



Draft - CISCO CONFIDENTIAL

CHAPTER 5

Configuring Real Servers and Server Farms

Revised Date: 2/18/09

This section provides an overview of server load balancing and procedures for configuring real servers and server farms for load balancing on an ACE.

Topics include:

- [Server Load Balancing Overview, page 5-1](#)
- [Configuring Real Servers, page 5-4](#)
- [Configuring Server Farms, page 5-12](#)
- [Health Monitoring, page 5-23](#)
- [Configuring Health Monitoring for Real Servers, page 5-25](#)
- [Managing Real Servers, page 5-6](#)

Server Load Balancing Overview

Server load balancing (SLB) is the process of deciding to which server a load-balancing device should send a client request for service. For example, a client request can consist of an HTTP GET for a Web page or an FTP GET to download a file. The job of the load balancer is to select the server that can successfully fulfill the client request and do so in the shortest amount of time without overloading either the server or the server farm as a whole.

Depending on the load-balancing algorithm or predictor that you configure, the ACE performs a series of checks and calculations to determine the server that can best service each client request. The ACE bases server selection on several factors, including the server with the fewest connections with respect to load, source or destination address, cookies, URLs, or HTTP headers.

to load, source or destination address, cookies, URLs, or HTTP headers.

The ANM allows you to configure load balancing using:

- Virtual servers—See [Configuring Virtual Servers, page 4-2](#).
- Real servers—See [Configuring Real Servers, page 5-4](#).
- Server farms—See [Configuring Server Farms, page 5-12](#).
- Sticky groups—See [Configuring Sticky Groups, page 6-7](#).
- Parameter maps—See [Configuring Parameter Maps, page 7-1](#).

Draft - CISCO CONFIDENTIAL

For information about SLB as configured and performed by the ACE, see:

- [Configuring Virtual Servers, page 4-2](#)
- [Load-Balancing Predictors, page 5-2](#)
- [Real Servers, page 5-3](#)
- [Server Farms, page 5-3](#)
- [Health Monitoring, page 5-23](#)
- [TCL Scripts, page 5-24](#)
- [Configuring Stickiness, page 6-1](#)

Load-Balancing Predictors

The ACE uses the following predictors to select the best server to satisfy a client request:

- **Roundrobin**—Selects the next server in the list of real servers based on server weight (weighted roundrobin). Servers with a higher weight value receive a higher percentage of the connections. This is the default predictor.
- **Leastconns**—Selects the server with the fewest number of active connections based on server weight. For the least connection predictor, you can configure a slow-start mechanism to avoid sending a high rate of new connections to servers that you have just put into service.
- **Hash_url**—Selects the server using a hash value based on the requested URL. You can specify a beginning pattern and an ending pattern to match in the URL. Use this predictor method to load-balance cache servers. Cache servers perform better with the URL hash method because you can divide the contents of the caches evenly if the traffic is random enough. In a redundant configuration, the cache servers continue to work even if the active ACE switches over to the standby ACE. For information about configuring redundancy, see [Configuring High Availability, page 10-1](#).
- **Hash_address**—Selects the server using a hash value based on either the source or destination IP address, or both. Use these predictors for firewall load balancing (FWLB).



Note FWLB allows you to scale firewall protection by distributing traffic across multiple firewalls on a per-connection basis. All packets belonging to a particular connection must go through the same firewall. The firewall then allows or denies transmission of individual packets across its interfaces. For more information about configuring FWLB on the ACE, see the *Cisco 4700 Series Application Control Engine Appliance Server Load-Balancing Configuration Guide*.

- **Hash_cookie**—Selects the server using a hash value based on a cookie name.
- **Hash_header**—Selects the server using a hash value based on the HTTP header name.
- **Least Loaded**—Selects the server with the lowest load as determined by information from SNMP probes.
- **Least Bandwidth**—Selects the server with the least amount of network traffic or a specified sampling period. Use this type for server farms with heavy traffic, such as downloading video clips.
- **Response**—Selects the server with the lowest response time for a specific response-time measurement.

**Note**

The different hash predictor methods do not recognize the weight value that you configure for real servers. The ACE uses the weight that you assign to real servers only in the round-robin and least-connections predictor methods.

Related Topic

[Configuring the Predictor Method for Server Farms, page 5-16](#)

Real Servers

To provide services to clients, you configure real servers on the ACE. Real servers are dedicated physical servers that you typically configure in groups called server farms. These servers provide client services such as HTTP or XML content, Web site hosting, FTP file uploads or downloads, redirection for Web pages that have moved to another location, and so on. You identify real servers with names and characterize them with IP addresses, connection limits, and weight values. The ACE also allows you to configure backup servers in case a server is taken out of service for any reason.

After you create and name a real server on the ACE, you can configure several parameters, including connection limits, health probes, and weight. You can assign a weight to each real server based on its relative importance to other servers in the server farm. The ACE uses the server weight value for the weighted round-robin and the least-connections load-balancing predictors. The load-balancing predictor algorithms (for example, roundrobin, least connections, and so on) determine the servers to which the ACE sends connection requests. For a listing and brief description of the load-balancing predictors, see [Load-Balancing Predictors, page 5-2](#).

The ACE uses traffic classification maps (class maps) within policy maps to identify traffic that meets defined criteria and to apply specific actions to that traffic based on the SLB configuration.

If a primary real server fails, the ACE takes that server out of service and no longer includes it in load-balancing decisions. If you configured a backup server for the real server that failed, the ACE redirects the primary real server connections to the backup server. For information about configuring a backup server, see the [Configuring Virtual Server Layer 7 Load Balancing, page 4-25](#).

The ACE can take a real server out of service for the following reasons:

- Probe failure
- ARP timeout
- Specifying Out of Service as the administrative state of a real server
- Specifying Inservice Standby as the administrative state of a real server

The Out of Service and Inservice Standby selections both provide the graceful shutdown of a server.

Related Topics

- [Configuring Real Servers, page 5-4](#)
- [Configuring Health Monitoring for Real Servers, page 5-25](#)

Server Farms

Typically, in data centers, servers are organized into related groups called *server farms*. Servers within server farms often contain identical content (referred to as mirrored content) so that if one server becomes inoperative, another server can take its place immediately. Also, having mirrored content

Draft - CISCO CONFIDENTIAL

allows several servers to share the load of increased demand during important local or international events, such as the Olympic Games. This phenomenon of a sudden large demand for content is called a *flash crowd*.

After you create and name a server farm, you can add existing real servers to it and configure other server farm parameters, such as the load-balancing predictor, server weight, backup server, health probe, and so on. For a listing and brief description of load-balancing predictors, see [Load-Balancing Predictors, page 5-2](#).

Related Topic

[Configuring Server Farms, page 5-12](#)

Configuring Real Servers

Real servers are dedicated physical servers that are typically configured in groups called server farms. These servers provide services to clients, such as HTTP or XML content, streaming media (video or audio), TFTP or FTP services, and so on. When configuring real servers, you assign names to them and specify IP addresses, connection limits, and weight values.

The ACE uses traffic classification maps (class maps) within policy maps to filter specified traffic and to apply specific actions to that traffic based on the load-balancing configuration. A load-balancing predictor algorithm (such as round-robin or least connections) determines the servers to which the ACE sends connection requests. For information about configuring class maps, see [Configuring Virtual Context Class Maps, page 11-6](#).

Use this procedure to configure load balancing on real servers.

Procedure

-
- Step 1** Select **Config > Devices > context > Load Balancing > Real Servers**. The Real Servers table appears.
- Step 2** Click **Add** to add a new real server, or select a real server you want to modify, then click **Edit**. The Real Servers configuration screen appears.
- Step 3** Configure the server using the information in [Table 5-1](#).

Table 5-1 Real Server Attributes

Field	Description
Name	Either accept the automatically incremented value in this field, or enter a unique name for this server. Valid entries are unquoted text strings with no spaces and a maximum of 64 characters.
Type	Select the type of server: <ul style="list-style-type: none"> Host—The real server provides content and services to clients. Redirect—The server redirects traffic to a new location.
Description	Enter a brief description for this real server. Valid entries are strings of up to 240 characters. Spaces and special characters are allowed.
IP Address	This field appears for only real servers specified as hosts. Enter a unique IP address in dotted-decimal format (such as 192.168.11.1). The IP address cannot be an existing virtual IP address (VIP).

Table 5-1 Real Server Attributes (continued)

Field	Description
Max Connections	Enter the maximum number of active connections allowed on this server. When the number of connections exceeds this value, the ACE appliance stops sending connections to this server until the number of connections falls below the Min Connections value. Valid entries are integers from 2 to 4000000, and the default is 4000000.
Min Connections	Enter the minimum number of connections to be allowed on this server before the ACE appliance starts sending connections again after it has exceeded the Max Connections limit. This value must be less than or equal to the Max Connections value. By default, this value is equal to the Max Connections value. Valid entries are integers from 2 to 4000000.
Weight	This field appears only for real servers identified as hosts. Enter the weight to be assigned to this real server in a server farm. Valid entries are integers from 1 to 100, and the default is 8.
State	Select the state of this real server: <ul style="list-style-type: none"> • In Service—The real server is in service. • Out of Service—The real server is out of service.
Probes	This field appears only for real servers identified as hosts. In the Probes field, select the probes to use for health monitoring in the Available Items list, then click Add . The probes appear in the Selected Items list. To remove probes that you do not want to use for health monitoring, select them in the Selected Items list, then click Remove . The probes appear in the Available Items list.
Webhost Redirection	URL string used to redirect requests to another server. This field appears only for real servers identified as redirect servers. Enter the URL and port used to redirect requests to another server. Valid entries are in the form http://host.com:port where host is the name of the server and port is the port to be used. Valid host entries are unquoted text strings with no spaces and a maximum of 255 characters. Valid port numbers are from 1 to 65535. The relocation string supports the following special characters: <ul style="list-style-type: none"> • %h—Inserts the hostname from the request Host header • %p—Inserts the URL path string from the request
Redirection Code	This field appears only for real servers identified as redirect servers. Select the appropriate redirection code: <ul style="list-style-type: none"> • N/A—The webhost redirection code is not defined. • 301—The requested resource has been moved permanently. For future references to this resource, the client should use one of the returned URIs. • 302—The requested resource has been found, but has been moved temporarily to another location. For future references to this resource, the client should use the request URI because the resource may be moved to other locations from time to time.

Draft - CISCO CONFIDENTIAL**Table 5-1 Real Server Attributes (continued)**

Field	Description
Rate Bandwidth	The bandwidth rate is the number of bytes per second and applies to the network traffic exchanged between the ACE and the real server in both directions. Specify the real server bandwidth limit in bytes per second. Valid entries are integers from 2 to 300000000. The default is 300000000.
Rate Connection	The connection rate is the number of connections per second received by the ACE and applies only to new connections destined to a real server. Specify the limit for connections per second. Valid entries are integers from 2 to 350000. The default is 350000.

Step 4 Click:

- **Deploy Now** to deploy this configuration.
- **Cancel** to exit the procedure without saving your entries and to return to the Real Servers table.
- **Next** to deploy your entries and to configure another real server.

Related Topics

- [Configuring Health Monitoring for Real Servers, page 5-25](#)
- [Configuring Server Farms, page 5-12](#)
- [Configuring Sticky Groups, page 6-7](#)

Managing Real Servers

The Real Servers table (**Config > Operations > Real Servers**) provides the following information by default for each server:

- Server name
- IP address
- Port
- Admin State (In Service, Out of Service, or In Service Standby)
- Operational state (See [Table 5-2](#) for descriptions of real server operational states.)
- Number of current connections
- Current server weight
- Associated server farm
- Associated virtual servers
- Device details
- Whether the server is part of a high availability pair

In the table, N/A indicates that either the information is not available from the database or that it is not being collected via SNMP. To identify any SNMP-related issues, select the real server's virtual context in the object selector. If there are problems with SNMP, SNMP status will appear in the upper right above the content pane.

The following options are available from the Real Servers table:

- [Activating Real Servers, page 5-7](#)
- [Suspending Real Servers, page 5-7](#)
- [Modifying Real Servers, page 5-8](#)
- [Viewing Real Servers, page 5-9](#)
- [Understanding CLI Commands Sent from Real Server Table, page 5-10](#)
- [Server Weight Ranges, page 5-12](#)

Activating Real Servers

Use this procedure to activate a real server.

Procedure

-
- Step 1** Select **Config > Operations > Real Servers**. The Real Servers table appears.
- Step 2** Select the servers that you want to activate, then click **Activate**. The Activate Server screen appears.
- Step 3** In the Reason field, enter a reason for this action. You might enter a trouble ticket, an order ticket, or a user message. **Do not enter a password in this field.**
- Step 4** Click:
- **OK** to activate the server and to return to the Real Servers table. The server appears in the table with the status Inservice.
 - **Cancel** to exit this procedure without activating the server and to return to the Real Servers table.
-

Related Topics

- [Managing Real Servers, page 5-6](#)
- [Suspending Real Servers, page 5-7](#)
- [Viewing Real Servers, page 5-9](#)

Suspending Real Servers

Use this procedure to suspend a real server.

Procedure

-
- Step 1** Select **Config > Operations > Real Servers**. The Real Servers table appears.
- Step 2** Select the server that you want to suspend, then click **Suspend**. The **Suspend Real Servers** window appears.

Draft - CISCO CONFIDENTIAL

Step 3 In the Reason field, enter the reason for this action. You might enter a trouble ticket, an order ticket, or a user message. **Do not enter a password in this field.**

Step 4 Select one of the following from the **Suspend Real Servers Type** pulldown menu:

- Graceful
- Suspend
- Suspend and Clear Connections to clear the existing connections to this server as part of the shutdown process.



Note Graceful suspend and suspend options vary by device type. For exact commands deployed by device type when these options are selected, see the [“Understanding CLI Commands Sent from Real Server Table”](#) section on page 5-10.

Step 5 Click:

- **Deploy Now** to suspend the server and to return to the Real Servers table. The server appears in the table with the status Out of Service.
- **Cancel** to exit this procedure without suspending the server and to return to the Real Servers table.

Related Topics

- [Managing Real Servers, page 5-6](#)
- [Activating Real Servers, page 5-7](#)
- [Viewing Real Servers, page 5-9](#)

Modifying Real Servers

Use this procedure to modify server weight and connection limits for real servers.

Procedure

Step 1 Select **Config > Operations > Real Servers**. The Real Servers table appears.

Step 2 Select the servers whose configuration you want to modify, then click **Change Weight** below the table to the right of **Activate** and **Suspend**. The **Change Weight Real Servers** window appears.

Step 3 Enter the following information for the selected server:

- Reason for change such as trouble ticket, order ticket or user message. **Do not enter a password in this field.**
- Weight (For allowable ranges for each device type, see [Table 5-4](#).)

Step 4 Click:

- **Deploy Now** to accept your entries and to return to the Real Servers table. The server appears in the table with the updated information.
- **Cancel** to exit this procedure without saving your entries and to return to the Real Servers table.

Related Topics

- [Managing Real Servers, page 5-6](#)
- [Activating Real Servers, page 5-7](#)
- [Viewing Real Servers, page 5-9](#)

Viewing Real Servers

To view all real servers, select **Config > Operations > Real Servers**. The Real Servers table displays the following information by default:

- Server name
- IP address
- Port
- Admin State (In Service, Out of Service, or In Service Standby)
- Operational state (See [Table 5-2](#) for descriptions of real server operational states.)
- Number of current connections
- Current server weight
- Associated server farm
- Associated virtual servers
- Device details
- Whether the server is part of a high availability pair

In the table, N/A indicates that either the information is not available from the database or that it is not being collected via SNMP. To identify any SNMP-related issues, select the real server's virtual context in the object selector. If there are problems with SNMP, SNMP status will appear in the upper right above the content pane.

Table 5-2 Real Server Operational States

State	Description
Failed	The server has failed and will not be retried for the amount of time specified by its retry timer.
Inband probe failed	The server has failed the inband Health Probe agent.
In service	The server is in use as a destination for server load balancing client connections.
Operation wait	The server is ready to become operational but is waiting for the associated redirect virtual server to be in service.
Out of service	The server is not in use by a server load balancer as a destination for client connections.
Probe failed	The server load-balancing probe to this server has failed. No new connections will be assigned to this server until a probe to this server succeeds.
Probe testing	The server has received a test probe from the server load balancer.
Ready to test	The server has failed and its retry timer has expired; test connections will begin flowing to it soon.

Draft - CISCO CONFIDENTIAL**Table 5-2 Real Server Operational States (continued)**

State	Description
Return code failed	The server has been disabled because it returned an HTTP code that matched a configured value.
Test wait	The server is ready to be tested. This state is applicable only when the server is used for HTTP redirect load balancing.
Testing	The server has failed and has been given another test connection. The success of this connection is not known.
Throttle: DFP	DFP has lowered the weight of the server to throttle level; no new connections will be assigned to the server until DFP raises its weight.
Throttle: max clients	The server has reached its maximum number of allowed clients.
Throttle: max connections	The server has reached its maximum number of connections and is no longer being given connections.
Unknown	The state of the server is not known.

Related Topics

- [Activating Real Servers, page 5-7](#)
- [Suspending Real Servers, page 5-7](#)
- [Modifying Real Servers, page 5-8](#)

Understanding CLI Commands Sent from Real Server Table

Table 5-3 displays the CLI commands dispatched to the device for a given Real Servers table option, and is sorted by device.

Table 5-3 CLI Commands Deployed from Real Servers Table

Command	Sample CLI Sent
ACE Modules and Appliances	
Real Server Activation	<pre>serverfarm host sf1 rserver rs1 80 inservice</pre>
Real Server Graceful Suspend	<pre>serverfarm host sf1 rserver rs1 80 inservice standby</pre>
Real Server Suspend	<pre>serverfarm host sf1 rserver rs1 80 no inservice</pre>
Real Server Suspend and Clear Connections	<pre>serverfarm host sf1 rserver rs1 80 no inservice clear conn rserver rs1 80 serverfarm sf1</pre>

Table 5-3 CLI Commands Deployed from Real Servers Table

Command	Sample CLI Sent
Real Server Change Weight	serverfarm host sf1 rserver rs1 80 weight 2
CSMs	
Real Server Activation	serverfarm host sf1 real 10.10.10.10 80 inservice
Real Server Graceful Suspend	serverfarm host sf1 real 10.10.10.10 80 weight 0
Real Server Suspend	serverfarm host sf1 real 10.10.10.10 80 no inservice
Real Server Suspend and Clear Connections	serverfarm host sf1 real 10.10.10.10 80 no inservice clear module contentSwitchingModule 3 connections real 10.10.10.10
Real Server Change Weight	serverfarm host sf1 rserver 10.10.10.10 80 weight 2
CSM Named Real Commands Sent	
Real Server Activation	serverfarm host sf1 real name rs1 80 inservice
Real Server Graceful Suspend	serverfarm host sf1 real name rs1 80 weight 0
Real Server Suspend	serverfarm host sf1 real name rs1 80 no inservice
Real Server Suspend and Clear Connections	serverfarm host sf1 real name rs1 80 no inservice clear module contentSwitchingModule 3 connections real 10.10.10.10
Real Server Change Weight	serverfarm host sf1 real name rs1 80 weight 2

Draft - CISCO CONFIDENTIAL**Table 5-3** CLI Commands Deployed from Real Servers Table

Command	Sample CLI Sent
CSS Devices	
Real Server Activation	service myReal7 active
Real Server Graceful Suspend	service myReal7 weight 0
Real Server Suspend	service myReal7 suspend
Real Server Suspend and Clear Connections	service myReal7 suspend
Real Server Change Weight	service myReal7 weight 2

Server Weight Ranges

Table 5-4 displays the allowable server weight ranges by device type.

Table 5-4 Real Servers Table Server Weight Ranges

Device Type	Valid Weight Configurations
ACE Appliances and Modules	1-100
CSMs	0-100
CSS Devices	0-10

Configuring Server Farms

Server farms are groups of networked real servers that contain the same content and that typically reside in the same physical location in a data center. Web sites often comprise groups of servers configured in a server farm. Load-balancing software distributes client requests for content or services among the real servers based on the configured policy and traffic classification, server availability and load, and other factors. If one server goes down, another server can take its place and continue to provide the same content to the clients who requested it.

Use this procedure to configure load balancing using server farms.

Procedure

- Step 1** Select **Config > Devices > context > Load Balancing > Server Farms**. The Server Farms table appears.
- Step 2** Click **Add** to add a new server farm, or select an existing server farm, then click **Edit**. The Server Farms configuration screen appears.
- Step 3** Configure the server farm using the information in [Table 5-5](#).

Table 5-5 Server Farm Attributes

Field	Description
Name	Either accept the automatically incremented value in this field, or enter a unique name for this server farm. Valid entries are unquoted text strings with no spaces and a maximum of 64 characters.
Type	Select the type of server farm: <ul style="list-style-type: none"> • Host—The server farm consists of real servers that provide content and services to clients. • Redirect—The server farm consists only of real servers that redirect client requests to alternate locations specified in the real server configuration. (See Configuring Real Servers, page 5-4.)
Description	Enter a brief description for this server farm. Valid entries are unquoted alphanumeric text strings with no spaces and a maximum of 240 characters.
Fail Action	Select the action the ACE is to take with respect to connections if any real server in the server farm fails: <ul style="list-style-type: none"> • N/A—The ACE is to take no action if any server in the server farm fails. • Purge—The ACE is to remove connections to a real server if that real server in the server farm fails. The ACE sends a reset command to both the client and the server that failed.
Transparent	This field appears only for host server farms. Specify whether network address translation from VIP address to server IP is to occur: <ul style="list-style-type: none"> • N/A—The default value is to be used; the default value is False. • False—Network address translation from VIP address to server IP address is not to occur. • True—Network address translation from VIP address to server IP address is to occur.
Partial Threshold Percentage	This field appears only for host server farms. Enter the minimum percentage of real servers in the primary server farm that must remain active for the server farm to stay up. If the percentage of active real servers falls below this threshold, the ACE takes the server farm out of service. Valid entries are integers from 0 to 99. The default is 0.
Back In Service	This field appears only for host server farms. Enter the percentage of real servers in the primary server farm that must be active again for the ACE to place the server farm back into service. Valid entries are integers from 0 to 99. The value in this field should be larger than the value in the Partial Threshold Percentage field. The default is 0.
Probes	This field appears only for host server farms. In the Available Items list, select the probes to use for health monitoring, then click Add . The selected probes appear in the Selected Items list. To remove probes that you do not want to use for health monitoring, select them in the Selected Items list, then click Remove . The selected probes appear in the Available Items list.

Draft - CISCO CONFIDENTIAL**Step 4** Click:

- **Deploy Now** to deploy this configuration on the ACE.
The screen refreshes with additional configuration options:
 - To add real servers to the server farm, see [Adding Real Servers to a Server Farm, page 5-14](#).
 - To specify a predictor method for the server farm, see [Configuring the Predictor Method for Server Farms, page 5-16](#).
 - To configure return code checking, see [Configuring Server Farm HTTP Return Error-Code Checking, page 5-21](#).
 - **Cancel** to exit the procedure without saving your entries and to return to the Server Farms table.
 - **Next** to deploy your entries and to configure another server farm.
-

Related Topics

- [Configuring Health Monitoring for Real Servers, page 5-25](#)
- [Configuring Real Servers, page 5-4](#)
- [Configuring Sticky Groups, page 6-7](#)
- [Configuring the Predictor Method for Server Farms, page 5-16](#)
- [Configuring Server Farm HTTP Return Error-Code Checking, page 5-21](#)

Adding Real Servers to a Server Farm

After adding a server farm, (see [Configuring Server Farms, page 5-12](#)), you can associate real servers with it and configure predictors and retcode maps. The options for these attributes appear after you have successfully added a new server farm.

Use this procedure to add real servers to a server farm.

Assumptions

- A server farm has been added to the ANM. (See [Configuring Server Farms, page 5-12](#).)
- At least one real server exists.

Procedure

-
- Step 1** Select **Config > Devices > context > Load Balancing > Server Farms**. The Server Farms table appears.
 - Step 2** Select the server farm you want to associate with real servers. The Real Servers table appears.
 - Step 3** Click **Add** to add a new entry to the Real Servers table, or select an existing server, then click **Edit** to modify it. The Real Servers configuration pane appears.

Step 4 Configure the real server using the information in [Table 5-6](#).

Table 5-6 Real Server Configuration Attributes

Field	Description
Name	Select the server that you want to associate with the server farm.
Port	Enter the port number to be used for server port address translation (PAT). Valid entries are integers from 1 to 65535.
Backup Server Name	Select the server that is to act as the backup server for the server farm. Leave this field blank to indicate that there is no designated backup server for the server farm.
Backup Server Port	If you select a backup server, enter the backup server port number. Valid entries are integers from 1 to 65535.
Max Connections	Enter the maximum number of active connections that can be sent to the server. When the number of connections exceeds this number, the ACE stops sending connections to the server until the number of connections falls below the number specified in the Min Connections field. For ACE 1.0 modules and ACE appliances, valid entries are integers from 2 to 4294967295. For ACE 2.0 modules, valid entries are integers from 2 to 4000000.
Min Connections	Enter the minimum number of connections that the number of connections must fall below before the ACE resumes sending connections to the server after it has exceeded the number in the Max Connections field. The number in this field must be less than or equal to the number in the Max Connections field. For ACE 1.0 modules and ACE appliances, valid entries are integers from 2 to 4294967295. For ACE 2.0 modules, valid entries are integers from 2 to 4000000.
Weight	Enter the weight to assign to the server. Valid entries are integers from 1 to 100, and the default is 8.
State	Select the state of this server: <ul style="list-style-type: none"> • In Service—The server is in service. • Out of Service—The server is out of service. • In Service Standby—The server is a backup server and remains inactive unless the primary server fails. If the primary server fails, the backup server becomes active and starts accepting connections.
Probes	Select the probes in the Available Items list that you want to apply to this server, then click Add . The selected probes appear in the Selected Items list. To remove probes you do not want to use, select the probes in the Selected Items list, then click Remove . The selected probes appear in the Available Items list.
Rate Bandwidth	The bandwidth rate is the number of bytes per second and applies to the network traffic exchanged between the ACE and the real server in both directions. Specify the bandwidth limit in bytes per second. Valid entries are integers from 2 to 300000000. The default is 300000000.
Rate Connection	The connection rate is the number of connections per second received by the ACE and applies only to new connections destined to a real server. Specify the limit for connections per second. Valid entries are integers from 2 to 350000. The default is 350000.

Draft - CISCO CONFIDENTIAL

- Step 5** When you finish configuring this server for this server farm, click:
- **Deploy Now** to deploy this configuration and to return to the Real Servers table.
 - **Cancel** to exit this procedure without saving your entries and to return to the Real Servers table.
 - **Next** to deploy your entries and to add another real server for this server farm.
-

Related Topics

- [Configuring Health Monitoring for Real Servers, page 5-25](#)
- [Configuring Real Servers, page 5-4](#)
- [Configuring Sticky Groups, page 6-7](#)
- [Configuring the Predictor Method for Server Farms, page 5-16](#)
- [Configuring Server Farm HTTP Return Error-Code Checking, page 5-21](#)

Configuring the Predictor Method for Server Farms

After adding a server farm, ([Configuring Server Farms, page 5-12](#)), you can associate real servers with it and configure the predictor method and retcode maps. The options for these attributes appear after you have successfully added a new server farm.

Use this procedure to configure the predictor method for a server farm. The predictor method specifies how the ACE is to select a server in the server farm when it receives a client request for a service.



Note You can configure only one predictor method per server farm.

Assumptions

- A server farm has been added to the ANM. (See [Configuring Server Farms, page 5-12](#).)
- At least one real server exists.

Procedure

-
- Step 1** Select **Config > Devices > context > Load Balancing > Server Farms**. The Server Farms table appears.
- Step 2** Select the server farm you want to configure the predictor method for, then click the Predictor tab. The Predictor configuration pane appears.
- Step 3** In the Type field, select the method that the ACE is to use to select a server in this server farm when it receives a client request. [Table 5-7](#) lists the available options and describes them.

Step 4 Enter the required information for the selected predictor method. See [Table 5-7](#).

Table 5-7 Predictor Method Attributes

Predictor Method	Description / Action
Round_Robin	The ACE selects the next server in the list of servers based on server weight. This is the default predictor method.
Least_Connections	<p>The ACE selects the server with the fewest number of connections.</p> <p>In the Slowstart Duration field, enter the slow-start value to be applied to this predictor method. Valid entries are integers from 1 to 65535, where 1 is the slowest ramp-up value.</p> <p>The slow-start mechanism is used to avoid sending a high rate of new connections to servers that you have just put into service.</p>
Hash_URL	<p>The ACE selects the server using a hash value based on the URL. Use this method to load balance firewalls.</p> <p>Enter values in one or both of the pattern fields:</p> <ul style="list-style-type: none"> In the URL Begin Pattern field, enter the beginning pattern of the URL and the pattern string to parse. In the URL End Pattern field, enter the ending pattern of the URL and the pattern string to parse. <p>Valid entries for these fields are unquoted text strings with no spaces and a maximum of 255 alphanumeric characters for each pattern you configure.</p>
Hash_Address	<p>The ACE selects the server using a hash value based on the source or destination IP address.</p> <p>To configure the hash address predictor method:</p> <ol style="list-style-type: none"> In the Mask Type field, indicate whether server selection is based on source IP address or the destination IP address: <ul style="list-style-type: none"> N/A—This option is not defined. Source—The server is selected based on the source IP address. Destination—The server is selected based on the destination IP address. In the IP Netmask field, select the subnet mask to apply to the address. If none is specified, the default is 255.255.255.255.
Hash_Cookie	<p>The ACE selects the server by using a hash value based on the cookie name.</p> <p>In the Cookie Name field, enter a cookie name in the form of an unquoted text string with no spaces and a maximum of 64 characters.</p>
Hash_Header	<p>The ACE selects the server by using a hash value based on the header name.</p> <p>In the Header Name field, select the HTTP header to be used for server selection:</p> <ul style="list-style-type: none"> To specify an HTTP header that is not one of the standard HTTP headers, select the first radio button and enter the HTTP header name in the Header Name field. Valid entries are unquoted text strings with no spaces and a maximum of 64 characters. To specify one of the standard HTTP headers, select the second radio button, then select one of the HTTP headers from the list.

Draft - CISCO CONFIDENTIAL**Table 5-7 Predictor Method Attributes (continued)**

Predictor Method	Description / Action
Hash_Content	<p>The ACE selects the server by using a hash value based on the specified content string of the HTTP packet body.</p> <ol style="list-style-type: none"> In the Begin Pattern field, enter the beginning pattern of the content string and the pattern string to match before hashing. If you do not specify a beginning pattern, the ACE starts parsing the HTTP body immediate following the offset byte. You cannot configure different beginning and ending patterns for different server farms that are part of the same traffic classification. Valid entries are unquoted text strings with no spaces and a maximum of 255 alphanumeric characters. The ACE supports regular expressions for matching string expressions. Table 11-35 lists the supported characters that you can use for matching string expressions. In the End Pattern field, enter the pattern that marks the end of hashing. If you do not specify either a length or an end pattern, the ACE continues to parse the data until it reaches the end of the field or the end of the packet, or until it reaches the maximum body parse length. You cannot configure different beginning and ending patterns for different server farms that are part of the same traffic classification. Valid entries are unquoted text strings with no spaces and a maximum of 255 alphanumeric characters. The ACE supports regular expressions for matching string expressions. Table 11-35 lists the supported characters that you can use for matching string expressions. In the Length field, enter the length in bytes of the portion of the content (starting with the byte after the offset value) that the ACE uses for sticking the client to the server. Valid entries are integers from 1 to 1000 bytes. The offset and length can vary from 0 to 1000 bytes. If the payload is longer than the offset but shorter than the offset plus the length of the payload, the ACE sticks the connection based on that portion of the payload starting with the byte after the offset value and ending with the byte specified by the offset plus the length. The total of the offset and the length cannot exceed 1000. <p>Note You cannot specify both the length and the end-pattern options for a Hash Content predictor.</p> <ol style="list-style-type: none"> In the HTTP Content Offset field, enter the portion of the content that the ACE uses to stick the client on a particular server by indicating the bytes to ignore starting with the first byte of the payload. Valid entries are integers from 0 to 999 bytes. The default is 0, which indicates that the ACE does not exclude any portion of the content.

Table 5-7 Predictor Method Attributes (continued)

Predictor Method	Description / Action
Hash_Layer4	<p>The ACE selects the server by using a Layer 4 generic protocol load-balancing method. Use this predictor to load balance packets from protocols that are not explicitly supported by the ACE.</p> <ol style="list-style-type: none"> In the Begin Pattern field, enter the beginning pattern of the Layer 4 payload and the pattern string to match before hashing. If you do not specify a beginning pattern, the ACE starts parsing the HTTP body immediate following the offset byte. You cannot configure different beginning and ending patterns for different server farms that are part of the same traffic classification. Valid entries are unquoted text strings with no spaces and a maximum of 255 alphanumeric characters. The ACE supports regular expressions for matching string expressions. Table 11-35 lists the supported characters that you can use for matching string expressions. In the End Pattern field, enter the pattern that marks the end of hashing. If you do not specify either a length or an end pattern, the ACE continues to parse the data until it reaches the end of the field or the end of the packet, or until it reaches the maximum body parse length. You cannot configure different beginning and ending patterns for different server farms that are part of the same traffic classification. Valid entries are unquoted text strings with no spaces and a maximum of 255 alphanumeric characters. The ACE supports regular expressions for matching string expressions. Table 11-35 lists the supported characters that you can use for matching string expressions. In the Length field, enter the length in bytes of the portion of the payload (starting with the byte after the offset value) that the ACE uses for sticking the client to the server. Valid entries are integers from 1 to 1000 bytes. The offset and length can vary from 0 to 1000 bytes. If the payload is longer than the offset but shorter than the offset plus the length of the payload, the ACE sticks the connection based on that portion of the payload starting with the byte after the offset value and ending with the byte specified by the offset plus the length. The total of the offset and the length cannot exceed 1000. <p>Note You cannot specify both the length and end-pattern options for a Hash Layer 4 predictor.</p> <ol style="list-style-type: none"> In the HTTP Content Offset field, enter the portion of the content that the ACE uses to stick the client on a particular server by indicating the bytes to ignore starting with the first byte of the payload. Valid entries are integers from 0 to 999 bytes. The default is 0, which indicates that the ACE does not exclude any portion of the content.
Least_Bandwidth	<p>The ACE selects the server with the least amount of network traffic over a specified sampling period.</p> <ol style="list-style-type: none"> In the Assess Time field, enter the number of seconds for which the ACE is to collect traffic information. Valid entries are integers from 1 to 10 seconds. In the Least Bandwidth Samples field, enter the number of samples over which you want to weight and average the results of the probe query to calculate the final load value. Valid entries are 1, 2, 4, 8, and 16 (integers from 1 to 16 that are also a power of 2).

Draft - CISCO CONFIDENTIAL

Table 5-7 Predictor Method Attributes (continued)

Predictor Method	Description / Action
Least Loaded	<p>The ACE selects the server with the lowest load based on information from SNMP probes.</p> <ol style="list-style-type: none"> 1. In the SNMP Probe Name field, select the name of the SNMP probe to use. 2. In the Auto Adjust field, configure the autoadjust feature to assign a maximum load value of 16000 to that server to prevent it from being flooded with new incoming connections. The ACE periodically adjusts this load value based on feedback from the server's SNMP probe and other configured options. Options include: <ul style="list-style-type: none"> - N/A—Indicates that this option is not defined. - Average—Instructs the ACE to apply the average load of the server farm to a real server whose load reaches zero. The average load is the running average of the load values across all real servers in the server farm. - Off—Overrides the default behavior of the ACE of setting the load value for a server with a load of zero to 16000. When you configure this parameter, the ACE sends all new connections to the server that has a load of zero until the next load update arrives from the SNMP probe for this server. There may be times when you want the ACE to send all new connections to a real server whose load is zero. 3. In the Weight Connection field, check the check box to instruct the ACE to use the current connection count in the final load calculation for a real server. When you configure this option, the ACE includes the current connection count in the total load calculation for each real server in a server farm. Clear the check box to reset the behavior of the ACE to the default of excluding the current connection count from the load calculation. <p>To instruct the ACE to select the server with the lowest load, use the predictor least-loaded command in server farm host or redirect configuration mode. With this predictor, the ACE uses SNMP probes to query the real servers for load parameter values (for example, CPU utilization or memory utilization). This predictor is considered adaptive because the ACE continuously provides feedback to the load-balancing algorithm based on the behavior of the real server.</p> <p>To use this predictor, you must associate an SNMP probe with it. The ACE queries user-specified OIDs periodically based on a configurable time interval. The ACE uses the retrieved SNMP load value to determine the server with the lowest load.</p> <p>The syntax of this predictor command is as follows:</p> <p style="padding-left: 40px;">predictor least-loaded probe <i>name</i></p> <p>The name argument specifies the identifier of the existing SNMP probe that you want the ACE to use to query the server. Enter an unquoted text string with no spaces and a maximum of 64 alphanumeric characters.</p> <p>For example, to configure the ACE to select the real server with the lowest load based on feedback from an SNMP probe called PROBE_SNMP, enter:</p> <pre>host1/Admin(config)# serverfarm SF1 host1/Admin(config-sfarm-host)# predictor least-loaded probe PROBE_SNMP host1/Admin(config-sfarm-host-predictor)#</pre> <p>To reset the predictor method to the default of round-robin, enter:</p> <pre>host1/Admin(config-sfarm-host)# no predictor</pre>

Table 5-7 Predictor Method Attributes (continued)

Predictor Method	Description / Action
Response	<p>The ACE selects the server with the lowest response time for a requested response-time measurement.</p> <ol style="list-style-type: none"> In the Response Type field, select the type of measurement to use: <ul style="list-style-type: none"> App-req-to-req—The response time from when the ACE sends an HTTP request to a server to the time that the ACE receives a response from the server for that request. Syn-to-close—The response time from when the ACE sends a TCP SYN to a server to the time that the ACE receives a CLOSE from the server. Syn-to-synack—The response time from when the ACE sends a TCP SYN to a server to the time that the ACE receives a SYN-ACK from the server. In the Response Samples field, enter the number of samples over which you want to average the results of the response-time measurement. Valid entries are 1, 2, 4, 8, and 16 (integers from 1 to 16 that are also a power of 2).

Step 5 Click **Deploy Now** to deploy this configuration.

Related Topics

- [Configuring Health Monitoring for Real Servers, page 5-25](#)
- [Configuring Real Servers, page 5-4](#)
- [Configuring Sticky Groups, page 6-7](#)
- [Adding Real Servers to a Server Farm, page 5-14](#)
- [Configuring Server Farm HTTP Return Error-Code Checking, page 5-21](#)

Configuring Server Farm HTTP Return Error-Code Checking

After adding a server farm, ([Configuring Server Farms, page 5-12](#)), you can associate real servers with it and configure the predictor method and retrace maps. These options appear after you have successfully added a server farm.

Use this procedure to configure HTTP return error-code checking (retrace map) for a server farm.



Note

This feature is available only for server farms configured as hosts. It is not available for server farms configured with the type Redirect.

Assumption

A host type server farm has been added to the ANM. (See [Configuring Server Farms, page 5-12](#).)

Procedure

Step 1 Select **Config > Devices > context > Load Balancing > Server Farms**. The Server Farms table appears.

Draft - CISCO CONFIDENTIAL

Step 2 Select the server farm you want to configure for return error-code checking, then click the Retcode Map tab. The Retcode Map table appears.

Step 3 Click **Add** to add a new entry to the table. The Retcode Map configuration pane appears.



Note You cannot modify an entry in the Retcode Map table. Instead, delete the existing entry, then add a new one.

Step 4 In the Lowest Retcode field, enter the minimum value for an HTTP return error code. Valid entries are integers from 100 to 599. This number must be less than or equal to the number in the Highest Retcode field.

Step 5 In the Highest Retcode field, enter the maximum number for an HTTP return error code. Valid entries are integers from 100 to 599. This number must be greater than or equal to the number in the Lowest Retcode field.

Step 6 In the Type field, specify the action to be taken and related options using the information in [Table 5-8](#).



Note For ACE 1.0 modules and ACE appliances, the only available option is Count.

Table 5-8 Return-Code Type Configuration Options

Option	Description
Count	The ACE tracks the total number of return codes received for each return code number that you specify.
Log	The ACE generates a syslog error message when the number of events reaches a specified threshold. <ol style="list-style-type: none"> In the Threshold field, enter the number of events that the ACE is to receive before generating a syslog error message. Valid entries are integers from 1 to 4294967295. In the Reset field, enter the time interval in seconds for which the ACE checks for the return code. Valid entries are integers from 1 to 4294967295 seconds.
Remove	The ACE generates a syslog error message when the number of events reaches a specified threshold and then removes the server from service. <ol style="list-style-type: none"> In the Threshold field, enter the number of events that the ACE is to receive before generating a syslog error message and removing the server from service. Valid entries are integers from 1 to 4294967295. In the Reset field, enter the time interval in seconds for which the ACE checks for the return code. Valid entries are integers from 1 to 4294967295 seconds. In the Resume Service field, enter the number of seconds that the ACE waits before it resumes service for the real server automatically after taking the real server out of service. Valid entries are 30 to 3600 seconds with a default of 300 seconds.

Step 7 Click:

- **Deploy Now** to deploy this configuration.
- **Cancel** to exit this procedure without saving your entries and to return to the Retcode Map table.
- **Next** to deploy your entries and to add another retcode map.

Related Topics

- [Using Virtual Contexts, page 3-1](#)
- [Configuring Virtual Context Class Maps, page 11-6](#)
- [Configuring Virtual Context Policy Maps, page 11-30](#)
- [Configuring Real Servers, page 5-4](#)
- [Configuring Sticky Groups, page 6-7](#)

Viewing All Server Farms

Use this procedure to view all server farms associated with a virtual context.

Procedure

-
- Step 1** Select **Config > Devices**. The Virtual Contexts table appears.
- Step 2** Select the virtual context with the server farms you want to view, then click **Load Balancing > Server Farms**. The Server Farms table appears with the following information:
- Server farm name
 - Server farm type (either host or redirect)
 - Description
 - Number of real servers associated with the server farm
 - Number of predictor methods for the server farm
 - Number of entries in the HTTP retcode map table

You can click on any of the entries in the last three columns to view specific information about those entries.

Related Topics

- [Configuring Server Farms, page 5-12](#)
- [Adding Real Servers to a Server Farm, page 5-14](#)
- [Configuring the Predictor Method for Server Farms, page 5-16](#)
- [Configuring Server Farm HTTP Return Error-Code Checking, page 5-21](#)

Health Monitoring

You can instruct the ACE to check the health of servers and server farms by configuring health probes (sometimes referred to as *keepalives*). After you create a probe, you assign it to a real server or a server farm. A probe can be one of many types, including TCP, ICMP, Telnet, HTTP, and so on. You can also configure scripted probes using the TCL scripting language (see [TCL Scripts, page 5-24](#)).

The ACE sends out probes periodically to determine the status of a server, verifies the server response, and checks for other network problems that may prevent a client from reaching a server. Based on the server response, the ACE can place the server in or out of service, and, based on the status of the servers in the server farm, can make reliable load-balancing decisions.

Draft - CISCO CONFIDENTIAL

Health monitoring on the ACE tracks the state of a server by sending out probes. Also referred to as out-of-band health monitoring, the ACE verifies the server response or checks for any network problems that can prevent a client to reach a server. Based on the server response, the ACE can place the server in or out of service, and can make reliable load balancing decisions.

The ACE identifies the health of a server in the following categories:

- Passed—The server returns a valid response.
- Failed—The server fails to provide a valid response to the ACE is unable to reach a server for a specified number of retries.

By configuring the ACE for health monitoring, the ACE sends active probes periodically to determine the server state.

The ACE supports 4000 unique probe configurations which includes ICMP, TCP, HTTP, and other predefined health probes. The ACE also allows the opening of 1000 sockets simultaneously.

Related Topics

- [Configuring Health Monitoring for Real Servers, page 5-25](#)
- [TCL Scripts, page 5-24](#)

TCL Scripts

The ACE supports several specific types of health probes (for example HTTP, TCP, or ICMP health probes) when you need to use a diverse set of applications and health probes to administer your network. The basic health probe types supported in the current ACE software release may not support the specific probing behavior that your network requires. To support a more flexible health-probing functionality, the ACE allows you to upload and execute TCL scripts on the ACE.

The TCL interpreter code in the ACE is based on Release 8.44 of the standard TCL distribution. You can create a script to configure health probes. Script probes operate similar to other health probes available in the ACE software. As part of a script probe, the ACE executes the script periodically, and the exit code that is returned by the executing script indicates the relative health and availability of specific real servers. For information on health probes, see [Configuring Health Monitoring for Real Servers, page 5-25](#).

For your convenience, the following sample scripts for the ACE are available to support the TCL feature and are supported by Cisco TAC:

- CHECKPORT_STD_SCRIPT
- ECHO_PROBE_SCRIPT
- FINGER_PROBE_SCRIPT
- FTP_PROBE_SCRIPT
- HTTP_PROBE_SCRIPT
- HTTPCONTENT_PROBE
- HTTPHEADER_PROBE
- HTTPPROXY_PROBE
- IMAP_PROBE
- LDAP_PROBE
- MAIL_PROBE

- POP3_PROBE
- PROBENOTICE_PROBE
- RTSP_PROBE
- SSL_PROBE_SCRIPT
- TFTP_PROBE

The ace_scripts.tgz zip file contains these scripts and is located at the URL:

<http://www.cisco.com/cgi-bin/tablebuild.pl/cat6000-intellother>

To load a script into memory on the ACE and enable it for use, use the **script file** command. For detailed information on uploading and executing Toolkit Command Language (TCL) scripts on the ACE, refer to the *Cisco 4700 Series Application Control Engine Appliance Routing and Bridging Configuration Guide*.

Configuring Health Monitoring for Real Servers

To check the health and availability of a real server, the ACE periodically sends a probe to the real server. Depending on the server response, the ACE determines whether to include the server in its load-balancing decision.

Use this procedure to establish monitoring of real servers to determine their viability in load-balancing decisions.

Procedure

-
- Step 1** Select **Config > Devices > context > Load Balancing > Health Monitoring**. The Health Monitoring table appears.
 - Step 2** Click **Add** to add a new health monitoring probe, or select an existing entry, then click **Edit** to modify it. The Health Monitoring screen appears.
 - Step 3** In the Name field, enter a name that identifies the probe and that associates the probe with the real server. Valid entries are text strings with a maximum of 64 characters.

Draft - CISCO CONFIDENTIAL

- Step 4** In the Type field, select the type of probe you want to use. The probe type determines what the probe sends to the real server. See [Table 5-9](#) for the types of probes and their descriptions.

Table 5-9 Probe Types

Probe Type	Description
DNS	Sends a request to a DNS server giving it a configured domain. To determine if the server is up, the ACE must receive the configured IP address for that domain.
ECHO-TCP	Sends a string to the server and compares the response with the original string. If the response string matches the original, the server is marked as passed. If not, the ACE retries as configured before the server is marked as failed.
ECHO-UDP	Sends a string to the server and compares the response with the original string. If the response string matches the original, the server is marked as passed. If not, the ACE retries as configured before the server is marked as failed.
FINGER	Sends a probe to the server to verify that a defined username is a username on the server.
FTP	Initiates an FTP session. By default, this probe is for an anonymous login with the option of configuring a user ID and password. The ACE performs an FTP GET or LS to determine the outcome of the problem. This probe supports only active connections.
HTTP	Sets up a TCP connection and issues an HTTP request. Any valid HTTP response causes the probe to mark the real server as passed.
HTTPS	Similar to an HTTP probe, but this probe uses SSL to generate encrypted data.
ICMP	Sends an ICMP request and listens for a response. If the server returns a response, the ACE marks the real server as passed. If there is no response and times out, or an ICMP standard error occurs, such as <code>DESTINATION_UNREACHABLE</code> , the ACE marks the real server as failed.
IMAP	Initiates an IMAP session, using a configured user ID and password. Then, the probe attempts to retrieve e-mail from the server and validates the result of the probe based on the return codes received from the server.
POP	Initiates a POP session, using a configured user ID and password. Then, the probe attempts to retrieve e-mail from the server and validates the result of the probe based on the return codes received from the server.
RADIUS	Connects to a RADIUS server and logs into it to determine if the server is up.
RTSP	Establishes a TCP connection and sends a request packet to the server. The ACE compares the response with the configured response code to determine whether the probe succeeded.
Scripted	Executes probes from a configured script to perform health probing. This method allows you to author specific scripts with features not present in standard probes. For ACE appliances, the script probe file name must first be established on the device.

Table 5-9 Probe Types (continued)

Probe Type	Description
SIP-TCP	Establishes a TCP connection and sends an OPTIONS request packet to the user agent on the server. The ACE compares the response with the configured response code or expected string, or both, to determine whether the probe has succeeded. If you do not configure an expected status code, any response from the server is marked as failed.
SIP-UDP	Establishes a UDP connection and sends an OPTIONS request packet to the user agent on the server. The ACE compares the response with the configured response code or expected string, or both, to determine whether the probe has succeeded. If you do not configure an expected status code, any response from the server is marked as failed.
SMTP	Initiates an SMTP session by logging into the server.
SNMP	Establishes a UDP connection and sends a maximum of eight SNMP OID queries to probe the server. The ACE weighs and averages the load information that is retrieved and uses it as input to the least-loaded algorithm for load-balancing decisions. If the retrieved value is within the configured threshold, the server is marked as passed. If the threshold is exceeded, the server is marked as failed.
TCP	Initiates a TCP handshake and expects a response. By default, a successful response causes the probe to mark the server as passed. The probe then sends a FIN to end the session. If the response is not valid, or if there is no response, the probe marks the real server as failed.
TELNET	Establishes a connection to the real server and verifies that a greeting from the application was received.
UDP	Sends a UDP packet to a real server. The probe marks the server as failed only if an ICMP Port Unreachable messages is returned.

Step 5 Enter health monitoring general attributes (see [Table 5-10](#)).

Table 5-10 Health Monitoring General Attributes

Field	Action
Description	Enter a description for this probe. Valid entries are unquoted alphanumeric text strings with no spaces and a maximum of 240 characters.
Probe Interval	Enter the number of seconds that the ACE is to wait before sending another probe to a server marked as passed. Valid entries are from 2 to 65535 with a default of 120.
Pass Detect Count	Enter the number of successful probe responses from the server before the server is marked as passed. Valid entries are integers from 1 to 65535 with a default of 3.
Pass Detect Interval	Enter the number of seconds that the ACE is to wait before sending another probe to a server marked as failed. Valid entries are integers from 2 to 65535 with a default of 300.
Receive Timeout	Enter the number of seconds the ACE is to wait for a response from a server that has been probed before marking the server as failed. Valid entries are integers from 1 to 65535 with a default of 10.

Draft - CISCO CONFIDENTIAL**Table 5-10 Health Monitoring General Attributes (continued)**

Field	Action
Fail Detect	Enter the consecutive number of times that an ACE must detect that probes have failed to contact a server before marking the server as failed. Valid entries are integers from 1 to 65535 with a default of 3.
Dest IP Address ¹	By default, the probe uses the IP address from the real or virtual server configuration for the destination IP address. To override the destination address that the probe uses, enter the preferred destination IP address in this field using dotted-decimal notation, such as 192.168.11.1.
Is Routed ²	Select the check box to indicate that the destination IP address is routed according to the ACE internal routing table. Clear the check box to indicate that the destination IP address is not routed according to the ACE internal routing table.

1. The Dest IP Address field does not appear for the Scripted and SNMP probe types.

2. The Is Routed field does not appear for the RTSP, Scripted, SIP-TCP, SIP-UDP, and SNMP probe types.

Step 6 Enter the attributes for the specific probe type selected:

- For DNS probes, see [Table 5-11](#).
- For Echo-TCP probes, see [Table 5-12](#).
- For Echo-UDP probes, see [Table 5-13](#).
- For Finger probes, see [Table 5-14](#).
- For FTP probes, see [Table 5-15](#).
- For HTTP probes, see [Table 5-16](#).
- For HTTPS probes, see [Table 5-17](#).
- There are no specific attributes for ICMP probes.
- For IMAP probes, see [Table 5-18](#).
- For POP probes, see [Table 5-19](#).
- For RADIUS probes, see [Table 5-20](#).
- For RTSP probes, see [Table 5-21](#).
- For Scripted probes, see [Table 5-22](#).
- For SIP-TCP probes, see [Table 5-23](#).
- For SIP-UDP probes, see [Table 5-24](#).
- For SMTP probes, see [Table 5-25](#).
- For SNMP probes, see [Table 5-26](#).
- For TCP probes, see [Table 5-27](#).
- For Telnet probes, see [Table 5-28](#).
- For UDP probes, see [Table 5-29](#).

Step 7 Click:

- **Deploy Now** to deploy this configuration on the ACE.
 - **Cancel** to exit this procedure without saving your entries and to return to the Health Monitoring table.
 - **Next** to deploy your entries and to configure another probe.
-

Related Topics

- [Configuring DNS Probe Expect Addresses, page 5-40](#)
- [Configuring Headers for HTTP and HTTPS Probes, page 5-41](#)
- [Configuring Health Monitoring Expect Status, page 5-42](#)
- [Configuring Real Servers, page 5-4](#)
- [Configuring Server Farms, page 5-12](#)
- [Configuring Sticky Groups, page 6-7](#)

Configuring Probe Attributes

Refer to the following topics to configure health monitoring probe-specific attributes:

- [DNS Probe Attributes, page 5-30](#)
- [Echo-TCP Probe Attributes, page 5-30](#)
- [Echo-UDP Probe Attributes, page 5-30](#)
- [Finger Probe Attributes, page 5-31](#)
- [FTP Probe Attributes, page 5-31](#)
- [HTTP Probe Attributes, page 5-31](#)
- [HTTPS Probe Attributes, page 5-33](#)
- [IMAP Probe Attributes, page 5-35](#)
- [POP Probe Attributes, page 5-35](#)
- [RADIUS Probe Attributes, page 5-36](#)
- [RTSP Probe Attributes, page 5-36](#)
- [Scripted Probe Attributes, page 5-37](#)
- [SIP-TCP Probe Attributes, page 5-38](#)
- [SIP-UDP Probe Attributes, page 5-38](#)
- [SMTP Probe Attributes, page 5-39](#)
- [SNMP Probe Attributes, page 5-39](#)
- [TCP Probe Attributes, page 5-39](#)
- [Telnet Probe Attributes, page 5-40](#)
- [UDP Probe Attributes, page 5-40](#)

Refer to the following topics for additional configuration options for health monitoring probes:

- [Configuring DNS Probe Expect Addresses, page 5-40](#)

Draft - CISCO CONFIDENTIAL

- [Configuring Headers for HTTP and HTTPS Probes, page 5-41](#)
- [Configuring Health Monitoring Expect Status, page 5-42](#)
- [Configuring an OID for SNMP Probes, page 5-43](#)

DNS Probe Attributes**Table 5-11** *DNS Probe Attributes*

Field	Action
Domain Name	Enter the domain name that the probe is to send to the DNS server. Valid entries are unquoted text strings with a maximum of 255 characters.
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.

To configure expect addresses for DNS probes, see [Configuring DNS Probe Expect Addresses, page 5-40](#).

Echo-TCP Probe Attributes**Table 5-12** *Echo-TCP Probe Attributes*

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.
Is Connection	Select the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 10.
Send Data	Enter the ASCII data that the probe is to send to the server. Valid entries are unquoted text strings with no spaces and a maximum of 255 characters.

Echo-UDP Probe Attributes**Table 5-13** *Echo-UDP Probe Attributes*

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.
Send Data	Enter the ASCII data that the probe is to send to the server. Valid entries are unquoted text strings with no spaces and a maximum of 255 characters.

Finger Probe Attributes

Table 5-14 *Finger Probe Attributes*

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.
Is Connection	Select the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 10.
Send Data	Enter the ASCII data that the probe is to send to the server. Valid entries are unquoted text strings with no spaces and a maximum of 255 characters.

FTP Probe Attributes

Table 5-15 *FTP Probe Attributes*

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.
Is Connection	Select the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 10.

To configure probe expect statuses for FTP probes, see [Configuring Health Monitoring Expect Status](#), page 5-42.

HTTP Probe Attributes

Table 5-16 *HTTP Probe Attributes*

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.
Is Connection	Select the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 10.
User Name	Enter the user identifier to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters.

Draft - CISCO CONFIDENTIAL**Table 5-16 HTTP Probe Attributes (continued)**

Field	Action
Password	Enter the password to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters. Reenter the password in the Confirm field.
Expect Regex	Enter the expected response data from the probe destination. Valid entries are text strings (quotes allowed) with a maximum of 255 characters.
Expect Regex Offset	Enter the number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regex field. Valid entries are integers from 1 to 4000.
Hash	Select the Hash check box to indicate that the ACE is to use an MD5 hash for an HTTP GET probe. Clear the Hash check box to indicate that the ACE should not use an MD5 hash for an HTTP GET probe.
Hash String	This field appears if the Hash check box is selected. Enter the 32-bit hash value that the ACE is to compare with the hash that is generated from the HTTP page sent by the server. If you do not provide this value, the ACE generates a value the first time it queries the server, stores this value, and matches this value with other responses from the server. A successful comparison causes the probe to maintain an Alive state. Enter the MD5 hash value as a quoted or unquoted hexadecimal string with 16 characters.
Request Method Type	Select the type of HTTP request method that is to be used for this probe: <ul style="list-style-type: none"> • N/A—This option is not defined. • Head—The server is to only get the header for the page. Using this method can prevent the ACE from assuming that the service is down due to changed content and therefore changed hash values. • Get—The HTTP request method is a GET with a URL of “/”. This request method directs the server to get the page, and the ACE calculates a hash value for the content of the page. If the page content information changes, the hash value no longer matches the original hash value and the ACE assumes the service is down. This is the default request method.
Request HTTP URL	This field appears if you select Head or Get in the Request Method Type field. Enter the URL path on the remote server. Valid entries are strings of up to 255 characters specifying the URL path. The default path is “/”.

To configure probe headers and expect statuses for HTTP probes, see:

- [Configuring Headers for HTTP and HTTPS Probes, page 5-41](#)
- [Configuring Health Monitoring Expect Status, page 5-42](#)

HTTPS Probe Attributes

Table 5-17 HTTPS Probe Attributes

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.
Is Connection	Select the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 10.
User Name	Enter the user identifier to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters.
Password	Enter the password to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters. Reenter the password in the Confirm field.
Expect Regex	Enter the expected response data from the probe destination. Valid entries are text strings (quotes allowed) with a maximum of 255 characters.
Expect Regex Offset	Enter the number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regex field. Value entries are integers from 1 to 4000.
Hash	Select the Hash check box to indicate that the ACE is to use an MD5 hash for an HTTP GET probe. Clear this check box to indicate that the ACE is not to use an MD5 hash for an HTTP GET probe.
Hash String	This field appears if the Hash check box is selected. Enter the 32-bit hash value that the ACE is to compare with the hash that is generated from the HTTP page sent by the server. If you do not provide this value, the ACE generates a value the first time it queries the server, stores this value, and matches this value with other responses from the server. A successful comparison causes the probe to maintain an Alive state. Enter the MD5 hash value as a quoted or unquoted hexadecimal string with 16 characters.
Request Method Type	Select the type of HTTP request method that is to be used for this probe: <ul style="list-style-type: none"> • N/A—This option is not defined. • Head—The server is to only get the header for the page. Using this method can prevent the ACE from assuming that the service is down due to changed content and therefore changed hash values. • Get—The HTTP request method is a GET with a URL of “/”. This request method directs the server to get the page, and the ACE calculates a hash value for the content of the page. If the page content information changes, the hash value no longer matches the original hash value and the ACE assumes the service is down. This is the default request method.

Draft - CISCO CONFIDENTIAL**Table 5-17** *HTTPS Probe Attributes (continued)*

Field	Action
Request HTTP URL	<p>This field appears if you select Head or Get in the Request Method Type field.</p> <p>Enter the URL path on the remote server. Valid entries are strings of up to 255 characters specifying the URL path. The default path is “/”.</p>
Cipher	<p>Select the cipher suite to be used with this HTTPS probe:</p> <ul style="list-style-type: none"> • RSA_ANY—The HTTPS probe accepts all RSA-configured cipher suites and that no specific suite is configured. This is the default action. • RSA_EXPORT1024_WITH_DES_CBC_SHA • RSA_EXPORT1024_WITH_RC4_56_MD5 • RSA_EXPORT1024_WITH_RC4_56_SHA • RSA_EXPORT_WITH_DES40_CBC_SHA • RSA_EXPORT_WITH_RC4_40_MD5 • RSA_WITH_3DES_EDE_CBC_SHA • RSA_WITH_AES_128_CBC_SHA • RSA_WITH_AES_256_CBC_SHA • RSA_WITH_DES_CBC_SHA • RSA_WITH_RC4_128_MD5 • RSA_WITH_RC4_128_SHA
SSL Version	<p>Select the version of SSL or TLS to be used in ClientHello messages sent to the server:</p> <ul style="list-style-type: none"> • SSLv2—The probe is to use SSL version 2. • SSLv3—The probe is to use SSL version 3. • TLSv1—The probe is to use TLS version 1. <p>By default, the probe sends ClientHello messages with an SSL version 3 header and a TLS version 1 message.</p>

To configure probe headers and expect statuses for HTTPS probes, see:

- [Configuring Headers for HTTP and HTTPS Probes, page 5-41](#)
- [Configuring Health Monitoring Expect Status, page 5-42](#)

IMAP Probe Attributes

Table 5-18 *IMAP Probe Attributes*

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.
Is Connection	Select the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 10.
User Name	Enter the user identifier to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 24 characters.
Password	Enter the password to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 24 characters. Reenter the password in the Confirm field.
Mailbox Name	Enter the user mailbox name from which to retrieve e-mail for this IMAP probe. Valid entries are unquoted text strings with a maximum of 64 characters.
Request Method	Enter the request method command for this probe. Valid entries are text strings with a maximum of 32 characters and no spaces.

POP Probe Attributes

Table 5-19 *POP Probe Attributes*

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.
Is Connection	Select the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 10.
User Name	Enter the user identifier to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters.
Password	Enter the password to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters. Reenter the password in the Confirm field.
Request Method	Enter the request method command for this probe. Valid entries are text strings with a maximum of 32 characters and no spaces.

Draft - CISCO CONFIDENTIAL**RADIUS Probe Attributes****Table 5-20 RADIUS Probe Attributes**

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.
User Secret	Enter the shared secret to be used to allow probe access to the RADIUS server. Valid entries are case-sensitive strings with no spaces and a maximum of 128 characters.
User Name	Enter the user identifier to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters.
Password	Enter the password to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters. Reenter the password in the Confirm field.
NAS IP Address	Enter the IP address of the Network Access Server (NAS) in dotted-decimal format, such as 192.168.11.1.


RTSP Probe Attributes**Table 5-21 RTSP Probe Attributes**

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.
Is Connection	Select the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 10.
RTSP Require Header Name	Enter the Require header for this probe.
RTSP Proxy-Require Header Name	Enter the Proxy-Require header for this probe.
RTSP Request Method Type	Select the request method type: <ul style="list-style-type: none"> N/A—No request method is selected. Describe—This probe is to use the Describe request type.

To configure probe expect statuses for RTSP probes, see [Configuring Health Monitoring Expect Status, page 5-42](#).

Scripted Probe Attributes

Table 5-22 Scripted Probe Attributes

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.
Script needs to be copied from remote location?	Select this check box to indicate that the file needs to be copied from a remote server. Clear this check box to indicate that the script resides locally.
Protocol	This field appears if the script is to be copied from a remote server. Select the protocol to be used for copying the script: <ul style="list-style-type: none"> • FTP—The script is to be copied using FTP. • TFTP—The script is to be copied using TFTP.
Username	This field appears if FTP is selected in the Protocol field. Enter the name of the user account on the remote server.
Password	This field appears if FTP is selected in the Protocol field. Enter the password for the user account on the remote server. Reenter the password in the Confirm field.
Source File Name	This field appears if the script is to be copied from a remote server. Enter the host IP address, path, and filename of the file on the remote server in the format <i>host-ip/path/filename</i> where: <ul style="list-style-type: none"> • <i>host-ip</i> represents the IP address of the remote server. • <i>path</i> represents the directory path of the file on the remote server. • <i>filename</i> represents the filename of the file on the remote server. For example, your entry might resemble 192.168.11.2/usr/bin/my-script.ext.
Script Name	Enter the local name that you want to assign to this file on the ACE. This file can reside in the disk0: directory or the probe: directory (if the probe: directory exists).  Note The script file must first be established on the ACE device and the name must be entered exactly as is appears on the device. Please refer to your ACE documentation for more details. <hr/> Valid entries are unquoted text strings with no spaces and a maximum of 255 characters.
Script Arguments	Valid arguments are unquoted text strings with no spaces; separate multiple arguments with a space. The field limit is 255 characters.

Draft - CISCO CONFIDENTIAL**SIP-TCP Probe Attributes****Table 5-23 SIP-TCP Probe Attributes**

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.
Is Connection	Select the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 10.
Expect Regex	Enter the expected response data from the probe destination. Valid entries are text strings with a maximum of 255 characters. This field accepts both single and double quotes. Double quotes are considered delimiters so they don't appear on the device. Single quotes will appear on the device.
Expect Regex Offset	Enter the number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regex field. Value entries are integers from 1 to 4000.

To configure probe expect statuses for SIP-TCP probes, see [Configuring Health Monitoring Expect Status](#), page 5-42.

SIP-UDP Probe Attributes**Table 5-24 SIP-UDP Probe Attributes**

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.
Expect Regex	Enter the expected response data from the probe destination. Valid entries are text strings (quotes allowed) with a maximum of 255 characters.
Expect Regex Offset	Enter the number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regex field. Value entries are integers from 1 to 4000.

To configure probe expect statuses for SIP-UDP probes, see [Configuring Health Monitoring Expect Status](#), page 5-42.

SMTP Probe Attributes

Table 5-25 *SMTP Probe Attributes*

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.
Is Connection	Select the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 10.

To configure probe expect statuses for SMTP probes, see [Configuring Health Monitoring Expect Status](#), page 5-42.

SNMP Probe Attributes

Table 5-26 *SNMP Probe Attributes*

Field	Action
SNMP Community String	Enter the SNMP community string. Valid entries are unquoted text strings with no spaces and a maximum of 32 characters.
SNMP Version Number	Select the SNMP version for this probe: <ul style="list-style-type: none"> N/A—No version is selected. SNMPv1—This probe is to use SNMP version 1. SNMPv2c—This probe is to use SNMP version 2c.

To configure the SNMP OID for SNMP probes, see [Configuring an OID for SNMP Probes](#), page 5-43.

TCP Probe Attributes

Table 5-27 *TCP Probe Attributes*

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.
Is Connection	Select the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 10.
Send Data	Enter the ASCII data that the probe is to send to the server. Valid entries are unquoted text strings with no spaces and a maximum of 255 characters.

Draft - CISCO CONFIDENTIAL**Table 5-27 TCP Probe Attributes (continued)**

Field	Action
Expect Regex	Enter the expected response data from the probe destination. Valid entries are text strings (quotes allowed) with a maximum of 255 characters.
Expect Regex Offset	Enter the number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regex field. Value entries are integers from 1 to 4000.

Telnet Probe Attributes**Table 5-28 Telnet Probe Attributes**

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.
Is Connection	Select the check box to indicate that connection parameters are configured. Clear the check box to indicate that connection parameters are not configured.
Open Timeout	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535, and the default value is 10.

UDP Probe Attributes**Table 5-29 UDP Probe Attributes**

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses the port number based on its type.
Send Data	Enter the ASCII data that the probe is to send to the server. Valid entries are unquoted text strings with no spaces and a maximum of 255 characters.
Expect Regex	Enter the expected response data from the probe destination. Valid entries are text strings (quotes allowed) with a maximum of 255 characters.
Expect Regex Offset	Enter the number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regex field. Value entries are integers from 1 to 4000.

Configuring DNS Probe Expect Addresses

When a DNS probe sends a domain name resolve request to the server, it verifies the returned IP address by matching the received IP address with the configured addresses.

Use this procedure to specify the IP address that the ACE expects to receive in response to a DNS request.

Assumption

A DNS probe has been configured. See [Configuring Health Monitoring for Real Servers, page 5-25](#) for more information.

Procedure

-
- Step 1** Select **Config > Devices > context > Load Balancing > Health Monitoring**. The Health Monitoring table appears.
- Step 2** Select the DNS probe that you want to configure with an expected IP address. The Expect Addresses table appears.
- Step 3** Click **Add** to add an entry to the Expect Addresses table. The Expect Address configuration pane appears.



Note You cannot modify an entry in the Expect Addresses table. Instead, delete the existing entry, then add a new one.

- Step 4** In the IP Address field, enter the IP address that the ACE is to expect as a server response to a DNS request. Valid entries are unique IP addresses in dotted-decimal notation, such as 192.168.11.1.
- Step 5** Click:
- **Deploy Now** to deploy this configuration.
 - **Cancel** to exit this procedure without saving your entry and to return to the Expect Addresses table.
 - **Next** to deploy your entry and to add another IP Address to the Expect Addresses table.
-

Related Topics

- [Configuring Health Monitoring for Real Servers, page 5-25](#)
- [DNS Probe Attributes, page 5-30](#)

Configuring Headers for HTTP and HTTPS Probes

Use this procedure to specify header fields for HTTP and HTTPS probes.

Assumption

An HTTP or HTTPS probe has been configured. See [Configuring Health Monitoring for Real Servers, page 5-25](#) for more information.

Procedure

-
- Step 1** Select **Config > Devices > context > Load Balancing > Health Monitoring**. The Health Monitoring table appears.
- Step 2** Select the HTTP or HTTPS probe that you want to configure with a header. The Probe Headers table appears.
- Step 3** Click **Add** to add an entry, or select an existing entry, then click **Edit** to modify it. The Probe Headers configuration pane appears.

Draft - CISCO CONFIDENTIAL

- Step 4** In the Header Name field, select the HTTP header the probe is to use.
- Step 5** In the Header Value field, enter the string to assign to the header field. Valid entries are text strings with a maximum of 255 characters. If the string includes spaces, enclose the string with quotes.
- Step 6** Click:
- **Deploy Now** to deploy this configuration.
 - **Cancel** to exit this procedure without saving your entry and to return to the Probe Headers table.
 - **Next** to deploy your entry and to add another header entry to the Probe Headers table.
-

Related Topics

- [Configuring Health Monitoring for Real Servers, page 5-25](#)
- [HTTP Probe Attributes, page 5-31](#)
- [HTTPS Probe Attributes, page 5-33](#)

Configuring Health Monitoring Expect Status

When the ACE receives a response from the server, it expects a status code to mark a server as passed. By default, there are no status codes configured on the ACE. If you do not configure a status code, any response code from the server is marked as failed.

Expect status codes can be configured for FTP, HTTP, HTTPS, RTSP, SIP-TCP, SIP-UDP, and SMTP probes.

Use this procedure to configure a single or range of code responses that the ACE expects from the probe destination.

Assumption

An FTP, HTTP, HTTPS, RTSP, SIP-TCP, SIP-UDP or SMTP probe has been configured. See [Configuring Health Monitoring for Real Servers, page 5-25](#) for more information.

Procedure

-
- Step 1** Select **Config > Devices > context > Load Balancing > Health Monitoring**. The Health Monitoring table appears.
- Step 2** Select the probe that you want to configure for expect status codes, then click the Expect Status tab. The Expect Status table appears.
- Step 3** Click **Add** to add an entry, or select an existing entry, then click **Edit** to modify it. The Expect Status configuration pane appears.
- Step 4** To configure a single expect status code:
- a. In the Min Expect Status Code field, enter the expect status code for this probe. Valid entries are integers from 0 to 999.
 - b. In the Max Expect Status code, enter the same expect status code that you entered in the Min Expect Status Code field.

- Step 5** To configure a range of expect status codes:
- In the Min Expect Status Code, enter the lower limit of the range of status codes. Valid entries are integers from 0 to 999.
 - In the Max Expect Status Code, enter the upper limit of a range of status codes. Valid entries are integers from 0 to 999. The value in this field must be greater than or equal to the value in the Min Expect Status Code field.
- Step 6** Click:
- Deploy Now** to deploy this configuration on the ACE.
 - Cancel** to exit this procedure without saving your entries and to return to the Expect Status table.
 - Next** to deploy your entries and to add another expect status code to the Expect Status table.
-

Related Topics

- [Configuring Health Monitoring for Real Servers, page 5-25](#)
- [FTP Probe Attributes, page 5-31](#)
- [HTTP Probe Attributes, page 5-31](#)
- [SMTP Probe Attributes, page 5-39](#)

Configuring an OID for SNMP Probes

When the ACE sends a probe with an SNMP OID query, the ACE uses the retrieved value as input to the least-loaded algorithm for load-balancing decisions. Least-loaded load balancing bases the server selection on the server with the lowest load value. If the retrieved value is within the configured threshold, the server is marked as passed. If the threshold is exceeded, the server is marked as failed.

The ACE allows a maximum of eight OID queries to probe the server.

Assumption

An SNMP probe has been configured. See [Configuring Health Monitoring for Real Servers, page 5-25](#) for more information.

Procedure

- Step 1** Select **Config > Devices > context > Load Balancing > Health Monitoring**. The Health Monitoring table appears.
- Step 2** Select the SNMP probe for which you want to specify an OID. The SNMP OID for Server Load Query table appears.
- Step 3** Click **Add** to add an entry, or select an existing entry, then click **Edit** to modify it. The SNMP OID configuration pane appears.
- Step 4** In the SNMP OID field, enter the OID that the probe is to use to query the server for a value. Valid entries are unquoted strings with a maximum of 255 alphanumeric characters in dotted-decimal notation, such as .1.3.6.1.4.2021.10.1.3.1. The OID string is based on the server type.
- Step 5** In the Maximum Absolute Server Load Value field, enter the OID value in the form of an integer and to indicate that the retrieved OID value is an absolute value instead of a percent. Valid entries are integers from 1 to 4294967295.

Draft - CISCO CONFIDENTIAL

When the ACE sends a probe with an SNMP OID query, the ACE uses the retrieved value as input to the least-loaded algorithm for load-balancing decisions. By default, the ACE assumes that the retrieved OID value is a percentile value. Use this option to specify that the retrieved OID value is an absolute value.

Step 6 In the Server Load Threshold Value field, specify the threshold at which the server is to be taken out of service:

- When the OID value is based on a percent, valid entries are integers from 0 to 100.
- When the OID is based on an absolute value, valid entries are from 0 to the value specified in the Maximum Absolute Server Load Value field.

Step 7 In the Server Load Weighting field, enter the weight to assign to this OID for the SNMP probe. Valid entries are integers from 0 to 16000.

Step 8 Click:

- **Deploy Now** to deploy this configuration.
 - **Cancel** to exit this procedure without saving your entries and to return to the SNMP OID table.
 - **Next** to deploy your entries and to add another item to the SNMP OID table.
-

Related Topics

- [Configuring Health Monitoring for Real Servers, page 5-25](#)
- [SNMP Probe Attributes, page 5-39](#)