Introducing the Global Site Selector

This chapter describes the Cisco Global Site Selector (GSS) and introduces you to the terms and concepts necessary to help you understand and operate the GSS device.

This chapter contains the following major sections:

GSS Overview

Server load-balancing devices, such as the Cisco Content Services Switch (CSS) and Cisco Content Switching Module (CSM) that are connected to a corporate LAN or the Internet, can balance content requests among two or more servers containing the same content. Server load-balancing devices ensure that the content consumer is directed to the host that is best suited to handle that consumer's request.

Organizations with a global reach or businesses that provide web and application hosting services require network devices that can perform complex request routing to two or more redundant, geographically dispersed data centers. These network devices need to provide fast response times and disaster recovery and failover protection through global server load balancing, or GSLB.

The Cisco Global Site Selector (GSS) platform allows you to leverage global content deployment across multiple distributed and mirrored data locations, optimizing site selection, improving Domain Name System (DNS) responsiveness, and ensuring data center availability.

The GSS is inserted into the traditional DNS routing hierarchy and is closely integrated with the Cisco CSS, Cisco CSM, or third-party server load balancers (SLBs) to monitor the health and load of the SLBs in your data centers. The GSS uses this information and user-specified routing algorithms to select the best-suited and least-loaded data center in real time.

The GSS can detect site outages, ensuring that web-based applications are always online and that customer requests to data centers that suddenly go offline are quickly rerouted to available resources.

The GSS offloads tasks from traditional DNS servers by taking control of the domain resolution process for parts of your domain name space, responding to requests at a rate of thousands of requests per second.

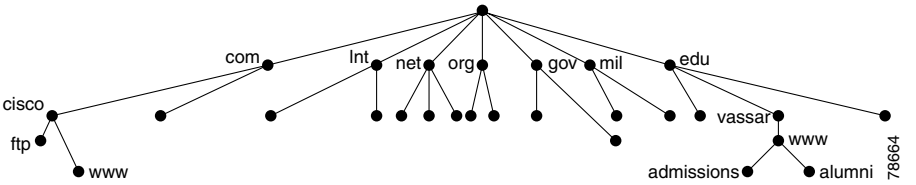
DNS Routing

This section explains some of the key DNS routing concepts behind the GSS.

Since the early 1980s, content routing on the Internet has been handled using the Domain Name System (DNS), a distributed database of host information that maps domain names to IP addresses. Almost all transactions that occur across the Internet rely on DNS, including electronic mail, remote terminal access such as Telnet, file transfers using FTP, and web surfing. DNS uses easy-to-remember alphanumeric host names instead of numeric IP addresses that bear no relationship to the content on the host.

With DNS, you can manage a nearly infinite number of host names referred to as the domain name space (Figure 1-1). DNS allows local administration of segments (individual domains) of the overall database, but allows for data in any segment to be available across the entire network. This process is referred to as *delegation*.

Figure 1-1 Domain Name Space



DNS Name Servers

Information about the domain name space is stored on name servers that are distributed throughout the Internet. Each server stores the complete information about its small part of the total domain name space. This space is referred to as a DNS *zone*. A zone file contains DNS information for one domain ("mycompany.com") or subdomain ("gslb.mycompany.com"). The DNS information is organized into lines of information called resource records.

End users requiring data from a particular domain or machine generate a recursive DNS request on their client that is sent first to the local name server (NS), also referred to as the *D-proxy*. The D-proxy returns the IP address of the requested domain to the end user.

The DNS structure is based on a hierarchical tree structure that is similar to common file systems. The key components in this infrastructure are as follows:

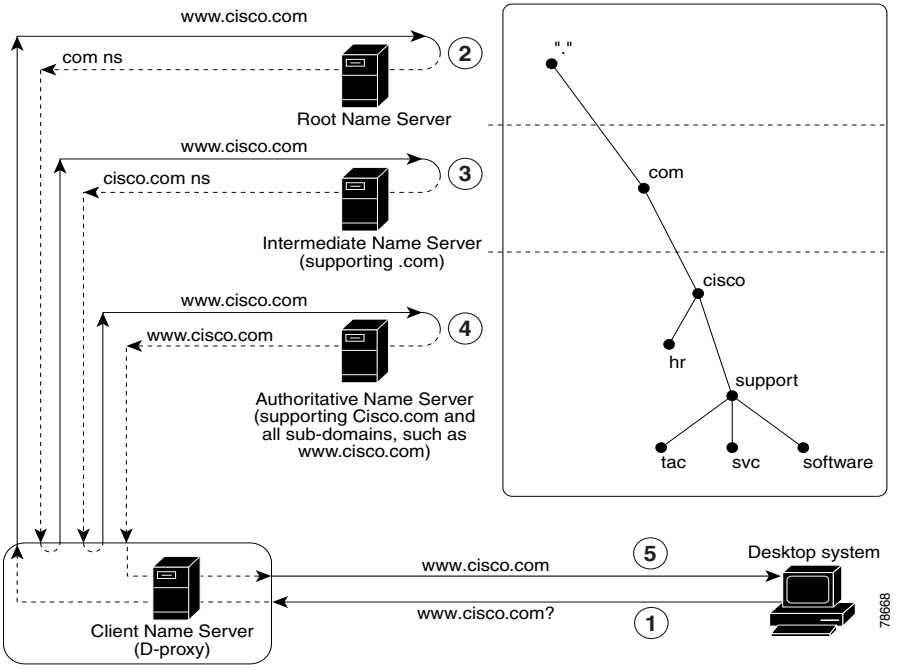
- **DNS Resolvers**—Clients that access client name servers.
- **Client Name Server**—A server running DNS software that has the responsibility of finding the requested web site. The client name server is also referred to as the client DNS proxy (D-proxy).
- **Root Name Servers**—A server residing at the top of the DNS hierarchy. The root name server knows how to locate every extension after the period "." in the host name. There are many top-level domains. The most common top-level domains include .org, .edu, .net, .gov, and .mil. Approximately 13 root servers worldwide handle all Internet requests.
- **Intermediate Name Server**—A server used for scaling purposes. When the root name server does not have the IP address of the authoritative name server, it sends the requesting client name server to an intermediate name server. The intermediate name server then refers the client name server to the authoritative name server.
- **Authoritative Name Server**—A server run by an enterprise or one that is outsourced to a service provider and is authoritative for the domain requested. The authoritative name server responds directly to the client name server (not to the client) with the requested IP address.

Request Resolution

If the local D-proxy does not have the information requested by the end user, it sends out iterative requests to the name servers that it knows are authoritative for the domains close to the requested domain. For example, a request for www.cisco.com causes the local D-proxy to check first for another name server that is authoritative for www.cisco.com.

The process outlined in Figure 1-2 summarizes the sequence performed by the DNS infrastructure to return an IP address when a client tries to access the www.cisco.com website.

Figure 1-2 DNS Request Resolution



1. The resolver (client) sends a query for www.cisco.com to the local client name server (D-proxy).
2. The local D-proxy does not have the IP address for www.cisco.com so it sends a query to a root name server (".") asking for the IP address. The root name server responds to the request by either:
   - Referring the D-proxy to the specific name server supporting the .com domain.
   - Sending the D-proxy to an intermediate name server that knows the address of the authoritative name server for www.cisco.com. This method is referred to as an *iterative query.*
3. The local D-proxy sends a query to the intermediate name server that responds by referring the D-proxy to the authoritative name server for cisco.com and all the associated subdomains.
4. The local D-proxy sends a query to the cisco.com authoritative name server that is the top-level domain. In this example, www.cisco.com is a sub-domain of cisco.com, so this name sever is authoritative for the requested domain and sends the IP address to the name server (D-proxy).
5. The name server (D-proxy) sends the IP address (172.16.56.76) to the client browser. The browser uses this IP address and initiates a connection to the www.cisco.com website.

Using the GSS for GSLB

The GSS addresses critical disaster recovery requirements by globally load-balancing distributed data centers. The GSS coordinates the efforts of geographically dispersed SLBs in a global network deployment for the following Cisco products:

- Cisco Content Services Switch 11500, 11000, or 11150
- Cisco Content Switching Module (CSM) for the Catalyst 6500 series switches
- Cisco LocalDirector
- Cisco IOS SLB
- Cisco router using the DRP agent for network proximity
- Any server that is capable of responding to HTTP HEAD, ICMP, or TCP requests
- Cisco router with cache modules
- Cisco Cache Engines

The GSS supports over 4000 separate virtual IP (VIP) addresses. It coordinates the activities of SLBs by acting as the authoritative DNS server for those devices under its control.

Once the GSS becomes responsible for GSLB services, the DNS process migrates to the GSS. The DNS configuration is the same process as described in the "Request Resolution" section. The only exception is that the NS-records point to the GSSs located at each data center. The GSS determines which data center site should receive the client traffic.

As the authoritative name server for a domain or subdomain, the GSS considers the following additional factors when responding to a DNS request:

- **Availability**—The servers that are online and available to respond to the query
- **Proximity**—The server that responded the fastest to a query
- **Load**—The type of traffic load handled by each server in the domain
- **Source of the Request**—The name server (D-proxy) requesting the content
- **Preference**—The first, second, or third choice of the load-balancing algorithm to use when responding to a query

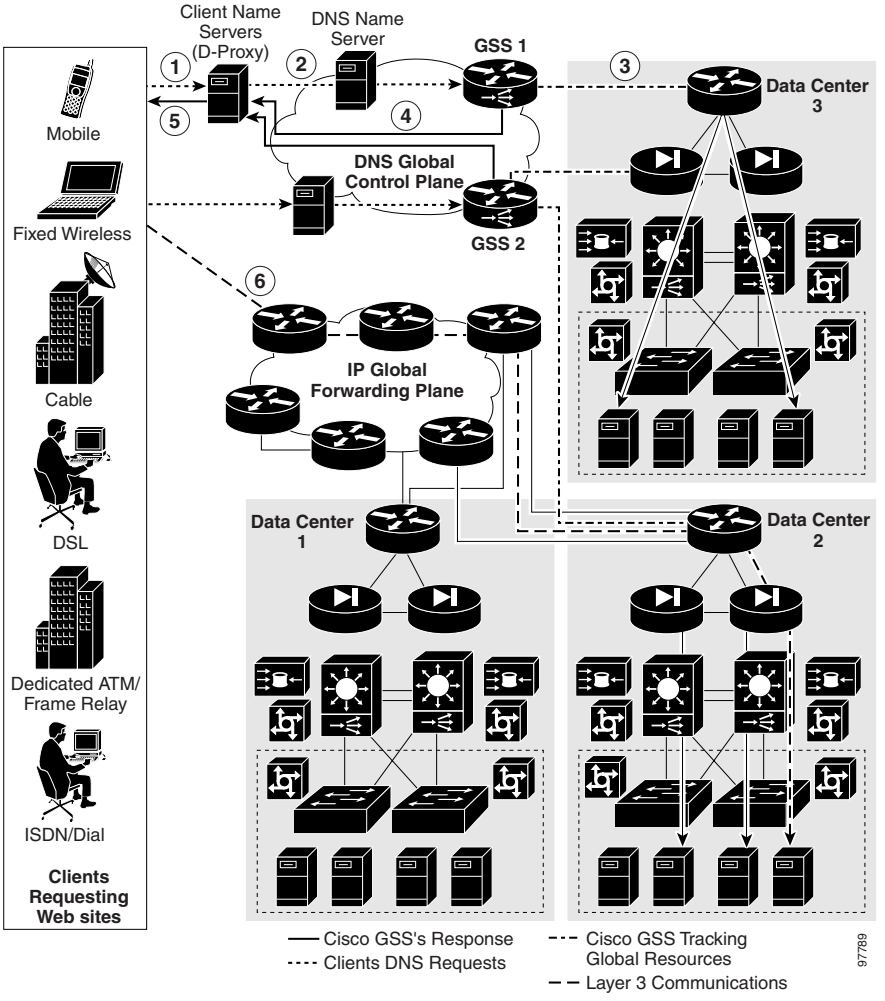
This type of global server load balancing helps to ensure that the end users are always directed to resources that are online, and that requests are forwarded to the most suitable device, resulting in faster response time for users.

When resolving DNS requests, the GSS performs a series of distinct operations that take into account the resources under its control and return the best possible answer to the requesting client's D-proxy.

The following process, illustrated in Figure 1-3, outlines how the GSS interacts with various clients as part of the website selection process to return the IP address of the requested content site.

1. A client starts to download an updated version of software from www.cisco.com and types **www.cisco.com** in the location or address field of the browser. This application is supported at three different data centers.
2. The DNS global control plane infrastructure processes the request and the request arrives at a GSS device.
3. The GSS sends to the client the IP address of the "best" server load balancer, in this case the SLB at Data Center 2.
4. The web browser processes the transmitted IP address.
5. The client is directed to the SLB at Data Center 2 by the IP control and forwarding plane.

Figure 1-3 GLSB Using the Cisco Global Site Selector



6. The Cisco GSS offloads the site selection process from the DNS global control plane. The request and site selection are based on the load and health information with customer-controlled load-balancing algorithms. The Cisco GSS, in real time, selects a data center that is available and not overloaded.

GSS Architecture

This section describes the following key components of a GSS deployment, including hardware and software, as well as GSS networking concepts:

- Global Site Selectors and Global Site Selector Managers
- DNS Rules
- Locations and Regions
- Owners
- Source Addresses and Source Address Lists
- Hosted Domains and Domain Lists
- Answers and Answer Groups
- Keepalives
- Balance Methods
- Traffic Management Load Balancing

Global Site Selectors and Global Site Selector Managers

The Global Site Selector solution relies on three distinct but closely related devices:

- Primary GSSM
- GSS
- Standby GSSM

All GSS devices in the network, including the primary GSSM and standby GSSM, are delegated authority for domains, respond to DNS queries and perform keepalives, and use their local CLI for basic network management. All GSS devices depend on the primary GSSM to provide centralized, shared global server load-balancing functionality.

Primary GSSM

The primary GSSM is a Cisco Global Site Selector running GSS software and performing content routing in addition to centralized management and shared global server load-balancing functions for the GSS network.

The primary GSSM serves as the organizing point of the GSS network, hosting the embedded GSS database that contains configuration information for all your GSS resources, such as individual GSSs and DNS rules. All connected GSS devices report their status to the primary GSSM.

On the primary GSSM, you monitor and administer GSS devices using either of the following methods:

- GUI (graphical user interface), as described in this document
- CLI commands, as described in the *Cisco Global Site Selector CLI-Based Global Server Load-Balancing Configuration Guide*

All configuration changes are communicated automatically to each device managed by the primary GSSM.

Any GSS device can serve as a the single, primary GSSM on a configured system.

GSS

The GSS is a Cisco Global Site Selector running GSS software and performing routing of DNS queries based on DNS rules and conditions configured using the primary GSSM.

Each GSS is known to and synchronized with the primary GSSM.

You manage each GSS individually through its Command Line Interface (CLI). Graphical user interface (GUI) support is not available on a GSS or on a standby GSSM.

Standby GSSM

As is the case for any GSS, the standby GSSM is a Global Site Selector running GSS software and performing routing of DNS queries based on DNS rules and conditions configured using the primary GSSM. Additionally, the standby GSSM is configured to function as the primary GSSM should the designated primary GSSM go offline or become unavailable to communicate with other GSS devices.

When the standby GSSM operates as the interim primary GSSM, it contains a duplicate copy of the embedded GSS database currently installed on the primary GSSM. Both CLI and GUI support are also available on the standby GSSM once you configure it as the interim primary GSSM. While operating as the primary GSSM, you can monitor GSS behavior and make configuration changes, as necessary.

Any configuration or network changes affecting the GSS network are synchronized between the primary and the standby GSSM so that the two devices are never out of sequence.

To enable the standby GSSM as the primary GSSM, use the **gssm standby-to-primary** CLI command. Ensure that your original primary GSSM is offline before you attempt to enable the standby GSSM as the new primary GSSM. Having two primary GSSMs active at the same time may result in the inadvertent loss of configuration changes for your GSS network. If this dual primary GSSM configuration occurs, the two primary GSSMs revert to standby mode and you must reconfigure one of the GSSMs as the primary GSSM.

The standby GSSM can temporarily assume the role as the primary GSSM in the event that the primary GSSM is unavailable (for example, you need to move the primary GSSM or you want to take it offline for repair or maintenance). Switching roles between the designated primary GSSM and the standby GSSM is intended to be a temporary GSS network configuration until the original primary GSSM can be brought back online. Once the original primary GSSM is available, reassign the two GSSMs to their original roles in the GSS network as described in the *Cisco Global Site Selector Administration Guide*.

DNS Rules

The GSS uses DNS rules that you configure at the primary GSSM to:

- Provide you with centralized command and control of how the GSS globally load balances a given hosted domain
- Define the IP addresses to send to the client's name server (D-proxy)
- Define the recovery method to use (using a maximum of three load-balance clauses)

Each DNS rule determines how the GSS responds to each query it receives by matching requests received from a known source, or D-proxy, to the most suitable member of a collection of name servers or virtual IP addresses (VIPs).

Each DNS rule takes into account the following variables:

- The source IP address of the requesting D-proxy
- The requested hosted domain
- An answer group, which is a group of resources considered for the response
- A balance method, which is an algorithm for selecting the best server, together with an answer group, makes up a clause
- Advanced traffic management load-balancing functions such as DNS sticky and network proximity

A DNS rule defines how a request is handled by the GSS by answering the following question:

> *When traffic arrives from a DNS proxy, querying a specific domain name, which resources should be considered for the response, and how should they be balanced?*

Each GSS network supports a maximum of 4000 DNS rules.

A maximum of three possible response answer group and balance method clauses are available for each DNS rule. Each clause specifies that a particular answer group serve the request and a specific balance method be used to select the best resource from that answer group. These clauses are evaluated in order, with parameters established to determine when a clause should be skipped if the first answer group and balance method specified does not yield an answer, and the next clause is to be used.

Refer to Chapter 7, Building and Modifying DNS Rules, for procedures about constructing the DNS rules that govern all global server load balancing on your GSS network.

Locations and Regions

As your GSS network expands, the job of organizing and administering your GSS resources—locations, regions, answers and answer groups, domain lists, and DNS rules—becomes more complex. The GSS provides the following features to help you organize your resources:

- **Locations**—Logical groupings for GSS resources that correspond to geographical areas such as a city, data center, or content site
- **Regions**—Higher-level geographical groupings that contain one or more locations

In addition to allowing you to easily sort and navigate long lists of answers and DNS rules, the use of logical groupings such as locations and regions makes it easier to perform bulk administration of GSS resources. For example, from the primary GSSM, you can suspend or activate all answers linked to a particular GSS data center, shutting down a site for scheduled maintenance and then bringing it back online with only a few mouse clicks.

Refer to Chapter 2, Configuring Resources, for information on configuring locations and regions.

Owners

An owner is an entity that owns web content and uses the GSS to manage access to the content. As locations and regions allow you to geographically configure your GSS network, owners allow you to organizationally configure your GSS network.

For example, a service provider using the GSS to manage multiple hosting sites might create an owner for each web- or application-hosting customer. With this organizational scheme, you can associate and manage the following elements through each owner: domain lists containing that owner's hosted content, DNS rules, answer groups, and source address lists that specify how traffic to those domains should be processed.

Deployed on a corporate intranet, you can configure owners to segregate GSS resources on a department-by-department basis, or to allocate specific resources to IT personnel. For example, you can create an owner for the finance, human resources, and sales departments so that resources corresponding to each can be viewed and managed together.

Refer to Chapter 2, Configuring Resources, for information on configuring owners.

Source Addresses and Source Address Lists

The term *source address* refers to the source of DNS queries received by the GSS. Source addresses typically point to an IP address or block of addresses that represent client D-proxies from which the queries originate.

Using a DNS rule, the GSS matches source addresses to domains hosted by the GSS using one of a number of different balance methods.

Source addresses are taken from the D-proxy (the local name server) to which a requesting client issued a recursive request. The D-proxy sends the client queries to multiple name servers, eventually querying the GSS, which matches the D-proxy source address against its list of configured source addresses.

DNS queries received by the GSS do not have to match a specific D-proxy to be routed; default routing can be performed on requests that do not emanate from a known source address. By default, the GSS provides a fail-safe "Anywhere" source address list. Incoming queries that do not match your configured source address lists are matched to this list.

In addition to specific IP addresses, source addresses can also be set up to represent address blocks using variable-prefix-length classless interdomain routing (CIDR) block masking. The following examples illustrate acceptable GSS source addresses:

```
192.168.1.110
192.168.1.110/32
192.168.1.0/24
192.168.0.0/16
```

Source addresses are grouped into lists, referred to as *source address lists*, for the purposes of routing requests. Source address lists can contain 1 to 30 source addresses or unique address blocks. Each GSS supports a maximum of 60 source address lists.

Refer to Chapter 3, Configuring Source Address Lists, for information on configuring source address lists.

Hosted Domains and Domain Lists

A hosted domain (HD) is any domain or subdomain that has been delegated to the GSS and configured using the primary GSSM GUI for DNS query responses. A hosted domain is a DNS domain name for which the GSS is authoritative.

All DNS queries must match a domain that belongs to a configured domain list, or the GSS denies the query. Queries that do not match domains on any GSS domain lists can also be forwarded by the GSS to an external DNS name server for resolution.

Hosted domains cannot exceed 128 characters in length. The GSS supports domain names that use wildcards. The GSS also supports POSIX 1003.2-extended regular expressions when matching wildcards.

The following examples illustrate domain or subdomain names configured on the GSS:

```
cisco.com
www.cisco.com
www.support.cisco.com
.*\.cisco\.com
```

Domain lists are groups of hosted domains that have been delegated to the GSS. Each GSS can support a maximum of 2000 hosted domains and 2000 hosted domain lists, with a maximum of 500 hosted domains supported for each domain list.

Domain lists are used by the GSS to match incoming DNS requests to DNS rules. After the query domain is found in a domain list and matched to a DNS rule, the balance method clauses of the DNS rule define how the GSS will choose the best *answer* (a VIP, for example) that can service the request.

Refer to Chapter 4, Configuring Domain Lists, for information on configuring domain lists.

Answers and Answer Groups

In a GSS network, the term *answers* refers to resources to which the GSS resolves DNS requests that it receives. The three types of possible answers on a GSS network are as follows:

- **VIP**—Virtual IP (VIP) addresses associated with an SLB such as the Cisco CSS, Cisco CSM, Cisco IOS-compliant SLB, Cisco LocalDirector, a web server, a cache, or any other geographically dispersed device in a global network deployment.
- **Name Server**—Configured DNS name server that can answer queries that the GSS cannot resolve.
- **CRA**—Content routing agents that use a resolution process called DNS race to send identical and simultaneous responses back to a user's D-proxy.

As with domains and source addresses, answers are configured using the primary GSSM GUI by identifying the IP address to which queries can be directed.

Once created, you group answers together as resource pools called *answer groups*. From the available answer groups, the GSS can use a maximum of three possible response answer group and balance method clauses in a DNS rule to select the most appropriate resource to serve a user request. Each balance method provides a different algorithm for selecting one answer from a configured answer group. Each clause specifies that a particular answer group serve the request and a specific balance method be used to select the best resource from that answer group.

Depending on the type of answer, further criteria can be applied to DNS queries to choose the best host. For example, a request that is routed to a VIP associated with a Cisco CSS is routed to the best resource based on load and availability, as determined by the CSS. A request that is routed to a content routing agent (CRA) is routed to the best resource based on proximity, as determined in a DNS race conducted by the GSS.

Refer to Chapter 6, Configuring Answers and Answer Groups, for information on configuring GSS answers and answer groups.

VIP Answers

SLBs use VIP answers to represent content hosted on one or more servers under their control. The use of VIP answers enables the GSS to balance traffic among multiple origin servers, application servers, or transaction servers in a way that results in faster response times for users and less network congestion for the host.

When queried by a client's D-proxy for a domain associated with a VIP answer type, the GSS responds with the VIP address of the SLB best suited to handle that request. The requesting client then contacts the SLB, which load balances the request to the server best suited to respond to the request.

Name Server Answers

A name server answer specifies the IP address of a DNS name server to which DNS queries are forwarded from the GSS.

Using the name server forwarding feature, queries are forwarded to an external (non-GSS) name server for resolution, with the answer passed back to the GSS name server, then on to the requesting D-proxy. A name server answer can act as a guaranteed fallback resource, a way to resolve requests that the GSS cannot resolve itself. The GSS may not be able to resolve such requests because:

- The requested content is unknown to the GSS
- The resources that typically handle such requests are unavailable

The external DNS name server answer forwarded by the GSS may be able to:

- Use DNS server features that are not supported by the GSS, such as mail exchanger (type MX) records
- Use a third-party content provider for failover and error recovery
- Provide access to a tiered DNS system

CRA Answers

The CRA answer relies on content routing agents and the GSS to choose a suitable answer for a given query based on the proximity of two or more possible hosts to the requesting D-proxy.

With the CRA answer, requests received from a particular D-proxy are served by the content server that responds first to the request. Response time is measured using a DNS race, coordinated by the GSS and content routing agents running on each content server. In the DNS race, multiple hosts respond simultaneously to an A-record request. The server with the fastest response time (the shortest network delay between itself and the client's D-proxy) is chosen to serve the content.

The GSS requires the following information before it can initiate a DNS race:

- The delay between the GSS and each of the CRAs in each data center. With this data, the GSS computes how much time to delay the race from each data center so that each CRA starts the race simultaneously.
- The online status of the CRA through the use of keepalives.

The boomerang balance method uses the DNS race to determine the best site. See the "DNS Race (Boomerang)" section for more information on this balance method.

Keepalives

In addition to specifying a resource, each answer also provides you with the option of specifying a *keepalive* for that resource. A keepalive is the method by which the GSS periodically checks to determine if a resource is still active. A keepalive is a specific interaction (handshake) between the GSS and another device using a commonly supported protocol. A keepalive is designed to test if a specific protocol on the device is functioning properly. If the handshake is successful, then the device is available, active, and able to receive traffic. If the handshake fails, then the device is considered to be unavailable and inactive. All answers are validated by configured keepalives and are not returned by the GSS to the D-proxy if the keepalive indicates that the answer is not viable.

The GSS uses keepalives to collect and track information from the online status of VIPs to services and applications running on a server. You can configure a keepalive to continually monitor the online status of a resource and report that information to the primary GSSM. Routing decisions involving that resource consider the reported online status information.

The GSS also supports the use of shared keepalives to minimize traffic between the GSS and the SLBs that it is monitoring. A shared keepalive identifies a common address or resource that can provide status for multiple answers. Shared keepalives are not used with name server or CRA answers.

When configuring a VIP-type answer, you have the option to configure one of several different keepalive types or multiple keepalive types to test for that answer. The primary GSSM supports the assignment of multiple keepalives and destination ports for a specific VIP answer. You can configure a maximum of five different keepalives for a VIP answer in a mix and match configuration of ICMP, TCP, HTTP HEAD, and KAL-AP VIP keepalive types. For TCP or HTTP HEAD keepalives, you may also specify different destination ports to a VIP server. The multi-port keepalive capability allows you monitor a single server and check responses from multiple ports

Note The primary GSSM supports only a single usage of a shared keepalive and a single KAL-AP keepalive when specifying multiple keepalive types.

The following sections explain the various keepalive types supported by the GSS:

- ICMP
- TCP
- HTTP HEAD
- KAL-AP
- CRA
- Name Server
- None
- Adjusting Failure Detection Time for Keepalives

Refer to Chapter 5, Configuring Keepalives, for information on modifying global keepalive parameters and creating shared keepalives.

ICMP

Use an ICMP keepalive when testing a GSS answer that is a VIP address, IP address, or a virtual server IP address. The Internet Control Message Protocol (ICMP) keepalive type monitors the health of resources by issuing queries containing ICMP packets to the configured VIP address (or a shared keepalive address) for the answer. Online status is determined by a response from the targeted address, indicating simple connectivity to the network. The GSS supports a maximum of 750 ICMP keepalives when using the standard detection method and a maximum of 150 ICMP keepalives when using the fast detection method. See the "Adjusting Failure Detection Time for Keepalives" section for details.

TCP

Use a TCP keepalive when testing a GSS answer that is a GSLB device that may be something other than a CSS or CSM. GSLB remote devices may include webservers, LocalDirectors, WAP gateways, and other devices that can be checked using a TCP keepalive. The TCP keepalive initiates a TCP connection to the remote device by performing the three-way handshake sequence.

Once the TCP connection is established, the GSS terminates the connection. You can choose to terminate the connection from two termination methods: Reset (immediate termination using a hard reset) or Graceful (standard three-way handshake termination).

The GSS supports up to 1500 TCP keepalives when using the standard detection method and up to 150 TCP keepalives when using the fast detection method. Refer to the "Adjusting Failure Detection Time for Keepalives" section for details.

HTTP HEAD

Use an HTTP HEAD keepalive when testing a GSS answer that is an HTTP web server acting as a standalone device or managed by an SLB device such as a Cisco CSS, Cisco CSM, Cisco IOS-compliant SLB, or Cisco LocalDirector. The HTTP HEAD keepalive type sends a TCP-formatted HTTP HEAD request to a web server at an address that you specify. The online status of the device is returned in the form of an HTTP Response Status Code of 200 (for example, HTTP/1.0 200 OK).

Once the HTTP HEAD connection is established, the GSS terminates the connection. There are two methods to terminate the connection: Reset (immediate termination using a hard reset) or Graceful (standard three-way handshake termination).

The GSS supports a maximum of 500 HTTP HEAD keepalives when using the standard detection method and a maximum of 100 HTTP HEAD keepalives when using the fast detection method. Refer to the "Adjusting Failure Detection Time for Keepalives" section for details.

KAL-AP

Use a KAL-AP (KeepAlive-Appliance Protocol) keepalive when testing a GSS answer that is a VIP associated with a Cisco CSS or a Cisco CSM. The KAL-AP keepalive type sends a detailed query to both a primary (master) and an optional secondary (backup) circuit address that you specify. The online status and load of each VIP that is specified in the KAL-AP keepalive are returned.

Depending on your GSS network configuration, you can use the KAL-AP keepalive to either query a VIP address directly (KAL-AP By VIP) or query an address with an alphanumeric tag (KAL-AP By Tag). Using a KAL-AP By Tag keepalive query can be useful in the following cases:

- You are attempting to determine the online status of a device that is located behind a firewall that is performing Network Address Translation (NAT).
- There are multiple content rule choices on the SLB.

The GSS supports a maximum of 128 primary and 128 secondary KAL-AP keepalives when using the standard detection method and a maximum of 40 primary and 40 secondary KAL-AP keepalives when using the fast detection method. See the "Adjusting Failure Detection Time for Keepalives" section for details.

CRA

Use the CRA keepalive when testing a CRA answer that responds to DNS race requests. The CRA keepalive type tracks the time required (in milliseconds) for a packet of information to reach the CRA and return to the GSS. The GSS supports a maximum of 200 CRA keepalives.

Name Server

Use the name server keepalive to send a query to the IP address of the name server for a specified query domain (for example, www.cisco.com). The online status for the name server answer is determined by the ability of the name server for the query domain to respond to the query and assign the domain to an address. The GSS supports a maximum of 100 name server keepalives.

None

With the keepalive set to None, the GSS assumes that the named answer is always online. Setting the keepalive type to None prevents your GSS from taking online status or load into account when routing. However, a keepalive of None can be useful under certain conditions, such as when adding devices to your GSS network that are not suited to other keepalive types. ICMP is a simple and flexible keepalive type that works with most devices. Using ICMP is often preferable to using the None option.

Adjusting Failure Detection Time for Keepalives

Failure detection time, for the GSS, is the amount of time between when a device fails (the answer resource goes offline) and when the GSS realizes the failure occurred. If a response packet fails to arrive back to the GSS within this window, the answer is marked offline.

The GSS supports two failure detection modes: standard and fast.

With standard mode, the failure detection time is typically 60 seconds before the GSS detects that a failure has occurred. Standard mode allows adjustment of the following parameters:

- **Response Timeout**—The length of time allowed before the GSS retransmits data to a device that is not responding to a request. The valid entries are 20 to 60 seconds. The default is 20 seconds.
- **Minimum Interval—**The minimum interval with which the GSS attempts to schedule a keepalive. The valid entries are 40 to 255 seconds. The default is 40 seconds.

With fast mode, the GSS controls the failure detection time by using the following keepalive transmission interval formula:

(*# Ack'd Packets* * (*Response TO* + (*Retry TO* * *# of Retries*))) + *Timed Wait*

where:

 # Ack'd Packets = Number of packets that require some form of acknowledgement

Response TO = Response Timeout, which is the length of time to wait for a reply for a packet that requires acknowledgement

Retry TO = Retry Timeout, which is the length of time to wait for a reply for a retransmitted packet

# of Retries = Number of Retries, which is the number of times the GSS retransmits packets to a potentially failed device before declaring the device offline

Timed Wait = Time for the remote side of the connection to close (TCP-based keepalive only)

Table 1-1 summarizes how the GSS calculates the fast keepalive transmission rates for a single keepalive per answer.

Table 1-1 Keepalive Transmission Rates for a Single Keepalive Per Answer

	# Ack'd Packets (Fixed Value)	Response TO (Fixed Value)	Retry TO (Fixed Value)	# of Retries (User Selectable)	Timed Wait (Fixed Value)	Transmission Interval
KAL-AP	1	2 seconds	2 seconds	1	0	4 seconds
ICMP	1	2 seconds	2 seconds	1	0	4 seconds
TCP (RST)	1	2 seconds	2 seconds	1	0	4 seconds
TCP (FIN)	2	2 seconds	1 second	1	2 seconds	10 seconds
HTTP HEAD (RST)	2	2 seconds	2 seconds	1	0	8 seconds
HTTP HEAD (FIN)	3	2 seconds	2 seconds	1	2 seconds	14 seconds

For a TCP (RST) connection, the default transmission interval for a TCP keepalive is as follows:

(1 * (2 + (2 * 1))) + 0 = 4 seconds

You can adjust the number of retries for the ICMP, TCP, HTTP HEAD, and KAL-AP keepalive types. The number of retries defines the number of times the GSS retransmits packets to a potentially failed device before declaring the device offline. The GSS supports a maximum of 10 retries, with a default of 1. As you adjust the number of retries, you change the detection time determined by the GSS. By increasing the number of retries, you increase the detection time. Reducing the number of retries decreases the detection time.

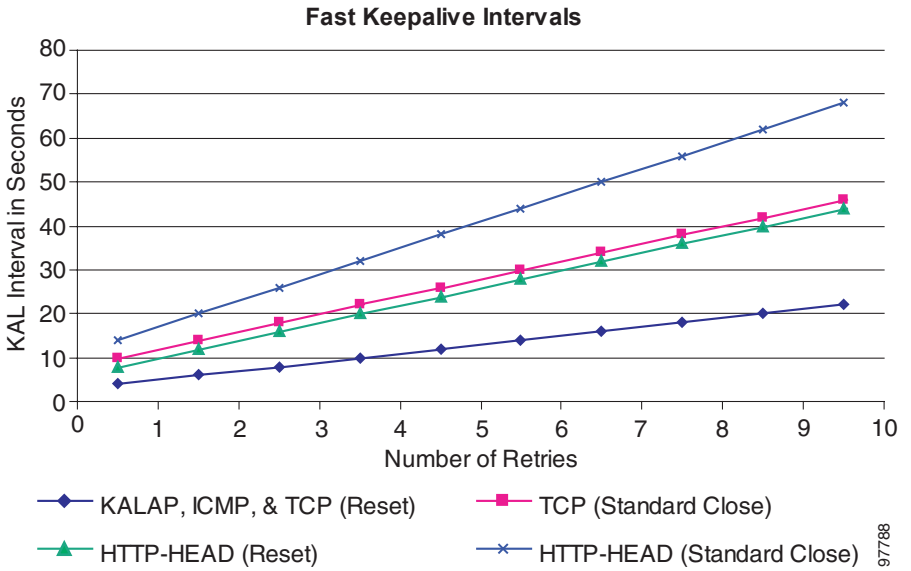
The GSS associates the number of retries value with every packet that requires some form of acknowledgement before continuing with a keepalive cycle (ICMP requests, TCP SYN, or TCP FIN). For example, to fully complete a TCP-based keepalive cycle, the TCP-based keepalive retries the SYN packet for the specified number of retries and then retries the FIN packet for the specified number of retries.

In the above example of a TCP (RST) connection, if you change the number of retries from the default value of 1 to a setting of 5, the transmission interval would be as follows:

(1 * (2 + (2 * 5))) + 0 = 12 seconds

Figure 1-4 illustrates the effect on the keepalive transmission interval as you increase the number of retries value.

Figure 1-4 Effect of the Number of Retries Value on the Keepalive Transmission Interval



You can also define the number of consecutive successful keepalive attempts (probes) that must occur before the GSS identifies that an offline answer is online. The GSS monitors each keepalive attempt to determine if the attempt was successful. The number of successful probes parameter identifies how many consecutive successful keepalive attempts the GSS must recognize before bringing an answer back online and reintroducing it back into the GSS network.

The primary GSSM allows you to assign multiple keepalives for a single VIP answer. You can configure a maximum of five different keepalives for a VIP answer in a mix and match configuration of ICMP, TCP, HTTP HEAD, and KAL-AP VIP keepalive types. In this configuration, the failure detection times are based on the calculated transmission levels identified for each of the different keepalives associated with an answer.

Balance Methods

The GSS supports six unique balance methods that allow you to specify how a GSS answer should be selected to respond to a given DNS query. Each balance method provides a different algorithm for selecting one answer from a configured answer group. The following sections explain the balance methods supported by the GSS:

- Ordered List
- Round-Robin
- Weighted Round-Robin
- Least Loaded
- Hashed
- DNS Race (Boomerang)

Ordered List

When the GSS uses the ordered list balance method, each resource within an answer group (for example, an SLB VIP or a name server) is assigned a number that corresponds to the rank of that answer within the group. The number you assign represents the order of the answer on the list. Subsequent VIPs or name servers on the list will only be used if preceding VIPs or name servers on the list are unavailable. The GSS supports gaps in numbering in an ordered list.

Note For answers that have the same order number in an answer group, the GSS will use only the first answer that contains the number. We recommend that you specify a unique order number for each answer in an answer group.

Using the ranking of each answer, the GSS tries each resource in the order that has been assigned, selecting the first available “live” answer to serve a user request. List members are given precedence and tried in order, and a member is not used unless all previous members fail to provide a suitable result.

The ordered list method is useful in managing resources across multiple content sites in which a deterministic method for selecting answers is required.

See the "Balance Method Options for Answer Groups" section for information on how the GSS determines which answer to select when using the ordered list balance method.

Round-Robin

When the GSS uses the round-robin balance method, each resource within an answer group is tried in turn. The GSS cycles through the list of answers, selecting the next answer in line for each request. In this way, the GSS can resolve requests by evenly distributing the load among possible answers.

The round-robin balance method is useful when balancing requests among multiple, active data centers that are hosting identical content; for example, between SLBs at a primary and at an active standby site that serves requests.

See the "Balance Method Options for Answer Groups" section for information on how the GSS determines which answer to select when using the round-robin balance method.

Weighted Round-Robin

As performed by the round-robin balance method, the weighted round-robin method also cycles through a list of defined answers to choose each available answer in turn. However, with weighted round-robin, an additional "weight" factor is assigned to each answer, biasing the GSS toward certain servers so that they are used more often.

See the "Balance Method Options for Answer Groups" section for information on how the GSS determines which answer to select when using the weighted round-robin balance method.

Least Loaded

When the GSS uses the least-loaded balance method, the GSS resolves requests to the least loaded of all resources, as reported by the KAL-AP keepalive process, which provides the GSS with detailed information on the SLB load and availability.

The least-loaded balance method resolves the request by determining the least number of connections on a CSM or the least-loaded CSS.

See the "Balance Method Options for Answer Groups" section for information on how the GSS determines which answer to select when using the least loaded balance method.

Hashed

When the GSS uses the hashed balance method, elements of the client's DNS proxy IP address and the requesting client's domain are extracted to create a unique value, referred to as a hash value. The unique hash value is attached to and used to identify a VIP that is chosen to serve the DNS query.

The use of hash values makes it possible to "stick" traffic from a particular requesting client to a specific VIP, ensuring that future requests from that client are routed to the same VIP. This type of continuity can be used to facilitate features, such as online shopping baskets, in which client-specific data is expected to persist even when client connectivity to a site is terminated or interrupted.

The GSS supports the following two hashed balance methods. You can apply one or both hashed balance methods to the specified answer group.

- **By Source Address—**The GSS selects the answer based on a hash value created from the source address of the request.
- **By Domain Name—**The GSS selects the answer based on a hash value created from the requested domain name.

DNS Race (Boomerang)

The GSS supports the DNS race (boomerang) method of proximity routing, which is a type of DNS resolution initiated by the GSS to load balance 2 to 20 sites.

The boomerang method is based on the concept that instantaneous proximity can be determined if a CRA within each data center sends an A-record (IP address) at the exact same time to the client's D-proxy. The DNS race method of DNS resolution gives all CRAs (Cisco content engines or content services switches) a chance at resolving a client request and allows for proximity to be determined without probing the client's D-proxy. The first A-record received by the D-proxy is, by default, considered to be the most proximate.

For the GSS to initiate a DNS race, it needs to establish the following information for each CRA:

- The delay between the GSS and each of the CRAs in each data center. With this data, the GSS computes the length of time to delay the race from each data center, so that each CRA starts the race simultaneously.
- The online status of the CRAs. With this data, the GSS knows not to forward requests to any CRA that is not responding.

The boomerang server on the GSS gathers this information by sending keepalive messages at predetermined intervals. The boomerang server uses this data, along with the IP addresses of the CRAs, to request the exact start time of the DNS race.

If the CRA response is to be accepted by the D-proxy, each CRA must spoof the IP address of the GSS to which the original DNS request was sent.

Balance Method Options for Answer Groups

For most balance methods supported by the GSS, there are additional configuration options when you group specific answers in an answer group. These configuration options ensure the GSS properly applies the balance method for answers, and that you receive the best possible results from your GSS device.

Table 1-2 describes the available answer group options for each answer type (VIP, CRA, or NS) and balance method combination.

Table 1-2 Answer Group Options

Answer Type	Balance Methods Used	Answer Group Options
VIP	Hashed	Order
	Least loaded	LT (Load Threshold)
	Ordered list	Weight
	Round-robin	
	Weighted round-robin	

Table 1-2 Answer Group Options (continued)

Answer Type	Balance Methods Used	Answer Group Options
Name server	Hashed	Order
	Ordered list	Weight
	Round-robin	
	Weighted round-robin	
CRA	Boomerang (DNS race)	None

The following sections explain each of the options available for the answers in an answer group.

Order

Use the Order option when the balance method for the answer group is Ordered List. Answers on the list are given precedence based upon their position in the list in responding to requests.

Weight

Use the Weight option when the balance method for the answer group is weighted round-robin or least loaded. You specify a weight by entering a value from 1 and 10. This value indicates the capacity of the answer to respond to requests. The weight creates a ratio that the GSS uses when directing requests to each answer. For example, if Answer A has a weight of 10 and Answer B has a weight of 1, Answer A receives 10 requests for every 1 request directed to Answer B.

When you specify a weight for the weighted round-robin balance method, the GSS creates a ratio of the number of times that the answer is used to respond to a request before trying the next answer on the list.

When you specify a weight for the least-loaded balance method, the GSS uses that value as the divisor for calculating the load number associated with the answer. The load number creates a bias in favor of answers with a greater capacity.

Load Threshold

Use the load threshold when the answer type is VIP and the keepalive method is KAL-AP to determine whether an answer is available, regardless of the balance method used. The load threshold is a number from *n* 2 and 254 that is compared to the load being reported by the answer device. If the reported load is greater than the specified threshold, the answer is considered offline and unavailable to serve further requests.

Traffic Management Load Balancing

The GSS includes DNS sticky and network proximity traffic management functions to provide advanced global server load-balancing capabilities in a GSS network.

DNS sticky ensures that e-commerce sites provide undisrupted services and remain open for business by supporting persistent sticky network connections between customers and e-commerce servers. Persistent network connections ensure that active connections are not interrupted and shopping carts are not lost before purchase transactions are completed.

Network proximity selects the closest or most proximate server based on measurements of round-trip time to the requesting client's D-proxy location, improving the efficiency within a GSS network. The proximity calculation is typically identical for all requests from a given location (D-proxy) as long as the network topology remains constant. This approach selects the best server based on a combination of site health (availability and load) and the network distance between a client and a server zone.

This section includes the following topics:

DNS Sticky GSLB

Stickiness, also known as persistent answers or answer caching, enables a GSS to remember the DNS response returned for a client D-proxy and to later return that same answer when the client D-proxy makes the same request. When you enable stickiness in a DNS rule, the GSS makes a best effort to always provide identical A-record responses to the requesting client D-proxy, assuming that the original VIP continues to be available.

DNS sticky on a GSS ensures that e-commerce clients remain connected to a particular server for the duration of a transaction even when the client's browser refreshes the DNS mapping. While some browsers allow client connections to remain for the lifetime of the browser instance or for several hours, other browsers impose a connection limit of 30 minutes before requiring a DNS re-resolution. This time may not be long enough for a client to complete an e-commerce transaction.

With local DNS sticky, each GSS device attempts to ensure that subsequent client D-proxy requests to the same domain name to the same GSS device will be "stuck" to the same location as the first request. DNS sticky guarantees that all requests from a client D-proxy to a particular hosted domain or domain list are given the same answer by the GSS for the duration of a user-configurable sticky inactivity time interval, assuming the answer is still valid.

With global DNS sticky enabled, each GSS device in the network shares answers with the other GSS devices in the network, operating as a fully connected peer-to-peer mesh. Each GSS device in the mesh stores the requests and responses from client D-proxies in its own local database, as well as shares this information with the other GSS devices in the network. As a result, subsequent client D-proxy requests to the same domain name to any GSS in the network causes the client to be "stuck".

The DNS sticky selection process is initiated as part of the DNS rule balance method clause.

Refer to Chapter 8, Configuring DNS Sticky, for information on configuring local and global DNS sticky for GSS devices in your network.

Network Proximity GSLB

The GSS responds to DNS requests with the most proximate answers (resources) relative to the requesting D-proxy. In this context, proximity refers to the distance or delay in terms of network topology (not geographical distance) between the requesting client's D-proxy and its answer.

To determine the most proximate answer, the GSS communicates with a probing device, a Cisco IOS-based router, located in each proximity zone to gather round-trip time (RTT) metric information measured between the requesting client's D-proxy and the zone. Each GSS directs client requests to an available server with the lowest RTT value

The proximity selection process is initiated as part of the DNS rule balance method clause. When a request matches the DNS rule and balance clause with proximity enabled, the GSS responds with the most proximate answer.

Refer to Chapter 9, Configuring Network Proximity, for information on configuring proximity for GSS devices in your network.

GSS Network Deployment

A typical GSS deployment may contain a maximum of eight GSS devices deployed on a corporate intranet or the Internet. At least one GSS must be configured as a primary GSSM. Optionally, a second GSS can be configured as a standby GSSM. The primary GSSM monitors the other GSS devices on the network and offers features for managing and monitoring request routing services using CLI commands or a GUI accessible through secure HTTP. Only one GSSM can be active at any time, with the second GSSM serving as a standby, or backup device.

The GSSM functionality is embedded on each GSS, and any GSS device can be configured to act as a primary GSSM or a standby GSSM.

You can configure additional GSS devices on the GSS network to respond to DNS requests and transmit periodic keepalives to provide resource state information about devices. The GSS devices do not perform primary GSSM network management tasks.

This section describes a typical network deployment of the GSS and includes the following topics:

Locating GSS Devices

Although your organization determines where your GSS devices are deployed in your network, you should observe the following guidelines when deploying these devices.

Because the GSS serves as the authoritative name server for one or more domains, each GSS must be publicly or privately addressable on your enterprise network. That way, the D-proxy clients that are requesting content can find the GSSs assigned to handle DNS requests.

Options are available for delegating responsibility for your domain to your GSS devices, depending on traffic patterns to and from your domain. For example, given a network containing five GSS devices, you might choose to modify your parent domain DNS servers so that all traffic sent to your domain is directed to your GSS network. Or you might choose to have a subset of your traffic delegated to one or more of your GSSs, with other devices handling other segments of your traffic.

Refer to Chapter 7, Building and Modifying DNS Rules for information on modifying your network's DNS configuration to accommodate the addition of GSS devices to your network.

Locating GSS Devices Behind Firewalls

Deploying a firewall can prevent unauthorized access to your GSS network and thwart common denial of service (DoS) attacks on your GSS devices. In addition to being deployed behind your corporate firewall, the GSS packet-filtering features can enable GSS administrators to permit and deny traffic to any GSS device.

When positioning your GSS behind a firewall or enabling packet filtering on the GSS itself, you must properly configure each device (the firewall and the GSS) to allow valid network traffic to reach the GSS device on specific ports. In addition to requiring HTTPS traffic to access the primary GSS graphical user interface, you may want to configure your GSSs to allow FTP, Telnet, and SSH access through certain ports. In addition, GSSs must be able to communicate their status to and receive configuration information from the GSSM. Also, primary and standby GSSMs must be able to communicate and synchronize with one another. Finally, if global DNS sticky enabled on the GSS network, all GSSs in the sticky mesh must be able to communicate with each other to share the sticky database.

Refer to the *Cisco Global Site Selector Administration Guide* for information about access lists to limit incoming traffic. See the "Deploying GSS Devices Behind Firewalls" section for information on which ports must be enabled and left open for the GSS to function properly.

Communication Between GSS Nodes

All GSS devices, including the primary GSSM and standby GSSM, respond to DNS queries and perform keepalives to provide global server load-balancing functionality. Additionally, the primary GSSM acts as the central management device and hosts the embedded GSS database that contains shared configuration information, such as DNS rules, for each GSS that it controls. Use the primary GSSM to make configuration changes, which are automatically communicated to each registered GSS device that the primary GSSM manages.

The standby GSSM performs GSLB functions for the GSS network. The standby GSSM can act as the interim primary GSSM for the GSS network should the designated primary GSSM suddenly goes offline or become unavailable to communicate with other GSS devices. If the primary GSS goes offline, the GSS network continues to function and does not impact global server load balancing.

The GSS performs routing of DNS queries based on the DNS rules and conditions created from the primary GSSM. Each GSS device on the network delegates authority to the parent domain GSS DNS server that serves the DNS requests.

Each GSS is known to and synchronized with the primary GSSM; unless global DNS sticky is enabled, individual GSSs do not report their presence or status to one another. If a GSS unexpectedly goes offline, the other GSSs on the network that are responsible for the same resources remain unaffected.

With both a primary and a standby GSSM deployed on your GSS network, device configuration information and DNS rules are automatically synchronized between the primary GSSM and a data store maintained on the standby GSSM.

Synchronization occurs automatically between the two devices whenever the GSS network configuration changes. Updates are packaged and sent to the standby GSSM using a secure connection between the two devices.

Refer to the *Cisco Global Site Selector Administration Guide* for instructions on enabling each GSS device in the GSS network and for details about changing the GSSM role in the GSS network.

Deployment Within Data Centers

A typical GSS network consists of multiple content sites, such as data centers and server farms. Access to a data center or server farm is managed by one or more SLBs, such as the Cisco CSS or Cisco CSM. One or more virtual IP addresses (VIPs) represent each SLB. Each VIP acts as the publicly addressable front end of the data center. Behind each SLB are transaction servers, database servers, and mirrored origin servers offering a wide variety of content, from websites to applications.

The GSS communicates directly with the SLBs representing each data center by collecting statistics on availability and load for each SLB and VIP. The GSS uses the data to direct requests to the most optimum data centers and the most available resources within each data center.

In addition to SLBs, a typical data center deployment may also contain DNS name servers that are not managed by the GSS. These DNS name servers can resolve requests, through name server forwarding, that the GSS is unable to resolve.

GSS Network Management

Management of your GSS network is divided into two types:

- CLI-Based GSS Management
- GUI-Based Primary GSSM Management

Certain GSS network management tasks require that you use the CLI (initial device setup, sticky and proximity group configuration, for example), other tasks require that you use the GUI (User Views and Roles, for example). In most cases, you have the option of using either the CLI or the GUI at the primary GSSM to perform GSLB configuration and monitoring.

Choosing when to use the CLI and when to use the GUI are also a matter of personal or organizational choice. Additionally, you have the option to create your GSLB configuration using one method, then modify the configuration using the alternate method.

CLI-Based GSS Management

You can use the CLI to configure the following installation, management, and global server load-balancing tasks for your GSS:

- Initial setup and configuration of GSS and GSSM (primary and standby) devices
- Software upgrades and downgrades on GSSs and GSSMs
- Database backups, configuration backups, and database restore operations
- Global server load-balancing configuration and DNS request handling by creating DNS rules and monitoring keepalives at the primary GSSM

In addition, you can use the CLI for the following network configuration tasks:

- Network address and host name configuration
- Network interface configuration
- Access control for your GSS devices, including IP filtering and traffic segmentation

You can also use the CLI for local status monitoring and logging for each GSS device.

Refer to the *Cisco Global Site Selector Command Reference* for an alphabetical list of all GSS CLI commands including syntax, options, and related commands.

GUI-Based Primary GSSM Management

The primary GSSM offers a single, centralized graphical user interface (GUI) for monitoring and administering your entire GSS network. You can use the primary GSSM GUI to perform the following tasks:

- Configure DNS request handling and global server load balancing through the creation of DNS rules and monitoring of keepalives
- Monitor GSS network resources
- Monitor request routing and GSS statistics

For more information about the GUI, see the "Understanding the Primary GSSM GUI" section.

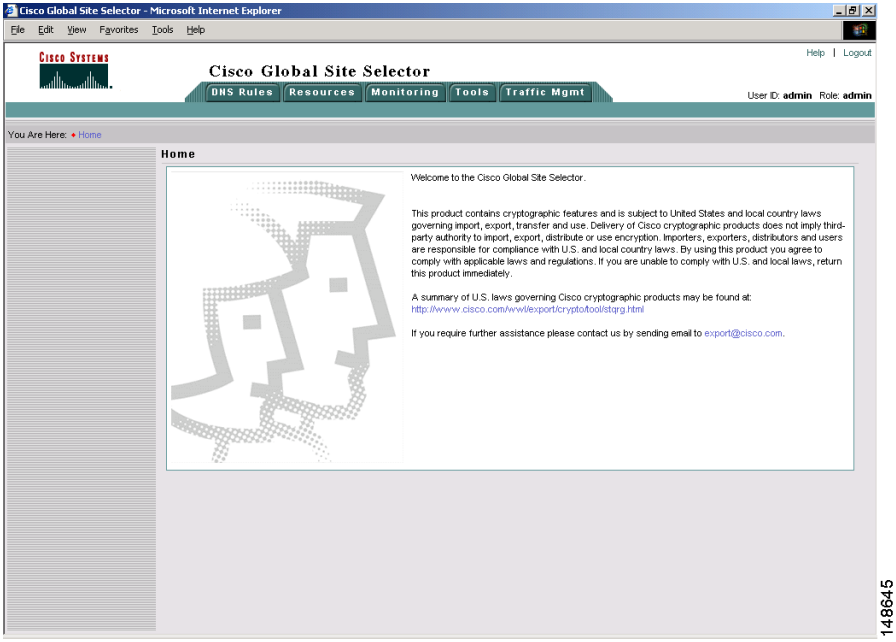
Understanding the Primary GSSM GUI

The primary GSSM graphical user interface is a web-based tool that you access using any standard web browser such as Microsoft Internet Explorer Version 5.0 and later releases and Netscape Navigator Version 4.79 or later releases. Basic authentication is used to restrict GUI access. All GUI traffic is encrypted using secure HTTP (HTTPS).

The primary GSSM GUI serves as a centralized management point for your entire GSS network. Using the primary GSSM GUI, you can add GSS devices to your network and build DNS rules that match groups of source addresses to hosted domains using one of a number of possible load-balancing methods. In addition, using the GSSM monitoring feature, you can obtain real-time statistics on the performance of your GSS network or of individual devices on that network.

After you log on to the primary GSSM GUI, the Welcome window (Figure 1-5) appears. The current login account information appears in the User ID (upper right) area of the Welcome window.

Figure 1-5 Primary GSSM Welcome Window



The following sections describe the organization and structure of the primary GSSM GUI:

- GUI Organization
- List Pages
- Details Pages
- Navigation
- Primary GSSM GUI Icons and Symbols
- Primary GSSM GUI Online Help

Review this information before using the primary GSSM GUI to define global load balancing for your GSS network.

GUI Organization

The primary GSSM GUI is organized into five main functional areas. Each area is divided by tabs, that you click to navigate to a particular section of the primary GSSM. These functional areas are as follows:

- **DNS Rules Tab**—Pages for creating and modifying DNS rules, including the creation of source address lists, (hosted) domain lists, answers, answer groups, and shared keepalives.
- **Resources Tab**—Pages for creating and modifying GSS network resources such as GSSs, locations, regions, and owners. You can also modify global keepalive properties from the Resources tab.
- **Monitoring Tab**—Pages for monitoring the performance of content routing on your GSS network, such as displays of hit counts organized by source address, domain, answer method, or DNS rule.
- **Tools Tab**—Pages for performing the administrative functions for the GSS network, such as creating login accounts, managing account passwords, and viewing system logs.
- **Traffic Mgmt Tab**—Pages for configuring the advanced global server load-balancing functions, DNS sticky, and network proximity

You access specific pages within each functional area by choosing from a series of navigation links in the upper left corner of the GUI. The navigation links vary according to the selected tab. Navigation links are available on all GUI pages.

Once you select a page, information is further organized into two areas: list pages and details page. List pages and details pages are described in the sections that follow.

List Pages

List pages provide you with a feature-specific overview. For example, you click the Answers tab (located on the DNS Rules tab) to display the Answers list page. This list page shows all of the answers currently configured on the listed GSS network.

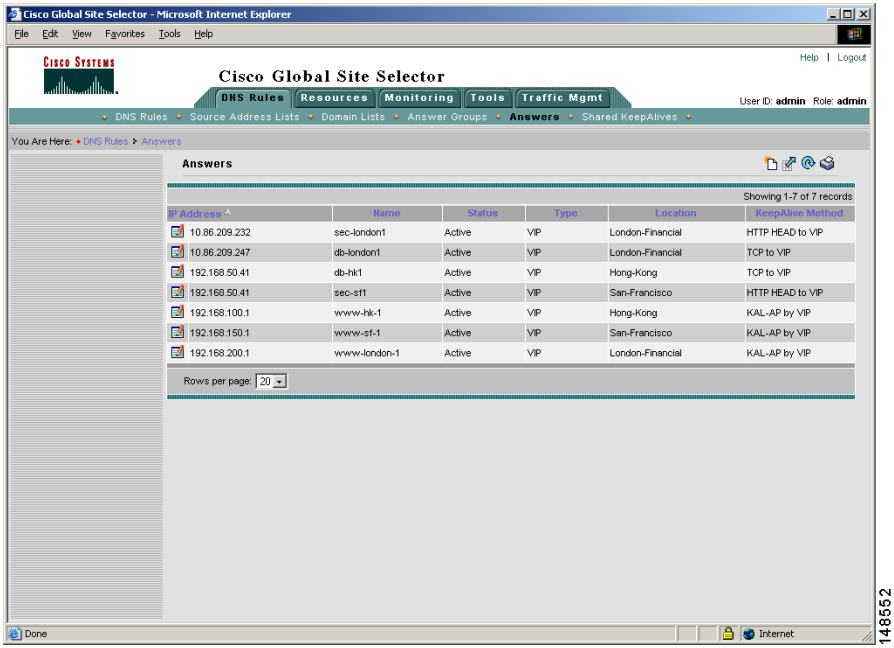
List pages display all data in tabular format to provide you with a detailed view of the resources available on your GSS network. In addition, you can use list pages to add new resources (for example, DNS rules or answer groups) or modify existing resources.

List pages enable you to sort resources by any one of a number of properties listed on the page. You can quickly locate a particular resource by using an identifying characteristic such as a name, owner, or type. You can sort information in ascending or descending order by any column. To sort the information in a list page, click the column header for the column containing the information that you wish to sort.

The GSS software temporarily retains information that you modify for a list page, allowing you to navigate to any of the details pages associated with the active list page while retaining the list page settings. The sort field, sort order, and rows per page are temporarily stored in memory for the active list page. Once you navigate to another list page, the GSS software discards the modifications for the previous list page.

Figure 1-6 shows an example of a primary GSSM Answers list page.

Figure 1-6 Answers List Page

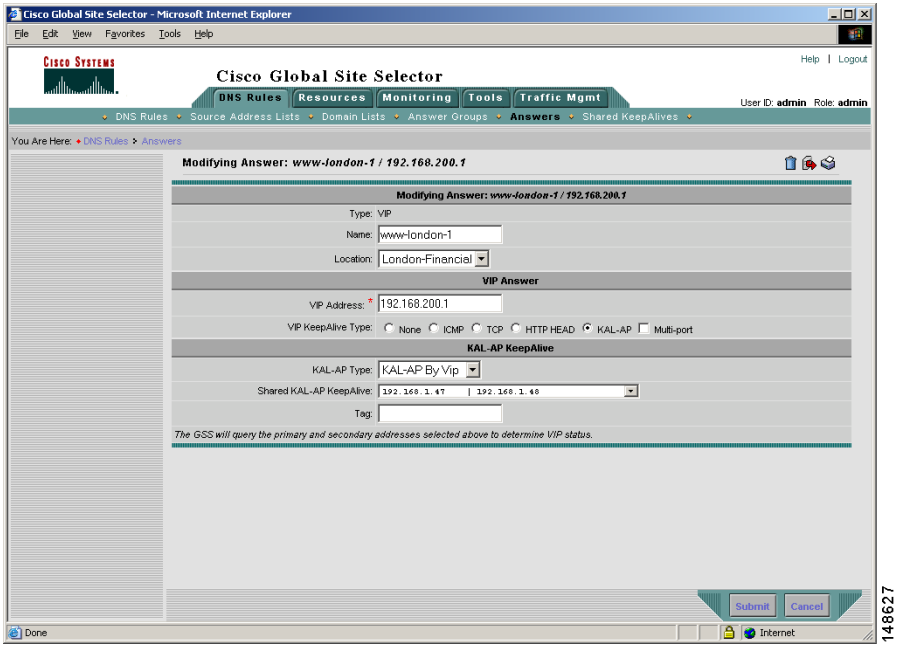


Details Pages

Details pages provide specific configuration information for a specific GSS function, enabling you to define or to modify those properties. You access a details page from a list page.

For example, click the Answers navigation link to display the Answers list page (see Figure 1-6). Next to each answer is an icon that show a pad and pencil, called the Modify icon. Click the Modify icon to display the details page for that answer (Figure 1-7). From the Modifying Answer details page, modify the properties of an answer or delete the answer.

Figure 1-7 Modifying Answer Details Page



Navigation

The primary GSSM GUI is viewed as a series of web pages using a standard browser. However, navigating between primary GSSM GUI pages is not the same as moving around different websites or even within a single site. Instead, you navigate from one content area of the GUI using the tabs for each of the major functional areas: DNS Rules, Resources, Monitoring, Tools, and Traffic Mgmt. Online Help is located as a navigation link at the top of each page.

Once you are located within a major content area, you can then access a particular feature or move between features using the navigation links. Choosing a feature from the navigation links immediately transfers you to that page in the graphical user interface. To move back from a details page to the corresponding list page, click another navigation link, or click either the **Submit** or **Cancel** buttons from the details page.

For example, to return to the Global Site Selectors list page after viewing the details for one of your GSS devices, click a different navigation link (or click the **Cancel** button). If you made configuration changes to a GSS that you wish to retain, click the **Submit** button. Any of these actions returns you to the Global Site Selectors list page.

Note Do not use your web browser Back or Forward buttons to move between pages in the primary GSSM GUI. Clicking **Back** cancels any unsaved changes in the primary GSSM.

Primary GSSM GUI Icons and Symbols

Table 1-3 lists and explains some common icons and graphical symbols in the primary GSSM GUI. These icons are referenced throughout this publication to explain how to use the features of the primary GSSM GUI.

Table 1-3 GSSM GUI Icons and Symbols

Icon or Symbol	Purpose	Location
	Modify icon. Opens the associated item for editing in a details page, displaying configuration settings on the details page.	List pages
	Sort icon. Indicates that the items listed in a list table are sorted in descending order according to the property listed in this column.	List pages
	Create icon/Open DNS Rules Builder icon. Opens the associated details page to accept user input for configuration.	List pages
	Print icon. When you view GSS resources or monitor GSS network activity, Print allows you to print data displayed in the page using your local or network printer.	List pages and Detail pages

Table 1-3 GSSM GUI Icons and Symbols (continued)

Icon or Symbol	Purpose	Location
	Export to CSV icon. When you view GSS resources or monitor GSS network activity, Export allows you to save data displayed in the window to a comma-delimited flat file for use in other applications.	List pages
	Refresh icon. When you view GSS resources or monitor GSS network activity, Refresh forces the GUI page to update its content.	List pages
	Run Wizard icon. Opens the associated DNS rule for editing using the DNS Rules Wizard.	DNS Rules list page
	Filter DNS Rule List icon. Provides filters that can be applied to your DNS rules, allowing you to view only those rules that have the properties in which you are interested.	DNS Rules list page
	Show All DNS Rules icon. Removes all filters, displaying a complete list of DNS rules for your GSS.	DNS Rules list page
*	Asterisk. Required field. Indicates that a value is required in the adjacent field before the item can be successfully saved.	Details pages

Table 1-3 GSSM GUI Icons and Symbols (continued)

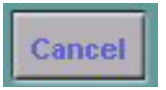
Icon or Symbol	Purpose	Location
	Submit icon. Saves the configuration information. When editing specific GSS system or device configuration information, Submit returns you to the associated list screen.	Detail pages
	Cancel icon. Cancels any configuration changes that were entered. When editing specific GSS system and device configuration information, Cancel returns you to the associated list screen.	Detail pages
	Delete icon. When you view configuration information for GSS resources, Delete allows you to delete the resource from the GSS network. **Note** Deletions of any kind cannot be undone in the primary GSSM GUI. If you might want to use the deleted data at a later point in time, we recommend performing a database backup of your GSSM. Refer to the *Cisco Global Site Selector Administration Guide* for details.	Detail pages

Table 1-3 GSSM GUI Icons and Symbols (continued)

Icon or Symbol	Purpose	Location
	Next icon. Moves forward to the next page in the DNS Rules Wizard. You can also use the links under the Wizard Contents table of contents to move between steps in the wizard.	DNS Rules wizard
	Back icon. Moves back to the previous page in the DNS Rules Wizard. You can also use the links under the Wizard Contents table of contents to move between steps in the wizard.	DNS Rules wizard
	Finish icon. Saves changes to the DNS rule. You return to the DNS Rules list page.	DNS Rules wizard
	Click to Add KAL icon. Adds multiple keepalives and/or destination ports to a single VIP-type answer.	Creating Answer and Modifying Answer details page
	Activate Answer icon. Reactivates a single suspended answer, all suspended answers in an answer group, all suspended answers associated with an owner, or all suspended answers associated with a location.	Modifying Answer, Modifying Answer Group, Modifying Owner, and Modifying Location detail page

Table 1-3 GSSM GUI Icons and Symbols (continued)

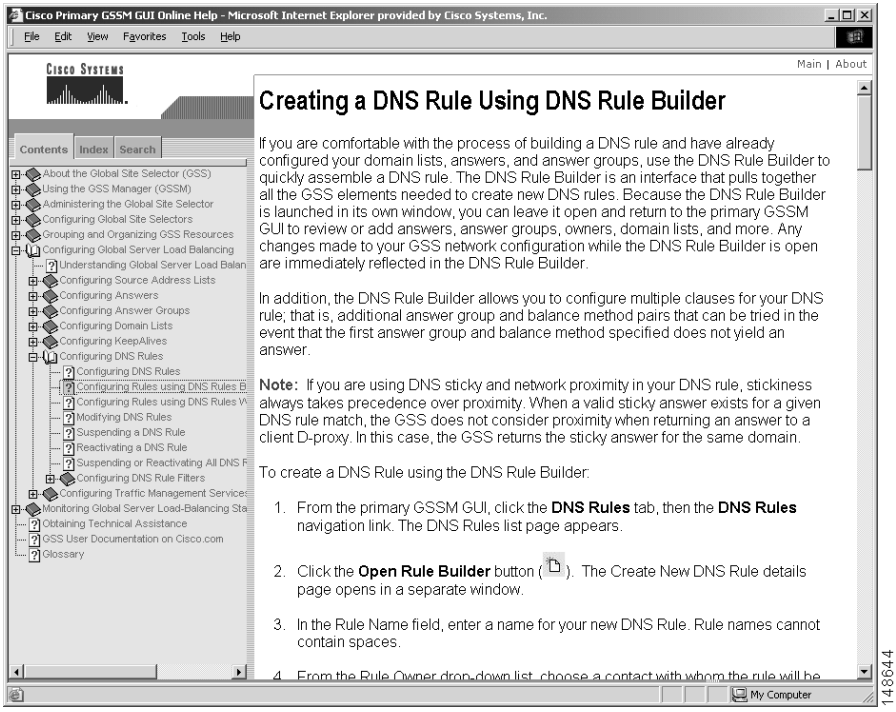
Icon or Symbol	Purpose	Location
	Suspend Answer icon. Temporarily stops the GSS from using a single answer, all answers in an answer group, all answers in all groups for an owner, or all answers in a location.	Modifying Answer, Modifying Answer Group, Modifying Owner, and Modifying Location detail page
	Activate DNS Rule icon. Reactivates a single suspended DNS Rule or all suspended DNS Rules associated with an Owner.	Modify DNS Rules and Modifying Owner detail pages
	Suspend DNS Rules icon. Temporarily stops requests from being processed by a single DNS rule or all suspended DNS rules associated with an owner on your GSS.	Modify DNS Rules and Modifying Owner detail pages
	Set Answers KAL ICMP icon. Disassociates all answers from a selected shared keepalive and sets the keepalive type of each of those answers to ICMP using the answer's associated VIP.	Modifying Shared Keepalive details page
	Set Answers KAL None icon. Disassociates all answers from a selected shared keepalive and sets the keepalive type for each answer to none. The GSS assumes that the answers are always alive.	Modifying Shared Keepalive details page

Primary GSSM GUI Online Help

The Help navigation link in the upper right corner of each primary GSSM GUI page launches the Online Help system (Figure 1-8), which contains information on using that page as well as the features of the primary GSSM GUI. The Online Help topic associated with the form displays in a separate child browser window.

Each page in the primary GSSM GUI has a context-sensitive online Help file associated with it. These Help files (in HTML format) contain detailed information related to the form that you are using. Online Help also includes a series of quick start procedures to assist you in navigating through the specific forms in the user interface and performing specific configuration procedures (for example, using the DNS Rules wizard to create a DNS rule).

Figure 1-8 Primary GSSM GUI Online Help



The GSS Online Help system contains several navigational aids to assist you in finding the information that you need quickly and easily. The navigation frame is contained in the left frame of each Help topic. The navigation frame contains the following three tabs:

- **Contents**—Displays all the topics in the GSSM Online Help system in a tiered format. Help topics are grouped into logical books by function. Books of Help topics may contain sub-books with additional topics. You can expand or collapse the contents to suit your needs. Note that the contents also automatically synchronizes with the Help topic that you are currently viewing.
- **Index**—Displays a list of terms that allows you to look up topics based on keywords similar to the index at the back of a book. If only one topic is associated with the Index entry, that topic displays immediately when you double-click the entry. If more than one topic is associated with an Index entry, the Help system displays a Topics Found dialog box that allows you to select the topic that you want to display from a list of topics.
- **Search**—Provides a full-text search tool that allows you to display a list of Help topics related to words that you enter in the text box. You can then select a topic and click **Display** to view that topic.

Global Server Load-Balancing Summary

After you create your GSSM (primary and standby) and GSS devices and configure them to connect to your network, you are ready to begin configuring request routing and global server load balancing for your GSS network. Refer to the *Cisco Global Site Selector Getting Started Guide* for procedures on getting your GSSM (primary and standby) and GSS devices set up, configured, and ready to perform global server load balancing.

You use the centralized GUI on the primary GSSM to configure global server load balancing for your GSS network. Using this interface, you configure keepalives to monitor the health of SLBs and servers on your network, and you create and manage DNS rules and the associated global server load-balancing configuration to process incoming DNS requests

Because you create DNS rules that route incoming DNS requests to the most available data centers and resources on your network, you must configure the elements that constitute your DNS rules before creating the rules themselves.

Use the following order when configuring your GSS devices and resources from the primary GSSM for global server load balancing:

1. Create regions, locations, and owners—Optional. Use these groupings to organize your GSS network resources by customer account, physical location, owner, or other organizing principle. Refer to Chapter 2, Configuring Resources for details.
2. Create one or more source address lists—Optional. Use these lists of IP addresses to identify the name servers (D-proxy) that forward requests for the specified domains. The default source address list is Anywhere to match any incoming DNS request to the domains. Refer to Chapter 3, Configuring Source Address Lists for details.
3. Create one or more domain lists—Establish lists of Internet domains, possibly using wildcards, that are managed by the GSS and queried by users. Refer to Chapter 4, Configuring Domain Lists for details.
4. Modify the default global keepalive settings or create any shared keepalives—Optional. These GSS network resources are regularly polled to monitor the online status of one or more GSS resources linked to the keepalive. Shared keepalives are required for any answer that uses the KAL-AP keepalive type. Refer to Chapter 5, Configuring Keepalives for details.
5. Create one or more answers and answer groups—Answers are resources that match requests to domains. Answer groups are collections of resources that balance requests for content. Refer to Chapter 6, Configuring Answers and Answer Groups for details.
6. Build the DNS rules that will control global server load balancing on your GSS network. Refer to Chapter 7, Building and Modifying DNS Rules for details.
7. If you plan to use DNS sticky for your global server load balancing, configure local or global DNS sticky for GSS devices in your network —Stickiness enables the GSS to remember the DNS response returned for a client D-proxy and to later return that answer when the client makes the same request. Refer to Chapter 8, Configuring DNS Sticky for details.
8. If you plan to use network proximity for your global server load balancing, configure proximity for GSS devices in your network—Proximity determines the best (most proximate) resource for handling global load-balancing requests. Refer to Chapter 9, Configuring Network Proximity for details.

Where to Go Next

Chapter 2, Configuring Resources describes how to organize resources on your GSS network as locations, regions, and owners.