns a management interface to all units in the cluster.

The following sections provide more detail about clustering concepts and implementation. See also Reference for Clustering, on page 48.

**Bootstrap Configuration**

When you deploy the cluster, the Firepower 4100/9300 chassis supervisor pushes a minimal bootstrap configuration to each unit that includes the cluster name, cluster control link interface, and other cluster settings. Some parts of the bootstrap configuration are user-configurable if you want to customize your clustering environment.

**Cluster Members**

Cluster members work together to accomplish the sharing of the security policy and traffic flows.

One member of the cluster is the *master* unit. The master unit is determined automatically. All other members are *slave* units.

You must perform all configuration on the master unit only; the configuration is then replicated to the slave units.

Some features do not scale in a cluster, and the master unit handles all traffic for those features. See Centralized Features for Clustering, on page 49.

**Master and Slave Unit Roles**

One member of the cluster is the master unit. The master unit is determined automatically. All other members are slave units.

You must perform all configuration on the master unit only; the configuration is then replicated to the slave units.
Some features do not scale in a cluster, and the master unit handles all traffic for those features. See Centralized Features for Clustering, on page 49.

### Cluster Control Link

The cluster-control link is an EtherChannel (port-channel 48) for unit-to-unit communication. For intra-chassis clustering, this link utilizes the Firepower 9300 backplane for cluster communications. For inter-chassis clustering, you need to manually assign physical interface(s) to this EtherChannel on the Firepower 4100/9300 chassis for communications between chassis.

For a 2-chassis inter-chassis cluster, do not directly-connect the cluster control link from one chassis to the other chassis. 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.

Cluster control link traffic includes both control and data traffic.

Control traffic includes:
- Master election.
- Configuration replication.
- Health monitoring.

Data traffic includes:
- State replication.
- Connection ownership queries and data packet forwarding.

### 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.

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 master 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 Network

The Firepower 4100/9300 chassis auto-generates the cluster control link interface IP address for each unit based on the chassis ID and slot ID: 127.2.chassis_id.slot_id. You can customize this IP address when you deploy the cluster. The cluster control link network cannot include any routers between units; only Layer 2 switching is allowed. For inter-site traffic, Cisco recommends using Overlay Transport Virtualization (OTV).

Cluster Interfaces

For intra-chassis clustering, you can assign both physical interfaces or EtherChannels (also known as port channels) to the cluster. Interfaces assigned to the cluster are Spanned interfaces that load-balance traffic across all members of the cluster.
For inter-chassis clustering, you can only assign data EtherChannels to the cluster. These Spanned EtherChannels include the same member interfaces on each chassis; on the upstream switch, all of these interfaces are included in a single EtherChannel, so the switch does not know that it is connected to multiple devices.

Individual interfaces are not supported, with the exception of a management interface.

**Connecting to a VSS or vPC**

We recommend connecting EtherChannels to a VSS or vPC to provide redundancy for your interfaces.

**Configuration Replication**

All units in the cluster share a single configuration. You can only make configuration changes on the master 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**

You must assign a Management type interface to the cluster. This interface is a special individual interface as opposed to a Spanned interface. The management interface lets you connect directly to each unit.

The Main cluster IP address is a fixed address for the cluster that always belongs to the current master unit. You also configure a range of addresses so that each unit, including the current master, can use a Local address from the range. The Main cluster IP address provides consistent management access to an address; when a master unit changes, the Main cluster IP address moves to the new master unit, so management of the cluster continues seamlessly.

For example, you can manage the cluster by connecting to the Main cluster IP address, which is always attached to the current master 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 master unit, uses the Local IP address to connect to the server.

**Master Unit Management Vs. Slave Unit Management**

All management and monitoring can take place on the master unit. From the master 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 slave units to the master unit.

You can monitor slave units directly if desired. Although also available from the master unit, you can perform file management on slave units (including backing up the configuration and updating images). The following functions are not available from the master 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 master unit, the key is replicated to all slave units. If you have an SSH session to the Main cluster IP address, you will be disconnected if the master unit fails. The new master 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 master 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 appears 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.

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.
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 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. 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:

- Configure Cluster Flow Mobility—Configure Cluster Flow Mobility, on page 32
- Enable Director Localization—Configure Basic ASA Cluster Parameters, on page 27
- Enable Site Redundancy—Configure Basic ASA Cluster Parameters, on page 27
Requirements and Prerequisites for Clustering on the Firepower 4100/9300 Chassis

Maximum Clustering Units Per Model

• Firepower 4100—16 chassis

• Firepower 9300—16 modules. For example, you can use 1 module in 16 chassis, or 2 modules in 8 chassis, or any combination that provides a maximum of 16 modules.

Hardware and Software Requirements for Inter-Chassis Clustering

All chassis in a cluster:

• For the Firepower 4100 series: All chassis must be the same model. For the Firepower 9300: All security modules must be the same type. For example, if you use clustering, all modules in the Firepower 9300 must be SM-40s. You can have different quantities of installed security modules in each chassis, although all modules present in the chassis must belong to the cluster including any empty slots.

• Must run the identical FXOS software except at the time of an image upgrade.

• Must include the same interface configuration for interfaces you assign to the cluster, such as the same Management interface, EtherChannels, active interfaces, speed and duplex, and so on. You can use different network module types on the chassis as long as the capacity matches for the same interface IDs and interfaces can successfully bundle in the same spanned EtherChannel. Note that all data interfaces must be EtherChannels in inter-chassis clustering. If you change the interfaces in FXOS after you enable clustering (by adding or removing interface modules, or configuring EtherChannels, for example), then perform the same changes on each chassis, starting with the slave units, and ending with the master. Note that if you remove an interface in FXOS, the ASA configuration retains the related commands so that you can make any necessary adjustments; removing an interface from the configuration can have wide effects. You can manually remove the old interface configuration.

• Must use the same NTP server. Do not set the time manually.

• ASA: Each FXOS chassis must be registered with the License Authority or satellite server. There is no extra cost for slave units. For permanent license reservation, you must purchase separate licenses for each chassis. For Firepower Threat Defense, all licensing is handled by the Firepower Management Center.

Switch Requirements

• Be sure to complete the switch configuration and successfully connect all the EtherChannels from the chassis to the switch(es) before you configure clustering on the Firepower 4100/9300 chassis.

• For supported switch characteristics, see Cisco FXOS Compatibility.
Licenses for Clustering on the Firepower 4100/9300 Chassis

Each Firepower 4100/9300 chassis must be registered with the License Authority or satellite server. There is no extra cost for slave units. For permanent license reservation, you must purchase separate licenses for each chassis.

The Strong Encryption license is automatically enabled for qualified customers when you apply the registration token. When using the token, each chassis must have the same encryption license. For the optional Strong Encryption (3DES/AES) feature license enabled in the ASA configuration, see below.

In the ASA license configuration, you can only configure smart licensing on the master unit. The configuration is replicated to the slave units, but for some licenses, they do not use the configuration; it remains in a cached state, and only the master unit requests the license. The licenses are aggregated into a single cluster license that is shared by the cluster units, and this aggregated license is also cached on the slave units to be used if one of them becomes the master unit in the future. Each license type is managed as follows:

- **Standard**—Only the master unit requests the Standard license from the server. Because the slave units have the Standard license enabled by default, they do not need to register with the server to use it.

- **Context**—Only the master unit requests the Context license from the server. The Standard license includes 10 contexts by default and is present on all cluster members. The value from each unit’s Standard license plus the value of the Context license on the master unit are combined up to the platform limit in an aggregated cluster license. For example:

  - You have 6 Firepower 9300 modules in the cluster. The Standard license includes 10 contexts; for 6 units, these licenses add up to 60 contexts. You configure an additional 20-Context license on the master unit. Therefore, the aggregated cluster license includes 80 contexts. Because the platform limit for one module is 250, the combined license allows a maximum of 250 contexts; the 80 contexts are within the limit. Therefore, you can configure up to 80 contexts on the master unit; each slave unit will also have 80 contexts through configuration replication.

  - You have 3 Firepower 4110 units in the cluster. The Standard license includes 10 contexts; for 3 units, these licenses add up to 30 contexts. You configure an additional 250-Context license on the master unit. Therefore, the aggregated cluster license includes 280 contexts. Because the platform limit for one unit is 250, the combined license allows a maximum of 250 contexts; the 280 contexts are over the limit. Therefore, you can only configure up to 250 contexts on the master unit; each slave unit will also have 250 contexts through configuration replication. In this case, you should only configure the master Context license to be 220 contexts.

- **Carrier**—Required for Distributed S2S VPN. This license is a per-unit entitlement, and each unit requests its own license from the server. This license configuration is replicated to the slave units.

- **Strong Encryption (3DES)** (for pre-2.3.0 Cisco Smart Software Manager satellite deployment, or for tracking purposes)—This license is a per-unit entitlement, and each unit requests its own license from the server.

If a new master unit is elected, the new master unit continues to use the aggregated license. It also uses the cached license configuration to re-request the master license. When the old master unit rejoins the cluster as a slave unit, it releases the master unit license entitlement. Before the slave unit releases the license, the master unit's license might be in a non-compliant state if there are no available licenses in the account. The retained license is valid for 30 days, but if it is still non-compliant after the grace period, you will not be able to make configuration changes to features requiring special licenses; operation is otherwise unaffected. The new active unit sends an entitlement authorization renewal request every 12 hours until the license is compliant. You
should refrain from making configuration changes until the license requests are completely processed. If a unit leaves the cluster, the cached master configuration is removed, while the per-unit entitlements are retained. In particular, you would need to re-request the Context license on non-cluster units.

**Licenses for Distributed S2S VPN**

A Carrier license is required for Distributed S2S VPN, on each member of the cluster.

Each VPN connection requires two Other VPN licensed sessions (the Other VPN license is part of the Base license), one for the active session and one for the backup session. The maximum VPN session capacity of the cluster can be no more than half of the licensed capacity due to using two licenses for each session.

**Clustering Guidelines and Limitations**

**Switches for Inter-Chassis Clustering**

- For the ASR 9006, if you want to set a non-default MTU, set the ASR 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 ASR IPv4 MTU.

- 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.

- When you see slow bundling of a Spanned EtherChannel on the switch, you can enable LACP rate fast for an individual interface on the switch. Note that some switches, such as the Nexus series, do not support LACP rate fast when performing in-service software upgrades (ISSUs), so we do not recommend using ISSUs with clustering.

- 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.

EtherChannels for Inter-Chassis Clustering

- FXOS EtherChannels have the LACP rate set to fast by default. Some switches, such as the Nexus series, do not support LACP rate fast when performing in-service software upgrades (ISSUs), so we do not recommend using ISSUs with clustering.

- 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 master 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 Clustering**

See the following guidelines for inter-site clustering:

- 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 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), 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 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 Firepower 4100/9300 chassis 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 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 connections; 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 interface, 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 cluster. These messages can result in some units of the cluster experiencing high CPU, which can affect performance. We recommend that you throttle ICMP error messages.

- We recommend connecting EtherChannels to a VSS or vPC for redundancy.

- Within a chassis, you cannot cluster some security modules and run other security modules in standalone mode; you must include all security modules in the cluster.
Defaults

• 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.

• Connection rebalancing is disabled by default. If you enable connection rebalancing, the default time between load information exchanges is 5 seconds.

• The cluster auto-rejoin feature for a failed cluster control link is set to unlimited attempts every 5 minutes.

• The cluster auto-rejoin feature for a failed data interface is set to 3 attempts every 5 minutes, with the increasing interval set to 2.

• Connection replication delay of 5 seconds is enabled by default for HTTP traffic.

Configure Clustering on the Firepower 4100/9300 Chassis

You can easily deploy the cluster from the Firepower 4100/9300 chassis supervisor. All initial configuration is automatically generated for each unit. This section describes the default bootstrap configuration and optional customization you can perform on the ASA. This section also describes how to manage cluster members from within the ASA. You can also manage cluster membership from the Firepower 4100/9300 chassis. See the Firepower 4100/9300 chassis documentation for more information.

Procedure

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>FXOS: Add an ASA Cluster, on page 18</td>
</tr>
<tr>
<td>Step 2</td>
<td>ASA: Change the Firewall Mode and Context Mode, on page 24</td>
</tr>
<tr>
<td>Step 3</td>
<td>ASA: Configure Data Interfaces, on page 25</td>
</tr>
<tr>
<td>Step 4</td>
<td>ASA: Customize the Cluster Configuration, on page 27</td>
</tr>
<tr>
<td>Step 5</td>
<td>ASA: Manage Cluster Members, on page 41</td>
</tr>
</tbody>
</table>

FXOS: Configure Interfaces

For a cluster, you need to configure the following types of interfaces:

• Add at least one Data type interface or EtherChannel (also known as a port-channel) before you deploy the cluster. See Add an EtherChannel (Port Channel), on page 17 or Configure a Physical Interface, on page 16.

For inter-chassis clustering, all data interfaces must be Spanned EtherChannels with at least one member interface. Add the same EtherChannels on each chassis. Combine the member interfaces from all cluster units into a single EtherChannel on the switch. See Clustering Guidelines and Limitations, on page 11 for more information about EtherChannels for inter-chassis clustering.

All interfaces are assigned to the cluster by default. For inter-chassis clustering, only EtherChannels are assigned; you cannot assign other interface types. You can also add data interfaces to the cluster after you deploy it.

• Add a Management type interface or EtherChannel. See Add an EtherChannel (Port Channel), on page 17 or Configure a Physical Interface, on page 16.
The management interface is required. Note that this management interface is not the same as the chassis management interface that is used only for chassis management (in FXOS, you might see the chassis management interface displayed as MGMT, management0, or other similar names).

For inter-chassis clustering, add the same Management interface on each chassis.

- For inter-chassis clustering, add a member interface to the cluster control link EtherChannel (by default, port-channel 48). See Add an EtherChannel (Port Channel), on page 17.

Do not add a member interface for intra-chassis clustering. If you add a member, the chassis assumes this cluster will be inter-chassis, and will only allow you to use Spanned EtherChannels, for example.

On the Interfaces tab, the port-channel 48 cluster type interface shows the Operation State as failed if it does not include any member interfaces. For intra-chassis clustering, this EtherChannel does not require any member interfaces, and you can ignore this Operational State.

Add the same member interfaces on each chassis. The cluster control link is a device-local EtherChannel on each chassis. Use separate EtherChannels on the switch per device. See Clustering Guidelines and Limitations, on page 11 for more information about EtherChannels for inter-chassis clustering.

**Configure a Physical Interface**

You can physically enable and disable interfaces, as well as set the interface speed and duplex. To use an interface, it must be physically enabled in FXOS and logically enabled in the application.

**Before you begin**

- Interfaces that are already a member of an EtherChannel cannot be modified individually. Be sure to configure settings before you add it to the EtherChannel.

**Procedure**

**Step 1** Choose Interfaces to open the Interfaces page.

The All Interfaces page shows a visual representation of the currently installed interfaces at the top of the page and provides a listing of the installed interfaces in the table below.

**Step 2** Click Edit in the row for the interface you want to edit to open the Edit Interface dialog box.

**Step 3** To enable the interface, check the Enable check box. To disable the interface, uncheck the Enable check box.

**Step 4** Choose the interface Type: Data, Mgmt, or Cluster.

Do not choose the Cluster type; by default, the cluster control link is automatically created on Port-channel 48.

**Step 5** (Optional) Choose the speed of the interface from the Speed drop-down list.

**Step 6** (Optional) If your interface supports Auto Negotiation, click the Yes or No radio button.

**Step 7** (Optional) Choose the duplex of the interface from the Duplex drop-down list.

**Step 8** Click OK.
Add an EtherChannel (Port Channel)

An EtherChannel (also known as a port channel) can include up to 16 member interfaces of the same type. The Link Aggregation Control Protocol (LACP) aggregates interfaces by exchanging the Link Aggregation Control Protocol Data Units (LACPDUs) between two network devices.

You can configure each physical Data interface in an EtherChannel to be:

- **Active**—Sends and receives LACP updates. An active EtherChannel can establish connectivity with either an active or a passive EtherChannel. You should use the active mode unless you need to minimize the amount of LACP traffic.

- **On**—The EtherChannel is always on, and LACP is not used. An “on” EtherChannel can only establish a connection with another “on” EtherChannel.

It may take up to three minutes for an EtherChannel to come up to an operational state if you change its mode from On to Active or from Active to On.

Non-data interfaces only support active mode.

LACP coordinates the automatic addition and deletion of links to the EtherChannel without user intervention. It also handles misconfigurations and checks that both ends of member interfaces are connected to the correct channel group. “On” mode cannot use standby interfaces in the channel group when an interface goes down, and the connectivity and configurations are not checked.

When the Firepower 4100/9300 chassis creates an EtherChannel, the EtherChannel stays in a **Suspended** state for Active LACP mode or a **Down** state for On LACP mode until you assign it to a logical device, even if the physical link is up. The EtherChannel will be brought out of this **Suspended** state in the following situations:

- The EtherChannel is added as a data or management interface for a standalone logical device
- The EtherChannel is added as a management interface or cluster control link for a logical device that is part of a cluster
- The EtherChannel is added as a data interface for a logical device that is part of a cluster and at least one unit has joined the cluster

Note that the EtherChannel does not come up until you assign it to a logical device. If the EtherChannel is removed from the logical device or the logical device is deleted, the EtherChannel will revert to a **Suspended** or **Down** state.

**Procedure**

**Step 1** Choose **Interfaces** to open the Interfaces page.

The **All Interfaces** page shows a visual representation of the currently installed interfaces at the top of the page and provides a listing of the installed interfaces in the table below.

**Step 2** Click **Add Port Channel** above the interfaces table to open the **Add Port Channel** dialog box.

**Step 3** Enter an ID for the port channel in the **Port Channel ID** field. Valid values are between 1 and 47.
Port-channel 48 is reserved for the cluster control link when you deploy a clustered logical device. If you do not want to use Port-channel 48 for the cluster control link, you can configure an EtherChannel with a different ID and choose the Cluster type for the interface. You can only add one cluster type EtherChannel. For intra-chassis clustering, do not assign any interfaces to the Cluster EtherChannel.

**Step 4** To enable the port channel, check the *Enable* check box. To disable the port channel, uncheck the *Enable* check box.

**Step 5** Choose the interface **Type**: Data, Mgmt, or Cluster.

You can only add one cluster type EtherChannel (by default, Port-channel 48).

**Step 6** Set the **Admin Speed** of the member interfaces from the drop-down list.

**Step 7** For Data interfaces, choose the LACP port-channel **Mode**: Active or On.

For non-Data interfaces, the mode is always active.

**Step 8** Set the **Admin Duplex**: Full Duplex or Half Duplex.

**Step 9** To add an interface to the port channel, select the interface in the **Available Interface** list and click **Add Interface** to move the interface to the Member ID list. You can add up to 16 interfaces of the same type and speed.

**Tip** You can add multiple interfaces at one time. To select multiple individual interfaces, click on the desired interfaces while holding down the Ctrl key. To select a range of interfaces, select the first interface in the range, and then, while holding down the Shift key, click to select the last interface in the range.

**Step 10** To remove an interface from the port channel, click the **Delete** button to the right of the interface in the Member ID list.

**Step 11** Click **OK**.

---

**FXOS: Add an ASA Cluster**

You can add a single Firepower 9300 chassis as an intra-chassis cluster, or add multiple chassis for inter-chassis clustering. For inter-chassis clustering, you must configure each chassis separately. Add the cluster on one chassis; you can then copy the bootstrap configuration from the first chassis to the next chassis for ease of deployment.

**Create an ASA Cluster**

You can easily deploy the cluster from the Firepower 4100/9300 chassis supervisor. All initial configuration is automatically generated for each unit.

For inter-chassis clustering, you must configure each chassis separately. Deploy the cluster on one chassis; you can then copy the bootstrap configuration from the first chassis to the next chassis for ease of deployment.

In a Firepower 9300 chassis, you must enable clustering for all 3 module slots, even if you do not have a module installed. If you do not configure all 3 modules, the cluster will not come up.

For multiple context mode, you must first deploy the logical device, and then enable multiple context mode in the ASA application.
When you deploy a cluster, the Firepower 4100/9300 chassis supervisor configures each ASA application with the following bootstrap configuration. You can later modify parts of the bootstrap configuration from the ASA, if desired (shown in **Bold** text).

```
interface Port-channel48
  description Clustering Interface
  cluster-group <service_type_name>
  key <secret>
    local-unit unit-<chassis#-module#>
    site-id <number>
    cluster-interface port-channel48 ip 127.2.<chassis#>.<module#> 255.255.255.0
    priority <auto>
    health-check holdtime 3
    health-check data-interface auto-rejoin 3 5 2
    health-check cluster-interface auto-rejoin unlimited 5 1
    enable
  ip local pool cluster_ipv4_pool <ip_address>-<ip_address> mask <mask>

interface <management_ifc>
  management-only individual
  nameif management
  security-level 0
  ip address <ip_address> <mask> cluster-pool cluster_ipv4_pool
  no shutdown

http server enable
http 0.0.0.0 0.0.0.0 management
route management <management_host_ip> <mask> <gateway_ip> 1
```

**Note**

The **local-unit** name can only be changed if you disable clustering.

**Before you begin**

- Download the application image you want to use for the logical device from Cisco.com, and then upload that image to the Firepower 4100/9300 chassis.
- Gather the following information:
  - Management interface ID, IP address, and network mask
  - Gateway IP address

**Procedure**

**Step 1** Configure interfaces. See FXOS: Configure Interfaces, on page 15.
**Step 2** Choose **Logical Devices**.
**Step 3** Click **Add > Cluster**, and set the following parameters:
Create an ASA Cluster

Add Cluster

- **I want to:** Create New Cluster
- **Device Name:** cluster1
- **Template:** Cisco Adaptive Security Appliance
- **Image Version:** 9.13.0.6
- **Instance Type:** Native

**OK**

You see the Provisioning - device name window. All interfaces are assigned to the cluster by default.

### Step 4
Click the device icon in the center of the screen.

A dialog box appears where you can configure initial bootstrap settings. These settings are meant for initial deployment only, or for disaster recovery. For normal operation, you can later change most values in the application CLI configuration.

### Step 5
On the **Cluster Information** page, complete the following.
a) For inter-chassis clustering, in the **Chassis ID** field, enter a chassis ID. Each chassis in the cluster must use a unique ID.

   This field only appears if you added a member interface to cluster control link Port-Channel 48.

b) For inter-site clustering, in the **Site ID** field, enter the site ID for this chassis between 1 and 8.

c) In the **Cluster Key** field, configure an authentication 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 key. This option does not affect datapath traffic, including connection state update and forwarded packets, which are always sent in the clear.

d) Set the **Cluster Group Name**, which is the cluster group name in the logical device configuration.

   The name must be an ASCII string from 1 to 38 characters.

e) Choose the **Management Interface**.
This interface is used to manage the logical device. This interface is separate from the chassis management port.

f) Choose the **Address Type** for the management interface.

This information is used to configure a management interface in the ASA configuration. Set the following information:

- **Management IP Pool**—Configure a pool of Local IP addresses, one of which will be assigned to each cluster unit for the interface, by entering the starting and ending addresses separated by a hyphen. Include at least as many addresses as there are units in the cluster. Note that for the Firepower 9300, you must include 3 addresses per chassis, even if you do not have all module slots filled. If you plan to expand the cluster, include additional addresses. The Virtual IP address (known as the Main cluster IP address) that belongs to the current master unit is *not* a part of this pool; be sure to reserve an IP address on the same network for the Main cluster IP address. You can use IPv4 and/or IPv6 addresses.

- **Network Mask** or **Prefix Length**

- **Network Gateway**

- **Virtual IP address**—Set the management IP address of the current master unit. This IP address must be on the same network as the cluster pool addresses, but not be part of the pool.

**Step 6**

On the **Settings** page, complete the following.

<table>
<thead>
<tr>
<th>Cisco: Adaptive Security Appliance - Bootstrap Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General Information</strong></td>
</tr>
<tr>
<td>Firewall Mode:</td>
</tr>
<tr>
<td>Password:</td>
</tr>
<tr>
<td>Confirm Password:</td>
</tr>
</tbody>
</table>

a) From the **Firewall Mode** drop-down list, choose **Transparent** or **Routed**.

In routed mode, the FTD is considered to be a router hop in the network. Each interface that you want to route between is on a different subnet. A transparent firewall, on the other hand, is a Layer 2 firewall that acts like a “bump in the wire,” or a “stealth firewall,” and is not seen as a router hop to connected devices.

The firewall mode is only set at initial deployment. If you re-apply the bootstrap settings, this setting is not used.

b) Enter and confirm a **Password** for the admin user and for the enable password.

The pre-configured ASA admin user is useful for password recovery; if you have FXOS access, you can reset the admin user password if you forget it.

**Step 7**

Click **OK** to close the configuration dialog box.

**Step 8**

Click **Save**.

The chassis deploys the logical device by downloading the specified software version and pushing the bootstrap configuration and management interface settings to the application instance. Check the **Logical Devices** page for the status of the new logical device. When the logical device shows its **Status** as **online**, you can start configuring the security policy in the application.
Step 9

For inter-chassis clustering, add the next chassis to the cluster:

a) On the first chassis Firepower Chassis Manager, click the Show Configuration icon ( ) at the top right; copy the displayed cluster configuration.
b) Connect to the Firepower Chassis Manager on the next chassis, and add a logical device according to this procedure.
c) Choose I want to: > Join an Existing Cluster.
d) Click OK.
e) In the Copy Cluster Details box, paste in the cluster configuration from the first chassis, and click OK.
f) Click the device icon in the center of the screen. The cluster information is mostly pre-filled, but you must change the following settings:
   • Chassis ID—Enter a unique chassis ID.
   • Site ID—Enter the correct site ID.
   • Cluster Key—(Not prefilled) Enter the same cluster key.
   Click OK.
g) Click Save.

Step 10

Connect to the master unit ASA to customize your clustering configuration.

Add More Cluster Members

Add or replace an ASA cluster member.

Note

This procedure only applies to adding or replacing a chassis; if you are adding or replacing a module to a Firepower 9300 where clustering is already enabled, the module will be added automatically.
Before you begin

- Make sure your existing cluster has enough IP addresses in the management IP address pool for this new member. If not, you need to edit the existing cluster bootstrap configuration on each chassis before you add this new member. This change causes a restart of the logical device.

- The interface configuration must be the same on the new chassis. You can export and import FXOS chassis configuration to make this process easier.

- For multiple context mode, enable multiple context mode in the ASA application on the first cluster member; additional cluster members will inherit the multiple context mode configuration automatically.

Procedure

**Step 1** On an existing cluster chassis Firepower Chassis Manager, choose **Logical Devices** to open the **Logical Devices** page.

**Step 2** Click the **Show Configuration** icon ( ) at the top right; copy the displayed cluster configuration.

**Step 3** Connect to the Firepower Chassis Manager on the new chassis, and click **Add > Cluster**.

**Step 4** Choose **I want to:** > **Join Existing Cluster**

**Step 5** For the **Device Name**, provide a name for the logical device.

**Step 6** Click **OK**.

**Step 7** In the **Copy Cluster Details** box, paste in the cluster configuration from the first chassis, and click **OK**.

**Step 8** Click the device icon in the center of the screen. The cluster information is mostly pre-filled, but you must change the following settings:

- **Chassis ID**—Enter a unique chassis ID.
- **Site ID**—Enter the correct site ID.
- **Cluster Key**—(Not prefilled) Enter the same cluster key.

Click **OK**.

**Step 9** Click **Save**.

ASA: Change the Firewall Mode and Context Mode

By default, the FXOS chassis deploys a cluster in routed firewall mode, and single context mode.
• Change the firewall mode—To change the mode after you deploy, change the mode on the master unit; the mode is automatically changed on all slave units to match. In multiple context mode, you set the firewall mode per context. See the ASA general operations configuration guide.

• Change to multiple context mode—To change to multiple context mode after you deploy, change the mode on the master unit; the mode is automatically changed on all slave units to match. See the ASA general operations configuration guide.

ASA: Configure Data Interfaces

This procedure configures basic parameters for each data interface that you assigned to the cluster when you deployed it in FXOS. For inter-chassis clustering, data interfaces are always Spanned EtherChannel interfaces.

Note

The management interface was pre-configured when you deployed the cluster. You can also change the management interface parameters in ASA, but this procedure focuses on data interfaces. The management interface is an individual interface, as opposed to a Spanned interface. See Management Interface, on page 6 for more information.

Before you begin

• 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.

• When using Spanned EtherChannels for inter-chassis clustering, 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

Select the interface, and click Edit. The Edit Interface dialog box appears.

Step 3

Set the following:

• (For EtherChannels) MIO Port-channel ID—Enter the same ID used in FXOS.

• Enable Interface (checked by default)

The rest of the fields on this screen are described later in this procedure.
Step 4

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 master unit. If you do not configure a MAC address, then if the master unit changes, the new master 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.

• In the ASA Cluster area, for inter-site clustering set Site specific MAC Addresses, and for routed mode, the 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.

Step 5

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

Step 6

(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 7

Click the General tab.

Step 8

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

Step 9

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

Step 10

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

Step 11

(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 12

(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 13**  
Click **OK** to return to the **Interfaces** screen.

**Step 14**  
Click **Apply**.

---

**ASA: Customize the Cluster Configuration**

If you want to change bootstrap settings after you deploy the cluster or configure additional options, such as clustering health monitoring, TCP connection replication delay, flow mobility, and other optimizations, you can do so on the master unit.

**Configure Basic ASA Cluster Parameters**

You can customize cluster settings on the master unit.

**Before you begin**

- For multiple context mode, complete this procedure in the system execution space on the master 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.

- The local-unit **Member Name** and several other options can only be set on the FXOS chassis, or they can only be changed on the ASA if you disable clustering, so they are not included in the following procedure.

**Procedure**

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

**Step 2**  
(Optional) Configure the following optional parameters:

- **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.

  - **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 master unit to the slave units.

- **Enable cluster load monitor**—You can monitor the traffic load for cluster members, including total connection count, CPU and memory usage, and buffer drops. If the load is too high, you can choose to manually disable clustering on the unit if the remaining units can handle the load, or adjust the load balancing on the external switch. This feature is enabled by default. For example, for inter-chassis clustering on the Firepower 9300 with 3 security modules in each chassis, if 2 security modules in a chassis leave the cluster, then the same amount of traffic to the chassis will be sent to the remaining module and potentially overwhelm it. You can periodically monitor the traffic load. If the load is too high, you can choose to manually disable clustering on the unit.

Set the following values:

- **Time Interval**—Sets the time in seconds between monitoring messages, between 10 and 360 seconds. The default is 20 seconds.

- **Number of Intervals**—Sets the number of intervals for which the ASA maintains data, between 1 and 60. The default is 30.

See Monitoring > ASA Cluster > Cluster Load-Monitoring to view the traffic load.

- **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.

- **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.

- **Replicate console output to the master’s console**—Enables console replication from slave units to the master 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, slave units send the console messages to the master 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 master unit to the slave units.

- **Enable Clustering Flow Mobility**. See Configure LISP Inspection, on page 33.

- **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.

- **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.

- **Enable config sync acceleration**—When a slave unit has the same configuration as the master unit, it will skip syncing the configuration and will join faster. This feature is enabled by default. This feature is configured on each unit, and is not replicated from the master to the slave.

**Note** Some configuration commands are not compatible with accelerated cluster joining; if these commands are present on the unit, even if accelerated cluster joining is enabled, configuration syncing will always occur. You must remove the incompatible configuration for accelerated cluster joining to work. Use the `show cluster info unit-join-acceleration incompatible-config` to view incompatible configuration.

**Step 3** In the **Cluster Control Link** area, you can configure the cluster control link MTU. Other options in this area cannot be configured on the ASA.

- **MTU**—Specifies the maximum transmission unit for the cluster control link interface. We suggest setting the MTU to the maximum of 9184; the minimum value is 1400 bytes.

**Step 4** (Optional) In the **Cluster LACP** area, you can enable static port priority. 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. Other options in this area cannot be configured on the ASA without disabling clustering.

- **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 master unit to the slave units.

**Step 5** (Optional) (Firepower 9300 only) In the **Parallel Join of Units Per Chassis** area, you can ensure that the security modules in a chassis join the cluster simultaneously, so that traffic is evenly distributed between the modules. If a module joins very much in advance of other modules, it can receive more traffic than desired, because the other modules cannot yet share the load.

- **Minimum Units Required to Join**—Specifies the minimum number of modules in the same chassis required to be ready before a module can join the cluster, between 1 and 3. The default is 1, meaning that a module will not wait for other modules to be ready before it joins the cluster. If you set the value to 3, for example, then each module will wait the maximum delay time or until all 3 modules are ready before joining the cluster. All 3 modules will request to join the cluster roughly simultaneously, and will all start receiving traffic around the same time.

- **Maximum Join Delay**—Specifies the maximum delay time in minutes before a module stops waiting for other modules to be ready before it joins the cluster, between 0 and 30 minutes. The default is 0, meaning the module will not wait for other modules to be ready before it joins the cluster. If you set the
minimum units to 1, then this value must be 0. If you set the minimum units to 2 or 3, then this value must be 1 or more. This timer is per module, but when the first module joins the cluster, then all other module timers end, and the remaining modules join the cluster.

For example, you set the minimum units to 3, and the maximum delay to 5 minutes. When module 1 comes up, it starts its 5 minute timer. Module 2 comes up 2 minutes later and starts its 5 minute timer. Module 3 comes up 1 minute later, therefore all modules will now join the cluster at the 4 minute mark; they will not wait for the timers to complete. If module 3 never comes up, then Module 1 will join the cluster at the end of its 5 minute timer, and Module 2 will also join, even though its timer still has 2 minutes remaining; it will not wait for its timer to complete.

Step 6  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 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.

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 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.

Procedure

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

Step 2  Select an interface in the Monitored Interfaces box, 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 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, Firepower 4100/9300 chassis, 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  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.

### 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.

- **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
Configure LISP Inspection

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

**Before you begin**

- Set the site ID for the chassis on Firepower 4100/9300 chassis supervisor.
- 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.

   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.

Configure Distributed Site-to-Site VPN

By default, the ASA cluster uses Centralized Site-to-Site VPN mode. To take advantage of the scalability of clustering, you can enable Distributed Site-to-Site VPN mode. In this mode, S2S IPsec IKEv2 VPN connections are distributed across members of an ASA cluster. Distributing VPN connections across the members of a cluster allows both the capacity and throughput of the cluster to be fully utilized, significantly scaling VPN support beyond Centralized VPN capabilities.
About Distributed Site-to-Site VPN

Distributed VPN Connection Roles

When running in Distributed VPN mode the following roles are assigned to the cluster members:

- **Active Session Owner** — The unit that initially receives the connection, or that has transitioned a backup session to an active session. The owner maintains state and processes packets for the complete session, including the IKE and IPsec tunnels and all traffic associated with them.

- **Backup Session Owner** — The unit that is handling the backup session for an existing active session. Depending on the backup strategy chosen, this may be a unit in the same chassis as the active session owner, or a unit in another chassis. If the active session owner fails, the backup session owner becomes the active session owner, and a new backup session is established on a different unit.

- **Forwarder** — If traffic associated with a VPN session is sent to a unit that does not own the VPN session, that unit will use the Cluster Control Link (CCL) to forward the traffic to the member which owns the VPN session.

- **Orchestrator** — The orchestrator (always the master node of the cluster) is responsible for calculating which sessions will move and where to when executing an Active Session Redistribution (ASR). It sends a request to the owner member X to move N sessions to member Y. Member X will respond back to the orchestrator when complete, specifying how many sessions it was able to move.

Distributed VPN Session Characteristics

Distributed S2S VPN Sessions have the following characteristics. Otherwise, VPN connections behave as they normally do if not on an ASA cluster.

- VPN sessions are distributed across the cluster at the session level. Meaning the same cluster member handles the IKE and IPsec tunnels, and all their traffic, for a VPN connection. If VPN session traffic is sent to a cluster member that does not own that VPN session, traffic is forwarded to the cluster member that owns the VPN session.

- VPN sessions have a Session ID that is unique across the cluster. Using the session ID, traffic is validated, forwarding decisions are made, and IKE negotiation is completed.

- In an S2S VPN hub and spoke configuration, when clients connect through the ASA cluster (called hair-pinning), the session traffic flowing in and the session traffic flowing out may be on different cluster members.

- You can require that the backup session to be allocated on a security module in another chassis; this provides protection against chassis failure. Or, you can choose to allocate backup sessions on any node in the cluster; this provides protection against node failure only. When there are two chassis in the cluster, remote-chassis backup is strongly recommended.

- Only IKEv2 IPsec S2S VPN is supported in Distributed S2S VPN mode, IKEv1 is not. IKEv1 S2S is supported in centralized VPN mode.

- Each security module supports up to 6K VPN sessions for a maximum of approximately 36K sessions across 6 members. The actual number of sessions supported on a cluster member is determined by platform capacity, allocated licenses, and per context resource allocation. When utilization is close to the limit, there may be cases where session creation fails, even though the maximum capacity has not been reached on each cluster unit. This is because active session allocation is determined by external switching, and
backup session allocation is determined by an internal cluster algorithm. Customers are encouraged to size their utilization accordingly and allow room for uneven distribution.

**Distributed VPN Handling of Cluster Events**

Table 2:

<table>
<thead>
<tr>
<th>Event</th>
<th>Distributed VPN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Member failure</td>
<td>For all active sessions on this failed member, the backup sessions (on another member) become active and backup sessions are reallocated on another unit according to the backup strategy.</td>
</tr>
<tr>
<td>Chassis failure</td>
<td>When a remote-chassis backup strategy is being used, for all active sessions on the failed chassis, the backup sessions (on a member in the other chassis) become active. When the units are replaced, backup sessions for these now active sessions will be reallocated on members in the replaced chassis. When a flat backup strategy is being used, if both the active and backup sessions are on the failed chassis, the connection will drop. All active sessions with backup sessions on a member in the other chassis, fallback to these sessions. New backup sessions will be allocated on another member in the surviving chassis.</td>
</tr>
<tr>
<td>Inactivate a cluster member</td>
<td>For all active sessions on the cluster member being inactivated, backup sessions (on another member) become active and reallocate backup sessions on another unit according to the backup strategy.</td>
</tr>
<tr>
<td>Cluster member join</td>
<td>If the VPN cluster mode is not set to distributed, the master unit will request a mode change. If, or once the VPN mode is compatible, the cluster member will be assigned active and backup sessions in the flow of normal operations.</td>
</tr>
</tbody>
</table>

**Unsupported Inspections**

The following types of inspections are not supported or are disabled in Distributed S2S VPN mode:

- CTIQBE
- DCERPC
- H323, H225, and RAS
- IPsec pass-through
- MGCP
- MMP
- NetBIOS
- PPTP
- RADIUS
- RSH
IPsecIKEv2 Modifications

IKEv2 is modified while in Distributed S2S VPN mode in the following ways:

- An identity is used in place of IP/port tuples. This allows for proper forwarding decisions on the packets, and cleanup of previous connections that may be on other cluster members.

- The (SPI) identifiers that identify a single IKEv2 session are locally generated, random 8-byte values that are unique across the cluster. An SPI embeds a time stamp and a cluster member ID. Upon receipt of an IKE negotiation packet, if the time stamp or cluster member ID check fails, the packet is dropped and a message is logged indicating the reason.

- IKEv2 processing has been modified to prevent NAT-T negotiations from failing by being split across cluster members. A new ASP classify domain, `cluster_isakmp_redirect`, and rules are added when IKEv2 is enabled on an interface.

Model Support

The only device supported for Distributed VPN is the Firepower 9300. Distributed VPN supports a maximum of 6 modules on up to 2 chassis. You can have different quantities of installed security modules in each chassis, although we recommend an equal distribution.

Inter-site clustering is not supported.

Firewall Mode

Distributed S2S VPN is supported in routed mode only.

Context Mode

Distributed S2S VPN operates in both single and multiple context modes. However, in multiple context mode, active session redistribution is done at the system level, not at the context level. This prevents an active session associated with a context from moving to a cluster member that contains active sessions associated with a different context, unknowingly creating an unsupportable load.

High Availability

The following capabilities provide resiliency against single failure of a security module or chassis:

- VPN Sessions that are backed up on another security module in the cluster, on any chassis, withstand security module failures.
• VPN Sessions that are backed up on another chassis withstand chassis failures.

• The cluster master can change without losing VPN S2S sessions.

If an additional failure occurs before the cluster has stabilized, connections may be lost if the both active and backup sessions are on the failed units.

All attempts are made to ensure no sessions are lost when a member leaves the cluster in a graceful manner such as disabling the VPN cluster mode, reloading a cluster member, and other anticipated chassis changes. During these types of operations, sessions will not be lost as long as the cluster is given time to re-establish session backups between operations. If a graceful exit is triggered on the last cluster member, it will gracefully tear down existing sessions.

Dynamic PAT

Is not available while in Distributed VPN mode.

CMPv2

The CMPv2 ID certificate and key pairs are synchronized across the cluster members. However, only the master in the cluster automatically renews and rekeys the CMPv2 certificate. The master synchronizes these new ID certificates and keys to all cluster members on a renewal. In this way, all members in the cluster utilize the CMPv2 certificates for authentication, and also any member is capable of taking over as master.

Enable Distributed S2S VPN

Enable Distributed Site-to-Site VPN to take advantage of the scalability of clustering for VPN sessions.

**Note**

Changing the VPN mode between centralized and distributed causes all existing sessions to be torn down. Changing the backup mode is dynamic and will not terminate sessions.

**Before you begin**

• You must have a Carrier License configured on all members of the cluster.

• Your S2S VPN configuration must be set.

**Procedure**

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

**Step 2** In the VPN Cluster Mode area, choose the VPN Mode for the cluster, Centralized or Distributed.

**Step 3** Choose the Backup Distribution Mode, Flat or Remote-chassis.

In flat backup mode, standby sessions are established on any other cluster member. This will protect users from blade failures, however, chassis failure protection is not guaranteed.

In remote-chassis backup mode standby sessions are established on a member of another chassis in the cluster. This will protect users from both blade failures and chassis failures.
If remote-chassis is configured in a single chassis environment (intentionally configured or the result of a failure), no backups will be created until another chassis joins.

---

**Redistribute Distributed S2S VPN Sessions**

Active Session Redistribution (ASR) redistributes the active VPN session load across the cluster members. Due to the dynamic nature of beginning and ending sessions, ASR is a best effort balancing of the sessions across all cluster members. Repeated redistribution actions will optimize the balance.

Redistribution can be run at any time, should be run after any topology change in the cluster, and is recommended after a new member joins the cluster. The goal of redistribution is to create a stable VPN cluster. A stable VPN cluster has an almost equal number of active and backup sessions across the nodes.

To move a session, the backup session becomes the active one and another node is selected to host a new backup session. Moving sessions is dependent on the location of the active session's backup and the number of active sessions already on that particular backup node. If the backup session node is unable to host the active session for some reason, the original node remains owner of the session.

In multiple-context mode, active session redistribution is done at the system level, not the individual context level. It is not done at the context level because an active session in one context could be moved to a member that contains many more active sessions in a different context, creating more load on that cluster member.

**Before you begin**

- Enable system logs if you would like to monitor redistribution activity.
- This procedure must be carried out on the master node of the cluster.

**Procedure**

**Step 1** Choose Monitoring > ASA Cluster > ASA Cluster > Cluster Summary > VPN Cluster Summary to view how active and backup sessions are distributed across the cluster.

Depending on the number of sessions to redistribute and the load on the cluster, this may take some time. Syslogs containing the following phrases (and other system details not shown here) are provided as redistribution activity occurs:

<table>
<thead>
<tr>
<th>Syslog Phrase</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>VPN session redistribution started</td>
<td>Master only</td>
</tr>
<tr>
<td>Sent request to move number sessions from orig-member-name to dest-member-name</td>
<td>Master only</td>
</tr>
<tr>
<td>Failed to send session redistribution message to member-name</td>
<td>Master only</td>
</tr>
<tr>
<td>Received request to move number sessions from orig-member-name to dest-member-name</td>
<td>Slave only</td>
</tr>
<tr>
<td>Moved number sessions to member-name</td>
<td>The number of active sessions moved to the named cluster.</td>
</tr>
<tr>
<td>Failed to receive session move response from dest-member-name</td>
<td>Master only</td>
</tr>
</tbody>
</table>
### FXOS: Remove a Cluster Member

The following sections describe how to remove members temporarily or permanently from the cluster.

#### Temporary Removal

A cluster member will be automatically removed from the cluster due to a hardware or network failure, for example. This removal is temporary until the conditions are rectified, and it can rejoin the cluster. You can also manually disable clustering.

To check whether a device is currently in the cluster, check the cluster status on the Firepower Chassis Manager Logical Devices page:

- Management Port: Ethernet1/4
- Status: online
- Cluster Operational Status: not-in-cluster
- FIREPOWER-MGMT-IP: 10.89.5.20
- CLUSTER-ROLE: none
- CLUSTER-IP: 127.2.1.1
- MGMT-URL: https://10.89.5.35/
- UUID: 8e459170-451d-11e9-6475-f2f68c32630

- Disable clustering in the application—You can disable clustering using the application CLI. Enter the `cluster remove unit name` command to remove any unit other than the one you are logged into. The bootstrap configuration remains intact, as well as the last configuration synced from the master unit, so you can later re-add the unit without losing your configuration. If you enter this command on a slave unit to remove the master unit, a new master unit is elected.

When a device becomes inactive, all data interfaces are shut down; only the Management 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 bootstrap configuration. However if you reload, and the unit is still inactive in the cluster (for example, you saved the configuration with clustering disabled), the Management interface is disabled.
To reenable clustering, on the ASA enter `cluster group name` and then `enable`.

- Disable the application instance—In Firepower Chassis Manager on the **Logical Devices** page, click the Disable slider (✓). You can later reenable it using the Enable slider (✓).

- Shut down the security module/engine—In Firepower Chassis Manager on the **Security Module/Engine** page, click the Power off icon (●).

- Shut down the chassis—In Firepower Chassis Manager on the **Overview** page, click the Shut down icon (○).

**Permanent Removal**

You can permanently remove a cluster member using the following methods.

- Delete the logical device—In Firepower Chassis Manager on the **Logical Devices** page, click the delete icon (🗑). You can then deploy a standalone logical device, a new cluster, or even add a new logical device to the same cluster.

- Remove the chassis or security module from service—If you remove a device from service, you can add replacement hardware as a new member of the cluster.

**ASA: Manage Cluster Members**

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

**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 Slave Member from the Master Unit

To deactivate a slave member, 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 slave that you want to remove, and click Delete.

The slave bootstrap configuration remains intact, so that you can later re-add the slave 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 slave unit from the master 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

Change the Master Unit

Caution

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

To change the master 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 **Change Master To** drop-down list, choose a slave unit to become master, and click **Make Master**.

**Step 3**  
You are prompted to confirm the master 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. (Note that alternatively there are show commands that you can enter on the master unit to view cluster-wide statistics.) 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:**

`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 master unit:

`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 memory` command shows memory information for each member in the cluster:

```plaintext
cluster exec show memory
unit=1-1(LOCAL):****************************************************
Free memory: 108724634538 bytes (92%)
Used memory: 9410087158 bytes ( 8%)
------------- ------------------
Total memory: 118111600640 bytes (100%)

unit=1-3:******************************************************************************
Free memory: 108749922170 bytes (92%)
Used memory: 9371097334 bytes ( 8%)
------------- ------------------
Total memory: 118111600640 bytes (100%)

unit=1-2:******************************************************************************
Free memory: 108426753537 bytes (92%)
Used memory: 9697869087 bytes ( 8%)
------------- ------------------
Total memory: 118111600640 bytes (100%)
```

ASA: Monitoring the ASA Cluster on the Firepower 4100/9300 chassis

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 master unit, which is then automatically enabled on all of the slave 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 > ASA Cluster > Cluster Load-Monitoring**
  This section includes the Load Monitor-Information and Load-Monitor Details panes. Load Monitor-Information shows the traffic load for cluster members for the last interval and also the average over total number of intervals configured (30 by default). Use the Load-Monitor Details pane to view the value for each measure at each interval.

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.

Monitoring Distributed S2S VPN

See the following screens for monitoring VPN cluster status:

- **Monitoring > ASA Cluster > ASA Cluster > Cluster Summary > VPN Cluster Summary**
  Displays the distribution of the session across the cluster and provides you with the ability to re-distribute the sessions.
Monitoring > VPN > VPN Statistics > Sessions

Both master and slave cluster members are listed. Click any member for details.

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.

Troubleshooting Distributed S2S VPN

Distributed VPN Notifications

You will be notified with messages containing the identified phrases when the following error situations occur on a cluster running distributed VPN:

<table>
<thead>
<tr>
<th>Situation</th>
<th>Notification</th>
</tr>
</thead>
<tbody>
<tr>
<td>If an existing or joining cluster slave is not in distributed VPN mode when attempting to join the cluster:</td>
<td>New cluster member <em>(member-name)</em> rejected due to vpn mode mismatch.</td>
</tr>
<tr>
<td></td>
<td>and</td>
</tr>
<tr>
<td></td>
<td>Master <em>(master-name)</em> rejects enrollment request from unit <em>(unit-name)</em> for the reason: the vpn mode capabilities are not compatible with the master configuration</td>
</tr>
<tr>
<td>If licensing is not properly configured on a cluster member for Distributed VPN:</td>
<td>ERROR: Master requested cluster vpn-mode change to distributed. Unable to change mode due to missing Carrier License.</td>
</tr>
<tr>
<td>If the time stamp or member ID is invalid in the SPI of a received IKEv2 packet:</td>
<td>Expired SPI received or Corrupted SPI detected</td>
</tr>
<tr>
<td>If the cluster is unable to create a backup session:</td>
<td>Failed to create the backup for an IKEv2 session.</td>
</tr>
<tr>
<td>IKEv2 Initial Contact (IC) processing error:</td>
<td>IKEv2 Negotiation aborted due to ERROR: Stale backup session found on backup</td>
</tr>
<tr>
<td>Redistribution problems:</td>
<td>Failed to send session redistribution message to <em>(member-name)</em></td>
</tr>
<tr>
<td></td>
<td>Failed to receive session move response from <em>(member-name)</em> (master only)</td>
</tr>
</tbody>
</table>
NotificationSituation

Cluster topology change detected. VPN session redistribution aborted.

If the topology changes during redistribution of the sessions:

You may be encountering one of the following situations:

- L2L VPN sessions are being distributed to only one of the chassis in a cluster when the N7K Switch is configured with L4port as a load balancing algorithm using the `port-channel load-balance src-dst l4port` command. An example of the cluster session allocation looks like below:

```
SSP-Cluster/slave(cfg-cluster)# show cluster vpn-sessiondb distribution
Member 0 (unit-1-3): active: 0
Member 1 (unit-2-2): active: 13295; backups at: 0(2536), 2(2769), 3(2495), 4(2835), 5(2660)
Member 2 (unit-2-3): active: 12174; backups at: 0(2074), 1(2687), 3(2207), 4(3084), 5(2122)
Member 3 (unit-2-1): active: 13416; backups at: 0(2419), 1(3013), 2(2712), 4(2771), 5(2501)
Member 4 (unit-1-1): active: 0
Member 5 (unit-1-2): active: 0
```

Since L2L IKEv2 VPN uses port 500 for both source and destination ports, IKE packets are only sent to one of the links in the port channel connected between the N7K and the chassis.

Change the N7K load balancing algorithm to IP and L4 port using the `port-channel load-balance src-dst ip-l4port` command. Then the IKE packets are sent to all the links and thus both Firepower9300 chassis.

For a more immediate adjustment, on the master of the ASA cluster execute: `cluster redistribute vpn-sessiondb` to redistribute active VPN sessions to the cluster members of the other chassis.

---

**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 master 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)
- IS-IS routing
- The following application inspections:
  - CTIQBE
  - H323, H225, and RAS
Centralized Features for Clustering

The following features are only supported on the master unit, and are not scaled for the cluster.

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

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

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

- The following application inspections:
  - DCERPC
  - NetBIOS
  - PPTP
  - RADIUS
  - RSH
  - SUNRPC
  - TFTPC
  - XDMCP
Features Applied to Individual Units

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

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 master with replication of the data structures to the cluster slaves. If a master is elected, the new master 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 master 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.

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 master unit.

### Identity Firewall and Clustering

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

### Multicast Routing and Clustering

The master unit handles all multicast routing packets and data packets until fast-path forwarding is established. After the connection is established, each slave can forward multicast data packets.

### 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 PAT with Port Block Allocation—This feature is not supported for the cluster.
- 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. But it will not use these blocks 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 master unit evenly pre-distributes addresses across the cluster. If a member receives a connection and they have no addresses left, then the connection is dropped even if other members still have addresses available. Make sure to include at least as many NAT addresses as there are units in the cluster to ensure that each unit receives an address. Use the `show nat pool cluster` command to see the address allocations.
• No round-robin—Round-robin for a PAT pool is not supported with clustering.

• Dynamic NAT xlates managed by the master unit—The master unit maintains and replicates the xlate table to slave units. When a slave unit receives a connection that requires dynamic NAT, and the xlate is not in the table, it requests the xlate from the master unit. The slave unit owns the connection.

• Per-session PAT feature—Although not exclusive to clustering, the per-session PAT feature improves the scalability of PAT and, for clustering, allows each slave unit to own PAT connections; by contrast, multi-session PAT connections have to be forwarded to and owned by the master 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

Dynamic Routing and Clustering

The routing process only runs on the master unit, and routes are learned through the master unit and replicated to secondaries. If a routing packet arrives at a slave, it is redirected to the master unit.

Figure 1: Dynamic Routing

After the slave members learn the routes from the master unit, each unit makes forwarding decisions independently.
The OSPF LSA database is not synchronized from the master unit to slave units. If there is a master unit switchover, the neighboring router will detect a restarting router; 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.

**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 master is elected, the poll to the new master unit will fail.

**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 master unit learns security group tag (SGT) information. The master unit then populates the SGT to slaves, and slaves can make a match decision for SGT based on the security policy.

**VPN and Clustering on the FXOS Chassis**

An ASA FXOS Cluster supports one of two mutually exclusive modes for S2S VPN, centralized or distributed:

- **Centralized VPN Mode**. The default mode. In centralized mode, VPN connections are established with the master of the cluster only.
VPN functionality is limited to the master unit and does not take advantage of the cluster high availability capabilities. If the master unit fails, all existing VPN connections are lost, and VPN connected users see a disruption in service. When a new master is elected, you must reestablish the VPN connections.

When you connect a VPN tunnel to a Spanned interface address, connections are automatically forwarded to the master unit. VPN-related keys and certificates are replicated to all units.

- **Distributed VPN Mode.** In this mode, S2S IPsec IKEv2 VPN connections are distributed across members of an ASA cluster providing scalability. Distributing VPN connections across the members of a cluster allows both the capacity and throughput of the cluster to be fully utilized, significantly scaling VPN support beyond Centralized VPN capabilities.

---

**Note**

Centralized VPN clustering mode supports S2S IKEv1 and S2S IKEv2.

Distributed VPN clustering mode supports S2S IKEv2 only.

Distributed VPN clustering mode is supported on the Firepower 9300 only.

Remote access VPN is not supported in centralized or distributed VPN clustering mode.

---

**Performance Scaling Factor**

When you combine multiple units into a cluster, you can expect the total cluster performance to be approximately:

- 80% of the combined TCP or CPS throughput
- 90% of the combined UDP throughput
- 60% of the combined Ethernet MIX (EMIX) throughput, depending on the traffic mix.

For example, for TCP throughput, the Firepower 9300 with 3 modules can handle approximately 135 Gbps of real world firewall traffic when running alone. For 2 chassis, the maximum combined throughput will be approximately 80% of 270 Gbps (2 chassis x 135 Gbps): 216 Gbps.

---

**Master Unit Election**

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

1. When you deploy the cluster, each unit broadcasts an election request every 3 seconds.
2. Any other units with a higher priority respond to the election request; the priority is set when you deploy the cluster and is not configurable.
3. If after 45 seconds, a unit does not receive a response from another unit with a higher priority, then it becomes master.
4. If a unit later joins the cluster with a higher priority, it does not automatically become the master unit; the existing master unit always remains as the master unless it stops responding, at which point a new master unit is elected.
You can manually force a unit to become the master. For centralized features, if you force a master unit change, then all connections are dropped, and you have to re-establish the connections on the new master unit.

High Availability Within the Cluster

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

Chassis-Application Monitoring

Chassis-application health monitoring is always enabled. The Firepower 4100/9300 chassis supervisor checks the ASA application periodically (every second). If the ASA is up and cannot communicate with the Firepower 4100/9300 chassis supervisor for 3 seconds, the ASA generates a syslog message and leaves the cluster. When the ASA is not responding for 45 seconds, it reloads the ASA. If the ASA cannot communicate with the supervisor, it removes itself from the cluster.

Unit Health Monitoring

The master unit monitors every slave unit by sending heartbeat messages over the cluster control link periodically (the period is configurable). Each slave unit monitors the master unit using the same mechanism. If the unit health check fails, the unit is removed from the cluster.

Interface Monitoring

Each unit monitors the link status of all hardware interfaces in use, and reports status changes to the master unit. For inter-chassis clustering, Spanned EtherChannels use the cluster Link Aggregation Control Protocol (cLACP). Each chassis monitors the link status and the cLACP protocol messages to determine if the port is still active in the EtherChannel, and informs the ASA application if the interface is down. When you enable health monitoring, all physical interfaces are monitored by default (including the main EtherChannel for EtherChannel interfaces). Only named interfaces that are in an Up state can be monitored. For example, all member ports of an EtherChannel must fail before a named EtherChannel is removed from the cluster (depending on your minimum port bundling setting). You can optionally disable monitoring per interface.

If a monitored interface fails on a particular unit, but it is active on other units, then the unit is removed from the cluster. The amount of time before the ASA removes a member from the cluster depends on whether the unit is an established member or is joining the cluster. 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 an established member, the unit is removed after 500 ms.

For inter-chassis clustering, if you add or delete an EtherChannel from the cluster, interface health-monitoring is suspended for 95 seconds to ensure that you have time to make the changes on each chassis.

Decorator Application Monitoring

When you install a decorator application on an interface, such as the Radware DefensePro application, then both the ASA and the decorator application must be operational to remain in the cluster. The unit does not join the cluster until both applications are operational. Once in the cluster, the unit monitors the decorator application health every 3 seconds. If the decorator application is down, the unit is removed from the cluster.
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 cluster link.

If the master unit fails, then another member of the cluster with the highest priority (lowest number) becomes the master 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 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. The unit attempts to rejoin the cluster every 5 seconds.

- Failed Chassis-Application Communication—When the ASA detects that the chassis-application health has recovered, the ASA tries to rejoin the cluster automatically.

- Failed decorator application—The ASA rejoins the cluster when it senses that the decorator application is back up.

- 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.

---

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 3: Features Replicated Across the Cluster

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

## How the 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.

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 master unit. If a multi-session PAT connection is initially received by a slave unit, then the slave unit forwards the connection to the master unit. By default, UDP (except for DNS UDP) and ICMP traffic use multi-session PAT, so these connections are always owned by the master 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. 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.
# History for ASA Clustering on the Firepower 4100/9300 Chassis

<table>
<thead>
<tr>
<th>Feature Name</th>
<th>Version</th>
<th>Feature Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initiator and responder information for Dead Connection Detection (DCD), and DCD support in a cluster.</td>
<td>9.13(1)</td>
<td>If you enable Dead Connection Detection (DCD), you can use the <code>show conn detail</code> command to get information about the initiator and responder. Dead Connection Detection allows you to maintain an inactive connection, and the <code>show conn</code> output tells you how often the endpoints have been probed. In addition, DCD is now supported in a cluster. No modified screens.</td>
</tr>
<tr>
<td>Monitor the traffic load for a cluster</td>
<td>9.13(1)</td>
<td>You can now monitor the traffic load for cluster members, including total connection count, CPU and memory usage, and buffer drops. If the load is too high, you can choose to manually disable clustering on the unit if the remaining units can handle the load, or adjust the load balancing on the external switch. This feature is enabled by default. New/Modified screens: • Configuration &gt; Device Management &gt; High Availability and Scalability &gt; ASA Cluster &gt; Cluster Configuration &gt; Enable Cluster Load Monitor check box</td>
</tr>
<tr>
<td>Accelerated cluster joining</td>
<td>9.13(1)</td>
<td>When a slave unit has the same configuration as the master unit, it will skip syncing the configuration and will join faster. This feature is enabled by default. This feature is configured on each unit, and is not replicated from the master to the slave. Note Some configuration commands are not compatible with accelerated cluster joining; if these commands are present on the unit, even if accelerated cluster joining is enabled, configuration syncing will always occur. You must remove the incompatible configuration for accelerated cluster joining to work. Use the <code>show cluster info unit-join-acceleration incompatible-config</code> to view incompatible configuration. New/Modified screens: Configuration &gt; Device Management &gt; High Availability and Scalability &gt; ASA Cluster &gt; Cluster Configuration &gt; Enable config sync acceleration check box</td>
</tr>
<tr>
<td>Per-site gratuitous ARP for clustering</td>
<td>9.12(1)</td>
<td>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 &gt; Device Management &gt; High Availability and Scalability &gt; ASA Cluster &gt; Cluster Configuration &gt; Site Periodic GARP field</td>
</tr>
<tr>
<td>Feature Name</td>
<td>Version</td>
<td>Feature Information</td>
</tr>
<tr>
<td>--------------</td>
<td>---------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Parallel cluster joining of units per Firepower 9300 chassis</td>
<td>9.10(1)</td>
<td>For the Firepower 9300, this feature ensures that the security modules in a chassis join the cluster simultaneously, so that traffic is evenly distributed between the modules. If a module joins very much in advance of other modules, it can receive more traffic than desired, because the other modules cannot yet share the load. New/modified screens: <strong>Configuration &gt; Device Management &gt; High Availability and Scalability &gt; ASA Cluster</strong> New/Modified options: <strong>Parallel Join of Units Per Chassis</strong> area</td>
</tr>
<tr>
<td>Cluster control link customizable IP Address for the Firepower 4100/9300</td>
<td>9.10(1)</td>
<td>By default, the cluster control link uses the 127.2.0.0/16 network. You can now set the network when you deploy the cluster in FXOS. The chassis auto-generates the cluster control link interface IP address for each unit based on the chassis ID and slot ID: 127.2.chassis_id.slot_id. However, some networking deployments do not allow 127.2.0.0/16 traffic to pass. Therefore, you can now set a custom /16 subnet for the cluster control link in FXOS except for loopback (127.0.0.0/8) and multicast (224.0.0.0/4) addresses. New/modified Firepower Chassis Manager screens: <strong>Logical Devices &gt; Add Device &gt; Cluster Information</strong> New/Modified options: <strong>CCL Subnet IP</strong> field</td>
</tr>
<tr>
<td>Cluster interface debounce time now applies to interfaces changing from a down state to an up state</td>
<td>9.10(1)</td>
<td>When an interface status update occurs, the ASA waits the number of milliseconds specified in the <strong>health-check monitor-interface debounce-time</strong> command or the ASDM <strong>Configuration &gt; Device Management &gt; High Availability and Scalability &gt; ASA Cluster</strong> 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.</td>
</tr>
<tr>
<td>Automatically rejoin the cluster after an internal failure</td>
<td>9.9(2)</td>
<td>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: <strong>Configuration &gt; Device Management &gt; High Availability and Scalability &gt; ASA Cluster &gt; Auto Rejoin</strong></td>
</tr>
<tr>
<td>Show transport related statistics for cluster reliable transport protocol messages</td>
<td>9.9(2)</td>
<td>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: <strong>show cluster info transport cp detail</strong></td>
</tr>
<tr>
<td>Feature Name</td>
<td>Version</td>
<td>Feature Information</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>---------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Improved chassis health check failure detection for the Firepower chassis  | 9.9(1)  | You can now configure a lower holdtime for the chassis health check: 100 ms. The previous minimum was 300 ms. Note that the minimum combined time \((\text{interval} \times \text{retry-count})\) cannot be less than 600 ms.  
   New or modified command: **app-agent heartbeat interval**  
   No ASDM support.                                                                 |
| 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** |
| Distributed Site-to-Site VPN with clustering on the Firepower 9300          | 9.9(1)  | An ASA cluster on the Firepower 9300 supports Site-to-Site VPN in distributed mode. Distributed mode provides the ability to have many Site-to-Site IPsec IKEv2 VPN connections distributed across members of an ASA cluster, not just on the master unit (as in centralized mode). This significantly scales VPN support beyond Centralized VPN capabilities and provides high availability. Distributed S2S VPN runs on a cluster of up to two chassis, each containing up to three modules (six total cluster members), each module supporting up to 6K active sessions (12K total), for a maximum of approximately 36K active sessions (72K total).  
   New or modified screens:  
   **Monitoring > ASA Cluster > ASA Cluster > VPN Cluster Summary**  
   **Monitoring > VPN > VPN Statistics > Sessions > Slave**  
   **Configuration > Device Management > High Availability and Scalability > ASA Cluster**  
   **Wizards > Site-to-Site**  
   **Monitoring > VPN > VPN Statistics > Sessions**  
   **Monitoring > ASA Cluster > ASA Cluster > VPN Cluster Summary**  
   **Monitoring > ASA Cluster > ASA Cluster > System Resource Graphs > CPU/Memory**  
   **Monitoring > Logging > Real-Time Log Viewer** |
| 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** |
<table>
<thead>
<tr>
<th>Feature Name</th>
<th>Version</th>
<th>Feature Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configurable debounce time to mark an interface as failed for the Firepower 4100/9300 chassis</td>
<td>9.8(1)</td>
<td>You can now configure 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. The default debounce time is 500 ms, with a range of 300 ms to 9 seconds. 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.</td>
</tr>
<tr>
<td>Inter-site clustering improvement for the ASA on the Firepower 4100/9300 chassis</td>
<td>9.7(1)</td>
<td>You can now configure the site ID for each Firepower 4100/9300 chassis when you deploy the ASA cluster. Previously, you had to configure the site ID within the ASA application; this new feature eases initial deployment. Note that you can no longer set the site ID within the ASA configuration. Also, for best compatibility with inter-site clustering, we recommend that you upgrade to ASA 9.7(1) and FXOS 2.1.1, which includes several improvements to stability and performance.</td>
</tr>
<tr>
<td>Director localization: inter-site clustering improvement for data centers</td>
<td>9.7(1)</td>
<td>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.</td>
</tr>
<tr>
<td>Support for 16 chassis for the Firepower 4100 series</td>
<td>9.6(2)</td>
<td>You can now add up to 16 chassis to the cluster for the Firepower 4100 series.</td>
</tr>
<tr>
<td>Support for the Firepower 4100 series</td>
<td>9.6(1)</td>
<td>With FXOS 1.1.4, the ASA supports inter-chassis clustering on the Firepower 4100 series for up to 6 chassis.</td>
</tr>
<tr>
<td>Support for site-specific IP addresses in Routed, Spanned EtherChannel mode</td>
<td>9.6(1)</td>
<td>For inter-site clustering in routed mode with Spanned EtherChannels, you can now configure site-specific IP addresses 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.</td>
</tr>
</tbody>
</table>

New or modified screen: **Configuration > Device Management > High Availability and Scalability > ASA Cluster**
### Feature Information

<table>
<thead>
<tr>
<th>Feature Name</th>
<th>Version</th>
<th>Feature Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-chassis clustering for 16 modules, and inter-site clustering for the Firepower 9300 ASA application</td>
<td>9.5(2.1)</td>
<td>With FXOS 1.1.3, you can now enable inter-chassis, and by extension inter-site clustering. You can include up to 16 modules. For example, you can use 1 module in 16 chassis, or 2 modules in 8 chassis, or any combination that provides a maximum of 16 modules. We did not modify any screens.</td>
</tr>
<tr>
<td>Site-specific MAC addresses for inter-site clustering support for Spanned EtherChannel in Routed firewall mode</td>
<td>9.5(2)</td>
<td>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: <strong>Configuration &gt; Device Management &gt; High Availability and Scalability &gt; ASA Cluster &gt; Cluster Configuration</strong></td>
</tr>
<tr>
<td>ASA cluster customization of the auto-rejoin behavior when an interface or the cluster control link fails</td>
<td>9.5(2)</td>
<td>You can now customize the auto-rejoin behavior when an interface or the cluster control link fails. We introduced the following screen: <strong>Configuration &gt; Device Management &gt; High Availability and Scalability &gt; ASA Cluster &gt; Auto Rejoin</strong></td>
</tr>
<tr>
<td>The ASA cluster supports GTPv1 and GTPv2</td>
<td>9.5(2)</td>
<td>The ASA cluster now supports GTPv1 and GTPv2 inspection. We did not modify any screens.</td>
</tr>
<tr>
<td>Cluster replication delay for TCP connections</td>
<td>9.5(2)</td>
<td>This feature helps eliminate the “unnecessary work” related to short-lived flows by delaying the director/backup flow creation. We introduced the following screen: <strong>Configuration &gt; Device Management &gt; High Availability and Scalability &gt; ASA Cluster Replication</strong></td>
</tr>
<tr>
<td>LISP Inspection for Inter-Site Flow Mobility</td>
<td>9.5(2)</td>
<td>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: <strong>Configuration &gt; Device Management &gt; High Availability and Scalability &gt; ASA Cluster Configuration</strong> <strong>Configuration &gt; Firewall &gt; Objects &gt; Inspect Maps &gt; LISP</strong> <strong>Configuration &gt; Firewall &gt; Service Policy Rules &gt; Protocol Inspection</strong> <strong>Configuration &gt; Firewall &gt; Service Policy Rules &gt; Cluster</strong> <strong>Monitoring &gt; Routing &gt; LISP-EID Table</strong></td>
</tr>
<tr>
<td>Carrier Grade NAT enhancements now supported in failover and ASA clustering</td>
<td>9.5(2)</td>
<td>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.</td>
</tr>
<tr>
<td>Feature Name</td>
<td>Version</td>
<td>Feature Information</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>---------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| 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. |
| Intra-chassis ASA Clustering for the Firepower 9300 | 94(1.150)| You can cluster up to 3 security modules within the Firepower 9300 chassis. All modules in the chassis must belong to the cluster.  
We introduced the following screen: Configuration > Device Management > High Availability and Scalability > ASA Cluster Replication |