Cisco GSS 4490 Global Site Selector

The Cisco® GSS 4490 Global Site Selector is a critical component of any business resilience strategy to optimize multisite deployments that involve globally distributed data centers. The Cisco GSS 4490 networking product globally load balances distributed data centers, and is the cornerstone of multisite disaster recovery plans in deployments of market-leading Cisco content switches.

Customers can benefit from the new levels of traffic management and centralized command and control provided by the Cisco GSS 4490—whether they are deploying new Cisco content switches (such as the Cisco CSS 11500 Series Content Services Switch [CSS] and the Content Switching Module [CSM] for Cisco Catalyst® 6500 Series switches), or have legacy switches in place (such as the Cisco CSS 11000 Series and Cisco LocalDirector devices).

The Cisco GSS 4490 delivers the following capabilities:

- Provides a scalable, dedicated hardware platform for industry-leading Cisco content switches to ensure Web-based applications are always available, by detecting site outages or site congestion and rerouting content requests

- Improves the global data center selection process by offering user-selectable global load-balancing algorithms

- Offloads Domain Name System (DNS) servers by taking over the domain resolution process, and transmits these requests at thousands of requests per second

- Scales to support hundreds of data centers or server load balancers (SLBs)

- Complements the existing DNS infrastructure by providing centralized domain management

- Tightly integrates with Cisco SLBs without sacrificing the ability to work in a heterogeneous environment of DNS-capable networking products

The Cisco GSS 4490 allows businesses to deploy global Internet and intranet applications with the confidence that Web application users will be quickly rerouted to a standby data center if a primary data center outage or overload occurs. The Cisco GSS 4490 traffic-management process continuously monitors the load and health of the SLBs within each data center. This information is used in conjunction with customer-controlled load-balancing algorithms to enable the Cisco GSS 4490 to select a data center that is available and not overloaded within user-definable load conditions—in real time.

By offloading the DNS server resolution process from traditional DNS servers, the Cisco GSS 4490 can optimize global site selection, boost DNS responsiveness, assure data center availability, and increase the scalability of Websites and data centers. The Cisco GSS 4490 is a critical component for enterprises and service providers deploying globally distributed data centers, or installing disaster recovery systems for Web-based applications.

Features and Benefits

The Cisco GSS 4490:

- Provides a resilient architecture that is critical for disaster recovery and multisite Web application deployments
- Offers flexible heterogeneous support for all Cisco SLBs and DNS-capable networking products
- Provides centralized command and control of the DNS resolution process for direct and precise control of the global load-balancing process
- Augments and offloads any DNS Berkeley Internet Name Domain (BIND) infrastructure to optimize content requests and delivery for all types of static and dynamic Web content
- Provides dedicated processing of DNS requests for greater performance and scalability
- Offers site persistence for e-commerce applications
- Offers a unique DNS race feature—the Cisco GSS 4490 can, in real time, direct content consumers to the closest data center
- Supports a Web-based GUI and DNS wizard to simplify DNS command and control

Cisco GSS 4490 Business Continuance and Global Service Load-Balancing Deployments

Breakthrough Concept—DNS Rules

At the core of the Cisco GSS 4490 is the "DNS rule," which provides centralized command and control of how the Cisco GSS 4490 globally load balances a given hosted domain, what IP addresses are sent to the client's name server (D-proxy), and what recovery method should be used if the preferred choice is unavailable. The Cisco GSS 4490 allows a DNS administrator to control how Web-based applications are globally load-balanced across multiple sites and under what condition clients should be directed to a backup site. This combination is ideal for DNS administrators seeking a centralized tool that gives them complete control of the DNS resolution process. The Cisco GSS 4490 provides the features needed by e-businesses looking for a real-time business continuance solution.

Using the Cisco GSS 4490 GUI, the user defines the domain that will be globally load-balanced, and can include the IP addresses of specific, client-named servers (D-proxy). This allows the user to "hard-wire" a single D-proxy or a group of D-proxies to a particular site, which is useful when the DNS administrator needs to specify a certain set of SLBs to support all the traffic coming from a specific Internet service provider (ISP) D-proxy.

On the same GUI screen, the DNS administrator defines what SLBs should support this domain. The DNS administrator has the option to assign multiple IP addresses (up to eight) and a time-to-live (TTL) value that will be used by the client D-proxy. The final configuration option is the recovery method that is used if the first method fails (in the GUI, these are called "DNS rules"). Each DNS rule can support up to three load-balance clauses. These clauses are used to establish the primary load-balancing method, the secondary load-balancing method, and the tertiary method, which could involve hard-wiring a global "sorry server."

Cisco GSS 4490 and Market-Leading Cisco Content Switches—The Ultimate Combination for Business-Resiliency Solutions

The Cisco GSS 4490, in combination with a local Cisco content switch, is critical for large enterprises and service providers planning to deploy highly reliable distributed data centers. The Cisco GSS 4490 selects the best site, based on the load and availability information supplied by the Cisco content switch. This allows the Cisco content switch to select the best local

server within the data center based on availability and local load. The Cisco GSS 4490 simplifies this network deployment architecture with its centralized command and control features. An example of the complete control is the fact the Cisco GSS 4490 can gracefully take a Cisco content switch out of rotation without affecting ongoing operations.

The Cisco GSS 4490 performs two major functions as part of the global site selection process:

- Takes an active role in the DNS infrastructure to connect the client to the SLB that supports the requested Website
- Continuously monitors the load and availability of these SLBs to select the SLB that is most capable of supporting the new client

In Figure 1, the Cisco GSS 4490 offloads the Website selection process from the DNS infrastructure. The Cisco GSS 4490 continuously monitors the load and health of up to 128 SLBs or 4000 virtual IP addresses (VIPs). These SLBs can be colocated, or located at remote and disparate data centers.

How the Cisco GSS 4490 interacts with the client in the Website selection process is summarized in the following six steps:

1. A client wants to access an application at foo.com. The client types www.foo.com in the browser. This application is supported at three different data centers.
2. The request is processed by the DNS global control plane infrastructure and arrives at the Cisco GSS 4490.
3. The Cisco GSS 4490 offloads the site selection process from the DNS global control plane. The request and site selection are based on the load and health information in conjunction with customer-controlled load-balancing algorithms. The Cisco GSS 4490, in real time, selects a data center that is available and not overloaded.
4. The Cisco GSS 4490 sends the IP address of the "best" server load balancer at a specific data center—in this case, the SLB at Data Center 2.
5. The browser processes this IP address.
6. When the hand-off for the DNS control plane is complete, the client is directed to the SLB at Data Center 2 by the IP control/forwarding plane.

Figure 1
Cisco GSS 4490 and Distributed Data Centers

The mechanism that is used to extract this load and health information is a specially designed keepalive (KAL). The Cisco GSS 4490 supports the following KALs:

- KAL-AP—This KAL was developed specifically to extract both load and availability from the Cisco CSS 11000/11500 and the Cisco CSM. When this detailed query is sent to the CSS or CSM, these SLBs respond with availability and load information about a hosted domain name or hosted virtual IP address of the SLB.
- HTTP—An HTTP "head request" method to a given origin server sends a request to an origin server and checks for a "200 OK. If the Cisco GSS 4490 receives a 200 OK, then it directs traffic to the virtual IP address supporting that server. The user can configure the host tag, the remote host IP, and the URL (including a path).

- TCP—This KAL is used when global SLB (GSLB) devices may be something other than a CSS or CSM. These GSLB remote devices include Web servers, LocalDirectors, Wireless Application Protocol (WAP) gateways, and other devices that can be checked using a TCP KAL. The TCP KAL initiates a TCP connection to the remote device by performing a three-way handshake.

- Internet Control Message Protocol (ICMP)—ICMP is a Layer 3 "ping" that indicates the status of a device based on connectivity to the network. This KAL is used with any device that responds to a ping request. Pings are sent once every 5 or 45 seconds. If there is no response, the ping is sent once every 5 seconds, up to three times. If there is still no response, the device (virtual IP or real server) is considered offline.

- Name server query—A simple DNS request is sent to a host (name or mail server, for example) to receive a resolved domain name to prove the system is "alive." In this case, the Cisco GSS 4490 sends an unused domain name, probing for a failure response, which proves that the DNS server is "alive." This is used in conjunction with the name server forwarding feature.

- Content routing agent (CRA)—This User Datagram Protocol (UDP)-based KAL is used with the DNS race feature; the GSS sends KAL requests to port 1304 to retrieve round-trip times between the GSS and agent (CSS or content engine).

Global Load-Balancing Algorithms—Complete Control of Site Selection

The Cisco GSS 4490 supports eight global load-balancing algorithms and gives administrators full flexibility in selecting the global load-balancing algorithm that matches their needs:

- Ordered list—This user-definable list specifies one or a group of IP addresses (corresponding to a virtual IP address or the IP address of a back-end server) that the Cisco GSS 4490 uses to respond to a DNS request for a specific domain. The Cisco GSS 4490 uses the first address in the list until it becomes unavailable or overloaded, and then moves to the next address in the list. This process is repeated for every subsequent entry in the list.

- Static based on client's DNS address—This algorithm is a variation of the order list that allows the administrator to map the IP address of the client's DNS name server to an available virtual IP address on a specific content switch. This feature is used when the administrator wants to allocate a specific community of users to a specific set of SLBs or back-end servers.

- Round robin—This algorithm cycles through available virtual IP addresses in order. 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 "active standby" site that serves requests.

- Weighted Round Robin (WRR)—The Cisco GSS 4490 cycles through the list of virtual IP addresses that are available as requests are received, but sends requests to a favored virtual IP address based on a user-assigned weighting value.

- Least loaded—The Cisco GSS 4490 can receive load values from either the CSS or CSM. The Cisco GSS 4490 monitors these load values to see if they exceed a threshold that is assigned by the administrator. If the load is above the specified threshold, the virtual IP on the CSS or CSM is considered offline and unavailable to serve further requests. If the load falls below the threshold, the GSS automatically starts sending requests to the virtual IP address.

- Source address and domain hash—When the Cisco GSS 4490 is using this balance method, the IP address of the client's DNS proxy and the requesting client's domain are used to create a unique "hash" value that is then "stuck" to the virtual IP that is chosen to serve the DNS query. This means every time that user requests that domain, the user will go to the virtual IP unless it is unavailable, and then the user is sent to an alternate virtual IP assigned by the administrator through configuration.

- DNS race—This algorithm initiates a race of A-record responses to the client's named server. It can achieve proximity without probing.
- Drop—This algorithm silently discards requests.

The administrator can configure up to three load-balancing clauses per DNS rule. Each clause can have a different answer group assigned to it. The multiple load-balancing clauses per DNS rule allow the user to control how the Cisco GSS 4490 responds if one of the load-balancing methods fails. Using the multiple clauses per DNS rule allows the network engineering staff to create a solution that ensures that a client always gets directed to the proper data center, based on the client's unique requirements.

For example, the preferred global load-balance method may be WRR as defined in the first load-balancing clause. If that method fails because of the loss of a local SLB or back-end server, the Cisco GSS 4490 tries the second method (for instance, ordered list as defined in the second load-balance clause). If that fails, the Cisco GSS 4490 tries the third method, which may be a static entry to a global "sorry server." Users can customize the clauses based on what is most advantageous to their own environments and global load-balancing objectives.

E-Commerce

The Cisco GSS 4490 provides unique capabilities that optimize e-business. The Cisco GSS 4490 ensures reliable transactions with a "sticky" DNS feature that guarantees site persistence, ensuring "sticky" connections between users and applications during secure and nonsecure portions of the transaction. When used with a Cisco content switch, these sticky connections can eliminate dropped shopping carts and allow e-businesses to prioritize users and transactions, thereby providing premium service for high-priority customers.

Name Server Forwarding

Name server forwarding plays a vital role in providing increased flexibility for the Cisco GSS 4490. This feature is used when errors in the DNS infrastructure send a request for a domain that cannot be processed by the Cisco GSS 4490. The Cisco GSS 4490 does not drop the request; instead, it forwards the request to a designated name server capable of responding to the request. The response of that name server is passed through the Cisco GSS 4490 so that it appears to have come from the GSS. This helps to ensure that any critical DNS request will be processed and not lost.

The Cisco GSS 4490 can forward the requests using either round robin or the order list load-balancing method. As the Cisco GSS 4490 processes the DNS requests, it can use one of the load-balancing methods to distribute these forward requests. A user can employ this feature to globally load balance e-mail servers. The following global load-balance methods can be used with the name server forwarding feature—WRR, ordered list, and hashing. If the Cisco GSS 4490 receives a request for an MX (mail) record, although the Cisco GSS 4490 cannot respond directly, it can forward the request to one of many DNS name servers deployed in the enterprise or service provider network.

DNS Race—Proximity Without Probing

The Cisco GSS 4490 supports the DNS race method of proximity. This feature allows for proximity to be achieved without probing the client name server (D-proxy). It is based on a simple concept—that instantaneous proximity can be found if a device within each data center sends an A-record (IP address), at the exact same time, to the client's named server. Whichever A-record is received first is the most proximate.

The DNS race method of DNS resolution is initiated by the Cisco GSS 4490 and is designed to load balance up to 20 sites. For the Cisco GSS 4490 to initiate a race, it needs to establish two pieces of information per CRA:

- The delay between the Cisco GSS 4490 and each of the CRAs in each data center. With this data, the Cisco GSS 4490 computes how much time to delay the race from each data center so that each CRA will start the race simultaneously.

- The "aliveness" of the CRAs. With this data, the Cisco GSS 4490 knows not to forward a request to any CRAs that are not responding. The Cisco GSS 4490 gathers this information by sending KAL messages at predetermined intervals. This data, along with the IP addresses of the CRA, is used to request the start of the race. When the Cisco GSS 4490 receives a DNS request, a race request is sent to each CRA at a predetermined time and the race is initiated from each data center. The first A-record received by the client's D-proxy is the winner and is the most proximate. The DNS race method on the Cisco GSS 4490 is what a DNS administrator wants as a way to reduce the time involved in the DNS resolution process.

Scalability

Highly scalable, the Cisco GSS 4490 meets the most demanding environments. Table 1 lists support information for the Cisco GSS 4490.

Table 1 Cisco GSS 4490 Performance and Scalability Metrics

Performance and Scalability Attributes	Metrics
DNS requests per second	4000
DNS rules	4000
Name server forwarding requests per second or DNS races per second	1500
Active SLBs	128
Hosted domains (maximum 1000 per SLB), 128 characters maximum per domain	2000
Hosted domain lists (500 members maximum per list)	250
Virtual IP addresses (no more than 500 with active ICMP keepalive)	2000 (4000 shared)
Name server addresses for name server forwarding (30 maximum per answer group)	100
CRA devices (20 maximum per race and answer group)	200
Source IP addresses configurable for DNS rules	500
Source address lists (30 members maximum per list)	60
Answer groups (100 members maximum per group)	500
KALs—100 fast, 500 standard per KAL types	

Manage Through Simple GUIs

For configuration and control, administrators have a Cisco IOS® Software-like command-line interface (CLI) and an intuitive, embedded GUI. The CLI is used for network startup, and the GUI is used for the configuration of all global load-balancing parameters and is supported directly by the Cisco GSS 4490. Users have a choice of using either an expert mode on the GUI or a wizard to configure all global load-balancing parameters at a single, centralized location. Configuration parameters are

shared with up to eight Cisco GSS 4490 selectors in a single deployment. The CLI is accessed through Telnet or Secure Shell (SSH) Protocol; the GUI is accessed through Secure Sockets Layer (SSL). Network managers can restrict console, FTP, Simple Network Management Protocol (SNMP), Telnet, and Web management access.

Monitoring the status of the Cisco GSS 4490 can be done using a GUI, SNMP, CLI, and log files. The information that can be accessed from the different methods includes:

- Displaying of configuration, software information, hardware information, system resources, user information, and current logins
- Booting configuration
- Configuring IP interfaces, IP routes, and host name
- The ability to see the operation status (active/inactive) of locally configured resources, with filtering by domain name or other parameters (GUI only)

Configuration State and File Management

- Show running configuration for all CLI-configured options
- Show startup configuration
- Archive and restore running configuration
- Archive and restore startup configuration
- Display archived configuration
- Copy running configuration to disk, FTP, or Trivial File Transfer Protocol (TFTP) server
- Clear running configuration
- Create startup configuration, including from text editor
- Copy startup-configuration to running- configuration

Logging

- Show logs, including real-time log
- Archive and restore log files

Cisco GSS 4490 Features—Quick Look

Table 2 gives features of the Cisco GSS 4490.

Table 2 Cisco GSS 4490 Features

Features	Description
Cisco GSS 4490	- Ports Two 10/100/1000 Fast Ethernet - DNS requests per second 4000, depending on configuration (~345 million DNS requests per day per Cisco GSS 4490; an entire system is capable of 2.7 billion DNS requests per day)

Table 2 Cisco GSS 4490 Features

Features	Description
Devices supported	• CSSs—Load and availability • CSMs—Load and availability • Cisco LocalDirector—Availability using HTTP head or TCP KAL • Cisco IOS SLB—Availability using HTTP head or TCP KAL • DNS name servers—Availability using name server request KAL • Origin servers—Availability using HTTP head or TCP KAL • Content engines—Availability using ICMP ping or TCP KAL • CRAs—Availability using CRA KAL
Network management	• Console port—CLI • Access to system through Telnet • Secure copy (SCP) or FTP • GUI—Secure HTTP (HTTPS) for Internet Explorer and Netscape Navigator
Network management MIBs	• Read-only monitoring of network and device status, including RFC 1213 • (MIB-II) and RFC 1514 (HOST-RESOURCES-MIB)
Physical	• One-rack-unit size chassis • Network management serial port 1 GB RAM; 2.4-Ghz PIV CPU • Storage One 80-GB hard drive • Power Integrated AC power (autosensing 110V/60 Hz)

Ordering Information

The Cisco GSS 4490 supports only AC power; there are no options. Orderable part numbers are given in Table 3.

Table 3 Part Numbers for Cisco GSS 4490

Product number	Description
Cisco GSS 4490-K9	Global Site Selector
SF-GSS-V1.1-K9	Global Site Selector Software

Environmental Specifications

Table 4 gives specifications of the Cisco GSS 4490.

Table 4 Specifications of Cisco GSS 4490

Feature	Descriptionl
Rack units	One rack unit
Ports Two	10/100/1000 Fast Ethernet autosensing, one console port
Storage One	80-GB hard drive
Software image	SF-GSS-V1.1-K9

Table 4 Specifications of Cisco GSS 4490

Feature	Descriptionl
Processor	2.4-GHz PIV CPU
Mechanical and Environmental	
Dimensions	(H x W x D): 1.75 x 16.54 x 16.69 in. (44 x 424 x 430 mm)
Weight	28 lb (12.7 kg)
Operational temperature	50 to 95 F (10 to 35 C)
Storage temperature	–40 to 140 F (–40 to 60 C)
Relative humidity	80 percent, nonoperational
Certification	
Safety	UL 60950; CAS –C22.2 No. 60950; EN60950; EMC FCC Part 15 Class A; ICES-003 Class A; EN55022 Class B with STP cables; EN50082-1; VCCI Class 1

CISCO SYSTEMS