



GLOSSARY

A

Application Control Engine (ACE) Network device that performs high-performance server load balancing (SLB) among groups of servers, server farms, firewalls, and other network devices, based on Layer 3 as well as Layer 4 through Layer 7 packet information. The ACE can also terminate and initiate SSL-encrypted traffic so that it can perform intelligent load balancing while ensuring secure end-to-end encryption. The ACE is available as a standalone appliance or as module for use in a Cisco Catalyst 6500 Series Switch or a Cisco 7600 Series Router.

answer Network resources that respond to user queries. As with domains and source addresses, answers are configured at the primary GSSM by identifying a resource of a particular type on your GSS network to which queries can be directed and which can provide your user's D-proxy with the address of a valid host to serve their request. The three types of possible Answers on a GSS network are as follows:

- Virtual IPs (VIPs)—IP addresses associated with an SLB like the Cisco CSS, CSM, or other Cisco IOS-compliant SLB
- Name Server—A configured DNS name server on your network
- CRA—Content routing agents associated with the GSS boomerang server

answer group Customer-defined set of virtual IP address (VIP), name server (NS), or content routing agent (CRA) addresses from which an individual answer is selected and used to reply to a content request. Answers are grouped together as resource pools. The GSS, using one of a number of available balance methods, can choose the most appropriate resource to serve each user request from the answers in an answer group.

B

balance method Algorithm for selecting the best server. It is used together with an answer group to makes up a clause in a DNS rule. Up to three possible response answer group and balance method clauses are available for each DNS rule.

boomerang Server load-balancing component of the GSS that uses calculations of network delay to select the site “closest” to the requesting D-proxy. Closeness is determined by conducting DNS races between content routing agents (CRAs) on each host server. The CRA that replies first to the requesting D-proxy is chosen to reply to the request.

C

client Content consumer, such as a web browser or multimedia stream player, that makes Domain Name System (DNS) requests for domains managed by the GSS.

content provider	Customer who deploys content on a Content Delivery Network (CDN) or purchases hosting services from a service provider or web hosting service.
content router	Machine that routes requests for content through Domain Name System (DNS) records.
content routing agent (CRA)	Software running on a Content Delivery Network (CDN) or server load-balancing device that provides information to a GSS for making content routing decisions and handles content routing requests from the GSS.
Content Services Switch (CSS)	Cisco server load-balancing appliance for Layer 4 through Layer 7 content.
Content Switching Module (CSM)	Server load-balancing component for the Catalyst 6500 series switches.
CRA (keepalive)	Keepalive type used when the GSS answer you are testing is a content routing agent (CRA) associated with the boomerang server component of your GSS, the CRA keepalive type pings a CRA at an address that you specify, returning the online status of the device.
customer	Cisco customer purchasing GSS hardware, software, or services. Typically, an Internet service provider (ISP), application service provider (ASP), or enterprise customer.

D

data center	Collection of centrally located devices (content servers, transaction servers, or web caches).
Distributed Denial of Service (DDoS)	<p>Type of attack designed to deny legitimate users access to specific computer or network resources. Such attacks send several thousand spoofed DNS requests to a target device. The target then treats these requests as valid and returns the DNS replies to the spoofed recipient (i.e., the victim).</p> <p>Since the target is busy replying to the attacks, it drops valid DNS requests from legitimate D-proxies. When the number of requests is in the thousands, the attacks can potentially generate a multi-gigabit flood of DNS replies, thus causing congestion in the network. To combat this, the GSS contains a DDoS detection and prevention module.</p>
DNS race	Balance method initiated by the Boomerang Server component of the GSS that is designed to balance between 2 and 20 sites. DNS race gives all possible CRA's a fair chance at resolving a DNS request using a "race" between sites.
DNS rule	Central configuration and routing concept of the GSS that allows specific request balance resources, methods, and options to be applied to source address and domain pairs.
domain list	One or more hosted domains logically grouped for administrative and routing purposes.

D-proxy Client's local name server, which makes iterative DNS queries on behalf of a client. A single recursive query from a client may result in many iterative queries from a D-proxy. Also referred to as local domain name server (LDNS).

DRP Director Response Protocol (DRP). The GSS uses DRP to communicate with the proximity probing agents, called DRP agents, in each zone. DRP is a general User Datagram Protocol (UDP)-based query and response information exchange protocol developed by Cisco Systems. You can use any Cisco router that is capable of supporting the DRP agent software and can measure ICMP echo-based RTT as the proximity probing agent in a zone. The GSS communicates with the Cisco IOS-based router using the DRP ICMP echo-based RTT query and response method.

F

fair weighted round-robin In the fair weighted round-robin balance method, the user assigns a fair weight number to each answer in the answer group that is based on a ratio of the total number of real servers behind each answer in an answer group. When the GSS uses fair weighted round-robin balance method, the GSS cycles through the list of answers that are available as the requests are received. The GSS divides the requests among the answers in a ratio of the number of active real servers behind these answers in a answer group. This ratio is determined by the fair weight value assigned to a particular answer and the KALAP LOAD value for that answer.

fully qualified domain name (FQDN) Domain name that specifies the named node's absolute location relative to the Domain Name System (DNS) root in the DNS hierarchy.

G

global server load balancing (GSLB) System based on the Content Services Switch that directs clients through the Domain Name System (DNS) to different sites based on load and availability. Two versions of GSLB currently exist:

- Rule-based GSLB
- Zone-based GSLB

Global Site Selector (GSS) Cisco content routing device that intelligently responds to Domain Name System (DNS) queries, selecting the "best" content locations to serve those queries based on DNS rules created by the customer.

Global Site Selector Manager (GSSM) Device that administers a GSS network, storing configuration information and statistics for GSS devices. GSS administrators can use CLI commands or the graphical user interface (GUI) to reconfigure or monitor the performance of their GSS network.

global sticky With global DNS sticky enabled, each GSS device in the network shares answers with the other GSS devices in the network, operating as a peer mesh. The individual GSS devices in the mesh each store the requests from client D-proxies in its own local database. When one GSS device in the mesh receives a query from the client for the same hosted domain or domain list, global sticky enables each GSS in the network to make a best effort attempt to return the same answer to the requesting client. This action is performed regardless of which GSS in the network is selected to answer the first and subsequent requests. The individual GSS devices work together to maintain a global sticky database across the network. Each GSS in the peer mesh receives updates from the other peers and sends local changes to its remote peers.

GSS network Set of Global Site Selectors (GSSs) in a scaled, redundant GSS deployment.

H

hosted domain Domain managed by the GSS. A minimum of two levels is required for delegation (for example, foo.com). Domain wildcards are supported.

Hosted Domain List (HDL) A grouping of one or more domains that are being fronted by the GSS. Domains are grouped for administrative and load-balancing purposes.

HTTP HEAD Used when the GSS answer that you are testing is a VIP associated with an SLB device such as a CSS or CSM. The HTTP HEAD keepalive type sends a TCP format HTTP HEAD request to a web server at an address that you specify, returning the online status of the device (in the form of a 200 response) as well as information on the web page status and content size.

HTTPS HEAD Used when the GSS answer that you are testing is a VIP associated with an SLB device such as a CSS or CSM. The HTTPS HEAD keepalive type sends a TCP format HTTPS HEAD request to a web server at an address that you specify, returning the online status of the device (in the form of a 200 response) as well as information on the web page status and content size.

I

ICMP Keepalive type used when the GSS answer that you are testing is a VIP associated with an SLB device such as a CSS, CSM, or ACE. The ICMP keepalive type pings the configured VIP address (or a shared keepalive address). Online status is determined by a response from the targeted address, indicating connectivity to the network.

K

KAL-AP Keepalive type used when the GSS answer that you are testing is a VIP associated with an SLB device such as a CSS, CSM, or ACE. The KAL-AP keepalive type sends a detailed query to both a primary (master) and secondary (backup) VIP address that you specify, returning the online status of each interface as well as information on load for whichever address is acting as the master VIP. Depending on your GSS network configuration, the KAL-AP keepalive can be used to either query a VIP address directly or to query an address by way of an alphanumeric tag (KAL-AP By Tag), which can be particularly useful when you are attempting to determine the online status of a device that is located behind a firewall that is performing Network Address Translation (NAT).

keepalive (KAL)	<p>Periodic testing of availability and status of a content service through the sending of intermittent queries to a specified address using one of a variety of methods.</p> <p>The GSS uses both primary keepalive and secondary keepalive IP addresses.</p> <p>See the keepalive method entry.</p>
keepalive method	<p>Protocol or strategy used to determine whether a device is online. Examples include ICMP, TCP, KAL-AP, HTTP HEAD, and CRA round-trip time.</p>
<hr/>	
L	
LDNS	<p>Local Domain Name Server for a client.</p>
load threshold	<p>Balance method option that is used with the VIP Answer type. Specifies a number between 0 and 255, which is compared to the load number being reported by the answer device. If the answer's load is above the specified threshold, the answer is deemed to be offline and unavailable to serve further requests.</p>
local sticky	<p>With local DNS sticky, the GSS device ensures that subsequent client D-proxy requests to the same domain name will be “stuck” to the same location as during the first request. DNS sticky guarantees that all requests from a client D-proxy to a particular host 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. Each GSS dynamically builds and maintains a local sticky database that is based on the answers that the GSS sends to the requesting client D-proxies. If a subsequent request comes from the same client D-proxy, and the answer is valid, the GSS returns the cached answer to the client D-proxy.</p>
location	<p>Grouping for devices with common geographical attributes, used for administrative purposes only, and similar to data center or content site.</p> <p>See the data center entry.</p>
<hr/>	
N	
name server (NS)	<p>Publicly or privately addressable Domain Name System (DNS) server that resolves DNS names to IP addresses. Name servers are used by the GSS for name server forwarding, in which queries that the GSS cannot resolve are forwarded to a designated name server that can resolve them.</p>
name server forwarding	<p>Although not an official balance method, Name Server Forwarding plays a vital role in server load balancing using the GSS. Used in instances where requests for domains cannot be handled by any of the name servers configured on the GSS network, the Name Server Forwarding feature passes on requests it cannot answer to a configured name server that does know. That name server's response is passed through the GSS so that it appears to have come from that device.</p>
None (keepalive)	<p>If the keepalive is set to None (using the GUI) or if no keepalive is specified for an answer (using the CLI), 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 requests. However, it enables you to expand the types of devices for which the GSS can perform load balancing, including remote caches, application servers, and SLBs.</p>

NS (keepalive) Keepalive that is used when the GSS answer that you are testing is a Name Server (NS). The NS keepalive type sends a query for a domain you specify to a name server at an address that you provide. The online status is determined by the ability of the name server to resolve the domain to an address.

O

order Balance method configuration option that is used when the balance method for the answer group is set to Ordered List. Answers on the list will be given precedence in responding to requests based upon their position in the list.

ordered list Balance method in which each resource within an answer group is assigned a number, from 1 to X—where X is the number of resources in the group. Each number corresponds to the rank of the device in the group, with devices that have lower numbers ranked above those with higher numbers. Using the rankings, the GSS tries each resource in an order established by the GSS administrator, selecting the first available answer to serve a user request. List members are preferred and tried in order. A member will not be used unless all previous members fail to provide a suitable result. The Ordered List method allows you to manage resources at a single content site, for example, in a standalone deployment, or a redundant deployment in which the standby SLBs remain passive and are not used to serve requests.

origin server Machine that serves original or replicated content provider content.

owner Internal department or resource or external customer associated with a group of GSS resources such as domain lists, answer groups, and so on.

P

PDB Proximity database (PDB) that provides the core intelligence for all proximity-based decisions of a GSS. Proximity lookup occurs when a DNS rule is matched and the associated clause has the proximity option enabled. When the GSS receives a request from a D-proxy and decides that a proximate answer should be provided, the GSS identifies the most proximate answer from the PDB that resides in GSS memory (the answer with the lowest RTT time) and sends the answer to the requesting D-proxy. If the PDB proximity process is unable to determine a proximate answer, the GSS collects the zone-specific RTT results, measured from proximity probing agents in every zone in the proximity network, and puts the results into the PDB in GSS memory. The GSS supports a maximum of 500,000 entries in the PDB.

probing Process of measuring RTT from one proximity probing agent (DRP agent) to a requesting D-proxy device. Probe management is the intelligence behind each GSS device's interaction with the proximity probing agent in a zone. Within each zone, there must be at least one proximity probing agent and, optionally, a backup proximity probing agent. If the primary proximity probing agent fails, the probes are redirected to the backup device. Once the primary proximity probing agent becomes available, probes are redirected back to the primary proximity probing agent. The GSS supports two probing methods, direct and refresh probing.

proximity Ability to answer DNS requests with the most proximate answers relative to the requesting D-proxy. 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 proximity probing agent, 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.

R

region Grouping of GSS locations with common geographic attributes used to organize GSS resources.

round-robin Balance method in which each resource within an answer group is listed, though in no particular order. As requests are received, the GSS cycles through the list of resources, selecting the first available answer from the group. The GSS is able to resolve requests by evenly distributing the load among possible answers at both local and remote content sites. This balance method allows you to balance requests among multiple, active data centers hosting identical content, for example, between SLBs at a primary and active standby site that serves requests.

RTT Round-trip time (RTT). The GSS transmits DRP queries to one or more proximity probing agents in the GSS network, instructing the DRP agent in the proximity probing agent to probe specific D-proxy IP addresses. Each proximity probing agent responds to the query by using a standard protocol, such as ICMP or TCP, to measure the RTT between the DRP agent in the zone and the IP address of the requesting client's D-proxy device. From the RTT values in the PDB, the GSS selects the zone with the smallest RTT value as the most proximate zone containing the answer for the client's D-proxy request.

S

Scripted Kal Keepalive type used when the GSS answer that you are testing is a VIP associated with an SLB device such as a CSS, CSM, or ACE. The Scripted Kal keepalive type is used to probe third-party devices and obtain the load information. Scripted Kal uses the SNMP get request to fetch the load information from the target device.

Secure Socket Layer (SSL) Industry-standard method for protecting and encrypting web communication.

server load balancer (SLB)	Network device that balances content requests to network resources based on content rules and real-time load and availability data collected from those devices. Server load balancers such as the Cisco Content Services Switch (CSS), the Content Switching Module (CSM), and LocalDirector provide publicly routable virtual IP addresses (VIPs) while front-ending content servers, firewalls, Secure Socket Layer (SSL) terminators, and caches. Third-party SLBs are supported in a GSS network through the use of Internet Message Control Protocol (ICMP), TCP, and HTTP HEAD keepalives.
service provider	Cisco customer that provides infrastructure for a Content Delivery Network (CDN). Also ISP (Internet service provider) and ASP (application service provider).
source address list	List of source IPs or source IP blocks that are logically grouped by the system administrator.
static proximity	Type of request routing in which incoming requests from specified D-proxies are routed to statically defined resources that have been identified as being in proximity to the source D-proxies.
sticky	Process of binding a client, via their D-Proxy, to a specific server for some amount of time in order to allow the client to complete a transaction. 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 provide identical A-record responses to the requesting client D-proxy, assuming that the original VIP continues to be available. This GSS supports local and global sticky operation.
sticky database	Database that provides the core intelligence for all DNS sticky-based decisions made by a GSS, on a local or global level. The GSS collects requests from the client D-proxies and stores these requests in memory as the sticky database. Requests may be the IP address of the client D-proxy or a database ID representing a list of D-proxy IP addresses (configured as a D-proxy group). The sticky database stores each hosted domain that the DNS rule matches, which may be a single hosted domain (including wildcard expressions) or a configured list of hosted domains. These components make up each sticky database key that the GSS uses for the lookup, storage, and persistence of stickiness for DNS responses. The GSS supports a maximum of 400,000 entries in the sticky database.
subscriber	Client or set of clients receiving a certain style of DNS routing. Subscribers often pay for application services from the GSS customer.

T

TCP	TCP keepalive is used when the GSS answer that you are testing is to GSLB devices other than a CSS or CSM. These GSLB remote devices can include web servers, 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.
Time To Live (TTL)	Length of time that a response is to be cached and considered valid by the requesting D-proxy.
transaction	Series of specific client and server interactions that are logically connected to a single activity, such as viewing a large VoD file or performing a secure financial transaction.

V

Video on Demand (VoD) Generic term for rich media content, including video, audio, presentations and program executables.

Virtual IP Address (VIP) Used by server load-balancing (SLB) devices such as the Cisco CSS and CSM to represent content hosted on one or more servers under their control. The use of VIPs requests for content is efficiently routed to the proper host without exposing that device's internal IP addresses to external users. When directed to a VIP by a GSS, the client's D-Proxy next queries the SLB device to a suitable host, and the A-record for that device is returned by the SLB device to the D-Proxy as an answer.

W

Web Cache Control Protocol (WCCP) Cisco IOS feature for packet interception.

Web Network Services (WebNS) VxWorks-based operating system and software that runs on the Content Services Switch (CSS).

weight Balance method used when the balance method for the answer group is set to Round-Robin or Least-Loaded. Specified by a number between 1 and 10, weights indicate the capacity of the Answer to respond to requests as follows:

- When used with a round-robin balance method, the number listed will be used by the GSS to create a ratio of the number of times the answer will be used to respond before trying the next answer on the list.
- When used with the least-loaded balance method, the number listed will be used by the GSS as the divisor in calculating the load number associated with the answer, which is used to create a bias in favor of answers with greater capacity.

weighted round robin Balance method that is similar to round robin in that the GSS cycles through a list of defined answers, choosing the first available answer based on the defined load threshold, and so on. However, using WRR, an additional weight factor is assigned to each answer, biasing the GSS toward certain servers so they are picked more often.

Z

zone Based on the arrangement of devices and network partitioned characteristics, a customer network can be logically partitioned into “zones.” A zone can be geographically related to data centers in a continent, a country, or a major city. All devices, such as web servers in a data center, that are located in the same zone have the same proximity value when communicating with other areas of the Internet. You can configure a GSS proximity network with up to 32 zones. Within each zone, an active proximity probing agent is configured to accept probing instructions from any GSS device. Probing refers to the process of measuring RTT from one proximity probing agent to a requesting D-proxy device.

