## Table of Contents

### Preface
- Who Should Use This Guide ................................................................. 12
- Summary of Contents ........................................................................... 13
- Appendices .......................................................................................... 13
- Related Documentation ....................................................................... 14
- More Information ................................................................................ 16
- Feedback .............................................................................................. 16

### Chapter 1
**Introduction to ClusterXL**
- The Need for Gateway Clusters ............................................................. 18
- Reliability through High Availability .................................................... 18
- Enhanced Reliability and Performance through Load Sharing ............ 18
- Check Point ClusterXL Gateway Clustering Solution ......................... 19
- The Cluster Control Protocol ................................................................. 20
- Installation, Licensing and Platform Support ....................................... 21
- Clock Synchronization in ClusterXL ..................................................... 21
- Clustering Definitions and Terms .......................................................... 22

### Chapter 2
**Synchronizing Connection Information Across the Cluster**
- The Need to Synchronize Cluster Information ...................................... 26
- The Check Point State Synchronization Solution .................................... 27
- Introduction to State Synchronization .................................................. 27
- The Synchronization Network ................................................................. 28
- How State Synchronization Works ....................................................... 29
- Non-Synchronized Services ................................................................. 30
- Choosing Services That Do Not Require Synchronization ................. 31
- Duration Limited Synchronization ....................................................... 32
- Non-Sticky Connections ....................................................................... 32
- Non-Sticky Connection Example: TCP 3-Way Handshake .................. 34
- Synchronizing Non-Sticky Connections .............................................. 35
- Synchronizing Clusters over a Wide Area Network ............................ 36
- Synchronized Cluster Restrictions ....................................................... 37
- Configuring State Synchronization ....................................................... 38
- Configuring State Synchronization ....................................................... 38
- Setting a Service to Non-Synchronized .............................................. 38
- Creating Synchronized and Non-Synchronized Versions .................. 39
- Configuring Duration Limited Synchronization ................................... 39

### Chapter 3
**Sticky Connections**
- Introduction to Sticky Connections .................................................... 42
- The Sticky Decision Function .............................................................. 42
- VPN Tunnels with 3rd Party Peers and Load Sharing .......................... 43
Chapter 4

High Availability and Load Sharing in ClusterXL

Introduction to High Availability and Load Sharing........................................... 50
Load Sharing........................................................................................................ 50
High Availability ............................................................................................ 51
Example ClusterXL Topology.......................................................................... 52
Defining the Cluster Member IP Addresses..................................................... 53
Defining the Cluster Virtual IP Addresses....................................................... 54
The Synchronization Network.......................................................................... 54
Configuring Cluster Addresses on Different Subnets......................................... 55
ClusterXL Modes............................................................................................. 56
Introduction to ClusterXL Modes..................................................................... 56
Load Sharing Multicast Mode.......................................................................... 57
Load Sharing Unicast Mode............................................................................ 59
New High Availability Mode........................................................................... 61
Mode Comparison Table ................................................................................ 63
Failover ............................................................................................................. 64
What is a Failover?........................................................................................... 64
When Does a Failover Occur?......................................................................... 65
What Happens When a Gateway Recovers?.................................................... 65
How a Recovered Cluster Member Obtains the Security Policy....................... 66
Implementation Planning Considerations....................................................... 67
High Availability or Load Sharing................................................................. 67
Choosing the Load Sharing Mode ................................................................. 67
IP Address Migration....................................................................................... 68
Hardware Requirements, Compatibility and Cisco Example............................ 69
ClusterXL Hardware Requirements................................................................. 69
ClusterXL Hardware Compatibility................................................................. 72
Example Configuration of a Cisco Catalyst Routing Switch........................... 73
Check Point Software Compatibility............................................................... 75
Operating System Compatibility................................................................. 75
Check Point Software Compatibility (excluding IPS)....................................... 75
ClusterXL Compatibility with IPS................................................................... 79
Forwarding Layer........................................................................................... 80
Configuring ClusterXL.................................................................................... 81
Configuring Routing for the Client Machines.................................................. 81
Preparing the Cluster Member Machines........................................................ 82
Choosing the CCP Transport Mode on the Cluster Members......................... 83
SmartDashboard Configuration ..................................................................... 84

Chapter 5

Working with OPSEC Certified Clustering Products

Introduction to OPSEC Certified Clustering Products...................................... 90
Configuring OPSEC Certified Clustering Products.......................................... 91
Preparing the Switches and Configuring Routing.......................................... 91
Chapter 8

ClusterXL Advanced Configuration

Working with VPNs and Clusters ............................................................... 170
Configuring VPN and Clusters ................................................................. 170
Defining VPN Peer Clusters with Separate Security Management Servers .... 171
Working with NAT and Clusters ............................................................... 172
Cluster Fold and Cluster Hide ................................................................. 172
Configuring NAT on the Gateway Cluster .............................................. 173
Configuring NAT on a Cluster Member .................................................. 173
Working with VLANs and Clusters ........................................................... 174
VLAN Support in ClusterXL ................................................................. 174
Connecting Several Clusters on the Same VLAN ........................................ 174
Monitoring the Interface Link State ....................................................... 179
Enabling Interface Link State Monitoring .............................................. 179
Working with Link Aggregation and Clusters ........................................... 180
Introduction to Working with Link Aggregation and Clusters .................... 180
Redundant Topologies ........................................................................... 181
Configuring Interface Bonds ................................................................. 187
Troubleshooting Bonded Interfaces ....................................................... 191
Advanced Cluster Configuration ............................................................ 197
How to Configure Gateway Configuration Parameters ............................ 197
How to Configure Gateway to Survive a Boot ...................................... 198
Controlling the Clustering and Synchronization Timers ............................ 199
Blocking New Connections Under Load ................................................. 200
Working with SmartView Tracker Active Mode ...................................... 201
Reducing the Number of Pending Packets .............................................. 202
Configuring Full Synchronization Advanced Options .............................. 202
Defining Disconnected Interfaces ........................................................... 204
Defining a Disconnected Interface on Unix .............................................. 204
Defining a Disconnected Interface on Windows ....................................... 204
Configuring Policy Update Timeout ....................................................... 205
Enhanced Enforcement of the TCP 3-Way Handshake ............................ 206
Configuring Cluster Addresses on Different Subnets ............................... 207
Introduction to Cluster Addresses on Different Subnets ............................ 207
Configuration of Cluster Addresses on Different Subnets .......................... 208
Example of Cluster Addresses on Different Subnets .................................. 209
Limitations of Cluster Addresses on Different Subnets ............................. 210
Moving from a Single Gateway to a ClusterXL Cluster .............................. 215
On the Single Gateway Machine ............................................................ 215
On Machine ‘B’ ....................................................................................... 215
In SmartDashboard, for Machine ‘B’ ....................................................... 216
On Machine ‘A’ ....................................................................................... 216
In SmartDashboard for Machine ‘A’ ....................................................... 217
Adding Another Member to an Existing Cluster ...................................... 218
Configuring ISP Redundancy on a Cluster .............................................. 219
Enabling Dynamic Routing Protocols in a Cluster Deployment .................. 220
## Components of the System

- Dynamic Routing in ClusterXL

## High Availability Legacy Mode

Appendix A

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction to High Availability Legacy Mode</td>
<td>224</td>
</tr>
<tr>
<td>Example of High Availability HA Legacy Mode Topology</td>
<td>225</td>
</tr>
<tr>
<td>Shared Interfaces IP and MAC Address Configuration</td>
<td>226</td>
</tr>
<tr>
<td>The Synchronization Interface</td>
<td>226</td>
</tr>
<tr>
<td>Implementation Planning Considerations for HA Legacy Mode</td>
<td>227</td>
</tr>
<tr>
<td>IP Address Migration</td>
<td>227</td>
</tr>
<tr>
<td>Security Management server Location</td>
<td>227</td>
</tr>
<tr>
<td>Routing Configuration</td>
<td>228</td>
</tr>
<tr>
<td>Switch (Layer 2 Forwarding) Considerations</td>
<td>228</td>
</tr>
<tr>
<td>Configuring High Availability Legacy Mode</td>
<td>229</td>
</tr>
<tr>
<td>Routing Configuration</td>
<td>229</td>
</tr>
<tr>
<td>SmartDashboard Configuration</td>
<td>230</td>
</tr>
<tr>
<td>Moving from High Availability Legacy with Minimal Effort</td>
<td>232</td>
</tr>
<tr>
<td>On the Gateways</td>
<td>232</td>
</tr>
<tr>
<td>From SmartDashboard</td>
<td>233</td>
</tr>
<tr>
<td>Moving from High Availability Legacy with Minimal Downtime</td>
<td>234</td>
</tr>
</tbody>
</table>

Appendix B

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example cphaprob Script</td>
<td>237</td>
</tr>
<tr>
<td>More Information</td>
<td>237</td>
</tr>
<tr>
<td>The clusterXL_monitor_process script</td>
<td>237</td>
</tr>
</tbody>
</table>
Preface

In This Chapter

Who Should Use This Guide  page 12
Summary of Contents  page 13
Related Documentation  page 14
More Information  page 16
Feedback  page 16
Who Should Use This Guide

This guide is intended for administrators responsible for maintaining network security within an enterprise, including policy management and user support.

This guide assumes a basic understanding of:

- System administration
- The underlying operating system
- Internet protocols (IP, TCP, UDP etc.)
## Summary of Contents

This guide contains the following chapters:

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chapter 1, “Introduction to ClusterXL”</strong></td>
<td>Describes the need for Gateway Clusters, introduces ClusterXL and the Cluster Control Protocol, specifies installation and licensing requirements, and lists clustering definitions and terms.</td>
</tr>
<tr>
<td><strong>Chapter 2, “Synchronizing Connection Information Across the Cluster”</strong></td>
<td>Describes State Synchronization, what not to synchronize, and how to configure State Synchronization.</td>
</tr>
<tr>
<td><strong>Chapter 3, “Sticky Connections”</strong></td>
<td>Describes the use of the Sticky Decision Function for Load Sharing connections.</td>
</tr>
<tr>
<td><strong>Chapter 4, “High Availability and Load Sharing in ClusterXL”</strong></td>
<td>Describes ClusterXL Load Sharing and High Availability modes.</td>
</tr>
<tr>
<td><strong>Chapter 5, “Working with OPSEC Certified Clustering Products”</strong></td>
<td>Describes the special considerations for working with OPSEC clustering products.</td>
</tr>
<tr>
<td><strong>Chapter 6, “UTM-1 Clustering”</strong></td>
<td>Procedures for configuring UTM-1 appliances to work with ClusterXL.</td>
</tr>
<tr>
<td><strong>Chapter 7, “Monitoring and Troubleshooting Gateway Clusters”</strong></td>
<td>Procedures for monitoring and troubleshooting a cluster, including the <code>cphaprob</code> and <code>fw ctl pstat</code> commands.</td>
</tr>
</tbody>
</table>

## Appendices

This guide contains the following appendices:

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Appendix A, “High Availability Legacy Mode”</strong></td>
<td>Describes High Availability Legacy Mode, and how to configure it.</td>
</tr>
<tr>
<td><strong>Appendix B, “Example cphaprob Script”</strong></td>
<td>An example of a script that can be used in conjunction with ClusterXL, using the <code>pnote</code> mechanism.</td>
</tr>
</tbody>
</table>
## Related Documentation

This release includes the following documentation

<table>
<thead>
<tr>
<th>Title</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internet Security Installation and Upgrade Guide</strong></td>
<td>Contains detailed installation instructions for Check Point network security products. Explains the available upgrade paths from versions R60-65 to the current version.</td>
</tr>
<tr>
<td><strong>High-End Installation and Upgrade Guide</strong></td>
<td>Contains detailed installation instructions for the Provider-1 and VSX products, including hardware and software requirements and licensing requirements. Explains all upgrade paths for Check Point products specifically geared towards upgrading to the current version.</td>
</tr>
<tr>
<td><strong>Security Management Server Administration Guide</strong></td>
<td>Explains Security Management solutions. This guide provides solutions for control over configuring, managing, and monitoring security deployments.</td>
</tr>
<tr>
<td><strong>Firewall Administration Guide</strong></td>
<td>Describes how to control and secure network access and VoIP traffic; how to use integrated web security capabilities; and how to optimize Application Intelligence with capabilities such as Content Vectoring Protocol (CVP) applications, URL Filtering (UFP) applications.</td>
</tr>
<tr>
<td><strong>IPS Administration Guide</strong></td>
<td>Describes how to use IPS to protect against attacks.</td>
</tr>
<tr>
<td><strong>Virtual Private Networks Administration Guide</strong></td>
<td>Describes the basic components of a VPN and provides the background for the technology that comprises the VPN infrastructure.</td>
</tr>
<tr>
<td>Title</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>Eventia Reporter Administration Guide</td>
<td>Explains how to monitor and audit traffic, and generate detailed or summarized reports in the format of your choice (list, vertical bar, pie chart etc.) for all events logged by Check Point Security Gateways, SecureClient and IPS.</td>
</tr>
<tr>
<td>SecurePlatform/SecurePlatform Pro Administration Guide</td>
<td>Explains how to install and configure SecurePlatform. This guide will also teach you how to manage your SecurePlatform machine and explains Dynamic Routing (Unicast and Multicast) protocols.</td>
</tr>
<tr>
<td>Provider-1/SiteManager-1 Administration Guide</td>
<td>Explains the Provider-1 security management solution. This guide provides details about a three-tier, multi-policy management architecture and a host of Network Operating Center oriented features that automate time-consuming repetitive tasks common in Network Operating Center environments.</td>
</tr>
</tbody>
</table>
More Information

- For additional technical information about Check Point products, consult Check Point's SecureKnowledge at http://support.checkpoint.com.
- To view the latest version of this document in the Check Point User Center, go to: http://support.checkpoint.com.

Feedback

Check Point is engaged in a continuous effort to improve its documentation. Please help us by sending your comments to:

cp_techpub_feedback@checkpoint.com
Chapter 1
Introduction to ClusterXL

In This Chapter

The Need for Gateway Clusters .................................................. page 18
Check Point ClusterXL Gateway Clustering Solution ......................... page 19
The Cluster Control Protocol ....................................................... page 20
Installation, Licensing and Platform Support ................................ page 21
Clock Synchronization in ClusterXL ............................................. page 21
Clustering Definitions and Terms ............................................... page 22
The Need for Gateway Clusters

Reliability through High Availability

Gateways and VPN connections are business critical devices. The failure of a Security Gateway or VPN connection can result in the loss of active connections and access to critical data. The gateway between the organization and the world must remain open under all circumstances.

High availability ensures gateway and VPN connection redundancy by providing transparent failover to a backup gateway in the event of failure.

Enhanced Reliability and Performance through Load Sharing

In a Load Sharing Gateway Cluster, all cluster members are active. Load Sharing also brings significant performance advantages. Using multiple gateways instead of a single gateway increases linear performance for CPU intensive applications such as VPNs, Security servers, Policy servers, and SmartDirectory (LDAP).
Check Point ClusterXL Gateway Clustering Solution

ClusterXL is a software-based Load Sharing and High Availability solution that distributes network traffic between clusters of redundant Security Gateways and provides transparent failover between machines in a cluster.

A cluster is a group of identical Check Point Security Gateways connected in such a way that if one fails, another immediately take its place.

Figure 1-1  A Firewalled Gateway Cluster

ClusterXL uses unique physical IP and MAC addresses for the cluster members and virtual IP addresses to represent the cluster itself. Virtual IP addresses do not belong to an actual machine interface (except in High Availability Legacy mode, explained later).

ClusterXL provides an infrastructure that ensures that data is not lost due to a failure, by ensuring that each cluster member is aware of connections passing through the other members. Passing information about connections and other Security Gateway states between the cluster members is known as State Synchronization.

Security Gateway Clusters can also be built using OPSEC certified High Availability and Load Sharing products. OPSEC certified clustering products use the same State Synchronization infrastructure as ClusterXL.
The Cluster Control Protocol

The Cluster Control Protocol (CCP) is the glue that links together the machines in the Check Point Gateway Cluster. CCP traffic is distinct from ordinary network traffic and can be viewed using any network sniffer.

CCP runs on UDP port 8116, and has the following roles:

- It allows cluster members to report their own states and learn about the states of other members by sending keep-alive packets (this only applies to ClusterXL clusters).
- State Synchronization.

Check Point's CCP is used by all ClusterXL modes as well as by OPSEC clusters. However, the tasks performed by this protocol and the manner in which they are implemented may differ between clustering types.

Note - There is no need to add a rule to the Security Policy Rule Base that accepts CCP.
ClusterXL must be installed in a distributed configuration in which the Security Management server and the cluster members are on different machines. ClusterXL is part of the standard Security Gateway installation.

To install a policy on a gateway cluster:

1. Install a license for each Check Point Security Gateway installed on at least one of the cluster members.

2. On other members, install a secondary Check Point Security Gateway license (The SKU has -HA as a suffix).

3. Install an additional Load Sharing add-on license is for the Security Management server managing each ClusterXL Load Sharing cluster. There are two Load Sharing license SKUs: CPMP-CXLS-U and CPMP-CXLS-500. ClusterXL High Availability and third-party clusters (both High Availability and Load Sharing) do not require additional licenses.

Both the plug and play and the evaluation licenses include the option to work with up to three ClusterXL Load Sharing clusters managed by the same Security Management server.

ClusterXL supported platforms are listed in the platform support matrix, which is available online at: http://support.checkpoint.com.

**Clock Synchronization in ClusterXL**

When using ClusterXL, ensure that you synchronize the clocks of all of the cluster members. You can synchronize the clocks manually or using a protocol such as NTP. Features such as VPN function properly only once the clocks of all of the cluster members are synchronized.
Clustering Definitions and Terms

Different vendors give different meanings to terms that relate to Gateway Clusters, High Availability and Load Sharing. Check Point uses the following definitions and terms when discussing clustering:

**Active Up**
When the *High Availability* machine that was Active and suffered a failure becomes available again, it returns to the cluster, not as the Active machine but as one of the standby machines in the cluster.

**Cluster**
A group of machines that work together to provide Load Sharing and/or High Availability.

**Critical Device**
A device that the Administrator has defined to be critical to the operation of the cluster member. A critical device is also known as a Problem Notification (pnote). Critical devices are constantly monitored. If a critical device stops functioning, this is defined as a failure. A device can be hardware or a process. The `fwd` and `cphad` processes are predefined by default as critical devices. The Security Policy is also predefined as a critical device. The Administrator can add to the list of critical devices using the `cphaprob` command.

**Failure**
A hardware or software problem that causes a machine to be unable to filter packets. A failure of an Active machine leads to a *Failover*.

**Failover**
A machine taking over packet filtering in place of another machine in the cluster that suffered a failure.

**High Availability**
The ability to maintain a connection when there is a failure by having another machine in the cluster take over the connection, without any loss of connectivity. Only the Active machine filters packets. One of the machines in the cluster is configured as the Active machine. If a failure occurs on the Active machine, one of the other machines in the cluster assumes its responsibilities.
Hot Standby

Also known as Active/Standby. It has the same meaning as *High Availability*.

Load Sharing

In a Load Sharing Gateway Cluster, all machines in the cluster filter packets. Load Sharing provides *High Availability*, gives transparent *Failover* to any of the other machines in the cluster when a failure occurs and provides enhanced reliability and performance. Load Sharing is also known as Active/Active.

Multicast Load Sharing

In ClusterXL’s Load Sharing Multicast mode, every member of the cluster receives all of the packets sent to the cluster IP address. A router or Layer 3 switch forwards packets to all of the cluster members using multicast. A ClusterXL decision algorithm on all cluster members decides which cluster member should perform enforcement processing on the packet.

Unicast Load Sharing

In ClusterXL’s Load Sharing Unicast mode, one machine (the Pivot) receives all traffic from a router with a unicast configuration and redistributes the packets to the other machines in the cluster. The Pivot machine is chosen automatically by ClusterXL.
Chapter 2

Synchronizing Connection Information Across the Cluster

In This Chapter

The Need to Synchronize Cluster Information page 26
The Check Point State Synchronization Solution page 27
Configuring State Synchronization page 38
The Need to Synchronize Cluster Information

A failure of a firewall results in an immediate loss of active connections in and out of the organization. Many of these connections, such as financial transactions, may be mission critical, and losing them will result in the loss of critical data. ClusterXL supplies an infrastructure that ensures that no data is lost in case of a failure, by making sure each gateway cluster member is aware of the connections going through the other members. Passing information about connections and other Security Gateway states between the cluster members is called State Synchronization.
The Check Point State Synchronization Solution

In This Section

Introduction to State Synchronization  page 27
The Synchronization Network  page 28
How State Synchronization Works  page 29
Non-Synchronized Services  page 30
Choosing Services That Do Not Require Synchronization  page 31
Duration Limited Synchronization  page 32
Non-Sticky Connections  page 32
Non-Sticky Connection Example: TCP 3-Way Handshake  page 34
Synchronizing Non-Sticky Connections  page 35
Synchronizing Clusters over a Wide Area Network  page 36
Synchronized Cluster Restrictions  page 37

Introduction to State Synchronization

State Synchronization enables all machines in the cluster to be aware of the connections passing through each of the other machines. It ensures that if there is a failure in a cluster member, connections that were handled by the failed machine will be maintained by the other machines.

Every IP based service (including TCP and UDP) recognized by the Security Gateway is synchronized.

State Synchronization is used both by ClusterXL and by third-party OPSEC-certified clustering products.

Machines in a ClusterXL Load Sharing configuration must be synchronized. Machines in a ClusterXL High Availability configuration do not have to be synchronized, though if they are not, connections will be lost upon failover.
The Synchronization Network

The Synchronization Network is used to transfer synchronization information about connections and other Security Gateway states between cluster members.

Because the synchronization network carries the most sensitive Security Policy information in the organization, it is important to make sure that it is secured against both malicious and unintentional interference. It is therefore recommended to secure the synchronization interfaces by:

- using a dedicated synchronization network, and
- connecting the physical network interfaces of the cluster members directly using a cross-cable. In a cluster with three or more members, use a dedicated hub or switch.

Note - It is possible to run synchronization across a WAN. For details, see “Synchronizing Clusters over a Wide Area Network” on page 36.

Following these recommendations guarantees the safety of the synchronization network because no other networks carry synchronization information.

It is possible to define more than one synchronization network for backup purposes. It is recommended that the backup be a dedicated network.

In Cluster XL, the synchronization network is supported on the lowest VLAN tag of a VLAN interface. For example, if three VLANs with tags 10, 20 and 30 are configured on interface eth1, interface eth1.10 may be used for synchronization.
How State Synchronization Works

Synchronization works in two modes:

- **Full sync** transfers all VPN kernel table information from one cluster member to another. It is handled by the `fwd` daemon using an encrypted TCP connection.

- **Delta sync** transfers changes in the kernel tables between cluster members. Delta sync is handled by the VPN kernel using UDP multicast or broadcast on port 8116.

Full sync is used for initial transfers of state information, for many thousands of connections. If a cluster member is brought up after being down, it will perform full sync. Once all members are synchronized, only updates are transferred via delta sync. Delta sync is much quicker than full sync.

State Synchronization traffic typically makes up around 90% of all Cluster Control Protocol (CCP) traffic. State Synchronization packets are distinguished from the rest of CCP traffic via an opcode in the UDP data header.

**Note** - The source MAC address for CCP packets can be changed. See “Connecting Several Clusters on the Same VLAN” on page 174.
Non-Synchronized Services

In a gateway cluster, all connections on all cluster members are normally synchronized across the cluster. However, not all services that cross a gateway cluster need necessarily be synchronized.

- It is possible to decide not to synchronize TCP, UDP and other service types. By default, all these services are synchronized.

- The VRRP and IP Clustering control protocols, as well as the IGMP protocol, are not synchronized by default (although you can choose to turn on synchronization for these protocols). Protocols that run solely between cluster members need not be synchronized. Although it is possible to synchronize them, no benefit will be gained if the cluster is configured to do so. The synchronization information is not relevant for this case because it will not help in case of a failover. Therefore the following protocols are not synchronized by default: IGMP, VRRP, IP clustering and some other OPSEC cluster control protocols.

- Broadcasts and multicasts are not synchronized, and cannot be synchronized.

It is possible to have both a synchronized service and a non-synchronized definition of a service, and to use them selectively in the Rule Base.
Choosing Services That Do Not Require Synchronization

Synchronization has some performance cost. You can decide not to synchronize a service if all the following conditions are true:

1. A significant proportion of the traffic crossing the cluster uses a particular service. Not synchronizing the service reduces the amount of synchronization traffic, thereby enhancing cluster performance.

2. The service usually opens short connections, whose loss may not be noticed. DNS (over UDP) and HTTP are typically responsible for most connections, and on the other hand frequently have very short life and inherent recoverability in the application level. Services which typically open long connections, such as FTP, should always be synchronized.

3. Configurations that ensure bi-directional stickiness for all connections do not require synchronization to operate (only to maintain High Availability). Such configurations include:
   - Any cluster in High Availability mode (for example, ClusterXL New HA or Nokia VRRP)
   - ClusterXL in a Load Sharing mode with clear connections (no VPN or static NAT)
   - OPSEC clusters that guarantee full stickiness (refer to the OPSEC cluster's documentation)

   VPN and Static NAT connections passing through a ClusterXL cluster in a Load Sharing mode (either multicast or unicast) may not maintain bi-directional stickiness; hence, State Synchronization must be turned on for such environments.

To configure a service so that it will not be synchronized, edit the Service object. See “Setting a Service to Non-Synchronized” on page 38.
Duration Limited Synchronization

Some TCP services (HTTP for example) are characterized by connections with a very short duration. There is no point in synchronizing these connections because every synchronized connection consumes gateway resources, and the connection is likely to have finished by the time a failover occurs.

For all TCP services whose Protocol Type (that is defined in the GUI) is HTTP or None, you can use this option to delay telling the Security Gateway about a connection, so that the connection will only be synchronized if it still exists x seconds after the connection is initiated. This feature requires a SecureXL device that supports “Delayed Notifications” and the current cluster configuration (such as Performance Pack with ClusterXL LS Multicast).

This capability is only available if a SecureXL-enabled device is installed on the Security Gateway through which the connection passes.

The setting is ignored if connection templates are not offloaded from the ClusterXL-enabled device. See the SecureXL documentation for additional information.

Non-Sticky Connections

A connection is called sticky if all packets of the connection are handled by a single cluster member. In a non-sticky connection, a reply packet may return through a different gateway than the original packet.

The synchronization mechanism knows how to properly handle non-sticky connections. In a non-sticky connection, a cluster member gateway can receive an out-of-state packet, which Security Gateway normally drops because it poses a security risk.

In Load Sharing configurations, all cluster members are active, and in Static NAT and encrypted connections, the source and destination IP addresses change. Therefore, Static NAT and encrypted connections through a Load Sharing cluster may be non-sticky. Non-stickiness may also occur with Hide NAT, but ClusterXL has a mechanism to make it sticky.

In High Availability configurations, all packets reach the Active machine, so all connections are sticky. If failover occurs during connection establishment, the connection is lost, but synchronization can be performed later.
If the other members do not know about a non-sticky connection, the packet will be out-of-state, and the connection will be dropped for security reasons. However, the Synchronization mechanism knows how to inform other members of the connection. The Synchronization mechanism thereby prevent out-of-state packets in valid, but non-sticky connections, so that these non-sticky connections are allowed.

Non-sticky connections will also occur if the network administrator has configured asymmetric routing, where a reply packet returns through a different gateway than the original packet.
Non-Sticky Connection Example: TCP 3-Way Handshake

The 3-way handshake that initiates all TCP connections can very commonly lead to a non-sticky (often called asymmetric routing) connection. The following situation may arise:

Client A initiates a connection by sending a SYN packet to server B (see Figure 2-1). The SYN passes through Gateway C, but the SYN/ACK reply returns through Gateway D. This is a non-sticky connection, because the reply packet returns through a different gateway than the original packet.

Gateway D is notified of the SYN packet via the synchronization network. If gateway D is updated before the SYN/ACK packet sent by server B reaches this machine, the connection is handled normally. If, however, synchronization is delayed, and the SYN/ACK packet is received on gateway D before the SYN flag has been updated, then the gateway will treat the SYN/ACK packet as out-of-state, and will drop the connection.

See “Enhanced Enforcement of the TCP 3-Way Handshake” on page 206 for additional information.

Figure 2-1 A Non-sticky (asymmetrically routed) connection
Synchronizing Non-Sticky Connections

The synchronization mechanism prevents out-of-state packets in valid, but non-sticky connections. The way it does this is best illustrated with reference to the 3-way handshake that initiates all TCP data connections. The 3-way handshake proceeds as follows:

1. SYN (client to server)
2. SYN/ACK (server to client)
3. ACK (client to server)
4. Data (client to server)

To prevent out-of-state packets, the following sequence (called “Flush and Ack”) occurs (The step numbers correspond to the numbers in Figure 2-1):

1. Cluster member receives first packet (SYN) of a connection.
2. Suspects that it is non-sticky.
3. Hold the SYN packet.
4. Send the pending synchronization updates to all cluster members (including all changes relating to this packet).
5. Wait for all the other cluster members to acknowledge the information in the sync packet.
7. All cluster members are ready for the SYN-ACK.
Synchronizing Clusters over a Wide Area Network

Organizations are sometimes faced with the need to locate cluster members in geographical locations that are distant from each other. A typical example is a replicated data center whose locations are widely separated for disaster recovery purposes. In such a configuration it is clearly impractical to use a cross cable as the synchronization network (as described in “The Synchronization Network” on page 28).

The synchronization network can be spread over remote sites, which makes it easier to deploy geographically distributed clustering. There are two limitations to this capability:

1. The synchronization network must guarantee no more than 100ms latency and no more than 5% packet loss.
2. The synchronization network may only include switches and hubs. No routers are allowed on the synchronization network, because routers drop Cluster Control Protocol packets.

To monitor and troubleshoot geographically distributed clusters, a command line is available. See “Troubleshooting Synchronization” on page 145.
Synchronized Cluster Restrictions

The following restrictions apply to synchronizing cluster members:

1. Only cluster members running on the same platform can be synchronized.
   For example, it is not possible to synchronize a Windows 2000 cluster member with a Secure Platform cluster member.

2. All cluster members must be of the same software version.
   For example, it is not possible to synchronize a Version NGX R65 cluster member with a version R70 cluster member.

3. A user-authenticated connection through a cluster member will be lost if the cluster member goes down. Other synchronized cluster members will be unable to resume the connection.
   However, a client-authenticated connection or session-authenticated connection will not be lost.

   The reason for these restrictions is that user authentication state is maintained on Security Servers, which are processes, and thus cannot be synchronized on different machines in the way that kernel data can be synchronized. However, the state of session authentication and client authentication is stored in kernel tables, and thus can be synchronized.

4. The state of connections using resources is maintained in a Security Server, so these connections cannot be synchronized for the same reason that user-authenticated connections cannot be synchronized.

5. Accounting information is accumulated in each cluster member and reported separately to the Security Management server, where the information is aggregated. In case of a failover, accounting information that was accumulated on the failed member but not yet reported to the Security Management server is lost. To minimize the problem it is possible to reduce the period in which accounting information is “flushed”. To do this, in the cluster object’s Logs and Masters > Additional Logging page, configure the attribute Update Account Log every:
Configuring State Synchronization

In This Section

- Configuring State Synchronization  page 38
- Setting a Service to Non-Synchronized  page 38
- Creating Synchronized and Non-Synchronized Versions  page 39
- Configuring Duration Limited Synchronization  page 39

Configuring State Synchronization

Configure State synchronization as part of the process of configuring ClusterXL and OPSEC certified clustering products. Configuring State synchronization involves:

- Setting up a synchronization network for the gateway cluster
- Installing Security Gateway and turning on the synchronization capability during the configuration phase.
- In SmartDashboard, ensuring State Synchronization is selected in ClusterXL page of the cluster object.

For configuration details, see:

- “Configuring ClusterXL” on page 81.
- “Configuring OPSEC Certified Clustering Products” on page 91.

Setting a Service to Non-Synchronized

For background information about configuring services so that they are not synchronized, see “Non-Synchronized Services” on page 30.

1. In the Services branch of the objects tree, double click the TCP, UDP or Other type service that you do not wish to synchronize.

2. In the Service Properties window, click Advanced to display the Advanced Services Properties window.

3. Deselect Synchronize connections on the cluster.
Creating Synchronized and Non-Synchronized Versions

It is possible to have both a synchronized and a non-synchronized definition of the service, and to use them selectively in the Security Rule Base.

1. Define a new TCP, UDP and Other type service. Give it a name that distinguishes it from the existing service.
2. Copy all the definitions from the existing service into the Service Properties window of the new service.
3. In the new service, click Advanced to display the Advanced Services Properties window.
4. Copy all the definitions from the existing service into the Advanced Service Properties window of the new service.
5. Set Synchronize connections on the cluster in the new service, so that it is different from the setting in the existing service.

Configuring Duration Limited Synchronization

For background information about the synchronization of services that have limited duration, see “Duration Limited Synchronization” on page 32.

1. In the Services branch of the objects tree, double click the TCP, UDP or Other type service that you wish to synchronize.
2. In the Service Properties window, click Advanced to display the Advanced Services Properties window.
3. Select Start synchronizing x seconds after connection initiation.

Note - As this feature is limited to HTTP-based services, the Start synchronizing - seconds after connection initiation checkbox is not displayed for other services.

4. In the seconds field, enter the number of seconds or select the number of seconds from the dropdown list, for which you want synchronization to be delayed after connection initiation.
Chapter 3

Sticky Connections

In This Chapter

- Introduction to Sticky Connections  page 42
- The Sticky Decision Function  page 42
- VPN Tunnels with 3rd Party Peers and Load Sharing  page 43
- Configuring Sticky Connections  page 46
Introduction to Sticky Connections

A connection is *sticky* when all of its packets are handled, in either direction, by a single cluster member. This is the case in High Availability mode, where all connections are routed through the same cluster member, and hence, sticky. This is also the case in Load Sharing mode when there are no VPN peers, static NAT rules or SIP.

In Load Sharing mode, however, there are cases where it is necessary to ensure that a connection that starts on a specific cluster member will continue to be processed by the same cluster member in both directions. To that end, certain connections can be made sticky by enabling the Sticky Decision Function.

**Note** - For the latest information regarding features that require sticky connections, refer to the Release Notes, available online at: [http://www.support.checkpoint.com](http://www.support.checkpoint.com).

The Sticky Decision Function

The Sticky Decision Function enables certain services to operate in a Load Sharing deployment. For example, it is required for L2TP traffic, or when the cluster is a participant in a site to site VPN tunnel with a third party peer.

The following services and connection types are now supported by enabling the Sticky Decision Function:

- VPN deployments with third-party VPN peers
- SecureClient/SecuRemote/SSL Network Extender encrypted connections, including SecureClient visitor mode

The Sticky Decision Function has the following limitations:

- Sticky Decision Function is not supported when employing either Performance Pack or a hardware-based accelerator card. Enabling the Sticky Decision Function disables these acceleration products.
- When the Sticky Decision Function is used in conjunction with VPN, cluster members are prevented from opening more than one connection to a specific peer. Opening another connection would cause another SA to be generated, which a third-party peer, in many cases, would not be able to process.
VPN Tunnels with 3\textsuperscript{rd} Party Peers and Load Sharing

Check Point provides interoperability with third-party vendor gateways by enabling them to peer with Security Gateway. A special case is when certain third-party peers (Microsoft LT2P, Nokia Symbian, and Cisco gateways and clients) attempt to establish VPN tunnels with ClusterXL Gateways in Load Sharing mode. These peers are limited in their ability to store SAs, which means that a VPN session that begins on one cluster member and, due to load sharing, is routed on the return trip through another, is unrecognized and dropped. Consider, for example, Figure 3-1:

**Figure 3-1** Third-party peers connected to ClusterXL in Load Sharing mode without Sticky Decision Function

In this scenario:

- A third-party peer (gateway or client) attempts to create a VPN tunnel.
- Cluster Members A and B belong to a ClusterXL Gateway in Load Sharing mode.

The third-party peers, lacking the ability to store more than one set of SAs, cannot negotiate a VPN tunnel with multiple cluster members, and therefore the cluster member cannot complete the routing transaction.

This issue is resolved for certain third-party peers or any gateways that can save only one set of SAs by making the connection sticky. Enabling the Sticky Decision Function sets all VPN sessions initiated by the same third-party gateway to be processed by a single cluster member. To enable the Sticky Decision Function, in SmartDashboard edit the cluster object > ClusterXL page > Advanced, and enable the property **Use Sticky Decision Function**.
Third-Party Gateways in Hub and Spoke Deployments

Another case where Load Sharing mode requires the Sticky Decision Function is when integrating certain third-party gateways into a hub and spoke deployment. Without the ability to store more than one set of SAs, a third-party gateway must maintain its VPN tunnels on a single cluster member in order to avoid duplicate SAs. The deployment is illustrated in Figure 3-2:

**Figure 3-2**  ClusterXL Supporting Star Topology VPN with Third-Party Gateway as Spoke

In this scenario:

- The intent of this deployment is to enable hosts that reside behind Spoke A to communicate with hosts behind Spoke B.
- The ClusterXL Gateway is in Load Sharing mode, is composed of Cluster Members A and B, and serves as a VPN Hub.
- Spoke A is a third-party gateway, and is connected by a VPN tunnel that passes through the Hub to Spoke B.
- Spoke B can be either another third-party gateway or a Check Point Security Gateway.
Spokes A and B must be set to always communicate using the same cluster member. Enabling the Sticky Decision Function solves half of this problem, in that all VPN sessions initiated by either third-party gateway are processed by a single cluster member. But how to make sure that all communications between Spokes A and B are always using the same cluster member? By making some changes to the user.def file, both third-party gateways can be set to always connect to the same cluster member, thereby preserving the integrity of the tunnel and circumventing this problem. For configuration instructions, see “Establishing a Third-Party Gateway in a Hub and Spoke Deployment” on page 46.
Configuring Sticky Connections

Configuring the Sticky Decision Function

The Sticky Decision Function is configurable in the SmartDashboard cluster object from the ClusterXL page, Advanced Load Sharing Configuration window.

Figure 3-3  Configuring the Sticky Decision Function

By default, the Sticky Decision Function is not enabled.

Establishing a Third-Party Gateway in a Hub and Spoke Deployment

To establish a third-party gateway as a spoke in a hub and spoke deployment, perform the following on the Security Management server:

1. Enable the Sticky Decision Function if not already enabled. In SmartDashboard, edit the cluster object > ClusterXL page > Advanced, and enable the property Use Sticky Decision Function.

2. Create a Tunnel Group to handle traffic from specific peers. Use a text editor to edit the file $FWDIR/lib/user.def, and add a line similar to the following:

   ```
   all@{member1,member2} vpn_sticky_gws = {<10.10.10.1;1>, <20.20.20.1;1>};
   ```
The elements of this configuration are as follows:

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>Stands for all the interfaces of the cluster Gateway</td>
</tr>
<tr>
<td>member1,member2</td>
<td>Names of the cluster members in SmartDashboard</td>
</tr>
<tr>
<td>vpn_sticky_gws</td>
<td>Name of the table</td>
</tr>
<tr>
<td>10.10.10.1</td>
<td>IP address of Spoke A</td>
</tr>
<tr>
<td>20.20.20.1</td>
<td>IP address of Spoke B</td>
</tr>
<tr>
<td>;1</td>
<td>Tunnel Group Identifier, which indicates that the traffic from these IP</td>
</tr>
<tr>
<td></td>
<td>addresses should be handled by the same cluster member</td>
</tr>
</tbody>
</table>

3. Other peers can be added to the Tunnel Group by including their IP addresses in the same format as shown above. To continue with the example above, adding Spoke C would look like this:

```plaintext
all@{member1,member2} vpn_sticky_gws = {<10.10.10.1;1>, <20.20.20.1;1>, <30.30.30.1;1>};
```

Note that the Tunnel Group Identifier ;1 stays the same, which means that the listed peers will always connect through the same cluster member.

**Note** - More tunnel groups than cluster members may be defined.

This procedure in essence turns off Load Sharing for the connections affected. If the implementation is to connect multiple sets of third-party gateways one to another, a form of Load Sharing can be accomplished by setting gateway pairs to work in tandem with specific cluster members. For instance, to set up a connection between two other spokes (C and D), simply add their IP addresses to the line and replace the Tunnel Group Identifier ;1 with ;2. The line would then look something like this:

```plaintext
all@{member1,member2} vpn_sticky_gws = {<10.10.10.1;1>, <20.20.20.1;1>, <192.168.15.5;2>, <192.168.1.4;2>,};
```

Note that there are now two peer identifiers: ;1 and ;2. Spokes A and B will now connect through one cluster member, and Spokes C and D through another.

**Note** - The tunnel groups are shared between active cluster members. In case of a change in cluster state (e.g., failover or member attach/detach), the reassignment is performed according to the new state.
Chapter 4

High Availability and Load Sharing in ClusterXL

In This Chapter

Introduction to High Availability and Load Sharing  page 50
Example ClusterXL Topology  page 52
ClusterXL Modes  page 56
Failover  page 64
Implementation Planning Considerations  page 67
Hardware Requirements, Compatibility and Cisco Example  page 69
Check Point Software Compatibility  page 75
Configuring ClusterXL  page 81
Introduction to High Availability and Load Sharing

ClusterXL is a software-based Load Sharing and High Availability solution that distributes network traffic between clusters of redundant Security Gateways.

ClusterXL provides:

- Transparent failover in case of machine failures
- Zero downtime for mission-critical environments (when using State Synchronization)
- Enhanced throughput (in Load Sharing modes)
- Transparent upgrades

All machines in the cluster are aware of the connections passing through each of the other machines. The cluster members synchronize their connection and status information across a secure synchronization network.

The glue that binds the machines in a ClusterXL cluster is the Cluster Control Protocol (CCP), which is used to pass synchronization and other information between the cluster members.

Load Sharing

ClusterXL Load Sharing distributes traffic within a cluster of gateways so that the total throughput of multiple machines is increased.

In Load Sharing configurations, all functioning machines in the cluster are active, and handle network traffic (Active/Active operation).

If any individual Check Point gateway in the cluster becomes unreachable, transparent failover occurs to the remaining operational machines in the cluster, thus providing High Availability. All connections are shared between the remaining gateways without interruption.
High Availability

High Availability allows organizations to maintain a connection when there is a failure in a cluster member, without Load Sharing between cluster members. In a High Availability cluster, only one machine is active (Active/Standby operation). In the event that the active cluster member becomes unreachable, all connections are re-directed to a designated standby without interruption. In a synchronized cluster, the standby cluster members are updated with the state of the connections of the active cluster member.

In a High Availability cluster, each machine is given a priority. The highest priority machine serves as the gateway in normal circumstances. If this machine fails, control is passed to the next highest priority machine. If that machine fails, control is passed to the next machine, and so on.

Upon gateway recovery, it is possible to maintain the current active gateway (Active Up), or to switch to the highest priority gateway (Primary Up). Note that in Active Up configuration, changing and installing the Security Policy may restart the ClusterXL configuration handshake on the members, which may lead to another member being chosen as the Active machine.
Example ClusterXL Topology

In This Section

- Defining the Cluster Member IP Addresses  page 53
- Defining the Cluster Virtual IP Addresses  page 54
- The Synchronization Network  page 54
- Configuring Cluster Addresses on Different Subnets  page 55

ClusterXL uses unique physical IP and MAC addresses for the cluster member, and virtual IP addresses to represent the cluster itself. Cluster interface addresses do not belong to any real machine interface.

Figure 4-1 shows a two-member ClusterXL cluster, and contrasts the virtual IP addresses of the cluster, and the physical IP addresses of the cluster members.

Each cluster member has three interfaces: one external interface, one internal interface, and one for synchronization. Cluster member interfaces facing in each direction are connected via a switch, router, or VLAN switch.

All cluster member interfaces facing the same direction must be in the same network. For example, there must not be a router between cluster members.

The Security Management Server can be located anywhere, and should be routable to either the internal or external cluster addresses.

Refer to the sections following Figure 4-1 for a description of the ClusterXL configuration concepts shown in the example.

Note

1. High Availability Legacy Mode uses a different Topology, and is discussed in the Appendix: "High Availability Legacy Mode" on page 223.

2. In the examples in this and subsequent sections, addresses in the range 192.168.0.0 to 192.168.255.255 which are RFC 1918 private addresses are used to represent routable (public) IP addresses.
Defining the Cluster Member IP Addresses

The guidelines for configuring each cluster member machine are as follows:

All machines within the cluster must have at least three interfaces:

- an interface facing the external cluster interface, which in turn faces the internet
- an interface facing the internal cluster interface, which in turn faces the internal network
- an interface to use for synchronization.

All interfaces pointing in a certain direction must be on the same network.
Defining the Cluster Virtual IP Addresses

For example, in the configuration in Figure 4-1, there are two cluster members, Member_A and Member_B. Each has an interface with an IP address facing the Internet through a hub or a switch. This is the External interface with IP address 192.168.10.1 on Member_A and 192.168.10.2 on Member_B, and is the interface that the cluster external interface sees.

**Note** - This release presents an option to use only two interfaces per member, one external and one internal and to run synchronization over the internal interface. However, this configuration is not recommended and should be used for backup only. For more information see Chapter 2, “Synchronizing Connection Information Across the Cluster”.

Defining the Cluster Virtual IP Addresses

In Figure 4-1, the IP address of the cluster is 192.168.10.100.

The cluster has one external virtual IP address and one internal virtual IP address. The external IP address is 192.168.10.100, and the internal IP address is 10.10.0.100.

The Synchronization Network

State Synchronization between cluster members ensures that if there is a failover, connections that were handled by the failed machine will be maintained. The synchronization network is used to pass connection synchronization and other state information between cluster members. This network therefore carries all the most sensitive security policy information in the organization, and so it is important to make sure the network is secure. It is possible to define more than one synchronization network for backup purposes.

To secure the synchronization interfaces, they should be directly connected by a cross cable, or in the case of a three of more member cluster, by means of a dedicated hub or switch.

Machines in a Load Sharing cluster must be synchronized because synchronization is used in normal traffic flow. Machines in a High Availability cluster do not have to be synchronized, though if they are not, connections may be lost upon failover. Figure 4-1 shows a synchronization interface with a unique IP address on each machine. 10.0.10.1 on Member_A and 10.0.10.2 on Member_B.
Configuring Cluster Addresses on Different Subnets

Only one routable IP address is required in a ClusterXL cluster, for the virtual cluster interface that faces the Internet. All cluster member physical IP addresses can be non-routable.

Configuring different subnets for the cluster IP addresses and the member addresses is useful in order to:

- Enable a multi-machine cluster to replace a single-machine gateway in a pre-configured network, without the need to allocate new addresses to the cluster members.
- Allow organizations to use only one routable address for the ClusterXL Gateway Cluster. This saves routable addresses.

For details, see “Configuring Cluster Addresses on Different Subnets” on page 207.
ClusterXL Modes

In This Section

- Introduction to ClusterXL Modes page 56
- Load Sharing Multicast Mode page 57
- Load Sharing Unicast Mode page 59
- New High Availability Mode page 61
- Mode Comparison Table page 63

Introduction to ClusterXL Modes

ClusterXL has four working modes. This section briefly describes each mode and its relative advantages and disadvantages.

- Load Sharing Multicast Mode
- Load Sharing Unicast Mode
- New High Availability Mode
- High Availability Legacy Mode

High Availability Legacy Mode is discussed in the Appendix chapter: “High Availability Legacy Mode” on page 223. It is recommended that you use High Availability New Mode to avoid problems with backward compatibility.

Note - All examples in the section refer to the ClusterXL configuration shown in Figure 4-1 on page 53.
Load Sharing Multicast Mode

Load Sharing enables you to distribute network traffic between cluster members. In contrast to High Availability, where only a single member is active at any given time, all cluster members in a Load Sharing solution are active, and the cluster is responsible for assigning a portion of the traffic to each member. This assignment is the task of a decision function, which examines each packet going through the cluster, and determines which member should handle it. Thus, a Load Sharing cluster utilizes all cluster members, which usually leads to an increase in its total throughput. See Figure 4-1 on page 53 for an example of a typical ClusterXL configuration.

It is important to understand that ClusterXL Load Sharing, when combined with State Synchronization, provides a full High Availability solution as well. When all cluster members are active, traffic is evenly distributed between the machines. In case of a failover event, caused by a problem in one of the members, the processing of all connections handled by the faulty machine is immediately taken over by the other members.

ClusterXL offers two separate Load Sharing solutions: Multicast and Unicast. The two modes differ in the way members receive the packets sent to the cluster. This section describes the Multicast mode. For a description of Unicast mode see “Load Sharing Unicast Mode” on page 59.

The Multicast mechanism, which is provided by the Ethernet network layer, allows several interfaces to be associated with a single physical (MAC) address. Unlike Broadcast, which binds all interfaces in the same subnet to a single address, Multicast enables grouping within networks. This means that it is possible to select the interfaces within a single subnet that will receive packets sent to a given MAC address.

ClusterXL uses the Multicast mechanism to associate the virtual cluster IP addresses with all cluster members. By binding these IP addresses to a Multicast MAC address, it ensures that all packets sent to the cluster, acting as a gateway, will reach all members in the cluster. Each member then decides whether it should process the packets or not. This decision is the core of the Load Sharing mechanism: it has to assure that at least one member will process each packet (so that traffic is not blocked), and that no two members will handle the same packets (so that traffic is not duplicated).

An additional requirement of the decision function is to route each connection through a single gateway, to ensure that packets that belong to a single connection will be processed by the same member. Unfortunately, this requirement cannot
always be enforced, and in some cases, packets of the same connection will be handled by different members. ClusterXL handles these situations using its State Synchronization mechanism, which mirrors connections on all cluster members.

**Example**

This scenario describes a user logging from the Internet to a web server behind the Firewall cluster that is configured in Load Sharing Multicast mode.

1. The user requests a connection from 192.168.10.78 (his computer) to 10.10.0.34 (the web server).
2. A router on the 192.168.10.x network recognizes 192.168.10.100 (the cluster's virtual IP address) as the gateway to the 10.10.0.x network.
3. The router issues an ARP request to 192.168.10.100.
4. One of the active members intercepts the ARP request, and responds with the Multicast MAC assigned to the cluster IP address of 192.168.10.100.
5. When the web server responds to the user requests, it recognizes 10.10.0.100 as its gateway to the Internet.
6. The web server issues an ARP request to 10.10.0.100.
7. One of the active members intercepts the ARP request, and responds with the Multicast MAC address assigned to the cluster IP address of 10.10.0.100.
8. All packets sent between the user and the web server reach every cluster member, which decides whether to handle or drop each packet.
9. When a failover occurs, one of the cluster members goes down. However, traffic still reaches all of the active cluster members, and hence there is no need to make changes in the network's ARP routing. All that changes is the cluster's decision function, which takes into account the new state of the members.
Load Sharing Unicast Mode

Load Sharing Unicast mode provides a Load Sharing solution adapted to environments where Multicast Ethernet cannot operate. In this mode a single cluster member, referred to as Pivot, is associated with the cluster's virtual IP addresses, and is thus the only member to receive packets sent to the cluster. The pivot is then responsible for propagating the packets to other cluster members, creating a Load Sharing mechanism. Distribution is performed by applying a decision function on each packet, the same way it is done in Load Sharing Multicast mode. The difference is that only one member performs this selection: any non-pivot member that receives a forwarded packet will handle it, without applying the decision function. Note that non-pivot members are still considered as "active", since they perform routing and Firewall tasks on a share of the traffic (although they do not perform decisions.).

Even though the pivot member is responsible for the decision process, it still acts as a Security Gateway that processes packets (for example, the decision it makes can be to handle a packet on the local machine). However, since its additional tasks can be time consuming, it is usually assigned a smaller share of the total load.

When a failover event occurs in a non-pivot member, its handled connections are redistributed between active cluster members, providing the same High Availability capabilities of New High Availability and Load Sharing Multicast. When the pivot member encounters a problem, a regular failover event occurs, and, in addition, another member assumes the role of the new pivot. The pivot member is always the active member with the highest priority. This means that when a former pivot recuperates, it will retain its previous role.

See Figure 4-1 on page 53 for an example of a typical ClusterXL configuration.
Example

In this scenario, we use a Load Sharing Unicast cluster as the gateway between the user’s computer and the web server.

1. The user requests a connection from 192.168.10.78 (his computer) to 10.10.0.34 (the web server).

2. A router on the 192.168.10.x network recognizes 192.168.10.100 (the cluster's virtual IP address) as the gateway to the 10.10.0.x network.

3. The router issues an ARP request to 192.168.10.100.

4. The pivot member intercepts the ARP request, and responds with the MAC address that corresponds to its own unique IP address of 192.168.10.1.

5. When the web server responds to the user requests, it recognizes 10.10.0.100 as its gateway to the Internet.

6. The web server issues an ARP request to 10.10.0.100.

7. The pivot member intercepts the ARP request, and responds with the MAC address that corresponds to its own unique IP address of 10.10.0.1.

8. The user's request packet reaches the pivot member on interface 192.168.10.1.

9. The pivot decides that the second member should handle this packet, and forwards it to 192.168.10.2.

10. The second member recognizes the packet as a forwarded one, and processes it.

11. Further packets are processed by either the pivot member, or forwarded and processed by the non-pivot member.

12. When a failover occurs on the pivot, the second member assumes the role of pivot.

13. The new pivot member sends gratuitous ARP requests to both the 192.168.10.x and the 10.10.0.x networks. These requests associate the virtual IP address of 192.168.10.100 with the MAC address that correspond to the unique IP address of 192.168.10.2, and the virtual IP address of 10.10.0.100 with the MAC address that correspond to the unique IP address of 10.10.0.2.

14. Traffic sent to the cluster is now received by the new pivot, and processed by the local machine (as it is currently the only active machine in the cluster).

15. When the first machine recovers, it re-assumes the role of pivot, by associating the cluster IP addresses with its own unique MAC addresses.
New High Availability Mode

The New High Availability Mode provides basic High-Availability capabilities in a cluster environment. This means that the cluster can provide Firewall services even when it encounters a problem, which on a stand-alone gateway would have resulted in a complete loss of connectivity. When combined with Check Point's State Synchronization, ClusterXL High Availability can maintain connections through failover events, in a user-transparent manner, allowing a flawless connectivity experience. Thus, High-Availability provides a backup mechanism, which organizations can use to reduce the risk of unexpected downtime, especially in a mission-critical environment (such as one involving money transactions over the Internet.)

To achieve this purpose, ClusterXL's New High Availability mode designates one of the cluster members as the active machine, while the rest of the members are kept in a stand-by mode. The cluster's virtual IP addresses are associated with the physical network interfaces of the active machine (by matching the virtual IP address with the unique MAC address of the appropriate interface). Thus, all traffic directed at the cluster is actually routed (and filtered) by the active member. The role of each cluster member is chosen according to its priority, with the active member being the one with the highest ranking. Member priorities correspond to the order in which they appear in the Cluster Members page of the Gateway Cluster Properties window. The top-most member has the highest priority. You can modify this ranking at any time.

In addition to its role as a Firewall gateway, the active member is also responsible for informing the stand-by members of any changes to its connection and state tables, keeping these members up-to-date with the current traffic passing through the cluster.

Whenever the cluster detects a problem in the active member that is severe enough to cause a failover event, it passes the role of the active member to one of the standby machines (the member with the currently highest priority). If State Synchronization is applied, any open connections are recognized by the new active machine, and are handled according to their last known state. Upon the recovery of a member with a higher priority, the role of the active machine may or may not be switched back to that member, depending on the user's configuration.

It is important to note that the cluster may encounter problems in standby machines as well. In this case, these machines are not considered for the role of active members, in the event of a failover.

See Figure 4-1, “Example ClusterXL Topology,” on page 53 for an example of a typical ClusterXL configuration.
**Example**

This scenario describes a user logging from the Internet to a web server behind the Firewall cluster.

1. The user requests a connection from 192.168.10.78 (his computer) to 10.10.0.34 (the web server).

2. A router on the 192.168.10.x network recognizes 192.168.10.100 (the cluster's virtual IP address) as the gateway to the 10.10.0.x network.

3. The router issues an ARP request to 192.168.10.100.

4. The active member intercepts the ARP request, and responds with the MAC address that corresponds to its own unique IP address of 192.168.10.1.

5. When the web server responds to the user requests, it recognizes 10.10.0.100 as its gateway to the Internet.

6. The web server issues an ARP request to 10.10.0.100.

7. The active member intercepts the ARP request, and responds with the MAC address that corresponds to its own unique IP address of 10.10.0.1.

8. All traffic between the user and the web server is now routed through the active member.

9. When a failover occurs, the standby member concludes that it should now replace the faulty active member.

10. The stand-by member sends gratuitous ARP requests to both the 192.168.10.x and the 10.10.0.x networks. These requests associate the virtual IP address of 192.168.10.100 with the MAC address that correspond to the unique IP address of 192.168.10.2, and the virtual IP address of 10.10.0.100 with the MAC address that correspond to the unique IP address of 10.10.0.2.

11. The stand-by member has now switched to the role of the active member, and all traffic directed through the cluster is routed through this machine.

12. The former active member is now considered to be “down”, waiting to recover from whatever problem that had caused the failover event.
**Mode Comparison Table**

Table 4-1 summarizes the similarities and differences between the ClusterXL modes.

<table>
<thead>
<tr>
<th></th>
<th>Legacy High Availability</th>
<th>New High Availability</th>
<th>Load Sharing Multicast</th>
<th>Load Sharing Unicast</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Availability</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Load Sharing</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Performance</td>
<td>Good</td>
<td>Good</td>
<td>Excellent</td>
<td>Very Good</td>
</tr>
<tr>
<td>Hardware Support</td>
<td>All</td>
<td>All</td>
<td>Not all routers are supported</td>
<td>All</td>
</tr>
<tr>
<td>SecureXL Support</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes, with Performance Pack or SecureXL Turbocard.</td>
<td>Yes</td>
</tr>
<tr>
<td>State Synchronization Mandatory</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>VLAN Tagging Support¹</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

¹ For further details, refer to the *Release Notes*, available online at: [http://support.checkpoint.com](http://support.checkpoint.com).
Failover

In This Section

What is a Failover? page 64
When Does a Failover Occur? page 65
What Happens When a Gateway Recovers? page 65
How a Recovered Cluster Member Obtains the Security Policy page 66

What is a Failover?

A failover occurs when a Gateway is no longer able to perform its designated functions. When this happens another Gateway in the cluster assumes the failed Gateway's responsibilities.

In a Load Sharing configuration, if one Security Gateway in a cluster of gateways goes down, its connections are distributed among the remaining Gateways. All gateways in a Load Sharing configuration are synchronized, so no connections are interrupted.

In a High Availability configuration, if one Gateway in a synchronized cluster goes down, another Gateway becomes active and "takes over" the connections of the failed Gateway. If you do not use State Synchronization, existing connections are closed when failover occurs, although new connections can be opened.

To tell each cluster member that the other gateways are alive and functioning, the ClusterXL Cluster Control Protocol maintains a heartbeat between cluster members. If a certain predetermined time has elapsed and no message is received from a cluster member, it is assumed that the cluster member is down and a failover occurs. At this point another cluster member automatically assumes the responsibilities of the failed cluster member.

It should be noted that a cluster machine may still be operational but if any of the above checks fail in the cluster, then the faulty member initiates the failover because it has determined that it can no longer function as a cluster member.

Note that more than one cluster member may encounter a problem that will result in a failover event. In cases where all cluster members encounter such problems, ClusterXL will try to choose a single member to continue operating. The state of the chosen member will be reported as Active Attention. This situation lasts until
another member fully recovers. For example, if a cross cable connecting the cluster members malfunctions, both members will detect an interface problem. One of them will change to the Down state, and the other to Active Attention.

**When Does a Failover Occur?**

A failover takes place when one of the following occurs on the active cluster member:

- Any critical device (such as fwd) fails. A critical device is a process running on a cluster member that enables the member to notify other cluster members that it can no longer function as a member. The device reports to the ClusterXL mechanism regarding its current state or it may fail to report, in which case ClusterXL decides that a failover has occurred and another cluster member takes over.
- An interface or cable fails.
- The machine crashes.
- The Security Policy is uninstalled. When the Security Policy is uninstalled the Gateway can no longer function as a firewall. If it cannot function as a firewall, it can no longer function as a cluster member and a failover occurs. Normally a policy is not uninstalled by itself but would be initiated by a user.

**What Happens When a Gateway Recovers?**

In a Load Sharing configuration, when the failed Gateway in a cluster recovers, all connections are redistributed among all active members.

In a High Availability configuration, when the failed Gateway in a cluster recovers, the recovery method depends on the configured cluster setting. The options are:

- **Maintain Current Active Gateway** means that if one machine passes on control to a lower priority machine, control will be returned to the higher priority machine only if the lower priority machine fails. This mode is recommended if all members are equally capable of processing traffic, in order to minimize the number of failover events.
- **Switch to Higher Priority Gateway** means that if the lower priority machine has control and the higher priority machine is restored, then control will be returned to the higher priority machine. This mode is recommended if one member is better equipped for handling connections, so it will be the default gateway.
How a Recovered Cluster Member Obtains the Security Policy

The administrator installs the security policy on the cluster rather than separately on individual cluster members. The policy is automatically installed on all cluster members. The policy is sent to the IP address defined in the General Properties page of the cluster member object.

When a failed cluster member recovers, it will first try to take a policy from one of the other cluster members. The assumption is that the other cluster members have a more up to date policy. If this does not succeed, it compares its own local policy to the policy on the Security Management server. If the policy on the Security Management server is more up to date than the one on the cluster member, the policy on the Security Management server will be retrieved. If the cluster member does not have a local policy, it retrieves one from the Security Management server. This ensures that all cluster members use the same policy at any given moment.
Implementation Planning Considerations

In This Section

- High Availability or Load Sharing... page 67
- Choosing the Load Sharing Mode... page 67
- IP Address Migration... page 68

High Availability or Load Sharing

Whether to choose a Load Sharing (Active/Active) or a High Availability (Active/Standby) configuration depends on the need and requirements of the organization. A High Availability gateway cluster ensures fail-safe connectivity for the organization. Load Sharing provides the additional benefit of increasing performance.

Note - When working on a sync network, it is recommended to use a NIC with the same bandwidth as the NICs that are used for general traffic.

Choosing the Load Sharing Mode

Load Sharing Multicast mode is an efficient way to handle a high load because the load is distributed optimally between all cluster members. However, not all routers can be used for Load Sharing Multicast mode. Load Sharing Multicast mode associates a multicast MAC with each unicast cluster IP address. This ensures that traffic destined for the cluster is received by all members. The ARP replies sent by a cluster member will therefore indicate that the cluster IP address is reachable via a multicast MAC address.

Some routing devices will not accept such ARP replies. For some routers, adding a static ARP entry for the cluster IP address on the routing device will solve the issue. Other routers will not accept this type of static ARP entry.

Another consideration is whether your deployment includes routing devices with interfaces operating in promiscuous mode. If on the same network segment there exists two such routers and a ClusterXL gateway in Load Sharing Multicast mode, traffic destined for the cluster that is generated by one of the routers could also be processed by the other router.
For these cases, use Load Sharing Unicast mode, which does not require the use of multicast for the cluster addresses.

For a list of supported hardware devices see “ClusterXL Hardware Compatibility” on page 72.

**IP Address Migration**

If you wish to provide High Availability or Load Sharing to an existing single gateway configuration, it is recommended to take the existing IP addresses from the current gateway, and make these the cluster addresses (cluster virtual addresses), when feasible. Doing so will avoid altering current IPSec endpoint identities, as well keep Hide NAT configurations the same in many cases.
Hardware Requirements, Compatibility and Cisco Example

In This Section

ClusterXL Hardware Requirements  page 69
ClusterXL Hardware Compatibility  page 72
Example Configuration of a Cisco Catalyst Routing Switch  page 73

ClusterXL Hardware Requirements

The Gateway Cluster is usually located in an environment having other networking devices such as switches and routers. These devices and the Gateways must interact to assure network connectivity. This section outlines the requirements imposed by ClusterXL on surrounding networking equipment.

In This Section

HA New and Load Sharing Unicast Modes  page 69
Load Sharing Multicast Mode  page 71

HA New and Load Sharing Unicast Modes

Multicast mode is the default Cluster Control Protocol (CCP) mode in High Availability New Mode and Load Sharing Unicast Mode (and also Load Sharing Multicast Mode). When using CCP in multicast mode, the following settings should be configured on the switch.
The following settings should be configured on the router:

### Table 4-2  Switch Setting for High Availability New Mode and Load Sharing

<table>
<thead>
<tr>
<th>Switch Setting</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGMP and Static CAMs</td>
<td>ClusterXL does not support IGMP registration (also known as IGMP Snooping) by default. Either disable IGMP registration in switches that rely on IGMP packets to configure their ports, or enable IGMP registration on ClusterXL. For instructions on how to enable IGMP snooping, refer to the ClusterXL IGMP Membership document at <a href="http://downloads.checkpoint.com/dc/download.htm?ID=6699">http://downloads.checkpoint.com/dc/download.htm?ID=6699</a>. In situations where disabling IGMP registration is not acceptable, it is necessary to configure static CAMs in order to allow multicast traffic on specific ports.</td>
</tr>
<tr>
<td>Disabling multicast limits</td>
<td>Certain switches have an upper limit on the number of broadcasts and multicasts that they can pass, in order to prevent broadcast storms. This limit is usually a percentage of the total interface bandwidth. It is possible to either turn off broadcast storm control, or to allow a higher level of broadcasts or multicasts through the switch. If the connecting switch is incapable of having any of these settings configured, it is possible, though less efficient, for the switch to use broadcast CCP (described in “Choosing the CCP Transport Mode on the Cluster Members” on page 83).</td>
</tr>
</tbody>
</table>

The following settings should be configured on the router:

### Table 4-3  Router Setting for High Availability New Mode and Load Sharing Unicast Mode

<table>
<thead>
<tr>
<th>Router Setting</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unicast MAC</td>
<td>When working in High Availability Legacy mode, High Availability New mode and Load Sharing Unicast mode, the Cluster IP address is mapped to a regular MAC address, which is the MAC address of the active member. The router needs to be able to learn this MAC through regular ARP messages.</td>
</tr>
</tbody>
</table>
Load Sharing Multicast Mode

When working in Load Sharing Multicast mode, the switch settings are as follows:

Table 4-4  Switch Configuration for Load Sharing Multicast Mode

<table>
<thead>
<tr>
<th>Switch Setting</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCP in Multicast mode</td>
<td>Multicast mode is the default Cluster Control Protocol mode in Load Sharing Multicast. For details of the required switch settings, see Table 4-2 on page 70.</td>
</tr>
<tr>
<td>Port Mirroring</td>
<td>ClusterXL does not support the use of unicast MAC addresses with Port Mirroring for Multicast Load Sharing solutions.</td>
</tr>
</tbody>
</table>

When working in Load Sharing Multicast mode, the router must support sending unicast IP packets with Multicast MAC addresses. This is required so that all cluster members will receive the data packets.

The following settings may need to be configured in order to support this mode, depending on the model of the router:

Table 4-5  Router Configuration for Load Sharing Multicast Mode

<table>
<thead>
<tr>
<th>Router Setting</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static MAC</td>
<td>Most routers can learn ARP entries with a unicast IP and a multicast MAC automatically using the ARP mechanism. If you have a router that is not able to learn this type of mapping dynamically, you'll have to configure static MAC entries.</td>
</tr>
<tr>
<td>IGMP and static cams</td>
<td>Some routers require disabling of IGMP snooping or configuration of static cams in order to support sending unicast IP packets with Multicast MAC addresses.</td>
</tr>
<tr>
<td>Disabling multicast limits</td>
<td>Certain routers have an upper limit on the number of broadcasts and multicasts that they can pass, in order to prevent broadcast storms. This limit is usually a percentage of the total interface bandwidth. It is possible to either turn off broadcast storm control, or to allow a higher level of broadcasts or multicasts through the router.</td>
</tr>
<tr>
<td>Disabling forwarding multicast traffic to the router</td>
<td>Some routers will send multicast traffic to the router itself. This may cause a packet storm through the network and should be disabled.</td>
</tr>
</tbody>
</table>
ClusterXL Hardware Compatibility

The following routers and switches are known to be compatible for all ClusterXL modes:

**Routers**
- Cisco 7200 Series
- Cisco 1600, 2600, 3600 Series

**Routing Switch**
- Extreme Networks Blackdiamond (Disable IGMP snooping)
- Extreme Networks Alpine 3800 Series (Disable IGMP snooping)
- Foundry Network Bigiron 4000 Series
- Nortel Networks Passport 8600 Series
- Cisco Catalyst 6500 Series (Disable IGMP snooping, Configure Multicast MAC manually)

**Switches**
- Cisco Catalyst 2900, 3500 Series
- Nortel BayStack 450
- Alteon 180e
- Dell PowerConnect 3248 and PowerConnect 5224
Example Configuration of a Cisco Catalyst Routing Switch

The following example shows how to perform the configuration commands needed to support ClusterXL on a Cisco Catalyst 6500 Series routing switch. For more details, or instructions for other networking devices, please refer to the device vendor documentation.

The example refers to the sample configuration described in Figure 4-2 on page 84.

**Disabling IGMP Snooping**

To disable IGMP snooping run:

```bash
no ip igmp snooping
```

**Defining Static Cam Entries**

To add a permanent multicast entry to the table for module 1, port 1, and module 2, ports 1, 3, and 8 through 12:

```bash
Console> (enable) set cam permanent 01-40-5e-28-0a-64 1/1,2/1,2/3,2/8-12
Permanent multicast entry added to CAM table.
```

Determining the MAC addresses which needs to be set is done by using the following procedure:

On a network that has a cluster IP address of x.y.z.w:

If y<=127, the multicast MAC address would be 01:00:5e:y:z:w. For example: 01:00:5e:5a:0a:64 for 192.90.10.100

If y>127, the multicast MAC address would be 01:00:5e:(y-128):z:w. For example: 01:00:5e:28:0a:64 for 192.168.10.100 (168-128=40 = 28 in hex).

For a network x.y.z.0 that does not have a cluster IP address, such as the sync, you would use the same procedure, and substitute fa instead of 0 for the last octet of the MAC.

For example: 01:00:5e:00:00:fa for the 10.0.0.X network.
Disabling Multicast Limits
To disable multicast limits run:
\texttt{no storm-control multicast level}

Configuring a Static ARP Entry on the Router
To define a static ARP entry, run:
\texttt{arp 192.168.10.100 01:00:5E:28:0A:64 arpa}

Determining the MAC address is done using the procedure described in \textit{Defining Static Cam Entries}.

Disabling Multicast Packets from Reaching the Router
To prevent multicast packets from reaching the router, run:
\texttt{set cam static 01:00:5E:28:0A:64 module/port}

Determining the MAC address is done using the procedure described in \textit{“Defining Static Cam Entries”}. 
Check Point Software Compatibility

In This Section

Operating System Compatibility  page 75
Check Point Software Compatibility (excluding IPS)  page 75
ClusterXL Compatibility with IPS  page 79
Forwarding Layer  page 80

Operating System Compatibility

The operating systems listed in Table 4-6 are supported by ClusterXL, with the limitations listed in the notes below. For details on the supported versions of these operating systems, see the Release Notes, available online at: http://support.checkpoint.com.

Table 4-6  ClusterXL Operating System Compatibility

<table>
<thead>
<tr>
<th>Operating System</th>
<th>Load Sharing</th>
<th>High Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Check Point SecurePlatform (1)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes
1. VLANs are supported on all interfaces.

Check Point Software Compatibility (excluding IPS)

Table 4-7 lists the products and features that are either not supported (marked as No), or are only partially supported with ClusterXL (marked as Yes, with a note). It does not apply to their use with OPSEC-certified clustering products.

Table 4-7  Products and features that are not fully supported with ClusterXL

<table>
<thead>
<tr>
<th>Feature or Product</th>
<th>Feature</th>
<th>Load Sharing</th>
<th>High Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Security Management</td>
<td>Authentication/Security</td>
<td>Yes (1.)</td>
<td>Yes (1.) (9.)</td>
</tr>
<tr>
<td></td>
<td>Servers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firewall</td>
<td>ACE servers and</td>
<td>Yes (7.)</td>
<td>Yes (7.)</td>
</tr>
<tr>
<td></td>
<td>SecurID</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Check Point Software Compatibility (excluding IPS)  page 75
ClusterXL Compatibility with IPS  page 79
Forwarding Layer  page 80
1. Since it requires per-packet state tracking, this feature cannot be guaranteed when a session starts on one cluster member and fails over to another.

2. Application Intelligence protocol inspection includes the general HTTP worm catcher, configuration of Optimized Protocol Enforcement, and Microsoft networks inspection.

3. Application Intelligence protocol inspection is supported when connections maintain unidirectional stickiness. Unidirectional stickiness means that packets in the client-to-server direction are handled by one cluster member, while packets in the server-to-client direction are handled by another cluster member. OPSEC cluster solutions must maintain at least unidirectional stickiness for all connections in order to qualify as OPSEC clusters.

Failover can break unidirectional stickiness for certain connections, and in that case, the Security Gateway will proactively reset these connections.

<table>
<thead>
<tr>
<th>Feature or Product</th>
<th>Feature</th>
<th>Load Sharing</th>
<th>High Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firewall</td>
<td>Application Intelligence protocol inspection (2.)</td>
<td>Yes (3.)</td>
<td>Yes</td>
</tr>
<tr>
<td>Firewall</td>
<td>Sequence Verifier</td>
<td>Yes (4.)</td>
<td>Yes (1.)</td>
</tr>
<tr>
<td>Firewall</td>
<td>UDP encapsulation</td>
<td>Yes (7.)</td>
<td>Yes</td>
</tr>
<tr>
<td>Firewall</td>
<td>SAM</td>
<td>Yes (8.)</td>
<td>Yes (8.)</td>
</tr>
<tr>
<td>Firewall</td>
<td>ISP Redundancy</td>
<td>Yes (11.) (12.)</td>
<td>Yes (11.) (13.)</td>
</tr>
<tr>
<td>VPN</td>
<td>Third party VPN peers</td>
<td>Yes (16.)</td>
<td>Yes</td>
</tr>
<tr>
<td>Endpoint Security Client</td>
<td>Software Distribution Server (SDS)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Endpoint Security Client</td>
<td>IP per user in Office Mode</td>
<td>Yes (10.)</td>
<td>Yes (10.)</td>
</tr>
<tr>
<td>SecureXL (hardware acceleration</td>
<td>Yes (11.) (15.)</td>
<td>Yes (11.)</td>
<td></td>
</tr>
<tr>
<td>SecureXL (hardware acceleration</td>
<td>or Performance Pack</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Check Point QoS</td>
<td></td>
<td>Yes (4.) (5.)</td>
<td>Yes</td>
</tr>
<tr>
<td>SmartProvisioning</td>
<td>SmartLSM Security Gateway</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Check Point Security Gateway</td>
<td></td>
<td>Yes (6.)</td>
<td>Yes</td>
</tr>
</tbody>
</table>
4. Supported when connections maintain bidirectional stickiness. Bidirectional stickiness is the situation where all packets of a connection, regardless of whether they are in the client-to-server direction or the server-to-client direction, are processed by a single cluster member.

5. Supported with bandwidth limits and guarantees that are manually divided between the members. With a 1.5 Mbps connection, and a three-member cluster, each member would have a bandwidth of 500 Kbps, and limits of 1/3 of the total. If a cluster member fails, the total bandwidth will not be automatically re-allocated among the remaining members.

6. Using OPSEC partners platform.

7. Configuration instructions for ACE server in Cluster environment:

   **High Availability:** To support failover scenarios, manually copy the secured file, which is created after the first authentication with the ACE server, from the initiating member to all other members.

   **Load Sharing:**
   - Every cluster member should be defined separately on the server with its unique IP address.
   - Add the following entry to the `tables.def` file on the Security Management server:
     ```
     no_hide_services_ports = {.., <5500, 17> };
     ```
     This forces the connection from the cluster members to the ACE server to go out with the member's IP address and not the Cluster address. Make sure the IP addresses of the cluster members are routable from the ACE server box, and then install the Security Policy.
   - In some cases the agent libraries (client side) will use the wrong interface IP address in the decryption, and the authentication will fail. To overcome this problem, place a new text file `sdopts.rec` in the same directory as the `dconf.rec` file, with the following line
     ```
     CLIENT_IP=x.x.x.x
     ```
     where x.x.x.x is the primary IP address, as defined on the server. This is the IP address of the interface to which the server is routed.

8. Works as two single gateways. SAM commands executed while a cluster is down are not enforced on this member.

9. In a High Availability configuration, client authentication Wait mode is not reliable. Use other client authentication modes instead.

10. The `ipassignment.conf` file must be copied manually.
11. ISP Redundancy is not supported if cluster addresses are configured on different subnets.

12. ISP redundancy works with ClusterXL in Load Sharing Unicast mode only if SecureXL is enabled.

13. Not supported in Legacy Mode.

14. For SecureXL hardware-based acceleration support consult the third party vendor.

15. Sticky Decision Function must be disabled.

16. If the VPN peer device supports only one Security Association (SA), the Sticky Decision Function must be enabled. Examples for such peers are Access VPN with Microsoft IPSec (L2TP), and Cisco VPN routers.
ClusterXL Compatibility with IPS

The IPS features listed in Table 4-8 are supported by ClusterXL, with the limitations listed in the notes.

Table 4-8  ClusterXL Compatibility with IPS

<table>
<thead>
<tr>
<th>Feature</th>
<th>Load Sharing</th>
<th>High Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragment Sanity Check</td>
<td>Yes (1, 3)</td>
<td>Yes (1)</td>
</tr>
<tr>
<td>Pattern Matching</td>
<td>Yes (2, 3)</td>
<td>Yes (2)</td>
</tr>
<tr>
<td>Sequence Verifier</td>
<td>Yes (2, 4)</td>
<td>Yes (2)</td>
</tr>
<tr>
<td>FTP, HTTP and SMTP Security Servers</td>
<td>Yes (2, 5)</td>
<td>Yes (2)</td>
</tr>
</tbody>
</table>

Notes
1. If there is a failover when fragments are being received, the packet will be lost.
2. Does not survive failover.
3. Requires unidirectional stickiness. This means that the same member must receive all external packets, and the same member must receive all internal packets, but the same member does not have to receive both internal and external packets.
4. Requires bidirectional connection stickiness.
5. Uses the forwarding layer, described in the next section.
The Forwarding Layer is a ClusterXL mechanism that allows a cluster member to pass packets to other members, after they have been locally inspected by the Firewall. This feature allows connections to be opened from a cluster member to an external host.

Packets originated by cluster members are hidden behind the cluster's virtual IP. Thus, a reply from an external host is sent to the cluster, and not directly to the source member. This can pose problems in the following situations:

- The cluster is working in New High Availability mode, and the connection is opened from the stand-by machine. All packets from the external host are handled by the active machine, instead.

- The cluster is working in a Load Sharing mode, and the decision function has selected another member to handle this connection. This can happen since packets directed at a cluster IP are distributed among cluster members as with any other connection.

If a member decides, upon the completion of the Firewall inspection process, that a packet is intended for another cluster member, it can use the Forwarding Layer to hand the packet over to that destination. This is done by sending the packet over a secured network (any subnet designated as a Synchronization network) directly to that member. It is important to use secured networks only, as encrypted packets are decrypted during the inspection process, and are forwarded as clear-text (unencrypted) data.

Packets sent on the Forwarding Layer use a special source MAC address to inform the receiving member that they have already been inspected by another Security Gateway. Thus, the receiving member can safely hand over these packets to the local Operating System, without further inspection. This process is secure, as Synchronization Networks should always be isolated from any other network (using a dedicated network).
Configuring ClusterXL

In This Section

- Configuring Routing for the Client Machines  page 81
- Preparing the Cluster Member Machines  page 82
- Choosing the CCP Transport Mode on the Cluster Members  page 83
- SmartDashboard Configuration  page 84

This procedure describes how to configure the Load Sharing Multicast, Load Sharing Unicast, and High Availability New Modes modes from scratch. Their configuration is identical, apart from the mode selection in SmartDashboard Gateway Cluster object or Gateway Cluster creation wizard. Figure 4-2 is used to illustrate the configuration steps.

Note - To configure High Availability Legacy Mode, see “High Availability Legacy Mode” on page 223

Configuring Routing for the Client Machines

1. Configure routing so that communication with the networks on the internal side of the cluster is via the cluster IP address on the external side of the cluster. For example, in Figure 4-2 on page 84, on the external router, configure a static route such that network 10.10.0.0 is reached via 192.168.10.100.

2. Configure routing so that communication with the networks on the external side of the cluster is via the cluster IP address on the internal side of the cluster. For example, in Figure 4-2 on page 84, define 10.10.0.100 as the default gateway on each machine on the internal side of the router.
Preparing the Cluster Member Machines


2. Define IP addresses for each interfaces on all cluster members. For example, in Figure 4-2 on page 84,
   - on Member_A configure the Int Interface with address 10.10.0.1, the Ext interface with address 192.168.10.1, and the SYNC interface with address 10.0.10.1
   - on Member_B configure the Int Interface with address 10.10.0.2, the Ext interface with address 192.168.10.2, and the SYNC interface with address 10.0.10.2

3. For a VPN cluster to properly function, the cluster member clocks must be accurately synchronized to within a second of each other. On cluster members that are constantly up and running it is usually enough to set the time once. More reliable synchronization can be achieved using NTP or some other time synchronization services supplied by the operating system. The cluster member clocks are not relevant for any other (non VPN) cluster capability.

4. Connect the cluster network machines, via the switches. For the Synchronization interfaces, use a cross cable, or a dedicated switch. Make sure that each network (internal, external, Synchronization, DMZ, and so on) is configured on a separate VLAN, switch or hub.

   **Note** - It is possible to run synchronization across a WAN. For details, see “Synchronizing Clusters over a Wide Area Network” on page 36.

5. Install Check Point Security Gateway on all cluster members.

6. During the configuration phase, enable ClusterXL and State Synchronization by selecting Enable cluster membership for this gateway on Unix machines, or This Gateway is part of a cluster on Windows.

   If you do not make this selection during installation, you can use the Check Point Configuration Tool at any time. Run the cpconfig utility from the command line, and select the option to turn on cluster capabilities on the gateway. Note that on some platforms you may be asked to reboot.
Choosing the CCP Transport Mode on the Cluster Members

If the connecting switch is incapable of forwarding multicast, it is possible, though less efficient, for the switch to use broadcast to forward traffic.

The ClusterXL Control Protocol (CCP) on the cluster members uses multicast by default, because it is more efficient than broadcast. To toggle the CCP mode between broadcast and multicast, use the following command on each cluster member:

cphaconf set_ccp broadcast/multicast
SmartDashboard Configuration

Figure 4-2 relates the physical cluster topology to the required SmartDashboard configuration.

When configuring a ClusterXL cluster in SmartDashboard, you use the Cluster object **Topology** page to configure the topology for both *cluster* and *cluster member*. The cluster IP addresses are virtual, in other words, they do not belong to any physical interface. One (or more) interfaces of each cluster member will be in the synchronization network.

Figure 4-2  Example ClusterXL topology and configuration

To define a new Gateway Cluster object, right click the **Network Objects** tree, and choose **New Check Point > Gateway Cluster…**. Configuration of the Gateway Cluster Object can be performed using

- **Simple Mode (Wizard)** which guides you step by step through the configuration process. See the online help for further assistance.
- **Classic Mode**, described below.
**Classic Mode Configuration**

1. In the General tab of the Gateway Cluster object, check ClusterXL as a product installed on the cluster.

2. Define the general IP address of the cluster. Define it to be the same as the IP address of one of the virtual cluster interfaces.

3. In the Cluster Members page, click Add > New Cluster Member to add cluster members to the cluster. Cluster members exist solely inside the Gateway Cluster object. For each cluster member:
   
   - In the Cluster Members Properties window General tab, define a Name and IP Address. Choose an IP address that is routable from the Security Management server so that the Security Policy installation will be successful. This can be an internal or an external address, or a dedicated management interface.
   
   - Click Communication, and Initialize Secure Internal Communication (SIC).
   
   - Define the NAT and VPN tabs, as required.

   You can also add an existing gateway as a cluster member by selecting Add > Add Gateway to Cluster in the Cluster Members page and selecting the gateway from the list in the Add Gateway to Cluster window.

   If you want to remove a gateway from the cluster, click Remove in the Cluster Members page and select Detach Member from Cluster or right-click on the cluster member in the Network Objects tree and select Detach from Cluster.

4. In the ClusterXL page (Figure 4-3), select either

   - **High Availability New Mode**, and specify the action Upon Gateway Recovery. See “What Happens When a Gateway Recovers?” on page 65 for additional information, OR
   
   - **Load Sharing**, Choose the Load Sharing mode (Multicast Mode or Unicast Mode) according to the capabilities of the router.
5. Choose whether to **Use State Synchronization**.

- Load Sharing configurations require synchronization between cluster members, and this option is checked, and grayed out.

- For High Availability New mode, this option is checked by default. If you uncheck this, the cluster members will not be synchronized, and existing connections on the failed gateway will be closed when failover occurs.

6. In the **Topology** page, define the virtual cluster IP addresses and at least one synchronization network. In the **Edit Topology** window:

- Define the topology for each cluster member interface. To automatically read all the predefined settings on the member interfaces, click **Get all members’ topology**.

- In the **Network Objective** column, define the purpose of the network by choose one of the options from the drop-down list (**Cluster, 1st Sync.,** etc.). The options are explained in the Online Help. To define a new network, click **Add Network**.

The **Edit Topology** window for the example in **Figure 4-2 on page 84** is as follows
7. Still in the **Topology** page, define the topology for each virtual cluster interface. In a virtual cluster interface cell, right click and select **Edit Interface**. The **Interface Properties** window opens.

   - In the **General** tab, **Name** the virtual interface, and define an **IP Address** (in Figure 4-2, 192.168.10.100 is one of the virtual interfaces).

   - In the **Topology** tab, define whether the interface is internal or external, and set up anti-spoofing.

   - In the **Member Networks** tab, define the member network and its netmask if necessary. This advanced option is explained in “Configuring Cluster Addresses on Different Subnets” on page 207.

8. Define the other pages in the cluster object as required (**NAT**, **VPN**, **Remote Access**, and so on).

Chapter 5

Working with OPSEC Certified Clustering Products

In This Chapter

- Introduction to OPSEC Certified Clustering Products  page 90
- Configuring OPSEC Certified Clustering Products  page 91
- CPHA Command Line Behavior in OPSEC Clusters  page 95
Introduction to OPSEC Certified Clustering Products

There are a number of OPSEC certified High Availability (sometimes called as Hot Standby) and Load Sharing (sometimes called Load Balancing) products. These products are used to build highly available Security Gateway clusters and to distribute traffic evenly among the clustered gateways.

Each OPSEC certified clustering application has its particular strengths and capabilities, whether it be monitoring, management, or performance. The role of these clustering applications is to:

1. Decide which cluster member will deal with each connection.
2. Perform health checks. This involves checking the status of a cluster member (for example, Active, Standby, or Down), and checking the status of the member interfaces.
3. Perform failover.

OPSEC certified clustering products use Check Point state synchronization mechanism (described in Chapter 2, “Synchronizing Connection Information Across the Cluster”) to exchange and update connection information and other states between cluster members.

This guide provides general guidelines for working with OPSEC certified clustering products. Configuration details vary for each clustering product. You are therefore urged to follow the instructions supplied with the OPSEC product.
Configuring OPSEC Certified Clustering Products

This procedure describes how to configure an OPSEC certified Security Gateway clustering solution.

Preparing the Switches and Configuring Routing

Follow the instructions in your clustering product documentation for:

- Preparing the switches and routers
- Configuring routing

Preparing the Cluster Member Machines

1. Define IP addresses for all interfaces on all the cluster members.
2. Connect the cluster network machines, via the switches. For the Synchronization interfaces, a cross-over cable or a dedicated switch is recommended.

   Note - It is possible to run synchronization across a WAN. For details, see “Synchronizing Clusters over a Wide Area Network” on page 36.

3. For Nokia clusters, configure VRRP or IP Clustering before installing Check Point Security Gateway. For other OPSEC certified clusters, follow the vendor recommendations.

   After the installation has finished, make sure that the Enable VPN-1/FW-1 monitoring is set to Enable in the Nokia configuration manager. This assures that IPSO will monitor changes in the status of the firewall. For VRRP and IP Clustering in IPSO 3.8.2 and above, the state of the firewall is reported to the Nokia cluster for failover purposes.

4. Install Check Point Security Gateway on all cluster members. During the configuration phase (or later, using the cpconfig Configuration Tool):
   - Install a license for Check Point Security Gateway on each cluster member. No special license is required to allow the OPSEC certified product to work with the Security Gateway.
SmartDashboard Configuration for OPSEC Clusters

During the configuration phase, enable State Synchronization by selecting Enable cluster membership for this gateway on Unix machines, or This Gateway is part of a cluster on Windows.

SmartDashboard Configuration for OPSEC Clusters

1. Using SmartDashboard, create the Gateway Cluster object. To define a new Gateway Cluster object, right click the Network Objects tree, and choose New Check Point > Gateway Cluster… Configuration of the Gateway Cluster Object can be performed using:

   • Simple Mode (Wizard) which guides you step by step through the configuration process. See the online help for further assistance.
   • Classic Mode, described below.

Classic Mode Configuration

2. In the General Properties page of the Gateway Cluster object, give the cluster a general IP address. In general, make it the external virtual IP address of the cluster.

   In the list of Check Point Products, ensure ClusterXL is not selected.

3. In the Cluster Members page, click Add > New Cluster Member to add cluster members to the cluster. Cluster members exist solely inside the Gateway Cluster object. For each cluster member:

   • In the Cluster Members Properties > General tab, define a name a Name and IP Address. Choose an IP address that is routable from the Security Management server so that the Security Policy installation will be successful. This can be an internal or an external address, or a dedicated management interface.

   • Click Communication, and Initialize Secure Internal Communication (SIC).

   • Define the NAT and VPN tabs, as required.

   You can also add an existing gateway as a cluster member by selecting Add > Add Gateway to Cluster in the Cluster Members page and selecting the gateway from the list in the Add Gateway to Cluster window.

   If you want to remove a gateway from the cluster, click Remove in the Cluster Members page and select Detach Member from Cluster or right-click on the cluster member in the Network Objects tree and select Detach from Cluster.

4. In the 3rd Party Configuration page, specify the cluster operating mode, and for the 3rd Party Solution, select OPSEC, and check Use State Synchronization.
5. The **Topology** page is used to define the virtual cluster IP addresses and cluster member addresses.

   For each cluster member, define the interfaces for the individual members.

   For OPSEC certified products, the configuration of virtual cluster IPs is mandatory in several products, while in others it is forbidden. Refer to your cluster product documentation for details.

   Define the synchronization networks. Depending on the OPSEC implementation, it might be possible to "get" the synchronization network from the OPSEC configuration if it is already defined. Refer to the OPSEC documentation to find out if this feature is implemented for a specific OPSEC product.

6. Now go back to the **3rd Party Configuration** page.

   A non-sticky connection is one in which packets from client to server and from server to client pass through different cluster members. Non-sticky connections are a problem because they can lead to out-of-state packets being received by the cluster member. The Security Gateway will reject out-of-state packets, even if they belong to a valid connection.

   Either the synchronization mechanism, or the OPSEC certified clustering product need to be able to identify valid non-sticky connections, so that the Security Gateway will allow those connections through the cluster.

   Find out whether or not the OPSEC certified clustering product can identify valid non-sticky connections.

   - If the clustering product cannot identify valid non-sticky connections, the synchronization mechanism can do so instead. In that case, check **Support non-sticky connections**.

   - If the clustering product can identify valid non-sticky connections, the synchronization mechanism does not have to take care of this. In that case, uncheck **Support non-sticky connections**. Usually it is safe to uncheck this option in High Availability solutions (not in Load Sharing). Unchecking this option will lead to a slight improvement in the connection establishment rate.

   If the **Hide Cluster Members' outgoing traffic behind the Clusters IP Address** option is checked, **Support non-sticky connections** should also be checked to support outgoing connections from a standby machine (unless specifically directed by OPSEC certified clustering product guide).

7. Many gateway clusters have a virtual cluster IP address that is defined in **Topology** page of the cluster object, in addition to physical cluster member interface addresses. The use of virtual cluster IP addresses affects the settings in the **3rd Party Configuration** page.
When a client behind the cluster establishes an \textit{outgoing} connection towards the Internet, the source address in the outgoing packets, is usually the physical IP address of the cluster member interface. If virtual cluster IP addresses are used, the clustering product usually changes the source IP address (using NAT) to that of the external virtual IP address of the cluster.

This corresponds to the default setting of \textbf{Hide Cluster Members’ outgoing traffic behind the Cluster's IP address} being checked.

When a client establishes an \textit{incoming} connection to the external virtual address of the cluster, the clustering product changes the destination IP address (using NAT) to that of the physical external address of one of the cluster members.

This corresponds to the default setting of \textbf{Forward Cluster’s incoming traffic to Cluster Members’ IP addresses} being checked. In the \textbf{Topology} page, define the interfaces for the individual members. In most OPSEC solutions, cluster IPs should not be added to the individual member’s \textbf{Topology} tab. Refer to your clustering product documentation for additional information.

8. Define the other pages in the cluster object as required (\textbf{NAT}, \textbf{VPN}, \textbf{Remote Access}, and so on).


\textbf{Note} - When defining a Nokia cluster (VRRP or IP clustering) of IPSO version 3.9 and later, the monitor \texttt{fw state} feature should be disabled before the first policy installation. Failing to do so impedes the setting of the cluster IP addresses, and consequently the \textbf{Get Interfaces} operation in the \textbf{Topology} section of the \textbf{Gateway Cluster Properties} window will fail. After policy installation, the monitor \texttt{fw state} feature can be re-enabled.
CPHA Command Line Behavior in OPSEC Clusters

In This Section

The cphastart and cphastop Commands in OPSEC Clusters page 95
The cphaprob Command in OPSEC Clusters page 96

This section describes the behavior of specific command lines in OPSEC clusters.

Note - For details of the cpha command lines see “Monitoring and Troubleshooting Gateway Clusters” on page 121.

The cphastart and cphastop Commands in OPSEC Clusters

The behavior of the cphastart and cphastop commands on ClusterXL clusters are described in “The cphastart and cphastop Commands” on page 138.

On OPSEC clusters, the cphastart command may not cause the cluster member to start working. On Nokia clusters the behavior is the same as with ClusterXL clusters.

The cphastop command may not cause failover on OPSEC clusters. On Nokia IP Clustering clusters (but not on VRRP clusters), the behavior is the same as with ClusterXL clusters.

As with ClusterXL clusters, these commands should only be run by the Security Gateway, and not directly by the user.
The cphaprob Command in OPSEC Clusters

Use the cphaprob command to verify that the cluster and the cluster members are working properly. This command is relevant only for Nokia IP clustering and Nokia VRRP.

In non-Nokia OPSEC clusters the command output is either empty or the command does not have any effect.

To produce a usage printout for cphaprob that shows all the available commands, type cphaprob at the command line and press Enter. The meaning of each of these commands is explained in the following sections.

cphaprob -d <device> -t <timeout(sec)> -s <ok|init|problem> [-p] register

example:

cphaprob -d <device> -t <timeout(sec)> -s <ok|init|problem> [-p] register

cphaprob -f <file> register

cphaprob -d <device> [-p] unregister

cphaprob -d <device> -s <ok|init|problem> report

cphaprob [-i[a]] [-e] list

cphaprob state

cphaprob [-a] if

Cphaprob state: When running this command, the machine state is only Check Point status and is not really a machine status. The command only monitors full sync success, and if a policy was successfully installed. For IP clustering, the state is accurate and also includes the status of the Nokia Cluster. For VRRP, the status is accurate for a firewall, but it does not correctly reflect the status of the Nokia machine (for example, it does not detect interface failure).

cphaprob [-a] if: Shows only the relevant information - interface name, if it is a sync interface or not. “Multicast”/“Broadcast” refers to the cluster control protocol and is relevant only for the sync interface. Note that the status of the interface is not printed since it is not monitored. (This also applies in the Nokia machine.)
Chapter 6

UTM-1 Clustering

In This Chapter:

- Overview page 98
- Configuring a Cluster on New Appliances page 99
- Adding an Existing UTM-1 Appliance to a Cluster page 112
- Removing a Cluster Member page 114
- Upgrading to a UTM-1 Cluster page 116
- Importing a Database to a Primary Cluster Member page 117
- Migrating a Database to a UTM-1 Cluster page 118
- Supported Logging Options for UTM-1 Clusters page 119
Overview

A pair of UTM-1 appliances can be clustered for high-availability. Each UTM-1 appliance becomes a single member in the cluster. High availability refers to the fact that both the gateway components and the Security Management server components are fully synchronized. If one component fails, a complete replication of the data exists on the member's peer in the cluster. Connections through the appliance continue uninterrupted.

Unlike between the gateway components, there is no fail-over between the Security Management server components. If the primary Security Management server goes down, the secondary Security Management server does not take over. However, the secondary Security Management server's database is fully synchronized with the Primary, so no data is lost. Before setting up a UTM-1 cluster, note that:

- The members of UTM-1 Cluster can either be configured together (both appliances are linked before the UTM-1 WebUI wizard is opened) or separately (the user chooses to figure a UTM-1 Cluster consisting of a single, primary member, and configure the secondary member at a later date).
- Even if you decide not to install a secondary cluster member, it is worthwhile to configure a cluster composed of a single primary member. A UTM-1 cluster is visible to the external network through its virtual IP addresses, not the actual physical addresses of its members. If at some point you do decide to add a secondary member, you will not have to alter the layer 3 topology of the network.
Configuring a Cluster on New Appliances

To configure a cluster on appliances with Messaging Security already installed:

1. Power up the UTM-1 appliance.
   The PWR/Status LED on the front panel starts blinking. When the LED ceases to blink, the appliance is ready for login.

2. Using the supplied Ethernet cable, connect the UTM-1 appliance’s internal interface to a PC.

3. Configure the PC to be on the same subnet as the UTM-1 appliance. For Windows XP operating systems:
   a. Click Start > Settings > Control Panel > Network Connections.
   b. Double-click Local Area Connection.
   c. On the General tab, click Properties.
   d. In the Local Area Connection Properties window, scroll down to and double-click Internet Protocol (TCP/IP).
   e. In the Internet Protocol (TCP/IP) Properties window that opens, select Use the following IP address.
   f. In the IP address field, enter 192.168.1.2.
   g. In the Subnet Mask field, enter 255.255.255.0.
   h. Leave the Default gateway settings empty.
i. Select **Use the following DNS server addresses**, and leave the **Preferred DNS server** and **Alternate DNS server** fields empty. The window should look like this:

![Internet Protocol (TCP/IP) Properties](image)

j. Click **OK**.

4. Launch your web browser, and connect to **https://192.168.1.2:4434**.
   
The UTM-1 Appliance login window appears.
   
If you receive a certificate error page notice, click **Continue to this Web site**. If you are using a popup block, allow all popups.
5. Log in with the default system administrator login name and password: **admin/admin**.
6. Change the default login name and password:

1. Download password recovery login token file.

   **Password recovery login token**

   Download the one time password recovery login token file, which you can use in case you forget your login password. It is recommended to save the password recovery login token file in a safe place.

   ![Download...]

2. Please change the default Login Name and Password to avoid unauthorized access to this appliance.

   **Change Default Login Name and Password**

   In order to protect your system, it is important to select a secure password. The following guidelines will help in selecting a secure password:
   - Choose a password between 6 and 32 characters.
   - It is recommended to use a combination of upper and lowercase letters with at least one number or non-alphanumeric character. Examples of non-alphanumeric characters include: !@# $%^&*()_+-=
   - Password may not be a common word found in a dictionary.

   ![Secure Password Schema]

   **Administrator Name:** admin

   **New Password:**

   **Confirm New Password:**

   ![Save and Login]

7. Click **Save and Login**.

   The first time configuration wizard opens.
8. Click **Next**.

The **Appliance Date and Time Setup** page opens.

![First Time Configuration Wizard - Appliance Date and Time Setup](image)

9. Set up the appliance date and time and click **Apply**.

The **Network Connections** page opens.

![First Time Configuration Wizard - Network Connections](image)

10. Add an interface for the cluster SYNC/LAN1 interface, and click **Next**.
The **Routing Table** page opens.

11. Add or delete routes as necessary, and click **Next**.

The **DNS and Domain Settings** page opens.

12. Configure a host name, domain name, and DNS servers, and click **Next**.
13. Select **Locally Managed**, and click **Next**.

The **UTM-1 High Availability Cluster** page opens.

14. Select **UTM-1 Primary cluster member**, and click **Next**.
If you are configuring the secondary Security Management Server, select **UTM-1 Secondary Cluster member** and click **Next**. The **Secure Internal Communication (SIC) setup** page opens:

Enter an activation key (one-time password), and make a note of it. You will need this activation key when initializing SIC in Security Management server. Click **Next**.

The **Web/SSH and GUI clients** page opens.

15. Add or remove remote machines from which Web, SSH, or SmartConsole Clients can connect to the Security Management server, and click **Next**.
16. Download SmartConsole to your PC, and click **Next**.

   A summary page is displayed.

   **Summary**
   
   - **Management Type**: Locally Managed
   - **Cluster Configuration**: UTM-1 Primary cluster member
   - **GUI Client**: Key
   - **Internal Interfaces**: 192.168.1.1/255.255.255.0
   - **Default Routers**: 192.168.1.1/24
   - **Hostnames**: Primary: UTM-1

   (In order to complete the configuration wizard, click on 'Finish' button. This will initiate the appliance. It will take several minutes to complete.)

17. Click **Finish**.

   The configuration process runs.
18. Wait until the completed message appears, and click OK.

19. Reboot the gateway.

20. Repeat the procedure for the secondary Security Management Server, but at step 14 on page 104, select UTM-1 Secondary cluster member, and enter an activation key when prompted.

   **Note** - Remember the activation key. You will need it later when configuring the cluster in SmartDashboard. Remember to configure the cluster SYNC interface on the same subnet as the SYNC interface on the Primary Security Management Server.

You now have two UTM-1 appliances, one configured as a primary Security Management Server, the other as secondary.

   **Note** - IP addresses need to be configured on both cluster members before opening SmartDashboard and running the first-time cluster configuration wizard, as described in step 23.

21. Using a cross cable, connect the SYNC/LAN1 ports.

22. Install the SmartDashboard you downloaded in step 16 on page 106.

23. Open SmartDashboard and connect to the IP address of the UTM-1 appliance that has been configured as the primary cluster member.

   The cluster configuration wizard starts automatically. If the UTM-1 Cluster wizard does not start automatically, double click the network cluster object in SmartDashboard and select the **Simple mode** option.

   Click **Next** on the welcome page. The Cluster’s general properties window opens:

   ![Cluster configuration wizard](image)

24. Enter the cluster name.
The name entered here will replace the provisional name that already appears in the network objects tree. (If you cancel the wizard at this point and examine the network objects tree, a cluster object is shown with only the primary member defined).

**Note** - If you choose to define the UTM-1 Cluster in classic mode, when entering the IP address of the cluster, enter the Virtual IP.

25. Click **Next**.

The **Secondary member's properties** window opens.

The UTM-1 appliance acting as the primary Security Management server has been defined automatically, so only the secondary Security Management server needs to be configured.

26. Enter:

   a. the Secondary cluster's name and IP address, the same name that you entered during the SecurePlatform first-time configuration wizard.
b. the activation key.

If you choose to define the secondary member later, then each time SmartDashboard is opened, the UTM-1 first-time wizard runs automatically.

27. Click **Next**.

Clicking **Next** initiates SIC. The Security Management server on the primary Security Management server retrieves a list of interfaces (and therefore networks) the secondary cluster member is connected to: the topology of the UTM-1 cluster.

Once SIC is initiated, the **Cluster topology** window opens:
Define member interfaces here, and also enter the Cluster interface IP address.

- **Configure Cluster topology for interface (SYNC, DMZ, EXT, INT, LANs)**
  
The LAN1 interface serves as the SYNC interface between cluster members. If not configured, SYNC interfaces are automatically set to 10.231.149.1 and 10.231.149.2. If these addresses are already in use, their values can be manually adjusted. If you manually adjust the default IP SYNC addresses, verify that both reside on the same subnet.

  **Note** - All interfaces in the cluster must have unique IP addresses. If the same IP address is used twice, policy installation will fail. A Load on gateway failed error message is displayed.

- **Configure Cluster Virtual Interface.**
  
  In addition to the UTM-1 cluster having its own unique IP address, which is visible to the network, each member interface also has a unique IP address. These unique IP addresses are used for internal communication between the cluster members as well as for accessing the cluster member directly. It is the Virtual IP address which makes the cluster visible to the external network, and populates the network routing tables — not the actual IP addresses of the individual members. If you chose to define a secondary member at a later date, but still configure the Virtual Cluster IP address, secondary members can later be added without changing the layer 3 topology of the network.

  **Note** - The external Virtual IP address of the cluster must be unique.

28. Click **Next**, and in the following windows configure IP addresses for the External, and Internal interfaces on each cluster member.

29. Click **Next**, and **Finish**.

  **Note** - After completing the first-time cluster wizard, the main IP address, as shown in the UTM-1 Cluster **General properties** window, will be the Virtual IP address of the cluster.
The new cluster of UTM-1 appliances appears in the network objects tree:
Adding an Existing UTM-1 Appliance to a Cluster

A single standalone UTM-1 appliance can easily be added to a UTM-1 cluster once a second UTM-1 device has been added to the network.

To add an existing UTM-1 device to a Cluster:
1. Open the SecurePlatform WebUI.
2. On the Product Configuration, Cluster page, select Make this Appliance the primary member of a UTM-1 High Availability Cluster.
3. Click Apply.
4. Reboot the appliance.
5. Using SmartDashboard, connect to the primary Security Management server.
   The first-time cluster configuration wizard opens.
6. Complete the wizard in order to configure the secondary cluster member.

In SmartDashboard, the network object representing the former standalone UTM-1 device is converted to a cluster object. If the standalone UTM-1 device appeared in the:

- INSTALL ON column of any rule (Security, QoS, NAT)
- Participating Gateways list for a VPN community
the new cluster object is now shown. In all other columns (Source, Destination, Groups) the standalone UTM-1 object needs to be manually changed to the cluster object.

Manually changing this network object may have implications for your Security, NAT, and QoS rules. To understand how and where the new primary cluster member is being used, right-click the cluster object and select Where used...

It is also recommended to use the Search > Query Rules... option on the File Menu.
1. Select the Stand Alone object and make sure it appears in the In List section.
2. Select the Explicit option.
3. Click Apply.

For the remaining rules that still contain the standalone object, replace by drag-and-dropping the new cluster object.

In addition, if you have a group that contains a standalone UTM-1 gateway, which has since been promoted to a primary cluster member, a policy install for that group will fail. To successfully install a policy, replace the standalone UTM-1 gateway with the object for the cluster.

Note - While the icon in SmartDashboard has changed to reflect the UTM-1 device's new status as a primary cluster member, the Name and UID of the object in the database remains unchanged.
Removing a Cluster Member

Although a UTM-1 Cluster consists of only two members, it is still possible to remove one of the members without erasing the cluster. A UTM-1 cluster can exist with only a single (primary) member until a new secondary member is added. This means that either member can be easily taken offline for maintenance.

A cluster member is removed in two stages: first in the UTM-1 WebUI, second on the command line of the appliance. To remove a cluster member:

1. Open the **UTM-1 WebUI > Product Configuration > Cluster** page, as shown in Figure 1:

![Figure 1](removepeer.png)

   **Figure 1**  
   Remove peer

   - **Remove Peer from UTM-1 Cluster:**
     - **Remove Peer**

   **Cluster Information**

   - **Type:** UTM-1 Primary cluster member
   - **Status:** Standby

2. Click **Remove Peer**.

   Clicking **Remove Peer** results in a UTM-1 Cluster with a single member.

   - If the current machine is the primary member of the cluster, the secondary member is deleted.
   - If the current machine is the secondary member, the secondary member is first promoted to a primary and then the peer is deleted.

   **Note** - Services running on the appliance are restarted.

3. On the appliance command line, run: `cp_conf fullha disable`.

   This command reverts the primary cluster member to a standalone configuration.

4. Reboot.
After reverting to a standalone configuration, the former cluster is shown in SmartDashboard as a locally managed gateway consisting of gateway and Security Management server.
Upgrading to a UTM-1 Cluster

If you have a single UTM-1 appliance in a production environment, and wish to upgrade it to the current version and configure it as a UTM-1 Cluster:

1. Perform a standard in-place upgrade the current version using the UTM-1 WebUI.

2. Using the UTM-1 WebUI, on the Cluster page, convert the appliance to a primary member.

3. Connect a second UTM-1 appliance to the network.
   a. If this second appliance is based on an earlier version, obtain the relevant upgrade package from the Download Center, save it to a USB stick, and reinstall the appliance as a secondary cluster member.
   b. If the second appliance is already R70, run the first-time wizard and select “secondary cluster member.”
Importing a Database to a Primary Cluster Member

To import an older Security Management server database to the primary member of a new UTM-1 cluster:

1. Install the primary as a standard Security Management server (not a cluster)
2. Import the older Security Management server database
3. Using the UTM-1 WebUI, convert the standard Security Management server to a primary cluster member.

Note - If the older database is from another primary member of a UTM-1 Cluster, the above procedure is not required.
Migrating a Database to a UTM-1 Cluster

To migrate a Security Management server database to a UTM-1 Cluster, go to the Check Point Support Center at:

http://support.checkpoint.com
Supported Logging Options for UTM-1 Clusters

The standard logging options for a regular cluster are also available in UTM-1. In UTM-1, the primary cluster member can send logs to the secondary, and the secondary to the primary. Both can also send logs to an external log server. However, while a regular cluster cannot save logs locally, a UTM-1 Cluster member also functions as a log server. When logs are saved locally on the UTM-1 cluster member, implications exist for High Availability and Load Sharing scenarios.

Recommended Logging Options for High Availability

In a High Availability scenario, one of the UTM-1 cluster members is active while the other cluster member remains in standby. Log files are not synchronized between the two UTM-1 cluster members. For this reason it is recommended to:

- Configure logging so that logs are always sent to the primary UTM-1 cluster member, but to the secondary UTM-1 cluster member when the primary is unreachable.
- Set scheduled log forwarding to the primary UTM-1 cluster member.

Alternatively:

- Configure logging so that logs are sent to both UTM-1 cluster members. (Eventia Analyzer and Eventia Reporter with standard reports should use only one of the cluster members as a source for log file correlation and consolidation). Or:
  - Use an external log server.

**Note** - These logging options can also be applied to new gateways.

Load Sharing

To enable load sharing on a UTM-1 Cluster, you must first deselect the **Save logs locally on each cluster member** option on the **Log Servers** page.
Load Sharing

When load sharing is enabled for a UTM-1 Cluster, a connection initiated to the primary member may terminate with the secondary. If logs were saved locally on each cluster member, only partial logs would be produced on each member for connections that were handled by both members. Saving logs locally would result in partial log data being displayed in Smartview Tracker.

If possible, when load sharing is enabled, configure log files to be sent to an external log server.
Chapter

Monitoring and Troubleshooting Gateway Clusters

In This Chapter

Verifying that a Cluster is Working Properly  page 122
Monitoring Cluster Status Using SmartConsole Clients  page 133
ClusterXL Configuration Commands  page 138
How to Initiate Failover  page 139
Monitoring Synchronization (fw ctl pstat)  page 141
Troubleshooting Synchronization  page 145
ClusterXL Error Messages  page 160
Member Fails to Start After Reboot  page 168
Verifying that a Cluster is Working Properly

In This Section

The cphaprob Command ................................................. page 122
Monitoring Cluster Status ............................................. page 123
Monitoring Cluster Interfaces ....................................... page 126
Monitoring Critical Devices .......................................... page 127
Registering a Critical Device ......................................... page 129
Registering Critical Devices Listed in a File ....................... page 130
Unregistering a Critical Device ....................................... page 131
Reporting Critical Device Status to ClusterXL ............... page 131
Example cphaprob Script .............................................. page 132

The cphaprob Command

Use the cphaprob command to verify that the cluster and the cluster members are working properly, and to define critical devices. A critical device is a process running on a cluster member that enables the member to notify other cluster members that it can no longer function as a member. The device reports to the ClusterXL mechanism regarding its current state or it may fail to report, in which case ClusterXL decides that a failover has occurred and another cluster member takes over. When a critical device (also known as a Problem Notification, or pnote) fails, the cluster member is considered to have failed.

There are a number of built-in critical devices, and the administrator can define additional critical devices. The default critical devices are:

- The cluster interfaces on the cluster members.
- Synchronization—full synchronization completed successfully.
- Filter—the Security Policy, and whether it is loaded.
- cphad—which follows the ClusterXL process called cphamcset.
- fwd—the VPN daemon.

These commands can be run automatically by including them in scripts.
To produce a usage printout for cphaprob that shows all the available commands, type cphaprob at the command line and press Enter. The meaning of each of these commands is explained in the following sections.

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>cphaprob -d &lt;device&gt; -t &lt;timeout(sec)&gt; -s &lt;ok</td>
<td>init</td>
</tr>
<tr>
<td>cphaprob -f &lt;file&gt; register</td>
<td>Register from file</td>
</tr>
<tr>
<td>cphaprob -d &lt;device&gt; [-p] unregister</td>
<td>Unregister device</td>
</tr>
<tr>
<td>cphaprob -d &lt;device&gt; -s &lt;ok</td>
<td>init</td>
</tr>
<tr>
<td>cphaprob [-i[a]] [-e] list</td>
<td>List devices</td>
</tr>
<tr>
<td>cphaprob state</td>
<td>Display state</td>
</tr>
<tr>
<td>cphaprob [-a] if</td>
<td>List interfaces</td>
</tr>
</tbody>
</table>

**Monitoring Cluster Status**

To see the status of a cluster member, and of all the other members of the cluster, run the following command on the cluster member:

```
cphaprob state
```

Do this after setting up the cluster, and whenever you want to monitor the cluster status. The following is an example of the output of `cphaprob state`:

```
cphaprob state
Cluster mode: Load sharing (Multicast)
Number Unique Address State
1 (local) 30.0.0.1 active
2 30.0.0.2 active
```

- **Cluster mode** can be
  - Load Sharing (Multicast).
  - Load Sharing (Unicast).
  - High Availability New Mode (Primary Up or Active Up).
  - High Availability Legacy Mode (Primary Up or Active Up).
  - For third-party clustering products: “Service”.

Refer to “Clustering Definitions and Terms” on page 22, for further information.
• The number of the member indicates the member ID for Load Sharing, and the Priority for High Availability.

• In Load sharing configuration, all machines in a fully functioning cluster should be Active. In High Availability configurations, only one machine in a properly functioning cluster must be Active, and the others must be in the Standby state.

Third-party clustering products show Active/Active even if one of the members is in standby state. This is because this command only reports the status of the full synchronization process. For Nokia VRRP, this command shows the exact state of the Firewall, but not the cluster member (for example, the member may not be working properly but the state of the Firewall is active).

When examining the state of the cluster member, you need to consider whether it is forwarding packets, and whether it has a problem that is preventing it from forwarding packets. Each state reflects the result of a test on critical devices. Table 7-1 lists and explains the possible cluster states, and whether or not they represent a problem.

### Table 7-1 Cluster States

<table>
<thead>
<tr>
<th>State</th>
<th>Meaning</th>
<th>Forwarding packets?</th>
<th>Is this state a Problem?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>Everything is OK.</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Active attention</td>
<td>A problem has been detected, but the cluster member is still forwarding packets because it is the only machine in the cluster or there is no other active machines in the cluster. In any other situation the state of the machine would be down.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Down</td>
<td>One of the critical devices is down.</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Monitoring Cluster Status

Table 7-1  Cluster States

<table>
<thead>
<tr>
<th>State</th>
<th>Meaning</th>
<th>Forwarding packets?</th>
<th>Is this state a Problem?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ready</td>
<td>Can occur in following scenarios:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. When a cluster is upgraded from one version of Check Point Security Gateway to another, and the cluster members have different versions of Check Point Security Gateway, the members with a new version have the \textit{ready} state and the members with the previous version have the \textit{active} state.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. Before a cluster member becomes \textit{active}, it sends a message to the rest of the cluster, and then expects to receive confirmations from the other cluster members agreeing that it will become active. In the period of time before it receives the confirmations, the machine is in the \textit{ready} state.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standby</td>
<td>Applies only to a High Availability configuration, and means the member is waiting for an active machine to fail in order to start packet forwarding.</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Initializing</td>
<td>An initial and transient state of the cluster member. The cluster member is booting up, and ClusterXL product is already running, but the Security Gateway is not yet ready.</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>ClusterXL inactive or machine is down</td>
<td>Local machine cannot hear anything coming from this cluster member.</td>
<td>Unknown</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Monitoring Cluster Interfaces

To see the state of the cluster member interfaces and the virtual cluster interfaces, run the following command on the cluster member:

```
cphaprob [-a] if
```

The output of this command must be identical to the configuration in the cluster object **Topology** page. For example:

```
cphaprob -a if

Required interfaces: 4
Required secured interfaces: 1

+-----------------+--------+---------------------------------------------+
| Interface       | Status | Description                                |
+-----------------+--------+---------------------------------------------+
| qfe4            | UP     | (secured, unique, multicast)               |
| qfe5            | UP     | (non secured, unique, multicast)           |
| qfe6            | DOWN   | 4810.2 secs (non secured, unique, multicast)|
| qfe7            | UP     | (non secured, unique, multicast)           |
+-----------------+--------+---------------------------------------------+

Virtual cluster interfaces: 2

qfe5 30.0.1.130
qfe6 30.0.2.130
```

The interfaces are ClusterXL critical devices. ClusterXL checks the number of good interfaces and sets a value of **Required interfaces** to the maximum number of good interfaces seen since the last reboot. If the number of good interfaces is less than the **Required** number, ClusterXL initiates failover. The same for **secured interfaces**, where only the good synchronization interfaces are counted.

An interface can be:

- **Non-secured** or **Secured.** A secured interface is a synchronization interface.
- **Shared** or **unique.** A shared interface applies only to High Availability Legacy mode.
- **Multicast** or **broadcast.** The Cluster Control Protocol (CCP) mode used in the cluster. CCP can be changed to use broadcast instead. To toggle between these two modes use the command `cphaconf set_ccp <broadcast|multicast>`

For third-party clustering products, except in the case of Nokia IP Clustering, `cphaprob -a if` should always show virtual cluster IP addresses.

When an interface is **DOWN**, it means that the interface cannot receive or transmit CCP packets, or both. This may happen when an interface is malfunctioning, is connected to an incorrect subnet, is unable to pick up Multicast Ethernet packets.
and so on. The interface may also be able to receive but not transmit CCP packets, in which case the status field is read. The displayed time is the number of seconds that have elapsed since the interface was last able to receive/transmit a CCP packet.

See “Defining Disconnected Interfaces” on page 204 for additional information.

**Monitoring Critical Devices**

When a critical device fails, the cluster member is considered to have failed. To see the list of critical devices on a cluster member, and of all the other machines in the cluster, run the following command on the cluster member:

```
cphaprob [-i[a]] [-e] list
```

There are a number of built-in critical devices, and the administrator can define additional critical devices. The default critical devices are:

- The cluster interfaces on the cluster members.
- Synchronization—full synchronization completed successfully.
- Filter—the Security Policy, and whether it is loaded.
- cphad—which follows the ClusterXL process called cphamcset.
- fwd—the VPN daemon.
For Nokia IP Clustering, the output is the same as for ClusterXL Load Sharing. For other third-party products, this command produces no output. The following example output shows that the fwd process is down:

cphaprob list

Built-in Devices:

Device Name: Interface Active Check
Current state: OK

Registered Devices:

Device Name: Synchronization
Registration number: 0
Timeout: none
Current state: OK
Time since last report: 15998.4 sec

Device Name: Filter
Registration number: 1
Timeout: none
Current state: OK
Time since last report: 15644.4 sec

Device Name: fwd
Registration number: 3
Timeout: 2 sec
Current state: problem
Time since last report: 4.5 sec
Registering a Critical Device

It is possible to add a user defined critical device to the default list of critical devices. Use this command to register `<device>` as a critical process, and add it to the list of devices that must be running for the cluster member to be considered active. If `<device>` fails, then the cluster member is considered to have failed.

If `<device>` fails to contact the cluster member in `<timeout>` seconds, `<device>` will be considered to have failed. For no timeout, use the value 0.

Define the status of the `<device>` that will be reported to ClusterXL upon registration. This initial status can be one of:

- **ok** — `<device>` is alive.
- **init** — `<device>` is initializing. The machine is down. This state prevents the machine from becoming active.
- **problem** — `<device>` has failed.

`[-p]` makes these changes permanent. After performing a reboot or after removing the Security Gateway (on Linux or IPSO for example) and re-attaching it, the status of critical devices that were registered with this flag will be saved.

```bash
phaprof -d <device> -t <timeout(sec)> -s <ok|init|problem> [-p] register
```
Registering Critical Devices Listed in a File

Register all the user defined critical devices listed in `<file>`. `<file>` must be an ASCII file, with each device on a separate line. Each line must list three parameters, which must be separated by at least a space or a tab, as follows:

- `<device>` — The name of the critical device. It must have no more than 15 characters, and must not include white spaces.
- `<timeout>` — If `<device>` fails to contact the cluster member in `<timeout>` seconds, `<device>` will be considered to have failed. For no timeout, use the value 0.
- `<status>` — can be one of
  - `ok` — `<device>` is alive.
  - `init` — `<device>` is initializing. The machine is down. This state prevents the machine from becoming active.
  - `problem` — `<device>` has failed.

```
cphaprob -f `<file>` register
```

```
<device> <timeout> <status>
```

Unregistering a Critical Device

```
cphaprob -d <device> [-p] unregister
```

Unregister a user defined <device> as a critical process. This means that this device is no longer considered critical. If a critical device (and hence a cluster member) was registered as “problem” before running this command, then after running this command the status of the cluster will depend only on the remaining critical devices.

[-p] makes these changes permanent. This means that after performing a reboot or after removing the kernel (on Linux or IPSO for example) and re-attaching it, these critical devices remain unregistered.

Reporting Critical Device Status to ClusterXL

```
cphaprob -d <device> -s <ok|init|problem> report
```

Use this command to report the status of a user defined critical device to ClusterXL.

<device> is the device that must be running for the cluster member to be considered active. If <device> fails, then the cluster member is considered to have failed.

The status to be reported. The status can be one of:

**ok** — <device> is alive

**init** — <device> is initializing. The machine is down. This state prevents the machine from becoming active.

**problem** — <device> has failed. If this status is reported to ClusterXL, the cluster member will immediately failover to another cluster member.

If <device> fails to contact the cluster member within the timeout that was defined when the it was registered, <device> and hence the cluster member, will be considered to have failed. This is true only for critical devices with timeouts. If a critical device is registered with the `-t 0` parameter, there will be no timeout, and until the device reports otherwise, the status is considered to be the last reported status.
**Example cphaprob Script**

Predefined cphaprob scripts are located on the location $FWDIR/bin. Two scripts are available:

- `clusterXL_monitor_ips`
- `clusterXL_monitor_process`

The `clusterXL_monitor_ips` script in the Appendix chapter “Example cphaprob Script” on page 237 has been designed to provide a way to check end-to-end connectivity to routers or other network devices and cause failover if the ping fails. The `clusterXL_monitor_process` script monitors the existence of given processes and causes failover if the processes die. This script uses the normal `pnote` mechanism. It can be found in the Appendix chapter “Example cphaprob Script” on page 237.
Monitoring Cluster Status Using SmartConsole Clients

In This Section

SmartView Monitor  
SmartView Tracker

SmartView Monitor

SmartView Monitor displays a snapshot of all ClusterXL cluster members in the enterprise, enabling real-time monitoring and alerting. For each cluster member, state change and critical device problem notifications are displayed. SmartView Monitor allows you to specify the action to be taken if the status of a cluster member changes. For example, the Security Gateway can issue an alert notifying you of suspicious activity.

Starting and Stopping ClusterXL Using SmartView Monitor

To stop ClusterXL on the machine and cause failover to another machine, open SmartView Monitor, click the cluster object, select one of the member gateway branches, right click a cluster member, and select Down.

To initiate a restart of ClusterXL, open SmartView Monitor, click the cluster object, select one of the member gateway branches, right click a cluster member, and select Up.

Note - SmartView Monitor does not initiate full synchronization, so that some connections may be lost. To initiate full synchronization, perform cpstart, or start the cluster member using the cphaprob command.
SmartView Tracker

Every change in status of a cluster member is recorded in SmartView Tracker according to the choice in the Fail-Over Tracking option of the cluster object ClusterXL page.

ClusterXL Log Messages

The following conventions are used in this section:

1. Square brackets are used to indicate place holders, which are substituted by relevant data when an actual log message is issued (for example, [NUMBER] will be replaced by a numeric value).

2. Angle brackets are used to indicate alternatives, one of which will be used in actual log messages. The different alternatives are separated with a vertical line (for example, <up|down> indicates that either “up” or “down” will be used).

3. The following place holders are frequently used:
   - ID: A unique cluster member identifier, starting from “1”. This corresponds to the order in which members are sorted in the cluster object’s GUI.
   - IP: Any unique IP address that belongs to the member.
   - MODE: The cluster mode (for example, New HA, LS Multicast, and so on).
   - STATE: The state of the member (for example, active, down, standby).
   - DEVICE: The name of a pnote device (for example, fwd, Interface Active Check).

General logs

Starting <ClusterXL|State Synchronization>.

Indicates that ClusterXL (or State Synchronization, for 3rd party clusters) was successfully started on the reporting member. This message is usually issued after a member boots, or after an explicit call to cphastart.

Stopping <ClusterXL|State Synchronization>.

Informs that ClusterXL (or State Synchronization) was deactivated on this machine. The machine will no longer be a part of the cluster (even if configured to be so), until ClusterXL is restarted.

Unconfigured cluster Machines changed their MAC Addresses. Please reboot the cluster so that the changes take affect.

This message is usually issued when a machine is shut down, or after an explicit call to cphastop.
State logs

Mode inconsistency detected: member [ID] ([IP]) will change its mode to [MODE]. Please re-install the security policy on the cluster.

This message should rarely happen. It indicates that another cluster member has reported a different cluster mode than is known to the local member. This is usually the result of a failure to install the security policy on all cluster members. To correct this problem, install the Security Policy again.

Note - The cluster will continue to operate after a mode inconsistency has been detected, by altering the mode of the reporting machine to match the other cluster members. However, it is highly recommended that the policy will be re-installed as soon as possible.

State change of member [ID] ([IP]) from [STATE] to [STATE] was cancelled, since all other members are down. Member remains [STATE].

When a member needs to change its state (for example, when an active member encounters a problem and needs to bring itself down), it first queries the other members for their state. If all other members are down, this member cannot change its state to a non-active one (or else all members will be down, and the cluster will not function). Thus, the reporting member continues to function, despite its problem (and will usually report its state as “active attention”).

member [ID] ([IP]) <is active|is down|is stand-by|is initializing> ([REASON]).

This message is issued whenever a cluster member changes its state. The log text specifies the new state of the member.

PNOTE logs

PNote log messages are issued when a pnote device changes its state.

• [DEVICE] on member [ID] ([IP]) status OK ([REASON]).

  The pnote device is working normally.

• [DEVICE] on member [ID] ([IP]) detected a problem ([REASON]).

  Either an error was detected by the pnote device, or the device has not reported its state for a number of seconds (as set by the “timeout” option of the pnote)

• [DEVICE] on member [ID] ([IP]) is initializing ([REASON]).

  Indicates that the device has registered itself with the pnote mechanism, but has not yet determined its state.

• [DEVICE] on member [ID] ([IP]) is in an unknown state ([STATE ID]) ([REASON]).

  This message should not normally appear. Contact Check Point Support.
Interface logs

- interface [INTERFACE NAME] of member [ID] ([IP]) is up.
  Indicates that this interface is working normally, meaning that it is able to receive and transmit packets on the expected subnet.

- interface [INTERFACE NAME] of member [ID] ([IP]) is down (receive <up|down>, transmit <up|down>.
  This message is issued whenever an interface encounters a problem, either in receiving or transmitting packets. Note that in this case the interface may still be working properly, as far as the OS is concerned, but is unable to communicate with other cluster members due to a faulty cluster configuration.

- interface [INTERFACE NAME] of member [ID] ([IP]) was added.
  Notifies users that a new interface was registered with the Security Gateway (meaning that packets arriving on this interface are filtered by the firewall). Usually this message is the result of activating an interface (such as issuing an `ifconfig up` command on Unix systems). The interface will now be included in the ClusterXL reports (such as in SmartView Monitor, or in the output of `cphaprobd -a if`). Note that the interface may still be reported as “Disconnected”, in case it was configured as such for ClusterXL.

- interface [INTERFACE NAME] of member [ID] ([IP]) was removed.
  Indicates that an interface was detached from the Security Gateway, and is therefore no longer monitored by ClusterXL.

SecureXL logs

- SecureXL device was deactivated since it does not support CPLS.
  This message is the result of an attempt to configure a ClusterXL in Load Sharing Multicast mode over Security Gateways using an acceleration device that does not support Load Sharing. As a result, acceleration will be turned off, but the cluster will work in Check Point Load Sharing mode (CPLS).
Reason Strings

- **member [ID] ([IP]) reports more interfaces up.**
  
  This text can be included in a pnote log message describing the reasons for a problem report: Another member has more interfaces reported to be working, than the local member does. This means that the local member has a faulty interface, and that its counterpart can do a better job as a cluster member. The local member will therefore go down, leaving the member specified in the message to handle traffic.

- **member [ID] ([IP]) has more interfaces - check your disconnected interfaces configuration in the <discntd.if file|registry>.**

  This message is issued when members in the same cluster have a different number of interfaces. A member having less interfaces than the maximal number in the cluster (the reporting member) may not be working properly, as it is missing an interface required to operate against a cluster IP address, or a synchronization network. If some of the interfaces on the other cluster member are redundant, and should not be monitored by ClusterXL, they should be explicitly designated as “Disconnected”. This is done using the file $FWDIR/conf/discntd.if (under Unix systems), or the Windows Registry.

- **[NUMBER] interfaces required, only [NUMBER] up.**

  ClusterXL has detected a problem with one or more of the monitored interfaces. This does not necessarily mean that the member will go down, as the other members may have less operational interfaces. In such a condition, the member with the highest number of operational interfaces will remain up, while the others will go down.
ClusterXL Configuration Commands

The cphaconf Command

Running this command is not recommended. It should be run only by the Security Gateway.

```
cphaconf [-i <machine id>] [-p <policy id>] [-b <db_id>] [-n <cluster num>] [-c <cluster size>] [-m <service >]
          [-t <secured IF 1>...] start

            cphaconf [-t <secured IF 1>...] [-d <disconnected IF 1>...] add

            cphaconf clear-secured
            cphaconf clear-disconnected
            cphaconf stop
            cphaconf init
            cphaconf forward <on/off>
            cphaconf debug <on/off>
            cphaconf set_ccp <broadcast/multicast>
            cphaconf mc_reload
            cphaconf debug_data
```

The cphastart and cphastop Commands

Running cphastart on a cluster member activates ClusterXL on the member. It does not initiate full synchronization. cpstart is the recommended way to start a cluster member.

Running cphastop on a cluster member stops the cluster member from passing traffic. State synchronization also stops. It is still possible to open connections directly to the cluster member. In High Availability Legacy mode, running cphastop may cause the entire cluster to stop functioning.

These commands should only be run by the Security Gateway, and not directly by the user.
How to Initiate Failover

In This Section

Stopping the Cluster Member
Starting the Cluster Member

The state of a cluster member can be manually controlled in order to take down the cluster member. This initiates failover to the other cluster member(s), in the case of Load Sharing, or failover to the next highest priority cluster member in the case of High Availability.

Stopping the Cluster Member

To stop ClusterXL on the machine and cause failover to another machine, do one of the following:

- Register a dummy critical device (faildevice for example) using the command `cphaprob -d faildevice -t 0 -s ok register`, and then run the following command to report to ClusterXL that the critical device faildevice has a problem: `cphaprob -d faildevice -s problem report`. Failover to another cluster member will immediately occur.
- Open SmartView Monitor, click the cluster object, select one of the member gateway branches, right click a cluster member, and then select Down.
Starting the Cluster Member

ClusterXL starts automatically when the Security Gateway is started on the cluster member (cpstart). To initiate a restart of ClusterXL, do one of the following:

- To reactivate a cluster member that was downed using the command cphaprob -d faildevice -s problem report, run either of the following commands:
  cphaprob -d faildevice -s ok report
cphaprob -d faildevice unregister

- Open SmartView Monitor, click the cluster object, select the member gateway branch that appears as "Down". Right click on that cluster member, and select Up.

**Note** - Starting the Cluster member from SmartView Monitor does not initiate full synchronization, so some connections may be lost. To initiate full synchronization, perform cpstart.
Monitoring Synchronization (fw ctl pstat)

To monitor the synchronization mechanism on ClusterXL or third-party OPSEC certified clustering products, run the following command on a cluster member:

```
fw ctl pstat
```

The output of this command is a long list of statistics for the Security Gateway. At the end of the list there is a section called “Synchronization” that applies per Gateway Cluster member. Many of the statistics are counters that can only increase. A typical output is as follows:

```
Version: new
Status: Able to Send/Receive sync packets
Sync packets sent:
  total : 3976, retransmitted : 0, retrans reqs : 58, acks : 97
Sync packets received:
  total : 4290, were queued : 58, dropped by net : 47
  retrans reqs : 0, received 0 acks
  retrans reqs for illegal seq : 0
Callback statistics: handled 3 cb, average delay : 1, max delay : 2
Delta Sync memory usage: currently using XX KB mem
Callback statistics: handled 322 cb, average delay : 2, max delay : 8
Number of Pending packets currently held: 1
Packets released due to timeout: 18
```

The meaning of each line in this printout is explained below.

```
Version: new
```

This line must appear if synchronization is configured. It indicates that new sync is working (as opposed to old sync from version 4.1).

```
Status: Able to Send/Receive sync packets
```

If sync is unable to either send or receive packets, there is a problem. Sync may be temporarily unable to send or receive packets during boot, but this should not happen during normal operation. When performing full sync, sync packet reception may be interrupted.
Sync packets sent:
  total: 3976, retransmitted: 0, retrans reqs: 58, acks: 97

The total number of sync packets sent is shown. Note that the total number of sync packets is non-zero and increasing.

The cluster member sends a retransmission request when a sync packet is received out of order. This number may increase when under load.

Acks are the acknowledgements sent for received sync packets, when an acknowledgement was requested by another cluster member.

Sync packets received:
  total: 4290, were queued: 58, dropped by net: 47

The total number of sync packets received is shown. The queued packets figure increases when a sync packet is received that complies with one of the following conditions:

1. The sync packet is received with a sequence number that does not follow the previously processed sync packet.
2. The sync packet is fragmented. This is done to solve MTU restrictions.

This figure never decreases. A non-zero value does not indicate a problem.

The dropped by net number may indicate network congestion. This number may increase slowly under load. If this number increases too fast, a networking error may be interfering with the sync protocol. In that case, check the network.

retrans reqs: 0, received 0 acks
retrans reqs for illegal seq: 0
Callback statistics: handled 3 cb, average delay: 1, max delay: 2

This message refers to the number of received retransmission requests, in contrast to the transmitted retransmission requests in the section above. When this number grows very fast, it may indicate that the load on the machine is becoming too high for sync to handle.

Acks refer to the number of acknowledgements received for the “cb request” sync packets, which are sync packets with requests for acknowledgments.
Retrans reqs for illegal seq displays the number of retransmission requests for packets which are no longer in this member’s possession. This may indicate a sync problem.

Callback statistics relate to received packets that involve Flush and Ack. This statistic only appears for a non-zero value.

The callback average delay is how much the packet was delayed in this member until it was released when the member received an ACK from all the other members. The delay happens because packets are held until all other cluster members have acknowledged reception of that sync packet.

This figure is measured in terms of numbers of packets. Normally this number should be small (1–5). Larger numbers may indicate an overload of sync traffic, which causes connections that require sync acknowledgements to suffer slight latency.

dropped updates as a result of sync overload: 0

In a heavily loaded system, the cluster member may drop synchronization updates sent from another cluster member.
Delta Sync memory usage: currently using XX KB mem

Delta Sync memory usage only appears for a non-zero value. Delta sync requires memory only while full sync is occurring. Full sync happens when the system goes up- after reboot for example. At other times, Delta sync requires no memory because Delta sync updates are applied immediately. For information about Delta sync see “How State Synchronization Works” on page 29.

Number of Pending packets currently held: 1
Packets released due to timeout: 18

Number of Pending packets currently held only appears for a non-zero value. ClusterXL prevents out-of-state packets in non-sticky connections. It does this by holding packets until a SYN-ACK is received from all other active cluster members. If for some reason a SYN-ACK is not received, the Security Gateway on the cluster member will not release the packet, and the connection will not be established.

Packets released due to timeout only appears for a non-zero value. If the Number of Pending Packets is large (more than 100 pending packets), and the number of Packets released due to timeout is small, you should take action to reduce the number of pending packets. To tackle this problem, see “Reducing the Number of Pending Packets” on page 202.
Introduction to cphaprob [-reset] syncstat

Heavily loaded clusters and clusters with geographically separated members pose special challenges. High connection rates, and large distances between the members can lead to delays that affect the operation of the cluster.

The cphaprob [-reset] syncstat command is a tool for monitoring the operation of the State Synchronization mechanism in highly loaded and distributed clusters. It can be used for both ClusterXL and third-party OPSEC certified clustering products.

The troubleshooting process is as follows:

1. Run the cphaprob syncstat command.
2. Examine and understand the output statistics.
3. Tune the relevant synchronization global configuration parameters.
4. Rerun the command, resetting the statistics counters using the -reset option:
   
   cphaprob -reset syncstat

5. Examine the output statistics to see if the problem is solved.

The section “Output of cphaprob [-reset] syncstat” on page 146 explains each of the output parameters, and also explains when the output represents a problem.

Any identified problem can be solved by performing one or more of the tips described in “Synchronization Troubleshooting Options” on page 157.
Output of cphaprob [-reset] syncstat

The output parameters of the cphaprob syncstat command are shown below. The values (not shown) give an insight into the state and characteristics of the synchronization network. Each parameter and the meaning of its possible values is explained in the following sections.

- "Sync Statistics (IDs of F&A Peers - 1):" on page 147
- "Other Member Updates:" on page 147
- "Sent Retransmission Requests" on page 147
- "Avg |Missing Updates per Request" on page 148
- "Old or too-new Arriving Updates" on page 148
- "Unsynced Missing Updates" on page 148
- "Lost Sync Connection (num of events)" on page 149
- "Timed out Sync Connection" on page 149
- "Local Updates" on page 149
- "Total Generated Updates" on page 150
- "Recv Retransmission requests" on page 150
- "Recv Duplicate Retrans request" on page 150
- "Blocking Scenarios" on page 151
- "Blocked Packets" on page 152
- "Max Length of Sending Queue" on page 152
- "Avg Length of Sending Queue" on page 153
- "Hold Pkts Events" on page 154
- "Unhold Pkt Events" on page 154
- "Not Held Due to no Members" on page 154
- "Max Held Duration (ticks)" on page 155
- "Avg Held Duration (ticks)" on page 155
- "Timers:" on page 156
- "Sync tick (ms)" on page 156
- "CPHA tick (ms)" on page 156
- "Queues:" on page 156
- "Sending Queue Size" on page 156
- "Receiving Queue Size" on page 156
Output of cphaprob [-reset] syncstat

**Sync Statistics (IDs of F&A Peers - 1):**

These statistics relate to the state synchronization mechanism. The F&A (Flush and Ack) peers are the cluster members that this member recognizes as being part of the cluster. The IDs correspond to IDs and IP addresses generated by the cphaprob state command.

**Other Member Updates:**

The statistics in this section relate to updates generated by other cluster members, or to updates that were not received from the other members. Updates inform about changes in the connections handled by the cluster member, and are sent from and to members. Updates are identified by sequence numbers.

**Sent Retransmission Requests**

The number of retransmission requests, which were sent by this member. Retransmission requests are sent when certain packets (with a specified sequence number) are missing, while the sending member already received updates with advanced sequences.

A high value can imply connectivity problems.

**Tip** - Compare the number of retransmission requests to the Total Regenerated Updates of the other members (see “Total Generated Updates” on page 150).

If its value is unreasonably high (more than 30% of the Total Generated Updates of other members), contact Technical Support equipped with the entire output and a detailed description of the network topology and configuration.
Output of cphaprob [-reset] syncstat

Avg | Missing Updates per Request

Each retransmission request can contain up to 32 missing consecutive sequences. The value of this field is the average number of requested sequences per retransmission request.

More than 20 missing consecutive sequences per retransmission request can imply connectivity problems.

**Tip** - If this value is unreasonably high, contact Technical Support, equipped with the entire output and a detailed description of the network topology and configuration.

Old or too-new Arriving Updates

The number of arriving sync updates where the sequence number is too low, which implies it belongs to an old transmission, or too high, to the extent that it cannot belong to a new transmission.

Large values imply connectivity problems.

**Tip** - See “Enlarging the Receiving Queue” on page 157 If this value is unreasonably high (more than 10% of the total updates sent), contact Technical Support, equipped with the entire output and a detailed description of the network topology and configuration.

Unsynced Missing Updates

The number of missing sync updates for which the receiving member stopped waiting. It stops waiting when the difference in sequence numbers between the newly arriving updates and the missing updates is larger than the length of the receiving queue.

This value should be zero. However, the loss of some updates is acceptable as long as the number of lost updates is less than 1% of the total generated updates.

**Tip** - To decrease the number of lost updates, expand the capacity of the Receiving Queue. See “Enlarging the Receiving Queue” on page 157
**Lost Sync Connection (num of events)**

The number of events in which synchronization with another member was lost and regained due to either Security Policy installation on the other member, or a large difference between the expected and received sequence number.

The value should be zero. A positive value indicates connectivity problems.

**Tip** - Allow the sync mechanism to handle large differences in sequence numbers by expanding the Receiving Queue capacity. See “Enlarging the Receiving Queue” on page 157

**Timed out Sync Connection**

The number of events in which the member declares another member as not connected. The member is considered as disconnected because no ACK packets were received from that member for a period of time (one second), even though there are Flush and Ack packets being held for that member.

The value should be zero. Even with a round trip time on the sync network as high as 100ms, one second should be enough time to receive an ACK. A positive value indicates connectivity problems.

**Tip** - Try enlarging the Sync Timer (see “Enlarging the Sync Timer” on page 158). However, you may well have to contact Technical Support equipped with the entire output and a detailed description of the network topology and configuration.

**Local Updates**

The statistics in this section relate to updates generated by the local cluster member. Updates inform about changes in the connections handled by the cluster member, and are sent from and to members. Updates are identified by sequence numbers.
Total Generated Updates

The number of sync update packets generated by the sync mechanism since the statistics were last reset. Its value is the same as the difference between the sequence number when applying the 

-reset 

option, and the current sequence number.

Can have any value.

Recv Retransmission requests

The number of received retransmission requests. A member requests retransmissions when it is missing specified packets with lower sequence numbers than the ones already received.

A large value can imply connectivity problems.

Tip - If this value is unreasonably high (more than 30% of the “Total Generated Updates” on page 150) contact Technical Support, equipped with the entire output and a detailed description of the network topology and configuration.

Recv Duplicate Retrans request

The number of duplicated retransmission requests received by the member. Duplicate requests were already handled, and so are dropped.

A large value may indicate network problem or storms on the sync network.

Tip - If this value is unreasonably high (more than 30% of the “Total Generated Updates” on page 150) contact Technical Support, equipped with the entire output and a detailed description of the network topology and configuration.
**Blocking Scenarios**

Under extremely heavy load conditions, the cluster blocks new connections. This parameter shows the number of times that the cluster member started blocking new connections due to sync overload.

The member starts to block connections when its Sending Queue has reached its capacity threshold. The capacity threshold is calculated as 80% of the difference between the current sequence number and the sequence number for which the member received an ACK from all the other operating members.

A positive value indicates heavy load. In this case, observe the “Blocked Packets” on page 152 to see how many packets we blocked. Each dropped packet means one blocked connection.

This parameter is only measured if the Block New Connections mechanism (described in “Blocking New Connections Under Load” on page 200) is active. To activate the Block New Connections mechanism, apply the following command on all the cluster members:

```
fw ctl set int fw_sync_block_new_conns 0
```

**Tip** - The best way to handle a severe blocking connections problem is to enlarge the sending queue. See “Enlarging the Sending Queue” on page 157.

Another possibility is to decrease the timeout after which a member initiates an ACK. See “Reconfiguring the Acknowledgment Timeout” on page 159. This updates the sending queue capacity more accurately, thus making the blocking process more precise.
**Blocked Packets**

The number of packets that were blocked because the cluster member was blocking all new connections (see “Blocking Scenarios” on page 151). The number of blocked packets is usually one packet per new connection attempt.

A value higher than 5% of the Sending Queue see “Avg Length of Sending Queue” on page 153) can imply a connectivity problem, or that ACKs are not being sent frequently enough.

This parameter is only measured if the Block New Connections mechanism (described in “Blocking New Connections Under Load” on page 200) is active. To activate the Block New Connections mechanism, apply the following command on all the cluster members:

```
fw ctl set int fw_sync_block_new_conns 0
```

**Tip** - The best way to handle a severe blocking connections problem is to enlarge the sending queue. See “Enlarging the Sending Queue” on page 157.

Another possibility is to decrease the timeout after which a member initiates an ACK. See “Reconfiguring the Acknowledgment Timeout” on page 159. This updates the sending queue capacity more accurately, thus making the blocking process more precise.

**Max Length of Sending Queue**

The size of the Sending Queue is fixed. By default it is 512 sync updates. As newer updates with higher sequence numbers enter the queue, older updates with lower sequence numbers drop off the end of the queue. An older update could be dropped from the queue before the member receives an ACK about that update from all the other members.

This parameter is the difference between the current sync sequence number and the last sequence number for which the member received an ACK from all the other members. The value of this parameter can therefore be greater than 512.

The value of this parameter should be less than 512. If larger than 512, there is not necessarily a sync problem. However, the member will be unable to answer retransmission request for updates which are no longer in its queue.

This parameter is only measured if the Block New Connections mechanism (described in “Blocking New Connections Under Load” on page 200) is active. To activate the Block New Connections mechanism, apply the following command on all the cluster members:

```
fw ctl set int fw_sync_block_new_conns 0
```
Output of `cphaprob [-reset] syncstat`

Tip - Enlarge the Sending Queue to value larger than this value. See "Enlarging the Sending Queue" on page 157.

**Avg Length of Sending Queue**

The average value of the “Max Length of Sending Queue” parameter, since reboot or since the Sync statistics were reset.

The value should be up to 80% of the size of the Sending Queue.

This parameter is only measured if the *Block New Connections* mechanism (described in "Blocking New Connections Under Load" on page 200) is active. To activate the Block New Connections mechanism, apply the following command on all the cluster members:

```
fw ctl set int fw_sync_block_new_conns 0
```

Tip - Enlarge the Sending Queue so that this value is not larger than 80% of the new queue size. See "Enlarging the Sending Queue" on page 157.
**Hold Pkts Events**

The number of occasions where the sync update required Flush and Ack, and so was kept within the system until an ACK arrived from all the other functioning members.

Should be the same as the number of **“Unhold Pkt Events”**.

**Tip** - Contact Technical Support equipped with the entire output and a detailed description of the network topology and configuration.

**Unhold Pkt Events**

The number of occasions when the member received all the required ACKS from the other functioning members.

Should be the same as the number of **“Hold Pkts Events”**.

**Tip** - Contact Technical Support equipped with the entire output and a detailed description of the network topology and configuration.

**Not Held Due to no Members**

The number of packets which should have been held within the system, but were released because there were no other operating members.

When the cluster has at least two live members, the value should be 0.

**Tip** - The cluster has a connectivity problem. Examine the values of the parameters: “Lost Sync Connection (num of events)” on page 149 and “Timed out Sync Connection” on page 149 to find out why the member thinks that it is the only cluster member.

You may also need to contact Technical Support equipped with the entire output and a detailed description of the network topology and configuration.
**Max Held Duration (ticks)**

The maximum time in ticks (one tick equals 100ms) for which a held packet was delayed in the system for Flush and Ack purposes.

It should not be higher than 50 (5 seconds), because of the pending timeout mechanism which releases held packets after a certain timeout. By default, the release timeout is 50 ticks. A high value indicates connectivity problem between the members.

**Tip** - Optionally change the default timeout by changing the value of the fwldbcast_pending_timeout global variable. See “Advanced Cluster Configuration” on page 197 and “Reducing the Number of Pending Packets” on page 202.

Also, examine the parameter “Timed out Sync Connection” on page 149 to understand why packets were held for a long time.

You may also need to contact Technical Support equipped with the entire output and a detailed description of the network topology and configuration.

**Avg Held Duration (ticks)**

The average duration in ticks (tick equals 100ms) that held packets were delayed within the system for Flush and Ack purposes.

The average duration should be about the round-trip time of the sync network. A larger value indicates connectivity problem.

**Tip** - If the value is high, contact Technical Support equipped with the entire output and a detailed description of the network topology and configuration in order to examine the cause to the problem.
**Timers:**
The Sync and CPHA timers perform sync and cluster related actions every fixed interval.

**Sync tick (ms)**
The *Sync timer* performs cluster related actions every fixed interval. By default, the Sync timer interval is 100ms. The base time unit is 100ms (or 1 *tick*), which is also the minimum value.

**CPHA tick (ms)**
The *CPHA timer* performs cluster related actions every fixed interval. By default, the CPHA timer interval is 100ms. The base time unit is 100ms (or 1 *tick*), which is also the minimum value.

**Queues:**
Each cluster member has two queues. The Sending Queue and the Receiving Queue.

**Sending Queue Size**
The Sending Queue on the cluster member stores locally generated sync updates. Updates in the Sending Queue are replaced by more recent updates. In a highly loaded cluster, updates are therefore kept for less time. If a member is asked to retransmit an update, it can only do so if the update is still in its Sending Queue. The default (and minimum) size of this queue is 512. Each member has one sending queue.

**Receiving Queue Size**
The Receiving Queue on the cluster member keeps the updates from each cluster member until it has received a complete sequence of updates. The default (and minimum) size of this queue is 256. Each member keeps a Receiving Queue for each of the peer members.
Synchronization Troubleshooting Options

The following options specify the available troubleshooting options. Each option involves editing a global system configurable parameter to reconfigure the system with different value than the default.

**Enlarging the Sending Queue**

The Sending Queue on the cluster member stores locally generated sync updates. Updates in the Sending Queue are replaced by more recent updates. In a highly loaded cluster, updates are therefore kept for less time. If a member is asked to retransmit an update, it can only do so if the update is still in its Sending Queue. The default (and minimum) size of this queue is 512. Each member has one sending queue.

To enlarge the sending queue size, change the value of the global parameter `fw_sync_sending_queue_size`. See “Advanced Cluster Configuration” on page 197. You must also make sure that the required queue size survives boot. See “How to Configure Gateway to Survive a Boot” on page 198.

Enlarging this queue allows the member to save more updates from other members. However, be aware that each saved update consumes memory. When changing this variable you should consider carefully the memory implications. Changes will only take effect after reboot.

**Enlarging the Receiving Queue**

The Receiving Queue on the cluster member keeps the updates from each cluster member until it has received a complete sequence of updates. The default (and minimum) size of this queue is 256. Each member keeps a Receiving Queue for each of the peer members.

To enlarge the receiving queue size, change the value of the global parameter `fw_sync_recv_queue_size`. See “Advanced Cluster Configuration” on page 197. You must also make sure that the required queue size survives boot. See “How to Configure Gateway to Survive a Boot” on page 198.

Enlarging this queue means that the member can save more updates from other members. However, be aware that each saved update consumes memory. When changing this variable you should carefully consider the memory implications. Changes will only take effect after reboot.
Enlarging the Sync Timer

The sync timer performs sync related actions every fixed interval. By default, the sync timer interval is 100ms. The base time unit is 100ms (or 1 tick), which is therefore the minimum value.

To enlarge the sync timer, change the value of the global parameter fwha_timer_sync_res. See “Advanced Cluster Configuration” on page 197. The value of this variable can be changed while the system is working. A reboot is not needed.

By default, fwha_timer_sync_res has a value of 1, meaning that the sync timer operates every base time unit (every 100ms). If you configure this variable to n, the timer will be operated every n*100ms.

Enlarging the CPHA Timer

The CPHA timer performs cluster related actions every fixed interval. By default, the CPHA timer interval is 100ms. The base time unit is 100ms (or 1 tick), which is also the minimum value.

If the cluster members are geographically separated from each other, set the CPHA timer to be around 10 times the round-trip delay of the sync network.

Enlarging this value increases the time it takes to detect a failover. For example, if detecting interface failure takes 0.3 seconds, and the timer is doubled to 200ms, the time needed to detect an interface failure is doubled to 0.6 seconds.

To enlarge the CPHA timer, change the value of the global parameter fwha_timer_cpha_res. See “Advanced Cluster Configuration” on page 197. The value of this variable can be changed while the system is working. A reboot is not needed.

By default, fwha_timer_cpha_res has a value of 1, meaning that the CPHA timer operates every base time unit (every 100ms). If you configure this variable to n, the timer will be operated every n*100ms.
Reconfiguring the Acknowledgment Timeout

A cluster member deletes updates from its Sending Queue (described in “Sending Queue Size” on page 156) on a regular basis. This frees up space in the queue for more recent updates.

The cluster member deletes updates from this queue if it receives an ACK about the update from the peer member.

The peer member sends an ACK in one of two circumstances—on condition that the Block New Connections mechanism (described in “Blocking New Connections Under Load” on page 200) is active:

- After receiving a certain number of updates.
- If it didn’t send an ACK for a certain time. This is important if the sync network has a considerable line delay, which can occur if the cluster members are geographically separated from each other.

To reconfigure the timeout after which the member sends an ACK, change the value of the global parameter `fw_sync_ack_time_gap`. See “Advanced Cluster Configuration” on page 197. The value of this variable can be changed while the system is working. A reboot is not needed.

The default value for this variable is 10 ticks (10 * 100ms). Thus, if a member didn’t send an ACK for a whole second, it will send an ACK for the updates it received.

Contact Technical Support

If the other recommendations do not help solve the problem, contact Technical Support for further assistance.
ClusterXL Error Messages

In This Section

General ClusterXL Error Messages page 160
SmartView Tracker Active Mode Messages page 162
Sync Related Error Messages page 163
TCP Out-of-State Error Messages page 165
Platform Specific Error Messages page 166

This section lists the ClusterXL error messages. For other, less common error messages, see SecureKnowledge solution sk23642 at http://supportcontent.checkpoint.com/solutions?id=sk23642.

General ClusterXL Error Messages

- FW-1: changing local mode from <mode1> to <mode2> because of ID <machine_id>

  This log message can happen if the working mode of the cluster members is not the same, for example, if one machine is running High Availability, and another Load Sharing Multicast or Unicast mode. In this case, the internal ClusterXL mechanism tries to synchronize the configuration of the cluster members, by changing the working mode to the lowest common mode. The order of priority of the working modes (highest to lowest) is: 1. Synchronization only 2. Load Sharing 3. High Availability (Active Up) 4. High Availability (Primary Up).

- CPHA: Received confirmations from more machines than the cluster size

  This log message can occur during policy installation on the cluster. It means that a serious configuration problem exists in that cluster. Probably some other cluster has been configured with identical parameters and both of them have common networks.

- fwldbcast_timer: peer X probably stopped...

  This is caused when the member that printed this message stops hearing certain types of messages from member X. Verify that cphaprob state shows all members as active and that fw ctl pstat shows that sync is configured correctly and working properly on all members. In such a case it is fair to assume that there was a temporary connectivity problem that was fixed in the meantime. There may be several connections that may suffer from connectivity
problems due to that temporary synchronization problem between the two members. On the other hand, this can indicate that the other member is really down.

• FW-1: fwha_notify_interface: there are more than 4 IPs on interface <interface name> notifying only the first ones

A member of the same cluster as the reporting machine has more than three virtual IP addresses defined on the same interface. This is not a supported configuration and will harm ClusterXL functionality.

• Sync could not start because there is no sync license

This is a license error message: If you have a basic Security Gateway license then sync is also licensed. Check the basic Security Gateway license using cplic print and cplic check.

• FW-1: h_slink: an attempt to link to a link kbuf id not found
fw_conn_post_inspect: fwconn_init_links failed

Several problems of this sort can happen during a full sync session when there are connections that are opened and closed during the full sync process. Full sync is automatic as far as possible, but it is not fully automatic for reasons of performance. A gateway continues to process traffic even when it is serving as a full sync server. This can cause some insignificant problems, such as a connection that is being deleted twice, a link to an existing link, and so forth. It should not affect connectivity or cause security issues.

• Error SEP_IKE_owner_outbound: other cluster member packet in outbound Cluster in not synchronized. Usually happens in OPSEC certified third-party load sharing products for which Support non-sticky connections is unchecked in the cluster object 3rd Party Configuration page.

• FW-1: fwha_pnote_register: too many registering members, cannot register

The critical device (also known as Problem Notification, or pnote) mechanism can only store up to 16 different devices. An attempt to configure the 17th device (either by editing the cphaprob.conf file or by using the cphaprob -d ... register command) will result in this message.

• FW-1: fwha_pnote_register: <NAME> already registered (# <NUMBER>)

Each device registered with the pnote mechanism must have a unique name. This message may happen when registering new pnote device, and means that the device <NAME> is already registered as with pnote number <NUMBER>.
SmartView Tracker Active Mode Messages

- FW-1: fwha_pnote_unregister: attempting to unregister an unregistered device <DEVICE NAME>
  Indicates an attempt to unregister a device which is not currently registered.

- FW-1: alert_policy_id_mismatch: failed to send a log
  A log indicating that there is a different policy id between the two or more members was not sent. Verify all cluster members have the same policy (using fw stat). It is recommended to re-install the policy.

- FW-1: fwha_receive_fwhap_msg: received incomplete HAP packet (read <number> bytes)
  This message can be received when ClusterXL hears CCP packets of clusters of version 4.1. In that case it can be safely ignored.

SmartView Tracker Active Mode Messages

The following error messages can appear in SmartView Tracker Active mode. These errors indicate that some entries may not have been successfully processed, which may lead to missing synchronization information on a cluster member and inaccurate reports in SmartView Tracker.

- FW-1: fwlddist_adjust_buf: record too big for sync. update Y for table <id> failed. fwlddist_state=<val>
  Indicates a configuration problem on a clustered machine. Either synchronization is misconfigured, or there is a problem with transmitting packets on the sync interface. To get more information on the source of the problem
  - Run fw ctl pstat (described in “Monitoring Synchronization (fw ctl pstat)” on page 141).
  - In ClusterXL clusters, run cphaprob -a if to get the statuses of the interfaces (see “Monitoring Cluster Interfaces” on page 126).

  To solve this problem, see “Working with SmartView Tracker Active Mode” on page 201.

- FW-1: fwldbcast_flush: active connections is currently enabled and due to high load it is making sync too slow to function properly. X active updates were dropped
  Indicates that a clustered machine has dropped SmartView Tracker Active mode updates in order to maintain sync functionality. To solve this problem, see “Working with SmartView Tracker Active Mode” on page 201.
Sync Related Error Messages

- **FW-1: fwldbcast_retreq**: machine `<MACHINE_ID>` sent a retrans request for seq `<SEQ_NUM>` which is no longer in my possession (current seq `<SEQ_NUM>`)

  This message appears when the local member receives a retransmission request for a sequence number which is no longer in its sending window. This message can indicate a sync problem if the sending member didn't receive the requested sequence.

- **FW-1: fwldist_save**: WARNING: this member will not be fully synchronized!
  
  FW-1: fwldist_save: current delta sync memory during full sync has reached the maximum of `<MEM_SIZE>` MB
  
  FW-1: fwldist_save: it is possible to set a different limit by changing `fw_sync_max_saved_buf_mem` value

  These messages may appear only during full sync. While performing full sync the delta sync updates are being saved and are applied only after the full sync process has finished. It is possible to limit the memory used for saving delta sync updates by setting the `fw_sync_max_saved_buf_mem` variable to this limit.

- **FW-1: fwldist_mode_change**: Failed to send trap requesting full sync

  This message may appear due to a problem starting the full sync process, and indicates a severe problem. Contact Technical Support.

- **FW-1: State synchronization is in risk. Please examine your synchronization network to avoid further problems!**

  This message could appear under extremely high load, when a synchronization update was permanently lost. A synchronization update is considered to be permanently lost when it cannot be retransmitted because it is no longer in the transmit queue of the update originator. This scenario does not mean that the Security Gateway will malfunction, but rather that there is a potential problem. The potential problem is harmless if the lost sync update was to a connection...
that runs only on a single member as in the case of unencrypted (clear) connections (except in the case of a failover when the other member needs this update).

The potential problem can be harmful when the lost sync update refers to a connection that is non-sticky (see “Non-Sticky Connections” on page 32), as is the case with encrypted connections. In this case the other cluster member(s) may start dropping packets relating to this connection, usually with a TCP out of state error message (see “TCP Out-of-State Error Messages” on page 165). In this case it is important to block new connections under high load, as explained in “Blocking New Connections Under Load” on page 200.

The following error message is related to this one.

• FW-1: fwldbcast_recv: delta sync connection with member <MACHINE_ID> was lost and regained. <UPDATES_NUM> updates were lost.
FW-1: fwldbcast_recv: received sequence <SEQ_NUM> (fragm <FRAG_NUM>, index <INDEX_NUM>), last processed seq <SEQ_NUM>

These messages appear when there was a temporary sync problem and some of the sync updates were not synchronized between the members. As a result some of the connections might not survive a failover.

The previous error message is related to this one.

• FW-1: The use of the non_sync_ports table is not recommended anymore. Refer to the user guide for configuring selective sync instead

Previous versions used a kernel table called non_sync_ports to implement selective sync, which is a method of choosing services that don't need to be synchronized. Selective sync can now be configured from SmartDashboard. See “Choosing Services That Do Not Require Synchronization” on page 31.
TCP Out-of-State Error Messages

When the synchronization mechanism is under load, TCP packet out-of-state error messages may appear in the Information column of SmartView Tracker. This section explains how to resolve each error.

- TCP packet out of state - first packet isn't SYN tcp_flags: FIN-ACK
  TCP packet out of state - first packet isn't SYN tcp_flags: FIN-PUSH-ACK

  These messages occur when a FIN packet is retransmitted after deleting the connection from the connection table. To solve the problem, in SmartDashboard Global properties for Stateful Inspection, enlarge the TCP end timeout from 20 seconds to 60 seconds. If necessary, also enlarge the connection table so it won't fill completely.

- SYN packet for established connection

  This message occurs when a SYN is received on an established connection, and the sequence verifier is turned off. The sequence verifier is turned off for a non-sticky connection in a cluster (or in SecureXL). Some applications close connections with a RST packet (in order to reuse ports). To solve the problem, enable this behavior to specific ports or to all ports. For example, run the command:
  
  `fw ctl set int fw_trust_rst_on_port <port>
  Which means that the Security Gateway should trust a RST coming from every port, in case a single port is not enough.`
Platform Specific Error Messages

Nokia Specific Error Messages

- FW-1: fwha_nok_get_mc_mac_by_ip: received a NULL query
  FW-1: fwha_nok_get_mc_mac_by_ip: nokcl_get_clustermac returned unknown type <TYPE>

  These messages mean that automatic proxy ARP entries for static NAT configuration might not be properly installed.

- FW-1: fwha_nokcl_sync_rx_f: received NULL mbuf from ipso. Packet dropped.
  FW-1: fwha_nokcl_sync_rx_f: received packet with illegal flag=<FLAG>. Drop packet.

  These messages mean that an illegal CPHA packet was received and will be dropped. If this happens more than few times during boot, the cluster malfunctions.

- FW-1: fwha_nokcl_reregister_rx: unregister old magic mac values with IPSO.
  FW-1: fwha_nokcl_reregister_rx: new magic mac values <MAC,FORWARD MAC> registered successfully with IPSO.

  A notification that the operation fw ctl set int fwha_magic_mac succeeded.

- FW-1: fwha_nokcl_reregister_rx: error in de-registration to the sync_rx (<ERR NUM>) new magic macs values will not be applied

  A notification that the operation fw ctl set int fwha_magic_mac failed. Previous MAC values will be retained.

- FW-1: fwha_nokcl_creation_f: error in registration ...
  FW-1: fwha_nok_init: NOT calling nokcl_register_creation since did not de-register yet.
  FW-1: fwha_nok_fini: failed nokcl_deregister_creation with rc=<ERROR NUM>

  These messages mean that an internal error in registration to the IPSO clustering mechanism has occurred. Verify that the IPSO version is supported by this the Security Gateway version and that the Nokia IP Clustering or VRRP cluster is configured properly.

- FW-1: successfully (dis)connected to Nokia Clustering

  A notification that should be normally received during Security Gateway initialization and removal.
• FW-1: fwha_pnote_register: noksr_register_with_status failed
FW-1: fwha_nokia_pnote_expiration: mismatch between nokia device to
ckp device <DEVICE NAME>
FW-1: fwha_nokia_pnote_expiration: can not find the device nokia
claims to be expired
FW-1: fwha_noksr_report_wrapper: attempting to report an unregistered
device <DEVICE NAME>

These messages may appear as a result of a problem in the interaction between
the Nokia and ClusterXL device monitoring mechanisms. A reboot should solve
this problem. Should this problem repeat itself contact Check Point Technical
support.
Member Fails to Start After Reboot

If a reboot (or `cpstop` followed by `cpstart`) is performed on a cluster member while the cluster is under severe load, the member may fail to start correctly. The starting member will attempt to perform a full sync with the existing active member(s) and may in the process use up all its resources and available memory. This can lead to unexpected behavior.

To overcome this problem, define the maximum amount of memory that the member may use when starting up for synchronizing its connections with the active member. By default this amount is not limited. Estimate the amount of memory required as follows:

Table 7-2 Memory required (MB) for Full Sync.

<table>
<thead>
<tr>
<th>Number of open Connections</th>
<th>100</th>
<th>1000</th>
<th>5000</th>
<th>10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>1.1</td>
<td>6.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10000</td>
<td>11</td>
<td>69</td>
<td>329</td>
<td></td>
</tr>
<tr>
<td>20000</td>
<td>21</td>
<td>138</td>
<td>657</td>
<td>1305</td>
</tr>
<tr>
<td>50000</td>
<td>53</td>
<td>345</td>
<td>1642</td>
<td>3264</td>
</tr>
</tbody>
</table>

Note - These figures were derived for cluster members using the Windows platform, with Pentium 4 processors running at 2.4 GHz.

For example, if the cluster holds 10,000 connections, and the connection rate is 1000 connections/sec you will need 69 MB for full sync.

Define the maximum amount of memory using the gateway global parameter:

```
fw_sync_max_saved_buf_mem
```

The units are in megabytes. For details, see “Advanced Cluster Configuration” on page 197.
Chapter

ClusterXL Advanced Configuration

In This Chapter

Working with VPNs and Clusters
Working with NAT and Clusters
Working with VLANS and Clusters
Monitoring the Interface Link State
Working with Link Aggregation and Clusters
Advanced Cluster Configuration
Defining Disconnected Interfaces
Configuring Policy Update Timeout
Enhanced Enforcement of the TCP 3-Way Handshake
Configuring Cluster Addresses on Different Subnets
Moving from a Single Gateway to a ClusterXL Cluster
Adding Another Member to an Existing Cluster
Configuring ISP Redundancy on a Cluster
Enabling Dynamic Routing Protocols in a Cluster Deployment
Working with VPNs and Clusters

In This Section

Configuring VPN and Clusters ........................................ page 170
Defining VPN Peer Clusters with Separate Security Management Servers ........................................ page 171

Configuring VPN and Clusters

Configuring a Security Gateway cluster using SmartDashboard is very similar to configuring a single Security Gateway. All attributes of the VPN are defined in the Gateway Cluster object, except for two attributes that are defined per cluster member.

1. Go to the Gateway Cluster Properties window, Cluster Members page. For each cluster member, in the Cluster member Properties window, configure the VPN tab:
   - **Office Mode for Remote access** — If you wish to use Office Mode for remote access, define the IP pool allocated to each cluster member.
   - **Hardware Certificate Storage List** — If your cluster member supports hardware storage for IKE certificates, define the certificate properties. In that case, Security Management server directs the cluster member to create the keys and supply only the required material for creation of the certificate request. The certificate is downloaded to the cluster member during policy installation.

2. In a VPN cluster, IKE keys are synchronized. In the Synchronization page of the Gateway Cluster Properties window, make sure that Use State Synchronization is selected, even for High Availability configurations.

3. In the Topology page of the Gateway Cluster Properties window, define the encryption domain of the cluster. Under VPN Domain, choose one of the two possible settings:
   - **All IP addresses behind cluster members based on topology information.** This is the default option.
   - **Manually Defined.** Use this option if the cluster IP address is not on the member network, in other words, if the cluster virtual IP address is on a different subnet than the cluster member interfaces. In that case, select a network or group of networks, which must include the virtual IP address of the cluster, and the network or group of networks behind the cluster.
Defining VPN Peer Clusters with Separate Security Management Servers

When working with a VPN peer that is a Check Point Gateway cluster, and the VPN peer is managed by a different Security Management server, do NOT define another cluster object. Instead, do the following:

1. In the objects tree, Network Objects branch, right click and select New Check Point Externally Managed Gateway.

2. In the Topology page, add the external and internal cluster interface addresses of the VPN peer. Do not use the cluster member interface addresses, except in the following cases:
   - If the external cluster is of version 4.1, add the IP addresses of the cluster member interfaces.
   - If the cluster is an OPSEC certified product (excluding Nokia), you may need to add the IP addresses of the cluster members.

   When adding cluster member interface IP addresses, in the interface Topology tab, define the interface as Internal, and the IP Addresses behind this interface as Not defined.

3. In the VPN Domain section of the page, define the encryption domain of the externally managed gateway to be behind the internal virtual IP address of the gateway. If the encryption domain is just one subnet, choose All IP addresses behind cluster members based on topology information. If the encryption domain includes more than one subnet, it must be Manually Defined.
Working with NAT and Clusters

In This Section

Cluster Fold and Cluster Hide  page 172
Configuring NAT on the Gateway Cluster  page 173
Configuring NAT on a Cluster Member  page 173

Cluster Fold and Cluster Hide

Network Address Translation (NAT) is a fundamental aspect of the way ClusterXL works.

When a cluster member establishes an outgoing connection towards the Internet, the source address in the outgoing packets, is the physical IP address of the cluster member interface. The source IP address is changed using NAT to that of the external virtual IP address of the cluster. This address translation is called “Cluster Hide”.

For OPSEC certified clustering products, this corresponds to the default setting in the 3rd Party Configuration page of the cluster object, of Hide Cluster Members' outgoing traffic behind the Cluster's IP address being checked.

When a client establishes an incoming connection to external (virtual) address of the cluster, ClusterXL changes the destination IP address using NAT to that of the physical external address of one of the cluster members. This address translation is called “Cluster Fold”.

For OPSEC certified clustering products, this corresponds to the default setting in the 3rd Party Configuration page of the cluster object, of Forward Cluster's incoming traffic to Cluster Members' IP addresses being checked.
Configuring NAT on the Gateway Cluster

Network Address Translation (NAT) can be performed on a Gateway Cluster, in the same way as it is performed on a Gateway. This NAT is in addition to the automatic “Cluster Fold” and “Cluster Hide" address translations.

To configure NAT, edit the Gateway Cluster object, and in the Gateway Cluster Properties window, select the NAT page. Do NOT configure the NAT tab of the cluster member object.

Configuring NAT on a Cluster Member

It is possible to perform Network Address Translation (NAT) on a non-cluster interface of a cluster member.

A possible scenario for this is if the non-Cluster interface of the cluster member is connected to another (non-cluster) internal Security Gateway, and you wish to hide the address of the non-Cluster interface of the cluster member.

Performing this NAT means that when a packet originates behind or on the non-Cluster interface of the cluster member, and is sent to a host on the other side of the internal Security Gateway, the source address of the packet will be translated.

Configure NAT on a non-cluster interface of a cluster member gateway as follows:

1. Edit the Gateway Cluster object.
2. In the Cluster Member page of the Gateway Cluster Properties window, edit the Cluster Member object.
3. In the Cluster Member Properties window, click the NAT tab.
4. Configure Static or Hide NAT as desired.
Working with VLANS and Clusters

In This Section

VLAN Support in ClusterXL page 174
Connecting Several Clusters on the Same VLAN page 174

VLAN Support in ClusterXL

A VLAN switch tags packets that originate in a VLAN with a four-byte header that specifies which switch port it came from. No packet is allowed to go from a switch port in one VLAN to a switch port in another VLAN, apart from ports (“global” ports) that are defined so that they belong to all the VLANs.

The cluster member is connected to the global port of the VLAN switch, and this logically divides a single physical port into many VLAN ports each associated with a VLAN tagged interface (VLAN interface) on the cluster member.

When defining VLAN tags on an interface, cluster IP addresses can be defined only on the VLAN interfaces (the tagged interfaces). Defining a cluster IP address on a physical interface that has VLANs is not supported. This physical interface has to be defined with the Network Objective Monitored Private.

Note - ClusterXL does not support VLANS on Windows 2000 or Windows 2003 Server.

Connecting Several Clusters on the Same VLAN

It is not recommended to connect the non-secured interfaces (the internal or external cluster interfaces, for example) of multiple clusters to the same VLAN. A separate VLAN, and/or switch is needed for each cluster.

Connecting the secured interfaces (the synchronization interfaces) of multiple clusters is also not recommended for the same reason. Therefore, it is best to connect the secured interfaces of a given cluster via a crossover link when possible, or to an isolated VLAN.

If there is a need to connect the secured or the non-secured interfaces of multiple clusters to the same VLAN you need to make changes to:
• The destination MAC address, to enable communication between the cluster and machines outside the cluster (for ClusterXL Load Sharing Multicast Mode clusters only).

• The source MAC address of the cluster, to enable Cluster Control Protocol communication between cluster members.

**Changes to the Destination MAC Address**

This section applies to ClusterXL Load Sharing Multicast Mode only.

**How the Destination Cluster MAC Address is Assigned in Load Sharing Multicast Mode**

When a machine that is outside the cluster wishes to communicate with the cluster, it sends an ARP query with the cluster (virtual) IP address. The cluster replies to the ARP request with a multicast MAC address, even though the IP address is a unicast address.

This destination multicast MAC address of the cluster is based on the unicast IP address of the cluster. The upper three bytes are \(01.00.5E\), and they identify a Multicast MAC in the standard way. The lower three bytes are the same as the lower three bytes of the IP address. An example MAC address based on the IP address 10.0.10.11 is shown in Figure 8-1.

**Figure 8-1** The Multicast MAC address of the cluster

![Figure 8-1](image)

**Duplicate Multicast MAC Addresses: The Problem**

When more than one cluster is connected to the same VLAN, the last three bytes of the IP addresses of the cluster interfaces connected to the VLAN must be different. If they are the same, then communication from outside the cluster that is intended for one of the clusters will reach both clusters, which will cause communication problems.
For example, it is OK for the cluster interface of one of the clusters connected to the VLAN to have the address 10.0.10.11, and the cluster interface of a second cluster to have the address 10.0.10.12. However, the following addresses for the interfaces of the first and second clusters will cause complications: 10.0.10.11 and 20.0.10.11.

**Duplicate Multicast MAC Addresses: The Solution**

The best solution is to change to the last three bytes of the IP address of all but one of the cluster interfaces that share the same last three bytes of their IP address.

If the IP address of the cluster interface cannot be changed, you must change the automatically assigned multicast MAC address of all but one of the clusters and replace it with a user-defined multicast MAC address. Proceed as follows:

1. In the ClusterXL page of the cluster object, select Load Sharing>Multicast Mode. In the Topology tab, edit the cluster interface that is connected to same VLAN as the other cluster.

2. In the Interface Properties window, General tab, click Advanced.

3. Change the default MAC address, and carefully type the new user defined MAC address. It must be of the form 01:00:5e:xy:yy:yy where x is between 0 and 7 and y is between 0 and f(hex).

**Changes to the Source MAC Address**

This section applies to all ClusterXL modes, both High Availability and Load Sharing, and to OPSEC certified clustering products.

**How the Source Cluster MAC Address is Assigned**

Cluster members communicate with each other using the Cluster Control Protocol (CCP). CCP packets are distinguished from ordinary network traffic by giving CCP packets a unique source MAC address.

- The first four bytes of the source MAC address are all zero: 00.00.00.00
The fifth byte of the source MAC address is a magic number. Its value indicates its purpose.

### Table 8-1

<table>
<thead>
<tr>
<th>Default value of fifth byte</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>0xfe</td>
<td>CCP traffic</td>
</tr>
<tr>
<td>0xfd</td>
<td>Forwarding layer traffic</td>
</tr>
</tbody>
</table>

The sixth byte is the ID of the sending cluster member.

### Duplicate Source Cluster MAC Addresses: The Problem

When more than one cluster is connected to the same VLAN, if CCP and forwarding layer traffic uses multicast, this traffic reaches only the intended cluster.

However, if broadcast is used for CCP and forwarding layer traffic (and in certain other cases), cluster traffic intended for one cluster is seen by all connected clusters, and is processed by the wrong cluster, which causes communication problems.
Duplicate Source Cluster MAC Addresses: The Solution

To ensure that the source MAC address in packets from different clusters that are connected to the same VLAN can be distinguished, change the MAC source address of the cluster interface that is connected to the VLAN in all but one of the clusters.

Use the following gateway configuration parameters to set more than one cluster on the same VLAN. These parameters apply to both ClusterXL and OPSEC certified clustering products.

Table 8-2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Default value</th>
</tr>
</thead>
<tbody>
<tr>
<td>fwha_mac_magic</td>
<td>0xfe</td>
</tr>
<tr>
<td>fwha_mac_forward_magic</td>
<td>0xfd</td>
</tr>
</tbody>
</table>

Changing the values of these gateway configuration parameters alters the fifth part of the source MAC address of Cluster Control Protocol and forwarded packets. Use any value as long as the two gateway configuration parameters are different. To avoid confusion, do not use the value 0x00.

For instruction about how to change these parameters, see "How to Configure Gateway Configuration Parameters" on page 197.
Monitoring the Interface Link State

Enabling Interface Link State Monitoring shortens the time it takes for ClusterXL to detect an interface failure. By monitoring the link state (i.e. the electrical state) of an interface, ClusterXL is immediately alerted to connectivity issues concerning a certain network interface, such as a disconnected cable, or an electrical failure (real or simulated) on a switch.

Interface Link State Monitoring requires an interface device driver that supports link state detection. The device driver reports the link state as either connected or disconnected.

Monitoring the interface link state is particularly useful in scenarios where a monitored interface (either a cluster interface or a monitored private interface) sends ICMP ECHO probe requests which are not answered by hosts or routers on the connected subnet.

When enabled, ClusterXL immediately detects when an interface goes down. When disabled, ClusterXL determines whether an interface is malfunctioning by watching subsecond timeout expiration.

Monitoring Interface Link State is disabled by default.

Note - Interface Link State Monitoring requires an interface device driver that supports link state detection, and is supported on Linux and SecurePlatform only.

Enabling Interface Link State Monitoring

To enable (or disable) Interface Link State Monitoring, set the global parameter fwha_monitor_if_link_state.

Usage:

```
fw ctl set int fwha_monitor_if_link_state <0|1>
```

Options:

- 0 – disables Interface Link State Monitoring. This is the default setting.
- 1 – enables Interface Link State Monitoring

For instructions on how to make these configuration parameters survive reboot, see SecureKnowledge sk26202 at http://supportcontent.checkpoint.com/solutions?id=sk26202.
Working with Link Aggregation and Clusters

In This Section

- Introduction to Working with Link Aggregation and Clusters  page 180
- Redundant Topologies  page 181
- Configuring Interface Bonds  page 187
- Troubleshooting Bonded Interfaces  page 191

Introduction to Working with Link Aggregation and Clusters

When dealing with mission-critical applications, an enterprise requires its network to be highly available. One way to build highly available networks is to provide reliability through redundancy in the network topology. By employing ClusterXL and bonded interfaces, the Security Gateway can now support a higher level of redundancy on the network level, termed Fully Meshed Topology. In this chapter, the concept of creating a Fully Meshed Topology is discussed in “Redundant Topologies”, and the steps necessary for building this topology are provided in the section “Configuring Interface Bonds”.

Note - The solutions provided here are only supported on the Check Point Operating System SecurePlatform running ClusterXL in New High Availability mode.
Redundant Topologies

In This Section

Simple Redundant Topology  page 181
Fully Meshed Redundancy via Interface Bonding  page 182
Bond Failover  page 184
Failover Support for VLANs  page 185

**Simple Redundant Topology**

In the case of switch or gateway failure, a High Availability cluster solution provides system redundancy. Figure 8-2 depicts a redundant system (two synchronized Security Gateway cluster members) deployed in a redundant topology.

**Figure 8-2** A High Availability topology

In this scenario:

- GW-1 and GW-2 are cluster members
- S-1 and S-2 are switches
- C-1 and C-2 are interconnecting networks

In Figure 8-2, cluster members GW-1 and GW-2 each have one external Network Interface Card (NIC) connected to an external switch (S-1 and S-2, respectively). In the event that the active cluster member GW-1 fails, the standby cluster member GW-2 becomes active, connecting to switch S-2 over network C-2.
**Fully Meshed Redundancy via Interface Bonding**

A Fully Meshed Topology further enhances the redundancy in the system by providing a backup to both the interface and the switch, essentially backing up the cable. Each cluster member has two external interfaces, one connected to each switch. This implementation is depicted in Figure 8-3, where both cluster members are connected to both external switches.

**Figure 8-3** A fully meshed topology

In this scenario:

- GW-1 and GW-2 are Security Gateway cluster members in New High Availability mode
- S-1 and S-2 are switches
- C-1, C-2, C-3 and C-4 are networks
On each gateway, only one external interface is active at any one time, with the other interface acting as a standby slave. This is accomplished by creating a bond, where the two interface cards are set to act as a single interface, using the same MAC and IP address. Interface bonding (also known as NIC teaming) allows each cluster member to be attached in an active manner to one switch, while at the same time be passively attached to another. Figure 8-4 depicts a bonded interface.

In this scenario:
- GW-1 is a Security Gateway cluster member
- S-1 and S-2 are switches
- eth0 and eth1 are bonded interfaces
- eth0 is the active interface, eth1 is the standby interface
- bond0 is the name of the bond

If GW-1 should lose connectivity with the currently active switch, it is able to detect the failure and initiate an internal failover to eth1.
**Bond Failover**

Failover can occur because of a failure in the link state, or a failure in the sending or receiving of ClusterXL Control Protocol (CCP) keep-alive packets. Either of these failures will induce a failover within the interface bond, or between cluster members, depending on the circumstances. The section below describes the two types of failover processes.

**Note** - The bond failover operation requires a network interface card that supports the Media-Independent Interface (MII) standard.

**Link State Initiated Failover**

1. The active bonded interface detects a link state of *down*, and notifies the bond interface.

2. The bond initiates an internal bond failover to the standby interface. (As this is a failover within the bond, the status of the other cluster member is not considered.)

3. If this interface should detect a link failure, and the initial interface is still down, ClusterXL initiates a failover to the other cluster member, as long as it is not in status *down*. 
CCP Initiated Failover

This type of failover occurs only when the other cluster member is not in status down.

1. ClusterXL detects a problem in the sending or receiving of CCP packets.
2. ClusterXL initiates an internal bond failover.
3. ClusterXL monitors CCP packet transmission and reception. If a problem is detected within three minutes, the system initiates a failover to the other cluster member.

See “Configuring Interface Bonds” on page 187 for configuration information.

Failover Support for VLANs

ClusterXL can now monitor VLAN IDs for connectivity failure or miscommunication, and initiate a failover when a failure is detected. In a VLAN-enabled switched environment, ClusterXL can be set to monitor either the lowest VLAN ID, or every VLAN on the active interface. The monitoring is conducted by sending ClusterXL Control Protocol (CCP) packets on round-trip paths at a set interval. Figure 8-5 illustrates an interface bond supporting multiple VLANs.

![Figure 8-5 VLANs on a bonded interface](image)

When a failure is detected, a log of the failure is recorded in SmartView Tracker.
Monitoring VLANs

VLAN monitoring can be set to monitor either the lowest VLAN ID.

- Monitoring the lowest VLAN ID

  The lowest VLAN ID indicates the status of the physical connection. This VLAN ID is always monitored, and a connectivity failure will initiate a failover. In most deployments this is the desired setting, as it supports the primary purpose of the feature (detecting a connectivity failure) and the traffic generated on the network is light. However, this setting will not detect a VLAN configuration problem on the switch.

  The default setting for monitoring VLANs is to monitor only the lowest VLAN ID. To modify this setting, see “Configuring Failover Mode” on page 189.

Failover Mode

When a VLAN failure is detected, the system can fail over to the other bonded slave interface, or to another cluster member, depending on the setting. See “Configuring Failover Mode” on page 189 for details.

To configure VLAN support on a bonded interface, see “Defining VLANs on a Bonded Interface (optional)” on page 189. For other command line utilities dealing with VLAN support, see “cphaprob -a if” on page 194.
Configuring Interface Bonds

In This Section

Creating a Bonded Interface  page 187
Configuring Failover Mode  page 189

**Creating a Bonded Interface**

There are four main steps to creating an interface bond on SecurePlatform.

1. “Setting Slave Interfaces as Disconnected”
2. “Removing IP Addresses from Slave Interfaces”
3. “Creating an Interface Bond”
4. “Verifying that the Bond is Functioning Properly”
5. When using VLANs, there is a fifth step as well: “Defining VLANs on a Bonded Interface (optional)”.

**Setting Slave Interfaces as Disconnected**

In a bond, both interfaces are configured to be disconnected slave interfaces. Disconnected interfaces are cluster member interfaces that are not monitored by the ClusterXL mechanism. If a disconnected interface fails, failover does not occur. To define a slave interface as disconnected in SecurePlatform, do the following:

1. In the directory $FWDIR/conf/, create a file named discntd.if.
2. Enter the name of each physical interface that will function as a slave/bond pair on a separate line.

**Removing IP Addresses from Slave Interfaces**

Slave interfaces cannot have IP addresses. The procedure for removing an IP address from an interface in SecurePlatform is detailed here using the sysconfig utility, however WebUI may also be used. To remove an IP address from a slave interface, do the following:

1. Start the SecurePlatform sysconfig utility.
2. Select configuration item 5) Network Connections.
3. Select configuration item 2) Configure Connection.
4. Select the relevant physical interface.
5. Select 3) Remove IP from interface.
6. Repeat for each slave interface.

Creating an Interface Bond

The procedure for creating an interface bond is detailed here using the sysconfig utility, however WebUI may also be used. For each interface bond, do the following:
1. Start the SecurePlatform sysconfig utility.
2. Select configuration item 5) Network Connections.
3. Select network connections configuration item 1) Add new connection.
4. Select connection type to add 4) Bond.
5. Select the interfaces to be enslaved under the bond, and then type n.
6. For bond argument to configure, type n.
7. For “Do you want to set a primary slave interface?” type n.
8. Enter the IP address and network mask of the new interface bond.
9. Press Enter to continue, and repeat steps 2 - 9 for each bond.
10. When finished creating bonds, type Q to exit sysconfig.

Verifying that the Bond is Functioning Properly

After installation or failover, it is recommended to verify that the bond is working.
1. Run the command cphaconf show_bond <bond-name>, and note which interface is active.
2. Test that the failover has succeeded with the command cphaconf failover_bond <bond-name>, and verify that the standby and active interfaces have switched.
Defining VLANs on a Bonded Interface (optional)

The procedure for defining a VLAN on a bonded interface is detailed here using the
sysconfig utility, however WebUI may also be used. For each VLAN, do the
following:
1. Start the SecurePlatform sysconfig utility.
2. Select configuration item 5) Network Connections.
3. Select network connections configuration item 1) Add new connection.
4. Select connection type to add 2) VLAN.
5. Select the bonded interface for the VLAN.
6. Enter the VLAN ID.
7. Enter the IP address and network mask of the VLAN.
8. Press Enter to continue, and repeat steps 2 - 8 for each VLAN.
9. When finished creating VLANs, type Q to exit sysconfig.

Configuring Failover Mode

There are a number of configurable settings regarding failover:

- "fw ctl set int fwha_manual_bond_failover" sets the failover mode:
  0 - fails over to the bond's other interface
  1 - fails over to the next cluster member, unless the command cphaconf enable_bond_failover is run, in which case the next failover will be to the bond's other interface

  In both modes, the next bond failover occurs in three minutes.

- "cphaconf enable_bond_failover" sets what happens during a failover after a bond has already failed over internally. It works only if fw ctl set int fwha_manual_bond_failover 1 was run previously.
fw ctl set int fwha_manual_bond_failover

The fwha_manual_bond_failover command is used to set the failover mode, either within the bond or to the next cluster member.

Usage:

```
fw ctl set int fwha_manual_bond_failover <0|1>
```

Options:

- 0 – set the system to fail over to the other bonded slave interface when an interface failure is detected. This is the default setting.
- 1 – set the system to fail over to another cluster member when an interface failure is detected.

There is no immediate output from this command, and it does not survive reboot. However, if added as a line to the file $FWDIR/boot/modules/fwkern.conf, it will survive reboot.

cphaconf enable_bond_failover

After a failover occurs within a bond, the next time a VLAN failure is detected the system automatically fails over to the other cluster member. An administrator can prevent this from occurring by first correcting the VLAN error that caused the failover, and then resetting the system to failover internally. The enable_bond_failover command directs the system to failover within the bond the next time a VLAN failure is detected.

This command should be run each time the system is reconfigured or restarted, after verifying that all VLANs are active.

Usage:

```
cphaconf enable_bond_failover <bondname>
```

Options:

- bondname – enter the name of the relevant bond

When successful, there is no immediate output from this command; however the words can failover appear in the output of cphaprobro -a if.
Troubleshooting Bonded Interfaces

In This Section

- Introduction to Troubleshooting Bonded Interfaces  page 191
- Bond Status in SmartView Tracker  page 192
- Check the Status of the Bond Interface  page 192
- Commands to be Used with Interface Bonds  page 193
- cphaprob -a if  page 194
- Connectivity Delays on Switches  page 195

Introduction to Troubleshooting Bonded Interfaces

This section presents troubleshooting methodology and the command line utilities for working with bonded interfaces on ClusterXL. This section is organized as follows:

1. The first step in troubleshooting bonded interfaces is to check the status of the bond. This can be done in a few ways:
   - Via the command `cphaconf show_bond`, detailed in “Verifying that the Bond is Functioning Properly” on page 188.
   - Via the command `cphaprob -a if`, detailed in “Check the Status of the Bond Interface” on page 192.

2. Further information regarding bond status and failovers may be found in SmartView Tracker. See “Bond Status in SmartView Tracker” on page 192.

3. Verify that slave interfaces in the bond are defined as disconnected. See “Check that all Bond Slaves Report as Disconnected” on page 193.

Usage and syntax of the relevant commands are discussed in “Commands to be Used with Interface Bonds” on page 193 and “cphaprob -a if” on page 194.
**Bond Status in SmartView Tracker**

A change in status of an interface bond is logged in SmartView Tracker when the failover is initiated by ClusterXL, and any the following occurs.

- When a bond’s status changes to only one slave interface running, the following is recorded:

  The bond interface bond0 has only one slave device up. Check the bond’s slave eth2 and repair it.

  In this example, bond0 is the bond interface, and eth2 is the problematic slave interface.

- When a CCP sending/receiving packet problem occurs, the following is recorded:

  Interface bond0.VID of member1 has done an internal failover. Please check the non standby slave and repair if needed.

  In this example, bond0.VID indicates the bond interface and the problematic VLAN ID, and member1 is the cluster member.

- When a bond’s status returns to normative, the following is recorded:

  The bond interface bond0 has two slave devices up. It can now failover.

  In this example, bond0 is the bond interface.

**Check the Status of the Bond Interface**

The status of the bond interface can be checked via the command `cphaprob -a if`.

- If the bond interface reports that it *can failover*, the bond is functioning properly.

- If the bond interface reports that it *cannot failover*, it could be for either of two reasons:
  - The link is down
  - One of the cluster members is down

**To Check if the Link is Down**

1. Run the command `cphaconf show_bond <bond-name>`
2. Look for a slave interface that reports the status of the link as *no*.
3. Check the cable connections and other hardware.
4. Check the port configuration on the switch
To Check if a Cluster Member is Down

Run the command `cphaprob state`. If any of the cluster members have a Firewall State other than active, see “Monitoring Cluster Status (cphaprob state)” in the ClusterXL User Guide for troubleshooting help.

Check that all Bond Slaves Report as Disconnected

1. Run the command `cphaconf show_bond <bond-name>`, and note the names of the slave interfaces.
2. Run the command `cphaprop -a if` to check that these interfaces report as disconnected. If not, refer to the section “Setting Slave Interfaces as Disconnected” on page 187. After following directions, reboot the machine.

Commands to be Used with Interface Bonds

In This Section

- `cphaconf show_bond` page 193
- `cphaconf failover_bond` page 194
- `cphaprop -a if` page 194

**cphaconf show_bond**

The `show_bond` command displays the status of each network interface card configured as a bonded slave.

**Usage:**

```
cphaconf show_bond <bond-name>
```

**Options:**

- `bond-name` – enter the name of the relevant bond

**Example:**

```
[Expert@GW-1]# cphaconf show_bond bond0
Slave Name    Status       Link
-------------------------
eth2            Active      Yes
eth3   Not Available  Yes
```
Troubleshooting Bonded Interfaces

Report Results

- **Status**
  - **Active** - displays the interface currently handling traffic
  - **Standby** - indicates the interface is ready, and can support internal bond failover
  - **Not Available** - indicates that either the physical link is broken, or that the Cluster member is in status *down*. The bond cannot failover in this state.

- **Link** - reports whether the physical link exists

**cphaconf failover_bond**

The `failover_bond` command provides the ability to initiate an internal failover within the bond.

**Usage:**

```
cphaconf failover_bond <bond-name>
```

**Options:**

- **bond-name** – enter the name of the relevant bond

**cphaprob -a if**

The `cphaprob -a if` command displays the status of all VLANs, and which interfaces can failover.

**Usage:**

```
cphaprob -a if
```
Example:

```
[Expert@GW-1]# cphaprof -a if
Required interfaces: 5
Required secured interfaces: 1
eth0   Disconnected  non sync(non secured), broadcast
eth1   Disconnected  non sync(non secured), broadcast
eth2   UP             sync(secure), broadcast
eth3   Disconnected  non sync(non secured), broadcast
eth4   UP             non sync(non secured), broadcast
eth5   Disconnected  non sync(non secured), broadcast
bond0  UP             non sync(non secured), broadcast, bond, can failover
bond0  UP             non sync(non secured), broadcast (bond0.22 )
bond0  UP             non sync(non secured), broadcast (bond0.23 )
bond0  UP             non sync(non secured), broadcast (bond0.24 )
Virtual cluster interfaces: 4
eth4    50.0.2.150
bond0.22 50.0.5.55
bond0.23 50.0.3.35
bond0.24 50.0.4.45
```

**Connectivity Delays on Switches**

When using certain switches, connectivity delays may occur during some internal bond failovers. With the various features that are now included on some switches, it can take close to a minute for a switch to begin servicing a newly connected interface. The following are suggestions for reducing the startup time after link failure.

1. Disable autonegotiation on the relevant interface.
2. On some Cisco switches, enable the PortFast feature, the configuration of which is demonstrated below.
Sample Configuration of PortFast Feature on a Cisco Switch

The following are the commands necessary to enable PortFast on a GigabitEthernet 1/0/15 interface of a Cisco 3750 switch running IOS.

1. Enter configuration mode:
   ```
   cisco-3750A#conf t
   ```

2. Specify the interface to configure:
   ```
   cisco-3750A(config)#interface gigabitethernet1/0/15
   ```

3. Set PortFast on this interface:
   ```
   cisco-3750A(config-if)#spanning-tree portfast
   ```

Warnings Regarding Use of PortFast

The PortFast feature should never be used on ports that connect to other switches or hubs. It is important that the Spanning Tree complete the initialization procedure in these situations. Otherwise, these connections may cause physical loops where packets are continuously forwarded (or even multiply) in such a way that network will ultimately crash.
Advanced Cluster Configuration

In This Section

- How to Configure Gateway Configuration Parameters  
  page 197
- How to Configure Gateway to Survive a Boot  
  page 198
- Controlling the Clustering and Synchronization Timers  
  page 199
- Blocking New Connections Under Load  
  page 200
- Working with SmartView Tracker Active Mode  
  page 201
- Reducing the Number of Pending Packets  
  page 202
- Configuring Full Synchronization Advanced Options  
  page 202

How to Configure Gateway Configuration Parameters

A number of synchronization and ClusterXL capabilities are controlled by means of Security Gateway configuration parameters. Run these commands on the Security Gateway as follows:

```
fw ctl set int Parameter <value>
```

*Parameter* is any of the parameters described in the following sections.

Changes to their default values must be implemented on all cluster members. Setting different values on cluster members can cause configuration problems and possibly connection failures.

All these gateway configuration parameters can be configured to survive a boot. The way to do this varies with the operating system.
How to Configure Gateway to Survive a Boot

Gateway configuration parameters that are changed using the `fw ctl set int` command do not survive reboot. The way to do make them survive a reboot varies with the operating system. In the following instructions, *Parameter* is any of the parameters described in the following sections.

**Linux/SecurePlatform**

1. Edit the file `$FWDIR/boot/modules/fwkern.conf`.
2. Add the line `Parameter=<value in hex>`.
3. Reboot.

**Windows**

1. Edit the registry.
2. Add a DWORD value named `Parameter` under the key 
   `HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\FW1\Parameters\Globals`.
3. Reboot.

**Nokia**

Run the command `modzap _Parameter $FWDIR/boot/modules/fwmod.o <value in hex>`.

Note that the underscore before `Parameter` is not a mistake.
Controlling the Clustering and Synchronization Timers

The following gateway configuration parameters are used to control the clustering and synchronization timers. Changing the default values is not recommended.

Table 8-3  Clustering and Synchronization timers

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>fwha_timer_cpha_res</td>
<td>The frequency of ClusterXL operations on the cluster. Operations occur every: 10 multiplied by fwha_timer_cpha_res multiplied by fwha_timer_base_res milliseconds</td>
<td>1</td>
</tr>
<tr>
<td>fwha_timer_sync_res</td>
<td>The frequency of sync flush operations on the cluster. Operations occur every: 10 multiplied by fwha_timer_sync_res multiplied by fwha_timer_base_res milliseconds</td>
<td>1</td>
</tr>
<tr>
<td>fwha_timer_base_res</td>
<td>Must be divisible by 10 with no remainders.</td>
<td>10</td>
</tr>
</tbody>
</table>
Blocking New Connections Under Load

The reason for blocking new connections is that new connections are the main source of new synchronization traffic, and synchronization may be put at risk if new traffic continues to be processed at this rate.

A related error message is: “FW-1: State synchronization is in risk. Please examine your synchronization network to avoid further problems!” on page 163.

Reducing the amount of traffic passing through the Security Gateway protects the synchronization mechanism.

- `fw_sync_block_new_conns` allows Security Gateway to detect heavy loads and start blocking new connections. Load is considered heavy when the synchronization transmit queue of the firewall starts to fill beyond the `fw_sync_buffer_threshold`.
  - To enable load detection, set to 0.
  - To disable load detection, set to -1 (the default).

Note that blocking new connections when sync is busy is only recommended for Load Sharing ClusterXL deployments. While it is possible to block new connections in High Availability mode, doing so does not solve inconsistencies in sync, as High Availability mode precludes that from happening. This parameter can be set to survive boot using the mechanism described in “How to Configure Gateway to Survive a Boot” on page 198.

- `fw_sync_buffer_threshold` is the maximum percentage of the buffer that may be filled before new connections are blocked. By default it is set to 80, with a buffer size of 512. By default, if more than 410 consecutive packets are sent without getting an ACK on any one of them, new connections are dropped. When blocking starts, `fw_sync_block_new_conns` is set to 1. When the situation stabilizes it is set back to 0.

- `fw_sync_allowed_protocols` is used to determine the type of connections that can be opened while the system is in a blocking state. Thus, the user can have better control over the system’s behavior in cases of unusual load. The `fw_sync_allowed_protocols` variable is a combination of flags, each specifying a different type of connection. The required value of the variable is the result of adding the separate values of these flags. For example, the default value of this
variable is 24, which is the sum of TCP_DATA_CONN_ALLOWED (8) and UDP_DATA_CONN_ALLOWED (16), meaning that the default allows only TCP and UDP data connections to be opened under load.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICMP_CONN_ALLOWED</td>
<td>1</td>
</tr>
<tr>
<td>TCP_CONN_ALLOWED</td>
<td>2 (except for data connections)</td>
</tr>
<tr>
<td>UDP_CONN_ALLOWED</td>
<td>4 (except for data connections)</td>
</tr>
<tr>
<td>TCP_DATA_CONN_ALLOWED</td>
<td>8 (the control connection should be established or allowed)</td>
</tr>
<tr>
<td>UDP_DATA_CONN_ALLOWED</td>
<td>16 (the control connection should be established or allowed)</td>
</tr>
</tbody>
</table>

Working with SmartView Tracker Active Mode

Active mode in SmartView Tracker shows connections currently open through any of the Security Gateways that are sending logs to the currently active Log File on the Security Management server.

Active mode tends to slow down synchronization. If that happens, the synchronization mechanism randomly drops Active connection updates in order to maintain synchronization. The drop will be accompanied by one of the error message described in “SmartView Tracker Active Mode Messages” on page 162.

Active mode view is not recommended on a heavily loaded cluster. To obtain a more accurate report of Active connections under load, two solutions are available. They apply both to a cluster and to a single Security Gateway:

1. Enlarge fwldist_buf_size

   The fwldist_buf_size parameter controls the size of the synchronization buffer in words. (Words are used for both synchronization and in SmartView Tracker Active mode. 1 word equals 4 Bytes). The default is 16k words. The maximum value is 64k words and the minimum value is 2k words.

   If changing this parameter, make sure that it survives boot, because the change is only applied after a reboot. Use the mechanism described in “How to Configure Gateway Configuration Parameters” on page 197.

2. Obtain a Hotfix from Technical Support

   Obtain a Check Point Technical Support Hotfix. This Hotfix has a variable that controls the rate at which Active connections are read by fwd on the gateway before being sent to the Security Management server. Note that this solution requires additional CPU resources.
Reducing the Number of Pending Packets

ClusterXL prevents out-of-state packets in non-sticky connections. It does this by holding packets until a Sync ACK is received from all other active cluster members. If for some reason a Sync ACK is not received, the Security Gateway on the cluster member will not release the packet, and the connection will not be established.

To find out if held packets are not being released, run the `fw ctl pstat` command. If the output of the command shows that the Number of Pending Packets is large under normal loads (more than 100 pending packets), and this value does not decrease over time, use the `fwldbcast_pending_timeout` parameter to reduce the number of pending packets.

Change the value of `fwldbcast_pending_timeout` from the default value of 50 to a value lower than 50.

The value is in ticks units, where each tick is equal to 0.1 sec, so that 50 ticks is 5 seconds.

The value represents the time after which packets are released even if Sync ACKs are not received.

Configuring Full Synchronization Advanced Options

When a cluster member comes up after being rebooted (or after `cpstart`), it has to perform Full Synchronization. As a first step in the Full Synchronization process, it performs a handshake with one of the other active cluster members. Only if this handshake succeeds does the cluster member continue with the Full Synchronization process.

The extended handshake that takes place (by default) exchanges information between cluster members. This information includes version information, information about the installed Check Point products, and can include information about which the VPN kernel tables are currently active. The extended handshake is unrelated to the exchange of kernel table information that happens later in the Full Synchronization.

All cluster members must have the same Check Point products and versions installed. The extended handshake identifies when different products are installed on the cluster members. When different products are installed, a console warning and a log message are issued.
In order to support backward compatibility, it is possible to change the behavior of the extended handshake by means of the following Gateway Configuration Parameters. How to edit these parameters is explained in “Advanced Cluster Configuration” on page 197:

- `fw_sync_no_ld_trans` has the default value of 1. Set to 0 in order to exchange kernel table information between members in the first phase of the Full Synchronization process.

- `fw_sync_no_conn_trans` has the default value of 0. Set to 1 in order not to exchange installed product information between members in the first phase of the Full Synchronization process.

- `fw_sync_fcu_ver_check` has the default value of 1. Set to 0 to allow Full Connectivity Upgrade for versions that do not comply with the version requirements. Read about these requirements in *The Upgrade Guide*. 
Defining Disconnected Interfaces

Disconnected interfaces are cluster member interfaces that are not monitored by the ClusterXL mechanism.

You may wish to define an interface as disconnected if the interface is down for a long time, and you wish the cluster member to continue to be active.

The processes listed below are equivalent to defining a non-monitored interface from the Topology page, with the exception that the GUI method works only for interfaces that have a defined IP address.

Defining a Disconnected Interface on Unix

Create a file under $FWDIR/conf/discntd.if and write the name of each interface that you do not want monitored by ClusterXL on a separate line.

Defining a Disconnected Interface on Windows

1. Open the regedit32 registry editor. Do not use regedit.
2. Under HKEY_LOCAL_MACHINES\System\CurrentControlSet\Services\CPHA create a new value with the following characteristics:
   Value Name : DisconnectedInterfaces
   Data Type : REG_MULTI_SZ
3. Add the interface name. To obtain the interface system name run the command:
   fw getifs
4. Add this name to the list of disconnected interfaces using the following format:
   \device\<System Interface Name>
5. Run cphastop and then cphastart to apply the change.
Configuring Policy Update Timeout

When policy is installed on a Gateway Cluster, the cluster members undertake a negotiation process to make sure all of them have received the same policy before they actually applying it. This negotiation process has a timeout mechanism which makes sure a cluster member does not wait indefinitely for responses from other cluster members, which is useful in cases when another cluster member goes down when policy is being installed (for example).

In configurations on which policy installation takes a long time (usually caused by a policy with a large number of rules), a cluster with more than two machines, and slow machines, this timeout mechanism may expire prematurely.

It is possible to tune the timeout by setting the following parameter:

fwha_policy_update_timeout_factor.

The default value is 1 which should be sufficient for most configurations. For configurations where the situation described above occurs, setting this parameter to 2 should be sufficient. Do NOT set this parameter to a value larger than 3.
Enhanced Enforcement of the TCP 3-Way Handshake

The standard enforcement on the 3-way handshake that initiates a TCP connection provides good security enforcement by guaranteeing one-directional stickiness. This means that it ensures that the SYN-ACK will always arrive after the SYN. However, it does not guarantee that the ACK will always arrive after the SYN-ACK, or that the first data packet will arrive after the ACK.

If you wish to have an extra strict policy that denies all out-of-state packets, it is possible to configure the synchronization mechanism so that all the TCP connection initiation packets arrive in the right sequence (SYN, SYN-ACK, ACK, followed by the data). The price to be paid for this extra security is a considerable slowdown in connection establishment.

To configured enhanced enforcement, use the Database Tool to change the global property `sync_tcp_handshake_mode` from the default value of `minimal_sync` to `complete_sync`. 
Configuring Cluster Addresses on Different Subnets

In this Section

Introduction to Cluster Addresses on Different Subnets ... page 207
Configuration of Cluster Addresses on Different Subnets ... page 208
Example of Cluster Addresses on Different Subnets ... page 209
Limitations of Cluster Addresses on Different Subnets ... page 210

Introduction to Cluster Addresses on Different Subnets

Cluster IPs are virtual IP addresses given to ClusterXL objects, which differ from the unique IPs of the individual cluster machines. These addresses enable the cluster to be seen as a single gateway, thus allowing it to serve as a router in a network that is unaware of the cluster’s internal structure and status.

In previous versions, cluster IP addresses had to be configured on the same subnets as those used by the unique addresses of the cluster members. Cluster IPs can reside on subnets other than those of the members. The advantage of this is that it

• Enables a multi-machine cluster to replace a single-machine gateway in a pre-configured network, without the need to allocate new addresses to the cluster members.

• Makes it possible to use only one routable address for the ClusterXL Gateway Cluster

Note - This capability is available only for ClusterXL Gateway Clusters. For details about OPSEC certified clusters, see the vendor documentation.

An important aspect of this is that packets sent from cluster members (as opposed to packets routed through the members) are hidden behind the cluster IP and MAC addresses. The cluster MAC is the:

• MAC of the active machine, in High Availability New mode.

• Multicast MAC, in Load Sharing Multicast mode.
• Pivot member MAC in Load Sharing Unicast mode.

This enables the members to communicate with the surrounding networks, but also has certain limitations, as described in “Limitations of Cluster Addresses on Different Subnets” on page 210.

Configuration of Cluster Addresses on Different Subnets

There are two major steps required in order for ClusterXL to function correctly with cluster IPs on different subnets.

The first step is to create static routes on each cluster member, which determine the interface connected to the cluster's network (the subnet to which the cluster IP belongs). Unless these entries are created, the OS cannot route packets to the cluster's network. No additional configuration is required for the cluster members. It is, however, important to note that the unique IPs given to the members must share common subnets on each “side” of the cluster (meaning, each interface on each machine must have an interface on every other machine using the same subnet).

The second step relates to the configuration of the cluster topology. Here the cluster IPs are determined, and associated with the interfaces of the cluster members (each member must have an interface responding to each cluster IP). Normally, cluster IPs are associated with an interface based on a common subnet. In this case these subnets are not the same. It must be explicitly specified which member subnet is associated with the cluster IP.

To specify the member network:

2. Click Edit Topology.
3. In the Edit Topology window, manually enter the IP address and subnet in the appropriate member interface fields.

Note that this interface actually refers to the cluster's virtual IP address, as determined in the cluster topology.
Example of Cluster Addresses on Different Subnets

In this example, a single-gateway firewall separating network 172.16.6.0 (Side “A”) from network 172.16.4.0 (Side “B”) is to be replaced with a ClusterXL cluster. The cluster members, however, will use networks 192.168.1.0 for Side “A”, 192.168.2.0 for Side “B” and 192.168.3.0 for the synchronization network (all network addresses given in this example are of class “C”). The addresses in italics are the cluster IP addresses. The resulting configuration is depicted in Figure 8-6: 

![Cluster addresses on different subnets](image)

**Configuring Static Routes on the Members**

Each member should be configured with two static routes:

- One setting its 192.168.1.x IP address as the gateway for network 172.16.6.0
- One setting its 192.168.2.x IP address as the gateway for network 172.16.4.0.

To configure a static route on SecurePlatform, run `sysconfig` from the command prompt, choose *Routing > Add New Network Route*, and follow the instructions.

**Configuring Cluster IP Addresses in SmartDashboard**

Configure the cluster interface IP addresses in this example as follows

1. In the Gateway cluster object *Topology > Edit Topology* window, edit a cluster interface, and open the *Interface Properties* window.
2. For each cluster interface, configure the **Interface Properties** window as follows:

<table>
<thead>
<tr>
<th></th>
<th>Cluster Interface A IP address</th>
<th>Cluster Interface B IP address</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General tab</strong></td>
<td>172.16.6.100</td>
<td>172.16.4.100</td>
</tr>
<tr>
<td><strong>Member Networks tab</strong></td>
<td>192.168.1.0</td>
<td>192.168.2.0</td>
</tr>
</tbody>
</table>

- All IP addresses have the Netmask 255.255.255.0

**Note** - Do not define Cluster IP addresses for the synchronization interfaces. The synchronization interfaces are also defined in the **Edit Topology** page of the Gateway Cluster object.

---

**Limitations of Cluster Addresses on Different Subnets**

In This Section

- Connectivity Between Cluster Members  page 210
- Load Sharing Multicast Mode with “Semi-Supporting” Hardware  page 212
- Manual Proxy ARP  page 212
- Connecting to the Cluster Members from the Cluster Network  page 213
- Default Gateway on SecurePlatform  page 213
- Anti-Spoofing  page 214

This new feature does not yet support all the capabilities of ClusterXL. Some features require additional configuration to work properly, while others are not supported.

**Connectivity Between Cluster Members**

Since ARP requests issued by cluster members are hidden behind the cluster IP and MAC, requests sent by one cluster member to the other may be ignored by the destination machine. To allow cluster members to communicate with each other, a static ARP should be configured for each cluster member, stating the MAC addresses of all other machines in the cluster. IP packets sent between members are not altered, and therefore no changes should be made to the routing table.
Limitations of Cluster Addresses on Different Subnets

Note - Static ARP is not required in order for the machines to work properly as a cluster, since the cluster synchronization protocol does not rely on ARP.
Limitations of Cluster Addresses on Different Subnets

Load Sharing Multicast Mode with “Semi-Supporting” Hardware

Although not all types of network hardware work with multicast MAC addresses, some routers can pass such packets, even though they are unable to handle ARP replies containing a multicast MAC address. Where a router semi-supports Load sharing Multicast mode, it is possible to configure the cluster MAC as a static ARP entry in the router's internal tables, and thus allow it to communicate with the cluster.

When different subnets are used for the cluster IPs, static ARP entries containing the router's MAC need to be configured on each of the cluster members. This is done because this kind of router will not respond to ARP requests containing a multicast source MAC. These special procedures are not required when using routers that fully support multicast MAC addresses.

Manual Proxy ARP

When using static NAT, the cluster can be configured to automatically recognize the hosts hidden behind it, and issue ARP replies with the cluster MAC address, on their behalf. This process is known as Automatic Proxy ARP.

However, if you use different subnets for the cluster IP addresses, this mechanism will not work, and you must configure the proxy ARP manually. To do so, in SmartDashboard, select Policy menu > Global Properties > NAT Network Address Translation, and disable Automatic ARP Configuration. Then create a file called local.arp in the firewall's configuration directory ($FWDIR/conf).

Each entry in this file is a triplet, containing the:
- host address to be published
- MAC address that needs to be associated with the IP address
- unique IP of the interface that responds to the ARP request.

The MAC address that should be used is the cluster's multicast MAC defined on the responding interface, when using multicast LS, or this interface's unique IP, for all other modes.

For example, if host 172.16.4.3 is to be hidden using the address 172.16.6.25, and the cluster uses Load Sharing Multicast mode, add the following line to the local.arp file of Member 1:

172.16.6.25 00:01:5e:10:06:64 192.168.1.1
The second parameter in this line is the multicast MAC address of cluster IP 172.16.6.100, through which ARP requests for 172.16.6.25 will be received. On Member 2, this line will be:

172.16.6.25 00:01:5e:10:06:64 192.168.1.2

If the cluster is in unicast LS mode, or in HA mode, the entries on Member 1 and 2 will be:

172.16.6.25 00:A0:C9:E8:C7:7F 192.168.1.1
- And -
172.16.6.25 00:A0:C9:E8:CB:3D 192.168.1.2

where the second entry in each line is the unique MAC address of the matching local interface.

**Connecting to the Cluster Members from the Cluster Network**

Since the unique IPs may be chosen arbitrarily, there is no guarantee that these addresses are accessible from the subnet of the cluster IP. In order to access the members through their unique IPs, you must configure routes on the accessing machine, such that the cluster IP is the gateway for the subnet of the unique IPs. Following the above example, 172.16.6.100 should be the gateway for subnet 192.168.1.0.

**Default Gateway on SecurePlatform**

Run `sysconfig > routing > add network route > add the routable network with its subnet`, and choose the correct physical interface in this direction.

Now go to `routing > add default gateway` and add the IP address of the default (routable) gateway. This will usually be the IP address of the router in one of the cluster IP’s subnet.

If you have the different subnets feature configured on more than one interface, repeat the addition of the network address (as above) for all these interfaces. (It is NOT required to define a default gateway for the other subnets as well.)
Anti-Spoofing

When the different subnets feature is defined on a non-external interface, the cluster IP in the Cluster Topology tab should not be defined with the Network defined by interface IP and Net Mask definition in the Topology tab of the Interface Properties window of the cluster interface. You must add a group of networks that contain both the routable network and the non-routable network, and define the Anti-spoofing for this interface as specific: network with this new group.

In the example shown in Figure 8-6 on page 209, suppose side “B” is the internal network, you must define a group which contains both 172.16.4.0 and 192.168.2.0 networks, and define the new group in the specific field of the Topology tab.
Moving from a Single Gateway to a ClusterXL Cluster

This procedure describes how to add a new gateway (Machine 'B') to a standalone Security Gateway (Machine 'A') to create a cluster.

As a prerequisite, there should be available IP addresses in a quantity equal to the number of new cluster members. If there are not, see: “Configuring Cluster Addresses on Different Subnets” on page 55.

On the Single Gateway Machine

If your single gateway installation uses the same machine for the Security Management server and the gateway:

1. Separate the Security Management server from the gateway, and place them on two machines.
2. Initialize SIC on the separated gateway (Machine 'A').

On Machine 'B'

1. Define an interface on machine 'B' for each proposed cluster interface and synchronization interface on machine 'A', with the same subnet. If the members exist on different subnets, see: “Configuring Cluster Addresses on Different Subnets” on page 55
2. Install the Security Gateway on the machine. During the installation you must enable ClusterXL.
In SmartDashboard, for Machine ‘B’

1. Create a ClusterXL object.
2. In the Cluster Members page, click Add, and select New Cluster Member.
3. Connect to machine 'B', and define its topology.
4. Define the Synchronization networks for the cluster.
5. Define the cluster topology. To avoid reconfiguring network devices, the cluster IP addresses should be the same as the addresses of machine 'A', on its proposed cluster interfaces.
6. Install the policy on the cluster, currently including member 'B' only.

On Machine 'A'

1. Disconnect all proposed cluster and Synchronization interfaces. New connections now open through the cluster, instead of through machine 'A'.
2. Change the addresses of these interfaces to some other unique IP address which is on the same subnet as machine B.
3. Connect each pair of interfaces of the same subnet using a dedicated network. Any hosts or gateways previously connected to the single gateway must now be connected to both machines, using a hub/switch.

Note - It is possible to run synchronization across a WAN. For details, see “Synchronizing Clusters over a Wide Area Network” on page 36.
**In SmartDashboard for Machine ‘A’**

1. Update the topology of gateway A, either manually or by clicking *Get Topology*.
   If the IP address of the management interface was changed, the *Get Topology* action will fail. If this happens, manually change the main IP address in the gateway object and save the policy prior to performing an automatic topology fetch.

2. In the **Cluster Members** page, click *Add*, and select *Add Gateway to Cluster*.

3. Select machine ‘A’ in the window.

4. In the **Edit Topology** page, determine which interface is a cluster interface, and which is an internal or an external interface.

5. Install the policy on the cluster.
Adding Another Member to an Existing Cluster

1. On the cluster member, run `cpconfig` to enable ClusterXL.
2. Change the IP addresses of the new cluster member to reflect the correct topology (either shared IP addresses or unique IP addresses, depending on the clustering solution).
3. Ensure that all required Check Point products are installed on the new cluster member.
4. In the Cluster Members page of the Gateway Cluster object, either create a new cluster member (if it is a new Security Gateway machine) with the appropriate properties, or convert an existing Gateway to a cluster member.
5. If this is a new Security Gateway machine, ensure that SIC is initialized. In the Edit Topology page, ensure that the topology is correctly defined.
6. If the Cluster Mode is Load Sharing or New HA, ensure that the proper interfaces on the new cluster member are configured as Cluster Interfaces.
7. Install the security policy on the cluster.
8. The new member is now part of the cluster.
Configuring ISP Redundancy on a Cluster

If you have a ClusterXL Gateway cluster, connect each cluster member to both ISPs via a LAN using two interfaces. The cluster-specific configuration is illustrated in Figure 8-7.

Note that the member interfaces must be on the same subnet as the cluster external interfaces. Configure ClusterXL in the usual way.

Figure 8-7  Gateway Cluster Connected to Two ISP links

---

Network 192.168.1.0
Objective: Cluster
Cluster IP: 192.168.1.100
Member A: 192.168.1.2
Topology: External

Network 172.16.2.0
Objective: Cluster
Cluster IP: 172.16.2.100
Member A: 172.16.2.2
Member B: 172.16.2.3
Topology: External

Network 10.10.0.0
Objective: Cluster
Cluster IP: 10.10.0.100
Member A: 10.10.0.1
Member B: 10.10.0.2
Topology: Internal

Network 10.0.10.0
Objective: 1st Sync
Member A: 10.0.10.1
Member B: 10.0.10.2
Topology: Internal

---

[Figure 8-7: Gateway Cluster Connected to Two ISP links]
Enabling Dynamic Routing Protocols in a Cluster Deployment

ClusterXL supports Dynamic Routing (Unicast and Multicast) protocols as an integral part of SecurePlatform. As the network infrastructure views the clustered gateway as a single logical entity, failure of a cluster member will be transparent to the network infrastructure and will not result in a ripple effect.

Components of the System

**Virtual IP Integration**

All cluster members use the cluster IP address(es).

**Routing Table Synchronization**

Routing information is synchronized among the cluster members using the Forwarding Information Base (FIB) Manager process. This is done to prevent traffic interruption in case of failover, and used for Load Sharing and High Availability modes. The FIB Manager is the responsible for the routing information.

The FIB Manager is registered as a critical device (Pnote), and if the slave goes out of sync, a Pnote will be issued, and the slave member will go down until the FIB Manager is synchronized.

**Failure Recovery**

Dynamic Routing on ClusterXL avoids creating a ripple effect upon failover by informing the neighboring routers that the router has exited a maintenance mode. The neighboring routers then reestablish their relationships to the cluster, without informing the other routers in the network. These restart protocols are widely adopted by all major networking vendors. The following table lists the RFC and drafts compliant with Check Point Dynamic Routing:

<table>
<thead>
<tr>
<th>Table 8-5</th>
<th>Compliant Protocols</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Protocol</strong></td>
<td><strong>RFC or Draft</strong></td>
</tr>
<tr>
<td>OSPF LLS</td>
<td>draft-ietf-ospf-lls-00</td>
</tr>
<tr>
<td>OSPF Graceful restart</td>
<td>RFC 3623</td>
</tr>
<tr>
<td>BGP Graceful restart</td>
<td>draft-ietf-idr-restart-08</td>
</tr>
</tbody>
</table>
Dynamic Routing in ClusterXL

The components listed above function “behind-the-scenes.” When configuring Dynamic Routing on ClusterXL, the routing protocols automatically relate to the cluster as they would to a single device.

When configuring the routing protocols on each cluster member, each member is defined identically, and uses the cluster IP address(es) (not the member’s physical IP address). In the case of OSPF, the router ID must be defined and identical on each cluster member. When configuring OSPF restart, you must define the restart type as signaled or graceful. For Cisco devices, use type signaled.

Use SecurePlatform’s command line interface to configure each cluster member. Figure 8-8 is an example of the proper syntax for cluster member A.

Figure 8-8  Enabling OSPF on cluster member A

```
---------- Launch the Dynamic Routing Module
[Expert@GWa]# router
localhost>enable
localhost#configure terminal
---------- Enable OSPF and provide an OSPF router ID
localhost(config)#router ospf 1
localhost(config-router-ospf)#router-id 192.168.116.10
localhost(config-router-ospf)#restart-type [graceful | signaled]
localhost(config-router-ospf)#redistribute kernel
---------- Define interfaces/IP addresses on which OSPF runs (Use the cluster IP address as defined in topology) and the area ID for the interface/IP address
localhost(config-router-ospf)#network 1.1.10.10 0.0.0.0 area 0.0.0.0
localhost(config-router-ospf)#network 1.1.10.20 0.0.0.0 area 0.0.0.0
---------- Exit the Dynamic Routing Module
localhost(config-router-ospf)#exit
localhost(config)#exit
---------- Write configuration to disk
localhost#write memory
IU0 999 Configuration written to '/etc/gated.ami'
```

The same configuration needs to be applied to each cluster member.

As the FIB Manager uses TCP 2010 for routing information synchronization, the Security Policy must accept all traffic on port TCP 2010 between cluster members.

For detailed information regarding Dynamic Routing, see the Check Point Advanced Routing Suite guide.
Dynamic Routing in ClusterXL
Appendix A

High Availability Legacy Mode

In This Appendix

- Introduction to High Availability Legacy Mode page 224
- Example of High Availability HA Legacy Mode Topology page 225
- Implementation Planning Considerations for HA Legacy Mode page 227
- Configuring High Availability Legacy Mode page 229
- Moving from High Availability Legacy with Minimal Effort page 232
- Moving from High Availability Legacy with Minimal Downtime page 234
Introduction to High Availability Legacy Mode

In High Availability configurations, only one machine is active at any one time. A failure of the active machine causes a failover to the next highest priority machine in the cluster.

In Legacy Mode the cluster members share identical IP and MAC addresses, so that the active cluster member receives from a hub or switch all the packets that were sent to the cluster IP address. A shared interface is an interface with MAC and IP addresses that are identical to those of another interface.

Moving from a single gateway configuration to a High Availability Legacy Mode cluster requires no changes to IP addresses, or routing, and any switch or hub can be used to connect interfaces. However, configuring this mode is complicated, and must be performed in a precise sequence in order to be successful. The Security Management server has to be connected to a non-shared cluster network, in other words, the synchronization network of the cluster, or to a dedicated management network.
Example of High Availability HA Legacy Mode Topology

In This Section

- Shared Interfaces IP and MAC Address Configuration  page 226
- The Synchronization Interface  page 226

Figure A-1 shows an example ClusterXL Topology for High Availability Legacy mode. The diagram relates the physical cluster topology to the required SmartDashboard configuration. It shows two cluster members: Member_A (the primary) and Member_B (the secondary) each with three interfaces. One for synchronization, one external shared interface, and one internal shared interface.

**Figure A-1**  Example High Availability Legacy Mode Topology

---

**Gateway Cluster**

- Name: Alaska_Gateway_Cluster
- General page settings:
  - IP Address: 192.168.10.100
- Topology page settings:
  - External IP Address: 192.168.10.100
  - Internal IP Address: 10.0.10.0
- Synchronization page settings:
  - Alaska_Sync_network
  - IP Address: 10.0.10.0

---

**Cluster Members**

- Topology tab settings:
  - Connected to Sync. Network or to a dedicated management network
  - Management Server

---

**Network Connections**

- Shared IF 172.20.0.1
- Secured IF 10.0.10.3
- Shared IF 152.168.10.1
- Sync Network switch/VLAN Cross cable
Shared Interfaces IP and MAC Address Configuration

High Availability Legacy mode uses identical IP and MAC addresses on all cluster members, on interfaces that face the same direction. Shared interfaces are configured with the same IP address, and they automatically obtain identical MAC addresses. One shared interface on each cluster member faces the Internet through a hub or switch, and one or more interfaces face the local networks through a hub or switch.

Only one cluster member is active at any given time, so that the outside world can see only the shared interfaces on one machine at any given time.

Figure A-1 shows the shared interfaces. The EXT interface, facing the Internet, has IP address 192.168.0.1 on both Member_A and Member_B. The INT interface facing the local network has IP address 172.20.10.1 on both Member_A and Member_B.

The Synchronization Interface

State Synchronization between cluster members ensures that if there is a failover, connections that were handled by the failed machine will be maintained. The synchronization network is used to pass connection synchronization and other state information between cluster members. This network therefore carries the most sensitive security policy information in the organization, and so it is important to make sure the network is secure. It is possible to define more than one synchronization network for backup purposes.

To secure the synchronization interfaces, they should be directly connected by a cross-cable, or in the case of three or more cluster members, by means of a dedicated hub, switch, or VLAN.

Machines in a High Availability cluster do not have to be synchronized, though if they are not, connections may be lost upon failover.

Figure A-1 shows a SYNC interface with a unique IP address on each machine. 10.0.10.1 on Member_A and 10.0.10.2 on Member_B.
Implementation Planning Considerations for HA Legacy Mode

In This Section

IP Address Migration .......................................................... page 227
Security Management server Location ......................... page 227
Routing Configuration ......................................................... page 228
Switch (Layer 2 Forwarding) Considerations .............. page 228

IP Address Migration

Many ClusterXL installations are intended to provide High Availability or Load Sharing to an existing single gateway configuration. In those cases, it is recommended to take the existing IP addresses from the current gateway, and make these the cluster addresses (cluster virtual addresses) when feasible. Doing so will avoid altering current IPSec endpoint identities, and in many cases will make it unnecessary to change Hide NAT configurations.

Security Management server Location

The Security Management server must be able to download a Security Policy to all cluster members. This is only possible if the Security Management server can “see” them all at any given time. Therefore, in High Availability Legacy mode, the Security Management server must be connected to a non-shared cluster network.

The Security Management server cannot be connected to any network that includes the cluster interfaces with shared IP addresses, because they are configured with identical IP and MAC addresses.

The Security Management server must therefore be connected to the cluster synchronization network of the cluster, because the SYNC interface on each cluster member must have a unique IP address, or to a dedicated management network attached to the cluster.
Routing Configuration

Configure routing so that communication with the opposite side of the cluster is via the cluster IP address on the near side of the cluster.

For example, in Figure A-1, configure routing as follows:

• On each machine on the internal side of the router, define 172.20.0.1 as the default gateway.

• On external router, configure a static route such that network 172.20.0.1 is reached via 192.168.10.1.

Switch (Layer 2 Forwarding) Considerations

The Cluster Control Protocol (CCP), used by both High Availability New Mode and Load Sharing configurations, makes use of layer two multicast. In keeping with multicast standards, this multicast address is used only as the destination, and is used in all CCP packets sent on “non-secured” interfaces.

A Layer 2 switch connected to non-secured interfaces, must be capable of forwarding multicast packets to ports of the switch, or within a VLAN, if it is a VLAN switch. It is acceptable that the switch forward such traffic to all ports, or to ports within the given VLAN. However, it is considered more efficient to forward to only those ports connecting cluster members.

Most switches support multicast by default. Please check your switch documentation for details.

If the connecting switch is incapable of forwarding multicast, CCP can be changed to use broadcast instead. To toggle between these two modes use the command:

'cphaconf set_ccp broadcast/multicast'
Configuring High Availability Legacy Mode

See Figure A-1 on page 225 for an example configuration.


2. Disconnect the machines that are to participate in the High Availability Legacy configuration from the hub/switch.

3. Define the same IP addresses for each machine participating in the High Availability Legacy configuration, only for the interfaces that will be shared. To avoid network conflicts due to the sharing of MAC addresses, define the IP addresses before connecting the machines into the High Availability Legacy topology.

4. Install the same version (and build number) of the Security Gateway on each cluster member. During the configuration phase, enable ClusterXL/State Synchronization. Do NOT reboot the machines after the configuration phase.

5. Connect (or reconnect) the machines participating in the High Availability Legacy configuration to the hub/switch. Make sure you connect the configured interfaces to the matching physical network outlet. Connect each network (internal, external, Synchronization, DMZ, etc.) to a separate VLAN, switch or hub. No special configuration of the switch is needed.

Routing Configuration

1. Configure routing so that communication with the networks on the internal side of the cluster is via the cluster IP address on the external side of the cluster. For example, in Figure A-1, on the external router, configure a static route such that network 10.255.255.100 is reached via 192.168.10.100.

2. Configure routing so that communication with the networks on the external side of the cluster is via the cluster IP address on the internal side of the cluster. For example, in Figure A-1, on each machine on the internal side of the router, define 10.255.255.100 as the default gateway.

3. Reboot the cluster members. MAC address configuration will take place automatically.
**SmartDashboard Configuration**

1. Using SmartDashboard, define the Gateway Cluster object. In the General Properties page of the Gateway Cluster object, assign the routable external IP address of the cluster as the general IP address of the cluster. Check ClusterXL as a product installed on the cluster.

2. In the Cluster Members page, click Add > New Cluster Member to add cluster members to the cluster. Cluster members exist solely inside the Gateway Cluster object. For each cluster member:
   - In the Cluster Members Properties window General tab, define a Name and IP Address. Choose an IP address that is routable from the Security Management server so that the Security Policy installation will be successful. This can be an internal or an external address, or a dedicated management interface.
   - Click Communication, and Initialize Secure Internal Communication (SIC).
   - Define the NAT and VPN tabs, as required.
   
   You can also add an existing gateway as a cluster member by selecting Add > Add Gateway to Cluster in the Cluster Members page and selecting the gateway from the list in the Add Gateway to Cluster window.

   If you want to remove a gateway from the cluster, click Remove in the Cluster Members page and select Detach Member from Cluster or right-click on the cluster member in the Network Objects tree and select Detach from Cluster.

3. In the ClusterXL page,
   - Check High Availability Legacy Mode,
   - Choose whether to Use State Synchronization. This option is checked by default. If you uncheck this, the cluster members will not be synchronized, and existing connections on the failed gateway will be closed when failover occurs.
   - Specify the action Upon Gateway Recovery (see “What Happens When a Gateway Recovers?” on page 65 for additional information).
   - Define the Fail-over Tracking method.

4. In the Topology page, define the cluster member addresses. Do not define any virtual cluster interfaces. If converting from another cluster mode, the virtual cluster interface definitions are deleted. In the Edit Topology window:
   - Define the topology for each cluster member interface. To automatically read all the predefined settings on the member interfaces, click Get all members’ topology.
• In the **Network Objective** column, define the purpose of the network by choosing one of the options from the drop-down list. Define the interfaces with shared IP addresses as belonging to a **Monitored Private** network, and define one (or more) interfaces of each cluster member as synchronization interface in a synchronization network (**1st Sync/2nd Sync/3rd Sync**). The options are explained in the Online Help. To define a new network, click **Add Network**.

5. Define the other pages in the Gateway Cluster object as required (**NAT, VPN, Remote Access**, etc.).


7. Reboot all the cluster members in order to activate the MAC address configuration on the cluster members.
Moving from High Availability Legacy with Minimal Effort

This procedure describes how to move from High Availability Legacy mode to Load Sharing Multicast mode or to High Availability New mode, when the consideration is simplicity of configuration, rather than the minimal downtime.

The shared internal and external interfaces become cluster interfaces. The general IP address of the cluster therefore stays as an external cluster IP address.

On the Gateways

1. Run `cpstop` on all members (all network connectivity will be lost).
2. Reconfigure the IP addresses on all the cluster members, so that unique IP addresses are used instead of shared (duplicate) IP addresses.
   
   **Note** - SecurePlatform only: These address changes delete any existing static routes. Copy them down for restoration in step 4.

3. Remove the shared MAC addresses by executing the command: `cphaconf uninstall_macs`

4. SecurePlatform cluster members only: Redefine the static routes deleted in step 2.

5. Reboot the members.
From SmartDashboard

In SmartDashboard, open the cluster object, select the ClusterXL tab, change the cluster mode from Legacy mode to new mode or to Load sharing mode. Then follow the Check Point Gateway Cluster Wizard. For manual configuration, proceed as follows:

1. In the Topology tab of the cluster object,
   - For each cluster member, get the interfaces which have changed since the IP addresses were changed. The interfaces which were previously shared interfaces should now be defined as Cluster interfaces.
   - Define the cluster IP addresses of the cluster. The cluster interfaces' names may be defined as you wish as they will be bound to physical interfaces according to the IP addresses.

   If the new IP addresses of the cluster members on a specific interface reside on a different subnet than the cluster IP address in this direction, the cluster members' network should be defined in the Members Network fields of the cluster interface (“Configuring Cluster Addresses on Different Subnets” on page 207).

2. Install the policy on the new cluster object (Security policy, QOS policy and so on).
Moving from High Availability Legacy with Minimal Downtime

This procedure describes how to move from Legacy Check Point High Availability to New Check Point High Availability or to Load Sharing while minimizing the downtime of the cluster.

The shared internal and external interfaces become the cluster interfaces. As the cluster members will need additional IP addresses these must be prepared in advance.

If downtime of the cluster during the change is not a major issue, it is recommended to use the easier process described in “Moving from High Availability Legacy with Minimal Effort” on page 232.

**Note -**
1. Make sure that you have all the IP addresses needed before you start implementing the changes described here.
2. Backup your configuration before starting this procedure, because this procedure deletes and recreates the objects in SmartDashboard.

In this procedure we use the example of machines 'A' and 'B', with the starting point being that machine 'A' is active, and machine 'B' is on standby.

1. Disconnect machine 'B' from all interfaces except the interface connecting it to the Security Management server (the management interface).
2. Run `cphastop` on machine 'B'.
3. Change the IP addresses of machine 'B' (as required by the new configuration).

**Note -** SecurePlatform only: These address changes delete any existing static routes. Copy them down for restoration in step 5.

4. Reset the MAC addresses on machine 'B' by executing `cphaconf uninstall_macs`. The Windows machine must be rebooted for the MAC address change to take affect.
5. SecurePlatform cluster members only: Redefine the static routes deleted in step 3.
6. In SmartDashboard, right-click member 'A' and select Detach from cluster.
7. In the Topology tab of the Cluster Member Properties window, define the topology of cluster member 'B' by clicking Get.... Make sure to mark the appropriate interfaces as Cluster Interfaces.
8. In the **Cluster Object**, define the new topology of the cluster (define the cluster interfaces in the cluster’s **Topology** tab).

9. In the **ClusterXL** page, change the cluster’s **High Availability** mode from **Legacy Mode** to **New Mode** or select **Load Sharing** mode.

10. Verify that the other pages in the **Cluster Object** (NAT, VPN, Remote Access and so on) are correct. In Legacy Check Point High Availability, the definitions were per cluster member, while now they are on the cluster itself.

11. Install the policy on the cluster, which now only comprises cluster member 'B'.

12. Reconnect machine 'B' (which you disconnected in step 1) to the networks.

13. In this example the cluster comprises only two members, but if the cluster comprises more then two members, repeat steps 1-9 for each cluster member.

14. For **Load Sharing Multicast** mode, configure the routers as described in Table 4-5 on page 71.

15. Disconnect machine 'A' from the all networks accept the management network. The cluster stops processing traffic.

16. Run **cphastop** on machine 'A'.

17. Run **cpstop** and then **cpstart** on machine 'B' (if there are more than two machines, run these commands on all machines except 'A').

18. Machine 'B' now becomes active and starts processing traffic.

19. Change the IP addresses of machine 'A' (as required by the new configuration).

20. Reset the MAC addresses of machine 'A' by executing `cphaconf uninstall_macs`. The Windows machine must be rebooted for the MAC address change to take affect.

21. In SmartDashboard, open the **Cluster Object** and select the **Cluster Members** page. Click **Add > Add Gateway to Cluster** and select member 'A' to re-attach it to the cluster.

22. Reconnect machine 'A' to the networks from which it was disconnected in step 13.

23. Install the security policy on the cluster.

24. Run **cpstop** and then **cpstart** on machine 'A'.

25. Redefine static routes

The cluster now operates in the new mode.
Appendix

Example cphaprob Script

The `clusterXL_monitor_process` script shown below has been designed to monitor the existence of given processes and cause failover if the processes die. It uses the normal `pnote` mechanism.

The `clusterXL_monitor_process` script is located in `$FWDIR/bin`.

More Information

- The `cphaprob` command is described in “Verifying that a Cluster is Working Properly” on page 122.
- Chapter 7, “Monitoring and Troubleshooting Gateway Clusters”.

The `clusterXL_monitor_process` script

```sh
#!/bin/sh
#
# This script monitors the existence of processes in the system. The process names should be written
# in the `$FWDIR/conf/cpha_proc_list` file one every line.
#
# USAGE:
# cpha_monitor_process X silent
# where X is the number of seconds between process probings.
# if silent is set to 1, no messages will appear on the console.
#
# We initially register a pnote for each of the monitored processes
```
The clusterXL_monitor_process script

# (process name must be up to 15 characters) in the problem notification mechanism.
# when we detect that a process is missing we report the pnote to be in "problem" state.
# when the process is up again - we report the pnote is OK.

if [ "$2" -le 1 ]
then
    silent=$2
else
    silent=0
fi
if [ -f $FWDIR/conf/cpha_proc_list ]
then
    procfile=$FWDIR/conf/cpha_proc_list
else
    echo "No process file in $FWDIR/conf/cpha_proc_list"
    exit 0
fi
arch=`uname -s`
for process in `cat $procfile`
do
    $FWDIR/bin/cphaprob -d $process -t 0 -s ok -p register > /dev/null 2>&1
done
while [ 1 ]
do
    result=1
    for process in `cat $procfile`
do
        ps -ef | grep $process | grep -v grep > /dev/null 2>&1
        status=$?
        if [ $status = 0 ]
        then
            if [ $silent = 0 ]
            then
                echo " $process is alive"
The clusterXL_monitor_process script

```bash
if
  echo "3, $FWDIR/bin/cphaprob -d $process -s ok report"
  $FWDIR/bin/cphaprob -d $process -s ok report
else
  if [ $silent = 0 ]
    then
      echo " $process is down"
      fi

  $FWDIR/bin/cphaprob -d $process -s problem report
  result=0
  fi
done

if [ $result = 0 ]
  then
    if [ $silent = 0 ]
      then
        echo " One of the monitored processes is down!"
      fi
    else
      if [ $silent = 0 ]
        then
          echo " All monitored processes are up "
        fi
  fi
  fi

if [ "$silent" = 0 ]
  then
    echo "sleeping"
  fi

  sleep $1
done
```
The clusterXL_monitor_process script