CHAPTER 7

Configuring Real Servers and Server Farms

This chapter describes how to configure real servers and server farms on the Cisco Application Control Engine (ACE) using Cisco Application Networking Manager (ANM).

Note When naming ACE objects (such as a real server, virtual server, parameter map, class map, health probe, and so on), enter an alphanumeric string of 1 to 64 characters, which can include the following special characters: underscore (_), hyphen (-), and dot (.). Spaces are not allowed.

If you are using ANM with an ACE module or ACE appliance and you configure a named object at the ACE CLI, keep in mind that ANM does not support all of the special characters that the ACE CLI allows you to use when configuring a named object. If you use special characters that ANM does not support, you may not be able to import or manage the ACE using ANM.

This chapter includes the following sections:

Information About Server Load Balancing

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.

ANM allows you to configure load balancing using:

- Virtual servers—See the "Configuring Virtual Servers" section on page 7-2.
- Real servers—See the "Configuring Real Servers" section on page 7-5.
- Dynamic Workload Scaling—See the "Configuring Dynamic Workload Scaling" section on page 7-27.
- Server farms—See the "Configuring Server Farms" section on page 7-31.
- Sticky groups—See the "Configuring Sticky Groups" section on page 8-12.
- Parameter maps—See the "Configuring Parameter Maps" section on page 9-1.

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

- Configuring Virtual Servers, page 7-2
- Load-Balancing Predictors, page 7-2
- Real Servers, page 7-3
- Dynamic Workload Scaling Overview, page 7-4
- Server Farms, page 7-5
- Configuring Health Monitoring, page 7-51
- TCL Scripts, page 7-51
- Configuring Stickiness, page 8-1

  This section includes the following topics:

  - Load-Balancing Predictors, page 7-2
  - Real Servers, page 7-3
  - Server Farms, page 7-5

Load-Balancing Predictors

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

- 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 Content— Selects the server by using a hash value based on the specified content string of the HTTP packet body
- 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.
- Hash Layer4—Selects the server using a Layer 4 generic protocol load-balancing method.
- 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 the "Configuring High Availability" section on page 12-1.

- 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.
- Least Connections—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.
- Least Loaded—Selects the server with the lowest load as determined by information from SNMP probes.
- Response—Selects the server with the lowest response time for a specific response-time measurement.
- Round Robin—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.

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 Topics

Configuring the Predictor Method for Server Farms, page 7-40

Real Servers

To provide services to clients, you configure real servers on the ACE. Real servers can be dedicated physical servers or VMware virtual machines (VMs) that you configure in groups called server farms.

Note VMs that you define as real servers can be VMs associated with a VMware vCenter Server that you import into ANM (see the "Importing VMware vCenter Servers" section on page 4-24) and VMs that the ACE recognizes when configured for Dynamic Workload Scaling (see the "Configuring Dynamic Workload Scaling" section on page 7-27).

Real servers provide client services such as HTTP or XML content, website 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 the "Load-Balancing Predictors" section on page 7-2.

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" section on page 7-29.

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

- Probe failure
- ARP timeout
- Neighbor Discovery (ND) failure (IPv6 only, which requires ACE module and ACE appliance software Version A5(1.0) or later)
- 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 7-5
- Configuring Health Monitoring for Real Servers, page 7-52

Dynamic Workload Scaling Overview

Note Dynamic Workload Scaling requires ACE module or appliance software Version A4(2.0) or later and a pair of the Cisco Nexus 7000 Series switches with Overlay Transport Virtualization (OTV) technology.

The ACE Dynamic Workload Scaling (DWS) feature permits on-demand access to remote resources, such as VMs, that you own or lease from an Internet service provider or cloud service provider. This feature uses Cisco Nexus 7000 Series switches with OTV to create a Data Center Interconnect (DCI) on a Layer 2 link over an existing IP network between geographically distributed data centers (see Figure 1-1). The local data center Cisco Nexus 7000 Series switch contains an OTV forwarding table that lists the MAC addresses of the Layer 2 extended virtual private network (VPN) and identifies the addresses as either local or remote.

When you configure the ACE for DWS, the ACE uses an XML query to poll the Cisco Nexus 7000 Series switch and obtain the OTV forwarding table information to determine the locality of the VMs (local or remote). The ACE also uses a health monitor probe that it sends to the local VMware vCenter Server to monitor the load of the local VMs based on CPU usage, memory usage, or both. When the average CPU and/or memory usage of the local VMs reaches its configured maximum threshold value, the ACE bursts traffic to the remote VMs. The ACE stops bursting traffic to the remote VMs when local VM usage drops below its configured minimum threshold value.

To use DWS, you configure the ACE to connect to the Data Center Interconnect device (Cisco Nexus 7000 Series switch) and the VMware Controller associated with the local and remote VMs. You also configure the ACE with the probe type VM to monitor a server farm's local VM CPU and memory usage, which determines when the ACE bursts traffic to the remote VMs (see the "Configuring Dynamic Workload Scaling" section on page 7-27).

For more details on this feature, see the *Cisco 4700 Series Application Control Engine Appliance Server Load-Balancing Configuration Guide*.

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 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 the "Load-Balancing Predictors" section on page 7-2.

Related Topics

Configuring Server Farms, page 7-31

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 the "Configuring Virtual Context Class Maps" section on page 13-6.

This section includes the following topics:

Configuring Load Balancing on Real Servers

You can configure load balancing on real servers.

Procedure

Step 1 Choose **Config > Devices >** *context* **> Load Balancing > Real Servers**.

The Real Servers table appears.

Step 2 In the Real Servers table, click **Poll Now** to instruct ANM to poll the devices and display the current values, and click **OK** when prompted if you want to poll the devices for data now.

Step 3 Click **Add** to add a new real server, or choose a real server you want to modify and click **Edit**.

The Real Servers configuration window appears.

Step 4 In the Real Servers configuration window, configure the server using the information in Table 7-1.

Note Fields and information related to IPv6 require ACE module and ACE appliance software Version A5(1.0) or later.

Table 7-1 Real Server Attributes

Field	Description
Name	Field that allows you to either enter a unique name for this server or accept the automatically incremented value in this field. Valid entries are unquoted text strings with no spaces and a maximum of 64 characters.
Type	Type of server: • **Host**—The real server provides content and services to clients. • **Redirect**—The server redirects traffic to a new location.
State	State of the real server: • **In Service**—The real server is in service. • **Out Of Service**—The real server is out of service.
Description	Brief description for this real server. Valid entries are strings of up to 240 characters. Spaces and special characters are allowed.
IP Address Type	Field that appears only for ACE module and ACE appliance software Version A5(1.0) or later, which supports IPv4 and IPv6. These selections appear only for real servers specified as hosts. Select the IP address type of this real server: • IPv6—The real server has an IPv6 address. • IPv4—The real server has an IPv4 address.
IPv6/IPv4 Address	For ACE module and ACE appliance software versions earlier than A5(1.0), this field does not include the IP version number. This field appears for only real servers specified as hosts. Enter a unique IP address as indicated by the IP Address Type field. The IP address cannot be of an existing virtual IP address (VIP), real server or interface in the context.

Table 7-1 Real Server Attributes (continued)

Field	Description
Fail-On-All	Field that appears only for real servers identified as host servers. By default, real servers with multiple probes configured for them have an OR logic associated with them, which means that if one of the real server probes fails, the real server fails and enters the PROBE-FAILED state. Check this checkbox to configure a real server to remain in the OPERATIONAL state unless all probes associated with it fail (AND logic). The Fail-On-All function is applicable to all probe types.
Min. Connections	Minimum number of connections to be allowed on this server before the ACE 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 from 2 to 4000000.
Max. Connections	Maximum number of active connections allowed on this server. When the number of connections exceeds this value, the ACE stops sending connections to this server until the number of connections falls below the Min. Connections value. Valid entries are from 2 to 4000000, and the default is 4000000.
Weight	Field that 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 from 1 to 100, and the default is 8.
Probes	Field that appears only as follows: • For all host real servers. The Available probe list contains all configured probe types. • For redirect real servers configured on ACE devices that use the following software versions: – ACE module: A2(3.x) and later releases – ACE appliance: A3(x) and later releases The redirect real server Available probe list contains only configured probes of the type Is Routed, which means that the ACE routes the probe address according to the ACE internal routing table (see the "Configuring Health Monitoring for Real Servers" section on page 7-52). In the Probes field, choose the probes to use for health monitoring in the Available Items list, and click **Add**. The probes appear in the Selected Items list. **Note** The probe must have the same IP address type (IPv6 or IPv4) as the real server. For example, you cannot configure an IPv6 probe to an IPv4 real server. IPv6 requires ACE module and ACE appliance software Version A5(1.0) or later. **Note** The list of available probes does not include VM probes used to monitor local VM usage. To remove probes that you do not want to use for health monitoring, choose them in the Selected Items list, and click **Remove**. The probes appear in the Available probe list.

Table 7-1 Real Server Attributes (continued)

Field	Description
Web Host 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: • %h—Inserts the hostname from the request Host header • %p—Inserts the URL path string from the request
Redirection Code	Field that appears only for real servers identified as redirect servers. Choose the appropriate redirection code: • **N/A**—Webhost redirection code is not defined. • **301**—Requested resource has been moved permanently. For future references to this resource, the client should use one of the returned URIs. • **302**—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.
Rate Bandwidth	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 from 2 to 300000000. The default is 300000000.
Rate Connection	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 from 2 to 350000. The default is 350000.

Step 5 Do one of the following:

- Click **Deploy Now** to deploy this configuration on the ACE and save your entries to the running-configuration and startup-configuration files.
- Click **Cancel** to exit the procedure without saving your entries and to return to the Real Servers table.
- Click **Next** to deploy your entries and to configure another real server.

Step 6 To display statistics and status information for an existing real server, choose a real server from the Real Servers table, then click **Details**. The **show rserver** *name* **detail** CLI command output appears. See the "Displaying Real Server Statistics and Status Information" section on page 7-9 for details.

Related Topics

- Managing Real Servers, page 7-9
- Configuring Health Monitoring for Real Servers, page 7-52
- Configuring Server Farms, page 7-31

- Configuring Sticky Groups, page 8-12

Displaying Real Server Statistics and Status Information

You can display statistics and status information for a particular real server.

Procedure

Step 1 Choose **Config > Devices >** *context* **> Load Balancing > Real Servers**.

The Real Servers table appears.

Step 2 In the Real Servers table, choose a real server from the Real Servers table, and click **Details**.

The **show rserver** *name* **detail** CLI command output appears. For details on the displayed output fields, see either the *Cisco ACE Module Server Load-Balancing Configuration Guide* or the *Cisco ACE 4700 Series Appliance Server Load-Balancing Configuration Guide*, Chapter 2, Configuring Real Servers and Server Farms.

Step 3 Click **Update Details** to refresh the output for the **show rserver** *name* **detail** CLI command. The new information appears in a separate panel with a new timestamp; both the old and the new real server statistics and status information appear side-by-side to avoid overwriting the last updated information.

Step 4 Click **Close** to return to the Real Servers table.

Related Topics

- Configuring Real Servers, page 7-5
- Managing Real Servers, page 7-9
- Displaying Real Servers, page 7-18

Managing Real Servers

This section shows how to display and manage the real servers from the Real Servers window (Config > Operations > Real Servers). This window provides you with information about each real server configured on ANM (see the "Displaying Real Servers" section on page 7-18) and provides access to function buttons that allow you to perform tasks such as activate or suspend a real server, display a real server topology map, or display connection statistics graphs.

Note The Real Servers window may not display the latest information if the periodic polling is disabled. To enable periodic polling, see the "Enabling Polling on All Devices" section on page 16-50.

Guidelines and Restrictions

The Real Servers window contains a Rows per page option that includes an All setting for displaying all configured real servers in one window. Use the All setting for viewing purposes only. ANM does not allow you to perform any operation from this window if you have more than 200 real servers selected. For example, if you use the All option to display and select more than 200 real servers and then attempt to perform the suspend operation, ANM cancels the request and displays an error message.

This section includes the following topics:

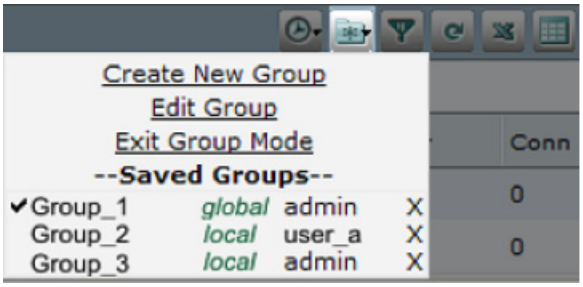
Managing Real Server Groups

This section describes how to organize real servers into groups, which allows you to display and manage a specific group of real servers without having to filter the real server display. When creating a group, you specify whether the group is available to just you or is available globally to all ANM users.

The real server group feature is available from the real servers operations window (Config > Operations > Real Servers), which contains the Groups option for managing object groups. Figure 7-1 shows the Groups icon with the following available options for managing object groups:

- Create New Group—Adds a new group.
- Edit Group—Modifies an existing group. This option displays only after you select a group to display in Group mode.
- Exit Group Mode—Changes the display from the specific group display to the display of all real servers. This option displays only after you select a group and the display enters the Group mode.
- Saved Groups—Lists the currently configured groups with each group's privilege level (local or global) and owner. From this view, you can choose a group to display or delete a group.

Figure 7-1 Object Grouping for Real Servers



Guidelines and Restrictions

This topic includes the following guidelines and restrictions:

- When you create a global group, other users can see the group if they have access to at least one object within the group. This rule does not apply to the admin user or a user with the anm-admin role because they have visibility to all global groups.
- To edit or delete a group, you must be the group owner, a user with the anm-admin role, or the admin user.
- When you delete a locally authenticated user from the ANM database, ANM deletes all the global and user-specific groups that the user created. However, when you delete a *remotely authorized* user from the remote AAA server database, ANM does not delete the groups that the user created. In this case, you must manually delete the user's groups.

This section includes the following topics:

- Creating a Real Server Group, page 7-11
- Editing or Copying a Real Server Group, page 7-12
- Displaying a Real Server Group, page 7-13
- Deleting a Real Server Group, page 7-14

Creating a Real Server Group

You can create a real server group.

Procedure

Step 1 Choose **Config > Operations > Real Servers**.

The Real Servers table appears.

Step 2 Click the **Groups** icon located above the Real Servers table.

The Groups menu appears below the icon (see Figure 7-1).

Step 3 From the Groups menu, choose **Create New Group**.

The display enters the edit mode and the Creating a New Group table appears with the list of the available real servers.

Step 4 From the Creating a New Group table, check the check box next to the real servers that you want to include in the group.

Step 5 (Optional) Check the **Hide unselected** check box to display only the real servers that you have chosen. Uncheck the check box to display all the available real servers.

Step 6 Do one of the following:

- Click **Save as** to save the group information. The Create Group popup window appears.

  From the popup window, do the following:

  a. In the Group Name text box, enter a name for the group. Enter 1 to 64 alphanumeric characters. Special characters and spaces are allowed.

  b. Choose the availability of the group by clicking one of the following radio buttons:

  – **This user only (local)**—Only you can view, modify, or delete the group.

– **All users (global)**—All ANM users can view the group if they have permission to view at least one of the real servers associated with the group. A user with the admin or anm-admin can view all groups and can also edit or delete any group.

c. Do one of the following:

– Click **Save** to save the group information. The Create Group popup window closes and the Viewing Group table appears, displaying the new group's name and associated real servers.

To exit Group mode and return to the Real Servers table, click the **Groups** icon and click **Exit Group Mode** from the Groups menu.

– Click **Cancel** to close the Create Group popup window without saving any information and to return to the Creating a New Group table.

- Click **Back to View** to exit the Group mode and return to the Virtual Servers table.

Related Topics

- Managing Real Server Groups, page 7-10
- Editing or Copying a Real Server Group, page 7-12
- Displaying a Real Server Group, page 7-13
- Deleting a Real Server Group, page 7-14

Editing or Copying a Real Server Group

You can edit a real server group or create a copy of a real server group under a different name.

Procedure

Step 1 Choose **Config > Operations > Real Servers**.

The Real Servers table appears.

Step 2 Click the **Groups** icon located above the Real Servers table.

The Groups menu appears below the icon (see Figure 7-1).

Step 3 From the Groups menu, choose the group that you want to edit.

The Viewing Group table appears, displaying the selected group's name and associated real servers.

Step 4 Click the **Groups** icon again and from the Groups menu, choose **Edit Group**.

The Editing Group table appears, displaying the complete list of available real servers with the real servers currently associated with the group highlighted and checked.

Step 5 Modify the group as needed by adding (check) or removing (uncheck) real servers as needed. Skip this step if you only want to save a copy of the current group under a different name.

Step 6 Do one of the following:

- Click **Save** to save the changes and return to the Viewing Group table, where you can view the changes.
- Click **Save as** to save the configuration under a new group name. The Create Group popup window appears.

From the popup window, do the following:

a. In the Group Name text box, enter a name for the group. Enter 1 to 64 alphanumeric characters. Special characters and spaces are allowed.

b. Choose the availability of the group by clicking one of the following radio buttons:

- **This user only (local)**—Only you can view, modify, or delete the group.
- **All users (global)**—All ANM users can view the group if they have permission view at least one of the real servers associated with the group. The admin user or a user with the anm-admin role can view all global groups and can also edit or delete these groups.

c. Do one of the following:

- Click **Save** to save the group information. The Create Group popup window closes and the Viewing Group table appears, displaying the new group's name and associated real servers.
- Click **Cancel** to close the Create Group popup window without saving any information and to return to the Creating a New Group table.

  Click **Back to View** to exit the edit mode and return to the Group mode.

Step 7 (Optional) To exit Group mode and return to the Real Servers table, click the **Groups** icon and click **Exit Group Mode** from the Groups menu.

Related Topics

- Managing Real Server Groups, page 7-10
- Creating a Real Server Group, page 7-11
- Displaying a Real Server Group, page 7-13
- Deleting a Real Server Group, page 7-14

Displaying a Real Server Group

You can display the list of real servers associated with a real server group.

Procedure

Step 1 Choose **Config > Operations > Real Servers**.

The Real Servers table appears.

Step 2 Click the **Groups** icon located above the Real Servers table.

The Groups menu appears below the icon (see Figure 7-1).

Step 3 From the Groups menu, choose the group that you want to display.

The Viewing Group table appears, displaying the selected group's name and associated real servers.

Step 4 (Optional) To exit Group mode and return to the Real Servers table, click the **Groups** icon and click **Exit Group Mode** from the Groups menu.

Related Topics

- Managing Real Server Groups, page 7-10

- Creating a Real Server Group, page 7-11
- Editing or Copying a Real Server Group, page 7-12
- Deleting a Real Server Group, page 7-14

Deleting a Real Server Group

You can delete a real server group. Deleting a real server group does not delete the group's associated real servers from the ANM database.

Procedure

Step 1 Choose **Config > Operations > Real Servers**.

The Real Servers table appears.

Step 2 Click the **Groups** icon located above the Real Servers table.

The Groups menu appears below the icon (see Figure 7-1).

Step 3 From the Groups menu, click **X** (delete) next to the group that you want to delete.

The Delete Group confirmation popup window appears.

Step 4 From the Delete Group confirmation popup window, do one of the following:

- Click **Delete** to removes the real server group.
- Click **Cancel** to ignore the deletion request.

Related Topics

- Managing Real Server Groups, page 7-10
- Creating a Real Server Group, page 7-11
- Editing or Copying a Real Server Group, page 7-12
- Displaying a Real Server Group, page 7-13

Activating Real Servers

You can activate a real server.

Note If you are using the ANM plug-in for vCenter Server to access ANM, see the "Activating Real Servers Using vSphere Client" section on page B-15.

Procedure

Step 1 Choose **Config > Operations > Real Servers**.

The Real Servers table appears.

Step 2 (Optional) To display only the real servers of a specific real server group, do the following:

a. Click the **Groups** icon located above the Real Servers table. The Groups menu appears below the icon (see Figure 7-1).

b. From the Groups menu, choose the group to display.

Step 3 From the Real Servers table, choose the servers that you want to activate, and click **Activate**.

The Activate Server window appears.

Step 4 In the Reason field of the Activate Server window, enter a reason for this action.

You might enter a trouble ticket, an order ticket, or a user message.

Note Do not enter a password in this field.

Step 5 Do one of the following:

- Click **OK** to activate the server and to return to the Real Servers table. The server appears in the table with the status Inservice.
- Click **Cancel** to exit this procedure without activating the server and to return to the Real Servers table.

Related Topics

- Managing Real Servers, page 7-9
- Managing Real Server Groups, page 7-10
- Suspending Real Servers, page 7-15
- Displaying Real Servers, page 7-18
- Using the Real Server Connection Statistics Graph, page 7-22
- Using the Real Server Topology Map, page 7-23

Suspending Real Servers

You can suspend a real server.

Note If you are using the ANM plug-in for vCenter Server to access ANM, see the "Suspending Real Servers Using vSphere Client" section on page B-16.

Procedure

Step 1 Choose **Config > Operations > Real Servers**.

The Real Servers table appears.

Step 2 (Optional) To display only the real servers of a specific real server group, do the following:

a. Click the **Groups** icon located above the Real Servers table. The Groups menu appears below the icon (see Figure 7-1).

b. From the Groups menu, choose the group to display.

Step 3 In the Real Servers table, choose the server that you want to suspend, and click **Suspend**.

The Suspend Real Servers window appears.

Step 4 In the Reason field of the Suspend Real Servers window, enter the reason for this action.

You might enter a trouble ticket, an order ticket, or a user message.

Note Do not enter a password in this field.

Step 5 From the Suspend Real Servers Type drop-down list, choose one of the following:

- **Graceful**—When executed on a primary server, the ACE gracefully shuts down the server with sticky connections as follows:
  - Tears down existing non-TCP connections to the server
  - Allows current TCP connections to complete
  - Allows new sticky connections for existing server connections that match entries in the sticky database
  - Load balances all new connections (other than the matching sticky connections mentioned above) to the other servers in the server farm

  When executed on a backup real server, the ACE places the backup server in service standby mode.

Note For the CSS, when the device is in the In Service admin state and you perform a graceful suspend operation, ANM saves the last known non-zero service (or real server) weight, and then sets the weight to zero. ANM references the saved weight when performing an Activate operation. If the current weight is zero, and a non-zero weight has been saved for that service (or real server), the Activate operation also sets the weight to the saved value.

To allow ANM to save and reset the weight value when gracefully suspending and then activating the CSS, you must have the device configured to permit SNMP traffic. For each device type, see the corresponding configuration guide to configure the device to permit SNMP traffic.

When the CSS is in the In Service Standby admin state and you perform a graceful suspend operation, ANM does not set the weight to zero.

Note Graceful suspend and suspend options vary by device type. For the commands deployed by the device type when these options are selected, see the "CLI Commands Sent from the Real Server Table" section on page 7-24.

- **Suspend**—The ACE resets all non-TCP connections to the server. For TCP connections, existing flows are allowed to complete before the ACE takes the real server out of service. No new connections are allowed. The ACE resets all Secure Sockets Layer (SSL) connections to the real server.
- **Suspend and Clear Connections**—Performs the tasks described for Suspend and clears the existing connections to this server.

Step 6 Do one of the following:

- 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.
- Click **Cancel** to exit this procedure without suspending the server and to return to the Real Servers table.

Related Topics

- Managing Real Servers, page 7-9
- Managing Real Server Groups, page 7-10
- Activating Real Servers, page 7-14
- Displaying Real Servers, page 7-18
- Using the Real Server Connection Statistics Graph, page 7-22
- Using the Real Server Topology Map, page 7-23

Modifying Real Server Weight Value

You can modify the weight value assigned to a real server that defines the connection capacity of the server in relation to the other real servers. The ACE uses the weight value that you specify for a server in the weighted round-robin and least-connections load-balancing predictors. Servers with a higher configured weight value have a higher priority with respect to connections than servers with a lower weight. For example, a server with a weight of 5 would receive five connections for every one connection for a server with a weight of 1.

Note If you are using the ANM plug-in for vCenter Server to access ANM, see the "Modifying Real Server Weight Value Using vSphere Client" section on page B-18.

Procedure

Step 1 Choose **Config > Operations > Real Servers**.

The Real Servers table appears.

Step 2 (Optional) To display only the real servers of a specific real server group, do the following:

a. Click the **Groups** icon located above the Real Servers table. The Groups menu appears below the icon (see Figure 7-1).

b. From the Groups menu, choose the group to display.

Step 3 In the Real Servers table, choose the servers whose configuration you want to modify, and click **Change Weight** below the table to the right of Activate and Suspend.

The Change Weight Real Servers window appears.

Step 4 In the Change Weight Real Servers window, enter the following information for the selected server:

- Reason for change such as trouble ticket, order ticket or user message.

Note Do not enter a password in this field.

- Weight value (for allowable ranges for each device type, see Table 7-5).

Step 5 Do one of the following:

- 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.
- Click **Cancel** to exit this procedure without saving your entries and to return to the Real Servers table.

Related Topics

- Managing Real Servers, page 7-9
- Managing Real Server Groups, page 7-10
- Activating Real Servers, page 7-14
- Displaying Real Servers, page 7-18
- Using the Real Server Connection Statistics Graph, page 7-22
- Using the Real Server Topology Map, page 7-23

Displaying Real Servers

You can display the list of real servers configured on ANM with information specific to each server.

Procedure

Step 1 Choose **Config** > **Operations** > **Real Servers**.

The Real Servers table appears, which contains the information described in Table 7-2.

Note In the table, N/A indicates that either the information is not available from the database or that it is not being collected using SNMP.

Table 7-2 Real Server Table Fields

Item	Description
Name	Real server name. For CSM real servers only, if you have the reverse DNS lookup feature enabled, ANM displays the DNS name of the CSM real server in this field. ANM learns and updates the DNS names during the following operations: • CSM import • CSM CLI synchronization • ANM restart By default, the reverse DNS lookup feature is disabled. You can enable it by modifying the ANM properties file and restarting ANM as follows: **a. echo "cisco.anm.enable-csm-dns-lookup=true" >> /opt/CSCOanm/etc/cs-config.properties** **b. /opt/CSCOanm/bin/anm-tool restart**
IP address	Real server IP address.
Port	Port used by the real server for communications.
VM	Virtual machine indicator that specifies if the real server is a VMware vCenter Server virtual machine (Yes) or is not a virtual machine (–). If the indicator state is Yes, you can click this link to open the Virtual Machine Details popup window to display statistical information about the VM. ANM polls the VM on a regular basis to update the displayed information. Click **OK** to close the popup window and return to the Real Servers table.
Vservers	Associated virtual servers.
HA	Indicators that display when the real server is part of a high availability pair. The indicators are as follows: • Asterisk (*)—The real server is associated with an HA pair and the HA configuration is complete. • Red dash (-)—The real server is associated with an HA pair; however, the HA configuration is incomplete. Typically, the HA pair are not properly configured for HA or only one of the devices has been imported into ANM. Ensure that both devices are imported into ANM and that they are configured as described in the "Configuring ACE High Availability" section on page 12-14. The table displays HA pair real servers together in the same row and they remain together no matter how you sort the information.
SLB Device	Name of the server load-balancing device.
Admin	Administrative state of the real server: In Service, Out Of Service, or In Service Standby.

Table 7-2 Real Server Table Fields (continued)

Item	Description
Oper	Operational state of the real server. Possible states are as follows: • Failed—Server has failed and is not retried for the amount of time specified by its retry timer. • Inband probe failed—Server has failed the inband Health Probe agent. • Inservice—Server is in use as a destination for server load-balancing client connections. • Inservice standby—Server is the backup real server, which remains inactive unless the primary real server fails. • Operation wait—Server is ready to become operational but is waiting for the associated redirect virtual server to be in service. • Out of service—Server is not in use by a server load balancer as a destination for client connections. • Probe failed—Server load-balancing probe to this server has failed. No new connections are assigned to this server until a probe to this server succeeds. • Probe testing—Server has received a test probe from the server load balancer. • Ready to test —Server has failed and its retry timer has expired; test connections will begin flowing to it soon. • Return code failed—Server has been disabled because it returned an HTTP code that matched a configured value. • Test wait—Server is ready to be tested. This state is applicable only when the server is used for HTTP redirect load balancing. • Testing—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 are assigned to the server until DFP raises its weight. • Throttle: max clients—Server has reached its maximum number of allowed clients. • Throttle: max connections —Server has reached its maximum number of connections and is no longer being given connections. • Unknown—State of the server is not known.  **Note** By default, the Details popup window feature is disabled. Click the value in this column to view the **show sticky database rserver** *name* **serverfarm** *name* **detail** command output in the Details popup window. This output is displayed only for ACE software version A5(1.0) or later. If you have the Details popup window feature enabled, click the value in this column to view both the **show rserver** *name* **detail** command and the **show sticky database rserver** *name* **serverfarm** *name* **detail** command output in the Details popup window. For information about enabling or disabling this feature, see the "Enabling the ACE Real Server Details Popup Window Option" section on page 17-64.
Conn	Number of current connections.
Wt	Current server weight.

Table 7-2 Real Server Table Fields (continued)

Item	Description
Locality	Item that pertains only to ACE software Version A4(2.0) or later releases on either device type (appliance or module). Locality also requires that you have the ACE configured for Dynamic Workload Scaling (see the "Configuring Dynamic Workload Scaling" section on page 7-27). Location of the real server, which must be a VM and not a physical server. Possible locality states are as follows: • N/A—Not available; the ACE cannot determine if the real server is local or remote. A possible cause for this issue is that Dynamic Workload Scaling is not configured correctly. • Local—The real server is located in the local network. • Remote—The real server is located in the remote network. The ACE bursts traffic to this server when the CPU and/or memory usage of the local real servers reaches the specified maximum threshold value.
Stat Age	Age of the statistical information.
Server Farm	Associated server farm.

Step 2 (Optional) To display only the real servers of a specific real server group, do the following:

a. Click the **Groups** icon located above the Real Servers table. The Groups menu appears below the icon (see Figure 7-1).

b. From the Groups menu, choose the group to display.

Step 3 (Optional) Use the function buttons located at the bottom of the window to activate or suspend a real server, change the weight assigned to a real server, and so forth. Table 7-3 describes the check box and function button options.

Table 7-3 Real Server Window Check Box and Function Button Options

Check Box/Function Button	Description
Poll Now	Function button that updates the displayed information. **Note** Even if the periodic polling is enabled, ANM polls all the devices thus ignoring the statistics defined during the periodic polling.
Activate	Function button that activates a suspended real server (see the "Activating Real Servers" section on page 7-14).
Suspend	Function button that suspends an active real server (see the "Suspending Real Servers" section on page 7-15).
Change Weight	Function button used to change the weight assigned to a real server (see the "Server Weight Ranges" section on page 7-26).
Graph	Function button that displays the statistics graph for a selected real server (see the "Using the Real Server Connection Statistics Graph" section on page 7-22).
Topology	Function button that displays the topology map for a selected real server (see the "Using the Real Server Topology Map" section on page 7-23).

Step 4 (Optional) To identify any SNMP-related issues, select the real server's virtual context in the object selector. If there are problems with SNMP, the SNMP status appears in the upper right above the content pane.

Related Topics

- Displaying Real Server Statistics and Status Information, page 7-9
- Using the Real Server Connection Statistics Graph, page 7-22
- Managing Real Server Groups, page 7-10
- Using the Real Server Topology Map, page 7-23
- Activating Real Servers, page 7-14
- Suspending Real Servers, page 7-15
- Modifying Real Server Weight Value, page 7-17
- Enabling the ACE Real Server Details Popup Window Option, page 17-64
- Filtering Entries, page 1-15

Using the Real Server Connection Statistics Graph

You can display real time and historical statistical information about the connections of a real server. ANM displays the information in graph or chart form. This feature also allows you to compare similar connection information across multiple real servers.

Procedure

Step 1 Choose **Config > Operations > Real Servers**.

The Real Servers table appears.

Step 2 (Optional) To display only the real servers of a specific real server group, do the following:

a. Click the **Groups** icon located above the Real Servers table. The Groups menu appears below the icon (see Figure 7-1).

b. From the Groups menu, choose the group to display.

Step 3 In the Real Servers table, check the check box next to server whose connection information you want to display, and click **Graph**.

You can choose up to four real servers if you want to compare statistical data.

The Real Server Graph window appears, displaying the default graph for each selected real server. For details about using the graph feature, see the "Configuring Historical Trend and Real Time Graphs for Devices" section on page 16-52.

Related Topics

- Managing Real Server Groups, page 7-10
- Activating Real Servers, page 7-14
- Suspending Real Servers, page 7-15

- Modifying Real Server Weight Value, page 7-17
- Displaying Real Servers, page 7-18
- Using the Real Server Topology Map, page 7-23

Using the Real Server Topology Map

You can display the nodes on your network based on the real server that you select.

Procedure

Step 1 Choose **Config > Operations > Real Servers**.

The Real Servers table appears.

Step 2 (Optional) To display only the real servers of a specific real server group, do the following:

a. Click the **Groups** icon located above the Real Servers table. The Groups menu appears below the icon (see Figure 7-1).

b. From the Groups menu, choose the group to display.

Step 3 In the Real Servers table, choose the server whose topology map you want to display, and click **Topology**.

The ANM Topology map appears. The map includes several tools for navigating the network map and zooming in and out. For details about using the map tools, see the "Displaying Network Topology Maps" section on page 16-72.

Step 4 Click **Exit** to return to the Real Server widow.

Related Topics

- Managing Real Server Groups, page 7-10
- Activating Real Servers, page 7-14
- Suspending Real Servers, page 7-15
- Modifying Real Server Weight Value, page 7-17
- Displaying Real Servers, page 7-18
- Using the Real Server Connection Statistics Graph, page 7-22

CLI Commands Sent from the Real Server Table

Table 7-4 displays the CLI commands dispatched to the device for a given Real Servers table option and is sorted by device type.

Table 7-4 CLI Commands Deployed from the Real Servers Table

Command	Sample CLI Sent
ACE Modules and Appliances	

Real Server Activation	`serverfarm host sf1`
	`rserver rs1 80`
	`inservice`

Real Server Graceful Suspend	`serverfarm host sf1`
	`rserver rs1 80`
	`inservice standby`

Real Server Suspend	`serverfarm host sf1`
	`rserver rs1 80`
	`no inservice`

Real Server Suspend and Clear Connections	`serverfarm host sf1`
	`rserver rs1 80`
	`no inservice`
	`clear conn rserver rs1 80 serverfarm sf1`

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`
	`inservice standby`

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`

Table 7-4 CLI Commands Deployed from the Real Servers Table (continued)

Command	Sample CLI Sent

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`
	`inservice standby`

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`

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 7-5 displays the allowable server weight ranges by device type.

Table 7-5 Real Servers Table Server Weight Ranges

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

Configuring Dynamic Workload Scaling

Note Dynamic Workload Scaling requires ACE software Version A4(2.0) or later release on either device type (appliance or module).

This section describes how to configure the ACE Dynamic Workload Scaling (DWS) feature, which enables an ACE to burst traffic to a remote pool of VMs when the average CPU and/or memory usage of the local VMs has reached a specified maximum threshold value. When the usage drops below a specified minimum threshold value, the ACE stops bursting traffic to the remote VMs.

Note To enable the ACE to use the VMs associated with DWS for load balancing, you must configure them as real servers on the ACE (see the "Configuring Real Servers" section on page 7-5).

For more information about DWS, see the "ANM Overview" section on page 1-1 and the "Dynamic Workload Scaling Overview" section on page 7-4.

Prerequisites

DWS requires the following configuration elements:

- An ACE with software Version A4(2.0) or later and configured with the following items:
  - Nexus 7000 Series switch—XML interface IP address of the local Cisco Nexus 7000 Series switch that the ACE polls to obtain VM location information (local or remote). You can define up to two switch profiles per Admin context depending on the ACE software version (see the "Guidelines and Restrictions" section on page 7-28 for this topic). For information about defining a switch profile, see the "Configuring and Verifying a Cisco Nexus 7000 Series Switch Connection" section on page 7-28.

    Note The Nexus 7000 Series switch must be configured for DCI/OTV in the local data center and in the remote data center. For details about configuring a Nexus 7000 for DCI/OTV, see the *Cisco Nexus 7000 NX-OS OTV Configuration Guide, Release 5.x*.

  - VM Controller—IP address of the VM Controller (also known as VMware vCenter Server) that the ACE sends a health probe to monitor usage of the local VMs associated with a server farm.
  - VM probe—Probe that the ACE sends to the VM Controller to monitor local VM usage based on CPU usage, memory usage, or both (see the "Configuring Health Monitoring" section on page 7-51).
  - Server Farms—Groups of networked real servers (physical servers and VMs) that provide content delivery (see the "Configuring Server Farms" section on page 7-31).V
- VMware vCenter Server 4.0 or later.
- Multiple local and remote VMs configured as real servers and associated with server farms configured on the ACE.
- ACE backend interface MTU set to 1430 or less to accommodate DCI encapsulation and the Don't Fragment (DF) bit is automatically set on the DCI link. For details about setting the ACE MTU, see the *Cisco 4700 Series Application Control Engine Appliance Routing and Bridging Configuration Guide*.

This section includes the following topics:

Configuring and Verifying a Cisco Nexus 7000 Series Switch Connection

Note This feature requires ACE software Version A4(2.0) or later release on either device type (appliance or module).

You can configure an ACE with the Cisco Nexus 7000 Series switch attributes required to allow the ACE to communicate with the switch using SSH. When configured for DWS, the ACE uses the Nexus 7000 Series switch to obtain VM location information (local or remote).

You can also use this procedure to edit the attributes of an existing Nexus 7000 Series switch profile or remove a switch profile.

Guidelines and Restrictions

The number of Nexus 7000 Series switch profiles that you can define per ACE Admin context is as follows:

- ACE software Version A4(2.0) to A5(1.1)—One switch profile only.
- ACE software Version A5(1.2) or later—Up to two switch profiles.

Procedure

Step 1 Choose **Config** > **Devices** > **Admin_context** > **Load Balancing** > **Dynamic Workload Scaling** > **Nexus 7000 Setup**.

The Nexus 7000 Setup pane appears.

Note If existing Nexus 7000 Series switch profiles already exist, the Name field lists their profile names in drop-down list on the right. Multiple switch profiles requires ACE software Version A5(1.2) or later.

Step 2 From the Nexus 7000 Setup pane, do one of the following:

- To define a new Nexus 7000 series switch profile, do the following:

  a. From the Name field, click the text box radio button if it is not already selected and enter a Nexus 7000 name with a maximum of 64 characters. See the Note at the beginning of this chapter for ACE object naming specifications.

  b. From the Primary IP filed, enter the Nexus 7000 XML interface IP address in dotted-decimal format (such as 192.168.11.1).

  c. From the User Name field, enter the username that the ACE uses for access and authentication on the Nexus 7000 Series switch. Valid entries are unquoted text strings with a maximum of 64 characters with no spaces.

Note The user must have either the vdc-admin or network-admin role to receive the Nexus 7000 Series switch output for the VM location information in XML format.

d. From the Password field, enter the password that the ACE uses for authentication on the Nexus 7000 Series switch. Valid entries are unquoted text strings with a maximum of 64 characters with no spaces.

e. From the Confirm field, reenter the password and go to Step 3.

- To edit an existing Nexus 7000 Series switch profile, do the following:

  a. From the Name field, click the radio button for the drop down list that contains the list of existing switch profile names.

  b. From the drop down list, choose the switch profile to edit. The current profile attributes display.

  c. Edit the profile fields as described in the procedure above for creating a new profile and go to Step 3.

Step 3 Click **Deploy Now** to deploy this configuration on the ACE and save your entries to the running-configuration and startup-configuration files.

Note Configuring the ACE for DWS also requires configuring the ACE with the VM Controller information (see the "Configuring and Verifying a VM Controller Connection" section on page 7-30) and configuring a VM health probe (see the "Configuring Health Monitoring" section on page 7-51).

Step 4 (Optional) Click **Details** to verify connectivity between the ACE and the Nexus 7000 Series switch.

The ACE **show nexus-device** *device_name* **detail** CLI command output displays in a popup window and includes information such as the device name, IP address, and connection information. For more information about the command output, see the *Cisco 4700 Series Application Control Engine Appliance Server Load-Balancing Configuration Guide*.

Step 5 (Optional) Click **Delete** to delete the currently configured Cisco Nexus 7000 series switch.

Caution If the ACE is currently configured for DWS, deleting the Nexus 7000 Series switch disables the feature.

Related Topics

- Configuring and Verifying a VM Controller Connection, page 7-30
- Configuring Health Monitoring, page 7-51
- Configuring Dynamic Workload Scaling, page 7-27
- Dynamic Workload Scaling Overview, page 7-4
- Configuring Real Servers, page 7-5
- Configuring Load Balancing Using Server Farms, page 7-32

Configuring and Verifying a VM Controller Connection

Note This feature requires ACE software Version A4(2.0) or later release on either device type (appliance or module).

You can configure an ACE with the VM Controller (VMware vCenter Server) attributes required to allow the ACE to communicate with the VM Controller to obtain local VM load information.

Guidelines and Restrictions

Configure only one VM Controller per ACE Admin context.

Prerequisites

The ACE is configured to communicate with the local Cisco Nexus 7000 Series switch that enables the ACE to discover the locality of the VM Controller VMs (see the "Configuring and Verifying a Cisco Nexus 7000 Series Switch Connection" section on page 7-28).

Procedure

Step 1 Choose **Config** > **Devices** > **Admin_context** > **Load Balancing** > **Dynamic Workload Scaling** > **VM Controller Setup**.

The VM Controller Setup pane appears.

Step 2 From the VM Controller Setup pane, define the VM Controller using the information in Table 7-6.

Table 7-6 VM Controller Setup

Field	Description
Name	VM Controller name (see the Note at the beginning of this chapter for ACE object naming specifications).
URL	IP address or URL for the VM Controller web services API agent. The URL must point to the VM Controller software development kit (SDK). For example, https://1.2.3.4/sdk. Enter up to 255 characters.
User Name	Username that the ACE uses for access and authentication on the VM Controller. The user must have a read-only role at least or a role with a read privilege. Valid entries are unquoted text strings with a maximum of 64 characters and no spaces.
Password	Password that the ACE uses for authentication on the VM Controller. Valid entries are unquoted text strings with a maximum of 64 characters and no spaces. Reenter the password in the Confirm field.

Step 3 Click **Deploy Now** to deploy this configuration on the ACE and save your entries to the running-configuration and startup-configuration files.

Note Configuring the ACE for Dynamic Workload Scaling also requires configuring the ACE with the Nexus 7000 information (see the "Configuring and Verifying a Cisco Nexus 7000 Series Switch Connection" section on page 7-28) and configuring a VM health probe (see the "Configuring Health Monitoring" section on page 7-51).

Step 4 (Optional) Click **Details** to verify connectivity between the ACE and the remote VM Controller.

The ACE **show vm-controller** *device_name* **detail** CLI command output displays in a popup window and includes information such as the VM Controller status, IP address, and connection information.

Step 5 (Optional) Click **Delete** to delete the currently configured VM Controller.

Note If the ACE is currently configured for Dynamic Workload Scaling, you must delete the associated VM health probe before you can delete the VM controller (see the "Configuring Health Monitoring" section on page 7-51).

Related Topics

- Configuring and Verifying a Cisco Nexus 7000 Series Switch Connection, page 7-28
- Configuring Health Monitoring, page 7-51
- Configuring Dynamic Workload Scaling, page 7-27
- Dynamic Workload Scaling Overview, page 7-4
- Configuring Real Servers, page 7-5
- Configuring Load Balancing Using Server Farms, page 7-32

Configuring Server Farms

You can configure load balancing using server farms, which are groups of networked real servers (physical servers and VMs) that contain the same content and that typically reside in the same physical location in a data center.

Websites often include 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.

Guidelines and Restrictions

- With Dynamic Workload Scaling configured on the ACE, the real servers that are VMs can also reside in a remote datacenter (see the "Configuring Dynamic Workload Scaling" section on page 7-27).
- A server farm can support a mix of IPv6 and IPv4 real servers, and can be associated with both IPv6 and IPv4 probes. IPv6 requires ACE module and ACE appliance software Version A5(1.0) or later.

This section includes the following topics:

- Configuring Load Balancing Using Server Farms, page 7-32
- Adding Real Servers to a Server Farm, page 7-38
- Configuring the Predictor Method for Server Farms, page 7-40
- Configuring Server Farm HTTP Return Error-Code Checking, page 7-47
- Displaying All Server Farms, page 7-49
- Displaying Server Farm Statistics and Status Information, page 7-50

Configuring Load Balancing Using Server Farms

Procedure

Step 1 Choose **Config > Devices >** *context* **> Load Balancing > Server Farms**.

The Server Farms table appears.

Step 2 In the Server Farms table, click **Poll Now** to instruct ANM to poll the devices and display the current values, and click **OK** when prompted if you want to poll the devices for data now.

Step 3 Click **Add** to add a new server farm, or choose an existing server farm and click **Edit**.

The Server Farms configuration window appears.

Step 4 In the Server Farms configuration window, configure the server farm using the information in Table 7-7.

Table 7-7 Server Farm Attributes

Field	Description
Name	Unique name for this server farm or accept the automatically incremented value in this field. Valid entries are unquoted text strings with no spaces and a maximum of 64 characters.
Type	Type of server farm as follows: • **Host**—Server farm consists of real servers that provide content and services to clients. • **Redirect**—Server farm consists only of real servers that redirect client requests to alternate locations specified in the real server configuration. (See the "Configuring Real Servers" section on page 7-5.)
Description	Brief description for this server farm. Valid entries are unquoted alphanumeric text strings with no spaces and a maximum of 240 characters.
Fail Action	Action that the ACE is to take with respect to connections if any real server in the server farm fails: • **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. • **Reassign**—The ACE is to reassign the existing server connections to the backup real server (if configured) if the real server fails after you enter this command. If a backup real server has not been configured for the failing server, this selection leaves the existing connections untouched in the failing real server.

Table 7-7 Server Farm Attributes (continued)

Field	Description
Failaction Reassign Across Vlans	Option that is available only for the ACE module A2(3.0), ACE appliance A4(1.0), and later releases of either device type. This field appears only when the Fail Action is set to Reassign. Check the check box to specify that the ACE reassigns the existing server connections to the backup real server on a different VLAN interface (commonly referred to as a bypass VLAN) if the real server fails. If a backup real server has not been configured for the failing server, this option has no effect and leaves the existing connections untouched in the failing real server. Note the following configuration requirements and restrictions when you enable this option: • Enable the Transparent option (see the next Field) to instruct the ACE not to use NAT to translate the ACE VIP address to the server IP address. The Failaction Reassign Across Vlans option is intended for use in stateful firewall load balancing (FWLB) on your ACE, where the destination IP address for the connection coming in to the ACE is for the end-point real server, and the ACE reassigns the connection so that it is transmitted through a different next hop. • Enable the MAC Sticky option on all server-side interfaces to ensure that packets that are going to and coming from the same server in a flow will traverse the same firewalls or stateful devices (see the "Configuring Virtual Context VLAN Interfaces" section on page 11-6). • Configure the Predictor Hash Address option after you add the serverfarm (see the "Configuring the Predictor Method for Server Farms" section on page 7-40). • You must configure identical policies on the primary interface and the backup-server interface. The backup interface must have the same feature configurations as the primary interface. • If you configure a policy on the backup-server interface that is different from the policies on the primary-server interface, that policy will be effective only for new connections. The reassigned connection will always have only the primary-server interface policies. • Interface-specific features (for example, NAT, application protocol inspection, outbound ACLs, or SYN cookie) are not supported. • You cannot reassign connections to the failed real server after it comes back up. This restriction also applies to same-VLAN backup servers. • Real servers must be directly connected to the ACE. This requirement also applies to same-VLAN backup server. • You must disable sequence number randomization on the firewall (see the "Configuring Connection Parameter Maps" section on page 9-3). • Probe configurations should be similar on both ACEs and the interval values should be low. For example, if you configure a high interval value on ACE-1 and a low interval value on ACE-2, the reassigned connections may become stuck because of the probe configuration mismatch. ACE-2 with the low interval value will detect the primary server failure first and will reassign all its incoming connections to the backup-server interface VLAN. ACE-1 with the high interval value may not detect the failure before the primary server comes back up and will still point to the primary server. To minimize packet loss, we recommend the following probe parameter values on both ACEs: Interval: 2, Faildetect: 2, Passdetect interval: 2, and Passdetect count: 5.

Table 7-7 Server Farm Attributes (continued)

Field	Description
Transparent	Field that appears only for host server farms. Specify whether network address translation from the VIP address to the server IP is to occur. Check the check box to indicate that network address translation from the VIP address to the server IP address is to occur. Uncheck the check box to indicate that network address translation from the VIP address to the server IP address is not to occur.
Dynamic Workload Scaling	Option that is available only for ACE software Version A4(2.0) or later release on either device type (appliance or module). Field that appears only for host server farms. Allows the ACE to burst traffic to remote VMs when the average CPU or memory usage of the local VMs has reached its specified maximum threshold value. The ACE stops bursting traffic to the remote VMs when the average CPU or memory usage of the local VMs has dropped below its specified minimum threshold value. This option requires that you have the ACE configured for Dynamic Workload Scaling using a Nexus 7000, VM Controller, and VM probe (see the "Configuring Dynamic Workload Scaling" section on page 7-27). Click one of the following radio button options: • N/A—Not applicable (default). • Local—Restricts the ACE to use of local VMs only for server load balancing. • Burst—Enables the ACE to burst traffic to remote VMs when needed. When you choose Burst, the VM Probe Name field displays along with a list of available VM probes. Choose an available VM probe or click **Add** to display the Health Monitoring popup window and create or edit a VM probe (see the "Configuring Health Monitoring" section on page 7-51).
Fail-On-All	Field that appears only for host server farms. By default, real servers that you configure in a server farm inherit the probes that you configure directly on that server farm. When you configure multiple probes on a server farm, the real servers in the server farm use an OR logic with respect to the probes, which means that if one of the probes configured on the server farm fails, all the real servers in that server farm fail and enter the PROBE-FAILED state. With AND logic, if one server farm probe fails, the real servers in the server farm remain in the operational state. If all the probes associated with the server farm fail, then all the real servers in that server farm fail and enter the PROBE-FAILED state. Check this check box to configure the real servers in a server farm to use AND logic with respect to multiple server farm probes. The Fail-On-All function is applicable to all probe types.

Table 7-7 Server Farm Attributes (continued)

Field	Description
Inband-Health Check	Option that is available only for the ACE module A4(1.0), ACE appliance A4(1.0), and later releases of either device type. Field that appears only for host server farms. By default, the ACE monitors the health of all real servers in a configuration through the use of ARPs and health probes. However, there is latency period between when the real server goes down and when the ACE becomes aware of the state. The inband health monitoring feature allows the ACE to monitor the health of the real servers in the server farm through the following connection failures: • For TCP, resets (RSTs) from the server or SYN timeouts. • For UDP, ICMP Host, Network, Port, Protocol, and Source Route unreachable messages. When you configure the failure-count threshold and the number of these failures exceeds the threshold within the reset-time interval, the ACE immediately marks the server as failed, takes it out of service, and removes it from load balancing. The server is not considered for load balancing until the optional resume-service interval expires. The Inband-Health Check attributes are as follows: • Count—Tracks the total number of TCP or UDP failures, and increments the counters. • Log—Logs a syslog error message when the number of events reaches the threshold value that you set for the Connection Failure Threshold Count attribute. • Remove—Logs a syslog error message when the number of events reaches the configured threshold and removes the real server from service.
Connection Failure Threshold Count	This field appears only when the Inband-Health Check is set to Log or Remove. Enter the maximum number of connection failures that a real server can exhibit in the reset-time interval before ACE marks the real server as failed. Valid entries are as follows: • ACE appliance—1 to 4294967295 • ACE module—4 to 4294967295
Reset Timeout (Milliseconds)	This field appears only when the Inband-Health Check is set to Log or Remove. Enter the number of milliseconds for the reset-time interval. Valid entries are integers from 100 to 300000. The default interval is 100. This interval starts when the ACE detects a connection failure. If the connection failure threshold is reached during this interval, the ACE generates a syslog message. If you configure the **remove** keyword, the ACE also removes the real server from service. Changing the setting of this option affects the behavior of the real server, as follows: • When the real server is in the OPERATIONAL state, even if several connection failures have occurred, the new reset-time interval takes effect the next time that a connection error occurs. • When the real server in the INBAND-HM-FAILED state, the new reset-time interval takes effect the next time that a connection error occurs after the server transitions to the OPERATIONAL state.

Table 7-7 Server Farm Attributes (continued)

Field	Description
Resume Service (Seconds)	Field that appears only when the Inband-Health Check is set to Remove. Enter the number of seconds after a server has been marked as failed to reconsider it for sending live connections. Valid entries are integers from 30 to 3600. The default setting is 0. The setting of this option affects the behavior of the real server in the inband failed state, as follows: • When this field is not configured and has the default setting of 0, the real server remains in the failed state until you manually suspend and then reactivate it. • When this field is not configured and has the default setting of 0 and then you configure this option with an integer between 30 and 3,600, the failed real server immediately transitions to the Operational state. • When you configure this field and then increase the value, the real server remains in the failed state for the duration of the previously-configured value. The new value takes effect the next time the real server transitions to the failed state. • When you configure this field and then decrease the value, the failed real server immediately transitions to the Operational state. • When you configure this field with an integer between 30 and 3,600 and then reset it to the default of 0, the real server remains in the failed state for the duration of the previously-configured value. The default setting takes effect the next time the real server transitions to the failed state. Then the real server remains in the failed state until you manually suspend and then reactivate it. • When you change this field within the reset-time interval the real server in the OPERATIONAL with several connection failures, the new threshold interval takes effect the next time that a connection error occurs, even if it occurs within the current reset-time interval.
Partial-Threshold Percentage	Field that 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 from 0 to 99. The default is 0.
Back Inservice	Field that 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 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.

Table 7-7 Server Farm Attributes (continued)

Field	Description
Probes	Field that appears only as follows: • For all host server farms. The Available probe list contains all probe types. • For redirect server farms configured on ACE devices that use the following software versions: – ACE module: A2(3.x) and later releases – ACE appliance: A3(x) and later releases The redirect server farm Available probe list contains only probes of the type Is Routed, which means that the ACE routes the probe address according to the ACE internal routing table (see the "Configuring Health Monitoring for Real Servers" section on page 7-52). In the Available Items list, choose the probes to use for health monitoring, and click **Add**. The selected probes appear in the Selected Items list. **Note** You can associate both IPv6 and IPv4 probes to a server farm. IPv6 requires ACE module and ACE appliance software Version A5(1.0) or later. **Note** The list of available probes does not include VM health monitoring probes. To choose a VM probe for monitoring local VM usage, see the Dynamic Workload Scaling field. To remove probes that you do not want to use for health monitoring, select them in the Selected Items list, and click **Remove**. The selected probes appear in the Available Items list.

Step 5 Do one of the following:

- Click **Deploy Now** to deploy this configuration on the ACE and save your entries to the running-configuration and startup-configuration files.

  The window refreshes with additional configuration options:

  – To add real servers to the server farm, see the "Adding Real Servers to a Server Farm" section on page 7-38.

  – To specify a predictor method for the server farm, see the "Configuring the Predictor Method for Server Farms" section on page 7-40.

  – To configure return code checking, see the "Configuring Server Farm HTTP Return Error-Code Checking" section on page 7-47.

- Click **Cancel** to exit the procedure without saving your entries and to return to the Server Farms table.
- Click **Next** to deploy your entries and to configure another server farm.

Step 6 (Optional) To display statistics and status information for an existing server farm, choose a server farm from the Server Farms table, and click **Details**.

The **show serverfarm** *name* **detail** CLI command output appears. See the "Displaying Server Farm Statistics and Status Information" section on page 7-50 for details.

Related Topics

- Configuring Health Monitoring for Real Servers, page 7-52
- Configuring Real Servers, page 7-5
- Configuring Sticky Groups, page 8-12
- Configuring the Predictor Method for Server Farms, page 7-40
- Configuring Server Farm HTTP Return Error-Code Checking, page 7-47
- Configuring Dynamic Workload Scaling, page 7-27

Adding Real Servers to a Server Farm

You can add real servers to a server farm. After adding a server farm (see the "Configuring Server Farms" section on page 7-31), 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.

Assumptions

This topic assumes the following:

- A server farm has been added to ANM (see the "Configuring Server Farms" section on page 7-31).
- At least one real server exists.

Consideration

A server farm can support a mix of IPv6 and IPv4 real servers. IPv6 requires ACE module and ACE appliance software Version A5(1.0) or later.

Procedure

Step 1 Choose **Config > Devices >** *context* **> Load Balancing > Server Farms**.

The Server Farms table appears.

Step 2 In the Server Farms table, choose the server farm that you want to associate with real servers.

The Real Servers table appears.

Step 3 In the Real Servers table, click **Add** to add a new entry, or select an existing server and click **Edit** to modify it.

The Real Servers configuration pane appears.

Step 4 In the Real Servers configuration pane, configure the real server using the information in Table 7-8.

Table 7-8 Real Server Configuration Attributes

Field	Description
Name	Server that you want to associate with the server farm.
Port	Port number to be used for server port address translation (PAT). Valid entries are from 1 to 65535.
Backup Server Name	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	Server port number. If you select a backup server, enter the backup server port number. Valid entries are from 1 to 65535.

Table 7-8 Real Server Configuration Attributes (continued)

Field	Description
Fail-On-All	Field that appears only for real servers identified as host servers. By default, real servers with multiple probes configured for them have an OR logic associated with them. This means that if one of the real server probes fails, the real server fails and enters the PROBE-FAILED state. Check this checkbox to configure a real server to remain in the OPERATIONAL state unless all probes associated with it fail (AND logic). The Fail-On-All function is applicable to all probe types.
State	State of this server as follows: • **In Service**—The server is in 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. • **Out Of Service**—The server is out of service.
Buddy Real Group Name	Create a buddy real server group or select an existing one to enable persistence to the same real server or group of real servers across multiple server farms (for more information, see the "Buddy Sticky Groups" section on page 8-6). **Note** This field appears only for ACE software Version A5(2.0) or later.
Min. Connections	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 appliances, valid entries are from 2 to 4294967295. For ACE modules, valid entries are from 2 to 4000000.
Max. Connections	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 appliances, valid entries are from 2 to 4294967295. For ACE modules, valid entries are from 2 to 4000000.
Weight	Weight to assign to the server. Valid entries are from 1 to 100. The default is 8.
Probes	Probes to apply to the server. Choose the probes in the Available Items list that you want to apply to this server, and click **Add**. The selected probes appear in the Selected Items list. To remove probes that you do not want to use, choose the probes in the Selected Items list, and click **Remove**. The selected probes appear in the Available Items list. **Note** The VM probe type does not display in the Available Items list even if you have one configured.

Table 7-8 Real Server Configuration Attributes (continued)

Field	Description
Rate Bandwidth	Bandwidth rate, which 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 from 2 to 300000000. The default is 300000000.
Rate Connection	Connection rate, which 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 from 2 to 350000. The default is 350000.

Step 5 When you finish configuring this server for this server farm, do one of the following:

- Click **Deploy Now** to deploy this configuration on the ACE and save your entries to the running-configuration and startup-configuration files. You return to the Real Servers table.
- Click **Cancel** to exit this procedure without saving your entries and to return to the Real Servers table.
- Click **Next** to deploy your entries and to add another real server for this server farm.

Related Topics

- Configuring Health Monitoring for Real Servers, page 7-52
- Configuring Real Servers, page 7-5
- Configuring Sticky Groups, page 8-12
- Configuring the Predictor Method for Server Farms, page 7-40
- Configuring Server Farm HTTP Return Error-Code Checking, page 7-47
- Configuring Dynamic Workload Scaling, page 7-27

Configuring the Predictor Method for Server Farms

You can 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. After adding a server farm (see the "Configuring Server Farms" section on page 7-31), 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.

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

Assumptions

This topic assumes the following:

- A server farm has been added to ANM (see the "Configuring Server Farms" section on page 7-31.)
- At least one real server exists.

Procedure

Step 1 Choose **Config > Devices >** *context* **> Load Balancing > Server Farms**.

The Server Farms table appears.

Step 2 In the Server Farms table, choose the server farm that you want to configure the predictor method for, and click the **Predictor** tab.

The Predictor configuration pane appears.

Step 3 In the Type field of the Predictor configuration pane, choose the method that the ACE is to use to select a server in this server farm when it receives a client request (see Table 7-9).

Step 4 Enter the required information for the selected predictor method (see Table 7-9).

Note Fields and information related to IPv6 require ACE module and ACE appliance software Version A5(1.0) or later.

Table 7-9 Predictor Method Attributes

Predictor Method	Description / Action
Hash Address	Server selection method that uses a hash value based on the source or destination IP address. To configure the hash address predictor method, do the following: **a.** In the Mask Type field, indicate whether server selection is based on source IP address or the destination IP address as follows: – **N/A**—This option is not defined. – **Destination**—The server is selected based on the destination IP address. – **Source**—The server is selected based on the source IP address. **Note** If you configure the server farm with IPv6 and IPv4 Hash Address predictors at the same time, both predictors must have the same mask type. IPv6 requires ACE module and ACE appliance software Version A5(1.0) or later. **b.** In the IP Netmask field, choose the subnet mask to apply to the address. If none is specified, the default is 255.255.255.255. **c.** In the IPv6 Prefix-Length field, enter the IPv6 prefix length. If none is specified, the default is 128. This field appears only for ACE module and ACE appliance software Version A5(1.0) or later.

Table 7-9 Predictor Method Attributes (continued)

Predictor Method	Description / Action
Hash Content	Server selection method that uses a hash value based on the specified content string of the HTTP packet body. Do the following: **a.** 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 13-35 lists the supported characters that you can use for matching string expressions. **b.** 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 13-35 lists the supported characters that you can use for matching string expressions. **c.** In the Length (Bytes) 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 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. **Note** You cannot specify both the length and the end-pattern options for a Hash Content predictor. **d.** In the HTTP Content Offset (Bytes) 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.
Hash Cookie	Server selection method that uses a hash value based on the cookie name. 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.
Hash Header	Server selection method that uses a hash value based on the header name. In the Header Name field, choose the HTTP header to be used for server selection as follows: • To specify an HTTP header that is not one of the standard HTTP headers, click 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, click the second radio button, and then choose one of the HTTP headers from the list.

Table 7-9 Predictor Method Attributes (continued)

Predictor Method	Description / Action
Hash Layer4	Layer 4 generic protocol load-balancing method. Use this predictor to load balance packets from protocols that are not explicitly supported by the ACE. **a.** 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 13-35 lists the supported characters that you can use for matching string expressions. **b.** 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 13-35 lists the supported characters that you can use for matching string expressions. **c.** In the Length (Bytes) 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 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. **Note** You cannot specify both the length and end-pattern options for a Hash Layer 4 predictor. **d.** In the HTTP Content Offset (Bytes) 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 from 0 to 999 bytes. The default is 0, which indicates that the ACE does not exclude any portion of the content.
Hash URL	Server selection method that uses a hash value based on the URL. Use this method to load balance firewalls. Enter values in one or both of the pattern fields as follows: • 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. Valid entries for these fields are unquoted text strings with no spaces and a maximum of 255 alphanumeric characters for each pattern that you configure. The following special characters are also allowed: @ # $

Table 7-9 Predictor Method Attributes (continued)

Predictor Method	Description / Action
Least Bandwidth	Server with the least amount of network traffic over a specified sampling period. Do the following: **a.** In the Assess Time (Seconds) field, enter the number of seconds for which the ACE is to collect traffic information. Valid entries are from 1 to 10 seconds. **b.** 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 (values from 1 to 16 that are also a power of 2).
Least Connections	Server with the fewest number of connections. In the Slow Start Duration (Seconds) field, enter the slow-start value to be applied to this predictor method. Valid entries are from 1 to 65535, where 1 is the slowest ramp-up value. The slow-start mechanism is used to avoid sending a high rate of new connections to servers that you have just put into service.

Table 7-9 Predictor Method Attributes (continued)

Predictor Method	Description / Action
Least Loaded	Least loaded server based on information from SNMP probes. Do the following: **a.** In the SNMP Probe Name field, choose the name of the SNMP probe to use. **b.** In the Auto Adjust field, configure the autoadjust feature to instruct the ACE to apply the maximum load of 16000 to a real server whose load reaches zero or override the default behavior. By default, the ACE applies the average load of the server farm to a real server whose load is zero. The ACE periodically adjusts this load value based on feedback from the server SNMP probe and other configured options. Options include the following: – **Average**—Instructs the ACE to apply the average load of the server farm to a real server whose load is zero. This setting allows the server to participate in load balancing, while preventing it from being flooded by new connections. This is the default setting. – **Maxload**—Instructs the ACE to apply the maximum load of the server farm to a real server whose load reaches zero. The maxload option requires the following ACE software versions: - ACE appliance—A3(2.7) or A4(1.0) or later - ACE module—A2(2.4), A2(3.2), or A4(1.0) or later If you choose the maxload option and the ACE does not support the option, ANM issues a command parse error message. – **Off**—Instructs the ACE to send 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. **c.** 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. Uncheck the check box to reset the behavior of the ACE to the default of excluding the current connection count from the load calculation. 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. 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. The syntax of this predictor command is as follows: **predictor least-loaded probe** *name* 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.

Table 7-9 Predictor Method Attributes (continued)

Predictor Method	Description / Action

Least Loaded (continued)	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 the following commands:
	`host1/Admin(config)# serverfarm SF1`
	`host1/Admin(config-sfarm-host)# predictor least-loaded probe PROBE_SNMP`
	`host1/Admin(config-sfarm-host-predictor)#`
	
	To reset the predictor method to the default of round-robin, enter the following command:
	`host1/Admin(config-sfarm-host)# no predictor`

Response	Server selection method based on the lowest response time for a requested response-time measurement. **a.** In the Response Type field, select the type of measurement to use as follows: – **App-Req-To-Resp**—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. **b.** 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 (values from 1 to 16 that are also a power of 2). **c.** 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. Uncheck the check box to reset the behavior of the ACE to the default of excluding the current connection count from the load calculation.
Round Robin	Server selection method in which The ACE selects the next server in the list of servers based on server weight. This method is the default predictor.

Step 5 Click **Deploy Now** to deploy this configuration on the ACE and save your entries to the running-configuration and startup-configuration files.

Related Topics

- Configuring Health Monitoring for Real Servers, page 7-52
- Configuring Real Servers, page 7-5
- Configuring Sticky Groups, page 8-12
- Adding Real Servers to a Server Farm, page 7-38
- Configuring Dynamic Workload Scaling, page 7-27

Configuring Server Farm HTTP Return Error-Code Checking

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

You can configure HTTP return error-code checking (retcode map) for a server farm. After adding a server farm (see the "Configuring Server Farms" section on page 7-31), you can associate real servers with it and configure the predictor method and retcode maps. These options appear after you have successfully added a server farm.

Assumption

A host type server farm has been added to ANM (see the "Configuring Server Farms" section on page 7-31).

Procedure

Step 1 Choose **Config > Devices >** *context* **> Load Balancing > Server Farms**.

The Server Farms table appears.

Step 2 In the Server Farms table, choose the server farm that you want to configure for return error-code checking, and click the **Retcode Map** tab.

The Retcode Map table appears.

Step 3 In the Retcode Map table, 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 of the Retcode Map configuration pane, enter the minimum value for an HTTP return error code.

Valid entries are 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 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 7-10.

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

Table 7-10 Return-Code Type Configuration Options

Option	Description
Count	Total number of return codes received for each return code number that you specify.

Table 7-10 Return-Code Type Configuration Options (continued)

Option	Description
Log	Syslog error message generated when the number of events reaches a specified threshold. **a.** In the Threshold field, enter the number of events that the ACE is to receive before generating a syslog error message. Valid entries are as follows: – ACE appliance (all) and ACE module pre A4(1.0)—1 to 4294967295. – ACE module A4(1.0)—4 to 4294967295. **b.** In the Reset (Seconds) field, enter the time interval in seconds for which the ACE checks for the return code. Valid entries are as follows: – ACE appliance or module pre A4(1.0)—1 to 4294967295 – ACE appliance or module A4(1.0) and later—1 to 2147483647
Remove	The ACE generates a syslog error message when the number of events reaches a specified threshold and then removes the server from service. **a.** 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 from 1 to 4294967295. **b.** In the Reset (Seconds) field, enter the time interval in seconds for which the ACE checks for the return code. Valid entries are from 1 to 4294967295 seconds. **c.** In the Resume Service (Seconds) 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. The default is 0 seconds. The setting of this field affects the behavior of the real server in the failed state, as follows: – When this field is not configured and has the default setting of 0, the real server remains in the failed state until you manually remove it from service and read it. – When this field is not configured and has the default setting of 0 and then you configure it with an integer between 30 and 3,600, the failed real server immediately transitions to the Operational state. – When you configure this field and then increase the value, the real server remains in the failed state for the duration of the previously-configured value. The new value takes effect the next time the real server transitions to the failed state. – When you configure this field and then decrease the value, the failed real server immediately transitions to the Operational state. – When you configure this field with an integer between 30 and 3,600 and then reset it to the default of 0, the real server remains in the failed state for the duration of the previously-configured value. The default setting takes effect the next time the real server transitions to the failed state. Then the real server remains in the failed state until you manually remove it from service and read it.

Step 7 Do one of the following:

- Click **Deploy Now** to deploy this configuration on the ACE and save your entries to the running-configuration and startup-configuration files.
- Click **Cancel** to exit this procedure without saving your entries and to return to the Retcode Map table.
- Click **Next** to deploy your entries and to add another retcode map.

Related Topics

- Information About Virtual Contexts, page 5-2
- Configuring Virtual Context Class Maps, page 13-6
- Configuring Virtual Context Policy Maps, page 13-32
- Configuring Real Servers, page 7-5
- Configuring Sticky Groups, page 8-12
- Configuring Dynamic Workload Scaling, page 7-27

Displaying All Server Farms

You can display all server farms associated with a virtual context.

Procedure

Step 1 Choose **Config > Devices**.

The Virtual Contexts table appears.

Step 2 In the Virtual Contexts table, choose the virtual context with the server farms you want to display, and 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

Step 3 (Optional) Do the following:

- Add or edit a server farm (see the "Configuring Server Farms" section on page 7-31)
- Choose a server farm and click **Buddy Group** to view a pop up window that displays the output of the **show buddy group** command. The pop up window displays the list of buddy groups configured in the virtual context (for more information, see the "Buddy Sticky Groups" section on page 8-6).

  Note This feature appears only for ACE software Version A5(2.0) or later.

- Click the **Real Servers** tab to display the real servers associated with the selected server farm. From this tab you can manage the server farm real servers (see the "Adding Real Servers to a Server Farm" section on page 7-38).
- Click the **Predictor** tab to display the predictor method associated with the selected server farm. From this tab you can choose the predictor method (see the "Configuring the Predictor Method for Server Farms" section on page 7-40).

- Click the **Retcode Map** tab to display the HTTP return error-code checking that has been configured for the selected server farm. From this tab you can manage the error-code checking (see the "Configuring Server Farm HTTP Return Error-Code Checking" section on page 7-47).

Related Topics

- Displaying Server Farm Statistics and Status Information, page 7-50
- Configuring Server Farms, page 7-31
- Adding Real Servers to a Server Farm, page 7-38
- Configuring the Predictor Method for Server Farms, page 7-40
- Configuring Server Farm HTTP Return Error-Code Checking, page 7-47
- Configuring Dynamic Workload Scaling, page 7-27

Displaying Server Farm Statistics and Status Information

You can display statistics and status information for a particular server farm.

Procedure

Step 1 Choose **Config > Devices >** *context* **> Load Balancing > Server Farms**.

The Server Farms table appears.

Step 2 In the Server Farms table, choose a server farm from the Server Farms table, and click **Details**.

The **show serverfarm** *name* **detail** CLI command output appears. For details about the displayed output fields, see the *Cisco ACE Module Server Load-Balancing Configuration Guide* or the *Cisco ACE 4700 Series Appliance Server Load-Balancing Configuration Guide*, Chapter 2, Configuring Real Servers and Server Farms.

Step 3 Click **Update Details** to refresh the output for the **show serverfarm** *name* **detail** CLI command.

The new information appears in a separate panel with a new timestamp; both the old and the new server farm statistics and status information appear side-by-side to avoid overwriting the last updated information.

Step 4 Click **Close** to return to the Server Farms table.

Related Topics

- Displaying All Server Farms, page 7-49
- Configuring Server Farms, page 7-31
- Adding Real Servers to a Server Farm, page 7-38
- Configuring the Predictor Method for Server Farms, page 7-40
- Configuring Server Farm HTTP Return Error-Code Checking, page 7-47
- Configuring Dynamic Workload Scaling, page 7-27

Configuring 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 the "TCL Scripts" section on page 7-51).

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, it can make reliable load-balancing decisions.

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 or 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.

This section includes the following topics:

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 Toolkit Command Language (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 the "Configuring Health Monitoring for Real Servers" section on page 7-52.

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

- 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

These scripts are located in the probe: directory and are accessible in both the Admin and user contexts. Note that the script files in the probe: directory are read-only, so you cannot copy or modify them. However, you can copy files from the probe: directory. For more information, see either the *Cisco Application Control Engine Module Administration Guide* or the *Cisco 4700 Series Application Control Engine Appliance Administration Guide*.

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 TCL scripts on the ACE, see either the *Cisco ACE Module Server Load-Balancing Configuration Guide* or the *Cisco ACE 4700 Series Appliance Server Load-Balancing Configuration Guide*.

Configuring Health Monitoring for Real Servers

You can establish monitoring of real servers to determine their viability in load-balancing decisions. 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 or not to include the server in its load-balancing decision.

Procedure

Step 1 Choose **Config > Devices >** *context* **> Load Balancing > Health Monitoring**.

The Health Monitoring table appears.

Step 2 In the Health Monitoring table, click **Add** to add a new health monitoring probe, or choose an existing entry and click **Edit** to modify it.

The Health Monitoring window appears.

Step 3 In the Name field of the Health Monitoring window, 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.

Step 4 In the Type field, choose the type of probe that you want to use.

The probe type determines what the probe sends to the real server. See Table 7-11 for the types of probes and their descriptions.

Table 7-11 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. **Note** This option is not available for the ACE NPE software version (see the "Information About the ACE No Payload Encryption Software Version" section on page 1-3).
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 DESTINATION_UNREACHABLE, 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 email 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 email 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 filename must first be established on the device.

Table 7-11 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 SMNP 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.
VM	This probe type requires the following: • The ACE appliance or module is using software Version A4(2.0) or a later release. • The ACE is configured with a VM Controller connection (see the "Configuring and Verifying a VM Controller Connection" section on page 7-30). Sends a probe to the VMware VM Controller to determine the average amount of both CPU and memory usage of its associated local VMs. The probe response determines whether the ACE load-balances traffic to the local VMs only or bursts traffic to the remote VMs due to high usage of the local VMs. **Note** You use a VM probe when you configure the ACE for Dynamic Workload Scaling (see the "Configuring Dynamic Workload Scaling" section on page 7-27).

Step 5 Enter health monitoring general attributes (see Table 7-12).

Note Fields and information related to IPv6 require ACE module and ACE appliance software Version A5(1.0) or later.

Note Click **More Settings** to access the additional general attributes for the selected probe type. By default, ANM hides the probe attributes with default values and the probe attributes that are not commonly used.

Table 7-12 Health Monitoring General Attributes

Field	Action
Description	Description for this probe. Valid entries are unquoted alphanumeric text strings with no spaces and a maximum of 240 characters.
Probe Interval (Seconds)	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 for all probe types except the VM probe, which has a range from 300 to 65535. The default values are as follows: • ACE appliance (all software versions)—Default is 15 seconds for all probe types except the VM probe, which has a default of 300 seconds. • ACE module: – Software Version A4(1.0) and later—Default is 15 seconds for all probe types except the VM probe, which has a default of 300 seconds. – All software versions before A4(1.0)—Default is 120 seconds. **Note** The VM probe type requires ACE software Version A4(2.0) or later on either device type.
Pass Detect Interval (Seconds)	Number of seconds that the ACE is to wait before sending another probe to a server marked as failed. Valid entries are from 2 to 65535. The default values are as follows: • ACE appliance (all software versions)—Default is 60 seconds. • ACE module: – Software Version A4(1.0) and later—Default is 60 seconds. – All software versions before A4(1.0)—Default is 300 seconds. **Note** This field is not applicable for the VM probe type.
Fail Detect	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 from 1 to 65535. The default is 3. **Note** This field is not applicable for the VM probe type.
More Settings (Not applicable for the VM probe type)	
Pass Detect Count	Number of successful probe responses from the server before the server is marked as passed. Valid entries are from 1 to 65535. The default is 3.
Receive Timeout (Seconds)	Number of seconds that the ACE is to wait for a response from a server that has been probed before marking the server as failed. Valid entries are from 1 to 65535. The default is 10.

Table 7-12 Health Monitoring General Attributes (continued)

Field	Action
Destination IPv4/IPv6 Address[1]	The IPv6 option requires ACE module and ACE appliance software Version A5(1.0) or later, which supports IPv4 and IPv6. Preferred destination IP address. 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. **Note** The following probes support IPv6 destination addresses: DNS, HTTP, HTTPS, ICMP, TCP, and UDP. **Note** When you assign a probe to a real server, they must be configured with the same IP address type (IPv6 or IPv4).
Is Routed [2]	Check box that indicates that the destination IP address is routed according to the ACE internal routing table. Uncheck the check box to indicate that the destination IP address is not routed according to the ACE internal routing table.
Port	By default, the precedence in which the probe inherits the port number is as follows: • The port number that you configure for the probe. • The configured port number from the real server in server farm. • The configured port number from the VIP in a Layer 3 and Layer 4 class map. • The default port number. Table 7-13 lists the default port number for each probe type. If you explicitly configure a default port, the ACE always sends the probe to the default port. The probe does not dynamically inherit the port number from the real server in a server farm or from the VIP specified in the class map.

1. The Dest IP Address field is not applicable to the Scripted probe type.
2. The Is Routed field is not applicable to the RTSP, Scripted, SIP-TCP, and SIP-UDP probe types.

Table 7-13 Default Port Numbers for Probe Types

Probe Type	Default Port Number
DNS	53
Echo	7
Finger	79
FTP	21
HTTP	80
HTTPS	443
ICMP	Not applicable
IMAP	143
POP3	110
RADIUS	1812
RTSP	554

Table 7-13 Default Port Numbers for Probe Types (continued)

Probe Type	Default Port Number
Scripted	1
SIP (both TCP and UDP)	5060
SMTP	25
SNMP	161
Telnet	23
TCP	80
UDP	53
VM	443

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

- For DNS probes, see Table 7-14.
- For Echo-TCP probes, see Table 7-15.
- For Echo-UDP probes, see Table 7-16.
- For Finger probes, see Table 7-17.
- For FTP probes, see Table 7-18.
- For HTTP probes, see Table 7-19.
- For HTTPS probes, see Table 7-20.
- There are no specific attributes for ICMP probes.
- For IMAP probes, see Table 7-21.
- For POP probes, see Table 7-22.
- For RADIUS probes, see Table 7-23.
- For RTSP probes, see Table 7-24.
- For Scripted probes, see Table 7-25.
- For SIP-TCP probes, see Table 7-26.
- For SIP-UDP probes, see Table 7-27.
- For SMTP probes, see Table 7-28.
- For SNMP probes, see Table 7-29.
- For TCP probes, see Table 7-30.
- For Telnet probes, see Table 7-31.
- For UDP probes, see Table 7-32.
- For VM probes, see Table 7-33.

Step 7 Do one of the following:

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

Step 8 (Optional) To display statistics and status information for a particular probe, choose the probe from the Health Monitoring table, and click **Details**.

The **show probe** *name* **detail** CLI command output appears. See the "Displaying Health Monitoring Statistics and Status Information" section on page 7-79 for details.

Related Topics

- Configuring DNS Probe Expect Addresses, page 7-75
- Configuring Headers for HTTP and HTTPS Probes, page 7-76
- Configuring Health Monitoring Expect Status, page 7-77
- Displaying Health Monitoring Statistics and Status Information, page 7-79
- Configuring Real Servers, page 7-5
- Configuring Server Farms, page 7-31
- Configuring Sticky Groups, page 8-12

Configuring Probe Attributes

You can configure health monitoring probe-specific attributes.

This section includes the following topics:

- SMTP Probe Attributes, page 7-71
- SNMP Probe Attributes, page 7-71
- TCP Probe Attributes, page 7-72
- Telnet Probe Attributes, page 7-73
- UDP Probe Attributes, page 7-73
- VM Probe Attributes, page 7-74

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

- Configuring DNS Probe Expect Addresses, page 7-75
- Configuring Headers for HTTP and HTTPS Probes, page 7-76
- Configuring Health Monitoring Expect Status, page 7-77
- Configuring an OID for SNMP Probes, page 7-78
- Displaying Health Monitoring Statistics and Status Information, page 7-79

DNS Probe Attributes

Table 7-14 lists the DNS probe attributes.

Note Click **More Settings** to access the additional attributes for the DNS probe type. By default, ANM hides the probe attributes with default values and the probe attributes that are not commonly used.

Table 7-14 DNS Probe Attributes

Field	Action
Domain Name	Domain name that the probe is to send to the DNS server. Valid entries are unquoted text strings with a maximum of 255 characters.
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.

To configure expect addresses for DNS probes, see the "Configuring DNS Probe Expect Addresses" section on page 7-75.

Echo-TCP Probe Attributes

Table 7-15 lists the Echo-TCP probe attributes.

Note Click **More Settings** to access the additional attributes for the Echo-TCP probe type. By default, ANM hides the probe attributes with default values and the probe attributes that are not commonly used.

Table 7-15 Echo-TCP Probe Attributes

Field	Action
Send Data	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.
More Settings	
TCP Connection Termination	Check box that when checked, configures the ACE to terminate TCP connections gracefully by sending a FIN to the server. Uncheck the check box to configure the ACE to terminate a TCP connection by sending an RST.
Open Timeout (Seconds)	Number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535. The default is as follows: • For ACE module version A2(3.x) and earlier, the default is 10 seconds. • For ACE module version A4(1.0) and later or ACE appliance version A3(1.x) and later, the default is 1 second.

Echo-UDP Probe Attributes

Table 7-16 lists the Echo-UDP probe attributes.

Note Click **More Settings** to access the additional attributes for the Echo-UDP probe type. By default, ANM hides the probe attributes with default values and the probe attributes that are not commonly used.

Table 7-16 Echo-UDP Probe Attributes

Field	Action
Send Data	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.
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.

Finger Probe Attributes

Table 7-17 lists the Finger probe attributes.

Note Click **More Settings** to access the additional attributes for the Finger probe type. By default, ANM hides the probe attributes with default values and the probe attributes that are not commonly used.

Table 7-17 Finger Probe Attributes

Field	Action
Send Data	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.
More Settings	

Table 7-17 Finger Probe Attributes (continued)

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
TCP Connection Termination	Check box that when checked, configures the ACE to terminate TCP connections gracefully by sending a FIN to the server. Uncheck the check box to configure the ACE to terminate a TCP connection by sending an RST.
Open Timeout (Seconds)	Number of seconds to wait when opening a connection with a real server. Valid entries are from 1 to 65535. The default is as follows: • For ACE module version A2(3.x) and earlier, the default is 10 seconds. • For ACE module version A4(1.0) and later or ACE appliance version A3(1.x) and later, the default is 1 second.

FTP Probe Attributes

Table 7-18 lists the FTP probe attributes.

Note Click **More Settings** to access the additional attributes for the FTP probe type. By default, ANM hides the probe attributes with default values and the probe attributes that are not commonly used.

Table 7-18 FTP Probe Attributes

Field	Action
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
TCP Connection Termination	Check box that when checked, configures the ACE to terminate TCP connections gracefully by sending a FIN to the server. Uncheck the check box to configure the ACE to terminate a TCP connection by sending an RST.
Open Timeout (Seconds)	Number of seconds to wait when opening a connection with a real server. Valid entries are integers from 1 to 65535. The default is as follows: • For ACE module version A2(3.x) and earlier, the default is 10 seconds. • For ACE module version A4(1.0) and later or ACE appliance version A3(1.x) and later, the default is 1 second.

To configure probe expect statuses for FTP probes, see the "Configuring Health Monitoring Expect Status" section on page 7-77.

HTTP Probe Attributes

Table 7-19 lists the HTTP probe attributes.

Note Click **More Settings** to access the additional attributes for the HTTP probe type. By default, ANM hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 7-19 HTTP Probe Attributes

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Request Method Type	Type of HTTP request method that is to be used for this probe. Choose one of the following: • **N/A**—This option is not defined. • **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. • **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.
Request HTTP URL	Field that appears if you chose 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 "/".
More Settings	
Append Port Host Tag	Check box that when checked, configures the ACE to append port information in the HTTP Host header when you configure a nondefault destination port for an HTTP probe. By default, the check box is unchecked and the ACE does not append this information. **Note** This feature requires ACE module software Version A2(3.4) and ACE appliance software Version A3(2.7) or later releases.
TCP Connection Termination	Check box that when checked, configures the ACE to terminate TCP connections gracefully by sending a FIN to the server. Uncheck the check box to configure the ACE to terminate a TCP connection by sending an RST.
Open Timeout (Seconds)	Number of seconds to wait when opening a connection with a real server. Valid entries are from 1 to 65535. The default is as follows: • For ACE module software Version A2(3.x) and earlier, the default is 10 seconds. • For ACE module software Version A4(1.0) and later or ACE appliance software Version A3(1.x) and later, the default is 1 second.
User Name	User identifier to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters.
Password	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.

Table 7-19 HTTP Probe Attributes (continued)

Field	Action
Expect Regular Expression	Expected response data from the probe destination. Valid entries are text strings (quotes allowed) with a maximum of 255 characters.
Expect Regex Offset	Number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regular Expression field. Valid entries are from 1 to 4000.
Hash	Check box that when checked, configures the ACE to use an MD5 hash for an HTTP GET probe. Uncheck the check box to configure the ACE not to use an MD5 hash for an HTTP GET probe.
Hash String	Field that 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.

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

- Configuring Headers for HTTP and HTTPS Probes, page 7-76
- Configuring Health Monitoring Expect Status, page 7-77

HTTPS Probe Attributes

Table 7-20 lists the HTTPS probe attributes.

Note Click **More Settings** to access the additional attributes for the HTTPS probe type. By default, ANM hides the probe attributes with default values and the probe attributes that are not commonly used.

Table 7-20 HTTPS Probe Attributes

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Request Method Type	Type of HTTP request method that is to be used for this probe. Choose one of the following: • **N/A**—This option is not defined. • **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. • **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 as a result changed hash values.

Table 7-20 HTTPS Probe Attributes (continued)

Field	Action
Request HTTP URL	Field that appears if you chose 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 "/'.
Cipher	Choose the cipher suite to be used with this HTTPS probe: • 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	Version of SSL or TLS to be used in ClientHello messages sent to the server as follows: • **All**—The probe is to use all SSL versions. • **SSLv3**—The probe is to use SSL version 3. • **TLSv1**—The probe is to use TLS version 1. By default, the probe sends ClientHello messages with an SSL version 3 header and a TLS version 1 message.
More Settings	
Append Port Host Tag	Check box that when checked, configures the ACE to append port information in the HTTPS Host header when you configure a nondefault destination port for an HTTPS probe. By default, the check box is unchecked and the ACE does not append this information. **Note** This feature requires ACE module software Version A2(3.4) and ACE appliance software Version A3(2.7) or later releases.
TCP Connection Termination	Check box that when checked, configures the ACE to terminate TCP connections gracefully by sending a FIN to the server. Uncheck the check box to configure the ACE to terminate a TCP connection by sending an RST.
Open Timeout (Seconds)	Number of seconds to wait when opening a connection with a real server. Valid entries are from 1 to 65535. The default is as follows: • For ACE module version A2(3.x) and earlier, the default is 10 seconds. • For ACE module version A4(1.0) and later or ACE appliance version A3(1.x) and later, the default is 1 second.

Table 7-20 HTTPS Probe Attributes (continued)

Field	Action
User Name	User identifier to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters.
Password	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 Regular Expression	Expected response data from the probe destination. Valid entries are text strings (quotes allowed) with a maximum of 255 characters.
Expect Regex Offset	Number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regular Expression field. Value entries are from 1 to 4000.
Hash	Check box that when checked, configures the ACE to use an MD5 hash for an HTTP GET probe. Uncheck the check box to configure the ACE not to use an MD5 hash for an HTTP GET probe.
Hash String	Field that 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.
Ignore Certificate Expiration	Check the Ignore Certificate Expiration check box to configure the probe to ignore the certificate expiration date so the probe does not affect ACE functionality when the certificate has expired. Uncheck the check box to configure the ACE not to ignore the certificate expiration date. **Note** This field appears only for ACE software Version A5(2.0) and later.

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

- Configuring Headers for HTTP and HTTPS Probes, page 7-76
- Configuring Health Monitoring Expect Status, page 7-77

IMAP Probe Attributes

Table 7-21 lists the IMAP probe attributes.

Note Click **More Settings** to access the additional attributes for the IMAP probe type. By default, ANM hides the probe attributes with default values and the probe attributes are not commonly used.

Table 7-21 IMAP Probe Attributes

Field	Action
User Name	User identifier to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters.
Password	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.
Mailbox Name	User mailbox name from which to retrieve email for this IMAP probe. Valid entries are unquoted text strings with a maximum of 64 characters.
Request Command	Request method command for this probe. Valid entries are text strings with a maximum of 32 characters and no spaces.
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
TCP Connection Termination	Check box that when checked, configures the ACE to terminate TCP connections gracefully by sending a FIN to the server. Uncheck the check box to configure the ACE to terminate a TCP connection by sending an RST.
Open Timeout (Seconds)	Number of seconds to wait when opening a connection with a real server. Valid entries are from 1 to 65535. The default is as follows: • For ACE module version A2(3.x) and earlier, the default is 10 seconds. • For ACE module version A4(1.0) and later or ACE appliance version A3(1.x) and later, the default is 1 second.

POP Probe Attributes

Table 7-22 lists the POP probe attributes.

Note Click **More Settings** to access the additional attributes for the POP probe type. By default, ANM hides the probe attributes with default values and the probe attributes which are not commonly used.

Table 7-22 POP Probe Attributes

Field	Action
User Name	User identifier to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters.
Password	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 Command	Request method command for this probe. Valid entries are text strings with a maximum of 32 characters and no spaces.
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.

Table 7-22 POP Probe Attributes (continued)

Field	Action
TCP Connection Termination	Check box that when checked, configures the ACE to terminate TCP connections gracefully by sending a FIN to the server. Uncheck the check box to configure the ACE to terminate a TCP connection by sending an RST.
Open Timeout (seconds)	Number of seconds to wait when opening a connection with a real server. Valid entries are from 1 to 65535. The default is as follows: • For ACE module version A2(3.x) and earlier, the default is 10 seconds. • For ACE module version A4(1.0) and later or ACE appliance version A3(1.x) and later, the default is 1 second.

RADIUS Probe Attributes

Table 7-23 lists the RADIUS probe attributes.

Note Click **More Settings** to access the additional attributes for the RADIUS probe type. By default, ANM hides the probe attributes with default values and the probe attributes that are not commonly used.

Table 7-23 RADIUS Probe Attributes

Field	Action
User Secret	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 64 characters.
User Name	User identifier to be used for authentication on the real server. Valid entries are unquoted text strings with a maximum of 64 characters.
Password	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.
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
NAS IP Address	IP address of the Network Access Server (NAS) in dotted-decimal format, such as 192.168.11.1.

RTSP Probe Attributes

Table 7-24 lists the RTSP probe attributes.

Note Click **More Settings** to access the additional attributes for the RTSP probe type. By default, ANM hides the probe attributes with default values and the probe attributes that are not commonly used.

Table 7-24 RTSP Probe Attributes

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
RTSP Require Header Value	Require header for the probe.
RTSP Proxy Require Header Value	Proxy-Require header for the probe.
RTSP Request Method Type	Request method type: • **N/A**—No request method is selected. • **Describe**—Probe is to use the Describe request type.
More Settings	
TCP Connection Termination	Check box that when checked, configures the ACE to terminate TCP connections gracefully by sending a FIN to the server. Uncheck the check box to configure the ACE to terminate a TCP connection by sending an RST.
Open Timeout (Seconds)	Number of seconds to wait when opening a connection with a real server. Valid entries are from 1 to 65535. The default is as follows: • For ACE module version A2(3.x) and earlier, the default is 10 seconds. • For ACE module version A4(1.0) and later or ACE appliance version A3(1.x) and later, the default is 1 second.

To configure probe expect statuses for RTSP probes, see the "Configuring Health Monitoring Expect Status" section on page 7-77.

Scripted Probe Attributes

Table 7-25 lists the HTTP probe attributes.

Note Click **More Settings** to access the additional attributes for the Scripted probe type. By default, ANM hides the probe attributes with default values and the probe attributes that are not commonly used.

Table 7-25 Scripted Probe Attributes

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Script Name	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. See your ACE documentation for more details. Valid entries are unquoted text strings with no spaces and a maximum of 255 characters.
Script Arguments	Valid arguments, which are unquoted text strings with no spaces; separate multiple arguments with a space. The field limit is 255 characters.
More Settings	
Script Needs To Be Copied From Remote Location?	Check box that indicates that the file needs to be copied from a remote server. Uncheck this check box to indicate that the script resides locally.
Protocol	Field that appears if the script is to be copied from a remote server. Choose the protocol to be used for copying the script: • **FTP**—The script is to be copied using FTP. • **TFTP**—The script is to be copied using TFTP.
User Name	Field that appears if FTP is selected in the Protocol field. Enter the name of the user account on the remote server.
Password	Field that 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	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 *host-ip/path/filename* where: • *host-ip* represents the IP address of the remote server. • *path* represents the directory path of the file on the remote server. • *filename* represents the filename of the file on the remote server. For example, your entry might be 192.168.11.2/usr/bin/my-script.ext.

SIP-TCP Probe Attributes

Table 7-26 lists the SIP-TCP probe attributes.

Note Click **More Settings** to access the additional attributes for the SIP-TCP probe type. By default, ANM hides the probe attributes with default values and the probe attributes that are not commonly used.

Table 7-26 SIP-TCP Probe Attributes

Field	Action
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
TCP Connection Termination	Check box that when checked, configures the ACE to terminate TCP connections gracefully by sending a FIN to the server. Uncheck the check box to configure the ACE to terminate a TCP connection by sending an RST.
Open Timeout (Seconds)	Number of seconds to wait when opening a connection with a real server. Valid entries are from 1 to 65535. The default is as follows: • For ACE module version A2(3.x) and earlier, the default is 10 seconds. • For ACE module version A4(1.0) and later or ACE appliance version A3(1.x) and later, the default is 1 second.
Expect Regular Expression	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	Number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regular Expression field. Value entries are from 1 to 4000.

To configure probe expect statuses for SIP-TCP probes, see the "Configuring Health Monitoring Expect Status" section on page 7-77.

SIP-UDP Probe Attributes

Table 7-27 lists the SIP-UDP probe attributes.

Note Click **More Settings** to access the additional attributes for the SIP-UDP probe type. By default, ANM hides the probe attributes with default values and the probe attributes that are not commonly used.

Table 7-27 SIP-UDP Probe Attributes

Field	Action
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Expect Regular Expression	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	Number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regular Expression field. Value entries are from 1 to 4000.

To configure probe expect statuses for SIP-UDP probes, see the "Configuring Health Monitoring Expect Status" section on page 7-77.

SMTP Probe Attributes

Table 7-28 lists the SMTP probe attributes.

Note Click **More Settings** to access the additional attributes for the SMTP probe type. By default, ANM hides the probe attributes with default values and the probe attributes that are not commonly used.

Table 7-28 SMTP Probe Attributes

Field	Action
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
TCP Connection Termination	Check box that when checked, configures the ACE to terminate TCP connections gracefully by sending a FIN to the server. Uncheck the check box to configure the ACE to terminate a TCP connection by sending an RST.
Open Timeout (Seconds)	Number of seconds to wait when opening a connection with a real server. Valid entries are from 1 to 65535. The default is as follows: • For ACE module version A2(3.x) and earlier, the default is 10 seconds. • For ACE module version A4(1.0) and later or ACE appliance version A3(1.x) and later, the default is 1 second.

To configure probe expect statuses for SMTP probes, see the "Configuring Health Monitoring Expect Status" section on page 7-77.

SNMP Probe Attributes

Table 7-29 lists the SNMP probe attributes.

Note Click **More Settings** to access the additional attributes for the SNMP probe type. By default, ANM hides the probe attributes with default values and the probe attributes that are not commonly used.

Table 7-29 SNMP Probe Attributes

Field	Action
SNMP Community	SNMP community string. Valid entries are unquoted text strings with no spaces and a maximum of 255 characters.
More Settings	

Table 7-29 SNMP Probe Attributes (continued)

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
SNMP Version	SNMP version for the probe: • **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 the "Configuring an OID for SNMP Probes" section on page 7-78.

TCP Probe Attributes

Table 7-30 lists the TCP probe attributes.

Note Click **More Settings** to access the additional attributes for the TCP probe type. By default, ANM hides the probe attributes with default values and the probe attributes that are not commonly used.

Table 7-30 TCP Probe Attributes

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Send Data	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.
Send Hex Data	Enter the data in hex format to be sent as part of probe request. The Hex data entered must be of even numbers and must be a single string consisting of alphanumeric within the range of 0-9,a-f or A-F, and a maximum of 254 characters. The conversion from Hex ASCII to Binary will happen when the probe data is sent out.
Data Format	Users can enter only one data format either in "send-hex-data" or in "send-data" format. Click the radio button "send-hex-data" or "send-data" to choose the format. **Expect Regex / Expect Hex Regex** and **Expect Regex Offset / Expect Hex Regex Offset** shall be displayed based on the radio button selection
More Settings	
TCP Connection Termination	Check box that when checked, configures the ACE to terminate TCP connections gracefully by sending a FIN to the server. Uncheck the check box to configure the ACE to terminate a TCP connection by sending an RST.
Open Timeout (Seconds)	Number of seconds to wait when opening a connection with a real server. Valid entries are from 1 to 65535. The default is as follows: • For ACE module version A2(3.x) and earlier, the default is 10 seconds. • For ACE module version A4(1.0) and later or ACE appliance version A3(1.x) and later, the default is 1 second.

Table 7-30 TCP Probe Attributes (continued)

Field	Action
Expect Regular Expression	Expected response data from the probe destination. Valid entries are text strings (quotes allowed) with a maximum of 255 characters.
Expect Regex Offset	Number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regular Expression field. Value entries are from 1 to 4000.
Expect Hex Regex	Enter the expected response data from the probe destination. The Hex data entered must be of even numbers and must be a single string consisting of alphanumeric within the range of 0-9,a-f or A-F, and a maximum of 255 characters.
Expect Hex Regex Offset	Enter the expected response data in Hex format. The Hex data entered must be of even numbered size and of maximum size of 254.

Telnet Probe Attributes

Table 7-31 lists the Telnet probe attributes.

Note Click **More Settings** to access the additional attributes for the Telnet probe type. By default, ANM hides the probe attributes with default values and the probe attributes that are not commonly used.

Table 7-31 Telnet Probe Attributes

Field	Action
More Settings	
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
TCP Connection Termination	Check box that when checked, configures the ACE to terminate TCP connections gracefully by sending a FIN to the server. Uncheck the check box to configure the ACE to terminate a TCP connection by sending an RST.
Open Timeout (Seconds)	Enter the number of seconds to wait when opening a connection with a real server. Valid entries are from 1 to 65535. The default is as follows: • For ACE module version A2(3.x) and earlier, the default is 10 seconds. • For ACE module version A4(1.0) and later or ACE appliance version A3(1.x) and later, the default is 1 second.

UDP Probe Attributes

Table 7-32 lists the UDP probe attributes.

Note Click **More Settings** to access the additional attributes for the UDP probe type. By default, ANM hides the probe attributes with default values and the probe attributes that are not commonly used.

Table 7-32 UDP Probe Attributes

Field	Action
Port	Enter the port number that the probe is to use. By default, the probe uses port inheritance to determine the port number. For more information, see the general attribute Port field description.
Send Data	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.
Send Hex Data	Enter the data in hex format to be sent as part of probe request. The Hex data entered must be of even numbers and must be a single string consisting of alphanumeric within the range of 0-9,a-f or A-F, and a maximum of 254 characters. The conversion from Hex ASCII to Binary will happen when the probe data is sent out.
Data Format	Users can enter only one data format either in “send-hex-data” or in“send-data” format. Click the radio button “send-hex-data” or “send-data” to choose the format. Expect Regex / Expect Hex Regex and Expect Regex Offset / Expect Hex Regex Offset shall be displayed based on the radio button selection.
More Settings	
Expect Regular Expression	Expected response data from the probe destination. Valid entries are text strings (quotes allowed) with a maximum of 255 characters.
Expect Regex Offset	Number of characters into the received message or buffer where the ACE is to begin looking for the string specified in the Expect Regular Expression field. Value entries are from 1 to 4000.
Expect Hex Regex	Enter the expected response data from the probe destination. The Hex data entered must be of even numbers and must be a single string consisting of alphanumeric within the range of 0-9,a-f or A-F, and a maximum of 255 characters.
Expect Hex Regex Offset	Enter the expected response data in Hex format. The Hex data entered must be of even numbered size and of maximum size of 254.

VM Probe Attributes

Note You use a VM probe when you configure the ACE for Dynamic Workload Scaling (see the “Configuring Dynamic Workload Scaling” section on page 7-27), which requires that the ACE appliance or module is using software Version A4(2.0) or a later release.

You configure the VM probe attributes to control when the ACE bursts traffic to remote VMs based on an average of local VM CPU usage, memory usage, or both. The ACE obtains the usage information by sending the VM probe to the specified VM Controller associated with the local VMs (see Figure 1-1). It calculates the average aggregate load information for all local VMs as a percentage of CPU usage or memory usage and uses either or both percentages to determine when to burst traffic to the remote data center. If the server farm consists of both physical servers and VMs, the ACE considers load information only from the VMs.

By default, the VM probe checks the percentage of usage for either the CPU or memory against the maximum threshold value. Whichever percentage reaches its maximum threshold value first causes the ACE to burst traffic to the remote data center. The default maximum burst threshold value of 99 percent instructs the ACE to always load balance traffic to the local VMs unless the load value is equal to 100 percent or the VMs are not in the Operational state. If you configure the maximum burst threshold value to 1 percent, the ACE always bursts traffic to the remote data center.

When the usage percentage is less than the minimum threshold value, the ACE stops bursting traffic to the remote data center and continues to load balance traffic to the local VMs. Any active connections to the remote data center are allowed to complete.

Table 7-33 lists the VM probe attributes.

Table 7-33 VM Probe Attributes

Field	Action
Max CPU Burst Threshold	Percentage of CPU usage by the local VMs at which the ACE begins to burst traffic to the remote VMs. Enter a value from 1 to 99. The default is 99.
Min CPU Burst Threshold	Percentage of CPU usage by the local VMs below which the ACE stops bursting traffic to the remote VMs. Enter a value from 1 to 99. The default is 99.
Max Memory Burst Threshold	Percentage of memory usage by the local VMs at which the ACE begins to burst traffic to the remote VMs. Enter a value from 1 to 99. The default is 99.
Min Memory Burst Threshold	Percentage of memory usage by the local VMs below which the ACE stops bursting traffic to the remote VMs. Enter a value from 1 to 99. The default is 99.
VM Controller Name	Identifier of the VM Controller that is associated with the local VMs and that you configured in the "Configuring and Verifying a VM Controller Connection" section on page 7-30. Click the radio button for the VM Controller.

To associate the VM probe with a server farm, see the "Configuring Server Farms" section on page 7-31.

Related Topics

- Configuring Dynamic Workload Scaling, page 7-27
- Configuring Server Farms, page 7-31
- Dynamic Workload Scaling Overview, page 7-4

Configuring DNS Probe Expect Addresses

You can specify the IP address that the ACE expects to receive in response to a DNS request. 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.

Assumption

A DNS probe has been configured. See the "Configuring Health Monitoring for Real Servers" section on page 7-52 for more information.

Procedure

Step 1 Choose **Config > Devices >** *context* **> Load Balancing > Health Monitoring**.

The Health Monitoring table appears.

Step 2 In the Health Monitoring table, choose the DNS probe that you want to configure with an expected IP address.

The Expect Addresses table appears.

Step 3 In the Expect Addresses table, 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 IPv4/IPv6 Address field, enter the IP address that the ACE appliance is to expect as a server response to a DNS request. You can enter multiple addresses in this field. However, you cannot mix IPv4 and IPv6 addresses.

Note IPv6 requires ACE module and ACE appliance software Version A5(1.0) or later.

Step 5 Do one of the following:

- Click **Deploy Now** to deploy this configuration on the ACE and save your entries to the running-configuration and startup-configuration files.
- Click **Cancel** to exit this procedure without saving your entry and to return to the Expect Addresses table.
- Click **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 7-52
- DNS Probe Attributes, page 7-59
- Displaying Health Monitoring Statistics and Status Information, page 7-79

Configuring Headers for HTTP and HTTPS Probes

You can specify header fields for HTTP and HTTPS probes.

Assumption

An HTTP or HTTPS probe has been configured. See the "Configuring Health Monitoring for Real Servers" section on page 7-52 for more information.

Procedure

Step 1 Choose **Config > Devices >** *context* **> Load Balancing > Health Monitoring**.

The Health Monitoring table appears.

Step 2 In the Health Monitoring table, choose the HTTP or HTTPS probe that you want to configure with a header.

The Probe Headers table appears.

Step 3 In the Probe Headers table, click **Add** to add an entry, or choose an existing entry and click **Edit** to modify it.

The Probe Headers configuration pane appears.

Step 4 In the Header Name field of the Probe Headers configuration pane, choose 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 Do one of the following:

- Click **Deploy Now** to deploy this configuration on the ACE and save your entries to the running-configuration and startup-configuration files.
- Click **Cancel** to exit this procedure without saving your entry and to return to the Probe Headers table.
- Click **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 7-52
- HTTP Probe Attributes, page 7-61
- HTTPS Probe Attributes, page 7-63
- Displaying Health Monitoring Statistics and Status Information, page 7-79

Configuring Health Monitoring Expect Status

You can configure a single or range of code responses that the ACE expects from the probe destination. 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.

Assumption

An FTP, HTTP, HTTPS, RTSP, SIP-TCP, SIP-UDP or SMTP probe has been configured. See the "Configuring Health Monitoring for Real Servers" section on page 7-52 for more information.

Procedure

Step 1 Choose **Config > Devices >** *context* **> Load Balancing > Health Monitoring**.

The Health Monitoring table appears.

Step 2 In the Health Monitoring table, choose the probe that you want to configure for expect status codes, and click the **Expect Status** tab.

The Expect Status table appears.

Step 3 In the Expect Status table, click **Add** to add an entry, or select an existing entry and click **Edit** to modify it.

The Expect Status configuration pane appears.

Step 4 In the Expect Status configuration pane, configure a single expect status code as follows:

a. In the Min. Expect Status Code field, enter the expect status code for this probe. Valid entries are 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 In the Expect Status configuration pane, configure a range of expect status codes as follows:

a. In the Min. Expect Status Code, enter the lower limit of the range of status codes. Valid entries are from 0 to 999.

b. In the Max. Expect Status Code, enter the upper limit of a range of status codes. Valid entries are 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 Do one of the following:

- Click **Deploy Now** to deploy this configuration on the ACE and save your entries to the running-configuration and startup-configuration files.
- Click **Cancel** to exit this procedure without saving your entries and to return to the Expect Status table.
- Click **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 7-52
- FTP Probe Attributes, page 7-61
- HTTP Probe Attributes, page 7-61
- SMTP Probe Attributes, page 7-71
- Displaying Health Monitoring Statistics and Status Information, page 7-79

Configuring an OID for SNMP Probes

You can configure OID queries to probe the server. 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 the "Configuring Health Monitoring for Real Servers" section on page 7-52 for more information.

Procedure

Step 1 Choose **Config > Devices >** *context* **> Load Balancing > Health Monitoring**.

The Health Monitoring table appears.

Step 2 In the Health Monitoring table, choose the SNMP probe for which you want to specify an OID.

The SNMP OID for Server Load Query table appears.

Step 3 In the SNMP OID for Server Load Query table, click **Add** to add an entry, or choose an existing entry and click **Edit** to modify it.

The SNMP OID configuration pane appears.

Step 4 In the SNMP OID field of the SNMP OID configuration pane, 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 Max. 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 from 1 to 4294967295.

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 as follows:

- When the OID value is based on a percent, valid entries are integers from 1 to 100.
- When the OID is based on an absolute value, valid entries are from 1 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 from 0 to 16000.

Step 8 Do one of the following:

- Click **Deploy Now** to deploy this configuration on the ACE and save your entries to the running-configuration and startup-configuration files.
- Click **Cancel** to exit this procedure without saving your entries and to return to the SNMP OID table.
- Click **Next** to deploy your entries and to add another item to the SNMP OID table.

Related Topics

- Configuring Health Monitoring for Real Servers, page 7-52
- SNMP Probe Attributes, page 7-71
- Displaying Health Monitoring Statistics and Status Information, page 7-79

Displaying Health Monitoring Statistics and Status Information

You can display statistics and status information for a particular probe.

Procedure

Step 1 Choose **Config > Devices >** *context* **> Load Balancing > Health Monitoring**.

The Health Monitoring table appears.

Step 2 In the Health Monitoring table, choose a probe from the Health Monitoring table, and click **Details**.

The **show probe** *name* **detail** CLI command output appears. For details on the displayed output fields, see the *Cisco ACE Module Server Load-Balancing Configuration Guide* or the *Cisco ACE 4700 Series Appliance Server Load-Balancing Configuration Guide*, Chapter 4, Configuring Health Monitoring.

Note For a DNS probe, the detailed probe results always identify a default DNS domain of www.Cisco.com.

Step 3 Click **Update Details** to refresh the output for the **show probe** *name* **detail** CLI command.

Step 4 Click **Close** to return to the Health Monitoring table.

Related Topics

- Configuring Health Monitoring for Real Servers, page 7-52

Configuring Secure KAL-AP

You can configure a secure keepalive-appliance protocol (KAL-AP) associated with a virtual context. A KAL-AP on the ACE enables communication between the ACE and a Global Site Selector (GSS), which sends KAL-AP requests to report the server states and loads for global-server load-balancing (GSLB) decisions. The ACE uses KAL-AP through a UDP connection to calculate weights and provide information for server availability to the KAL-AP device. The ACE acts as a server and listens for KAL-AP requests. When KAL-AP is initialized on the ACE, the ACE listens on the standard 5002 port for any KAL-AP requests. You cannot configure any other port.

The ACE supports secure KAL-AP for MD5 encryption of data between it and the GSS. For encryption, you must configure a shared secret as a key for authentication between the GSS and the ACE context.

When configuring a KAL-AP, you can use the wildcard KAL-AP GSS IP address (0.0.0.0) to establish a secure communications channel between the ACE and multiple GSS devices that use the same MD5 encryption secret.

Guidelines and Restrictions

Use the following guidelines and restrictions when using the 0.0.0.0 wildcard KAL-AP GSS IP address:

- The wildcard GSS IP address feature requires ACE software Version A5(2.0) or later.
- Use the wildcard IP address when both the following conditions exist:
  - All GSS devices in the cluster use a secure channel for KAL-AP message exchange with ACE. Do not use the wildcard IP address if any GSS in the cluster uses an unsecure channel.
  - All or a set of GSS devices in the cluster use the same MD5 secret.

  Note You can only use the wildcard VIP address for one set of GSS devices that use the same MD5 secret. You must configure all other GSS devices individually for KAL-AP.

- When removing a KAL-AP IP address, using the wildcard IP address removes only those GSS IP addresses that use the secret associated with the wildcard value. KAL-AP IP addresses that were defined using a specific GSS IP addresses remain and must be removed individually.

Assumptions

- You have created a virtual context that specifies the Keepalive Appliance Protocol over UDP.
- You have enabled KAL-AP on the ACE by configuring a management class map and policy map, and apply it to the appropriate interface.

Procedure

Step 1 Choose **Config > Devices >** *context* **> Load Balancing > Secure KAL-AP**.

The Secure KAL-AP table appears.

Step 2 In the Secure KAL-AP table, click **Add** to configure secure KAL-AP for MD5 encryption of data.

The Secure KAL-AP configuration window appears.

Step 3 In the IP Address field of the Secure KAL-AP configuration window, enable secure KAL-AP by configuring the VIP address for the GSS.

Using dotted-decimal notation (for example, 192.168.11.1), enter the IP address of a specific GSS device or enter the wildcard value (0.0.0.0) if all GSS devices in the cluster use the same MD5 encryption secret (see the "Guidelines and Restrictions" section on page 7-80).

Step 4 In the Hash Key field, enter the MD5 encryption method shared secret between the KAL-AP device and the ACE.

Enter the shared secret as a case-sensitive string with no spaces and a maximum of 31 alphanumeric characters. The ACE supports the following special characters in a shared secret:

, . / = + - ^ @ ! % ~ # $ * ( )

Step 5 Do one of the following:

- Click **Deploy Now** to deploy this configuration on the ACE and save your entries to the running-configuration and startup-configuration files. The ACE validates the secure KAL-AP configuration and deploys it.
- Click **Cancel** to exit this procedure without accepting your entries and to return to the Secure KAL-AP table.
- Click **Next** to accept your entries.

Related Topics

- Creating Virtual Contexts, page 5-2
- Setting Match Conditions for Layer 3/Layer 4 Management Traffic Class Maps, page 13-12