RADIUS Server Load Balancing

Information About RADIUS Server Load Balancing

RADIUS Server Load Balancing Overview

By default, if two RADIUS servers are configured in a server group, only one is used. The other server acts as standby, if the primary server is declared as dead, the secondary server receives all the load.

If you need both servers to perform transactions actively, you need to enable Load Balancing.

✎

Note   By default, load balancing is not enabled on the RADIUS server group.

If you enable load balancing in a RADIUS server group with two or more RADIUS servers, the Server A and Server B receives a AAA transaction. The transaction queues are checked in Server A and Server B. The server with less number of outstanding transactions are assigned the next batch of AAA transaction.

Load balancing distributes batches of transactions to RADIUS servers in a server group. Load balancing assigns each batch of transactions to the server with the lowest number of outstanding transactions in its queue. The process of assigning a batch of transactions is as follows:

1. The first transaction is received for a new batch.

2. All server transaction queues are checked.

3. The server with the lowest number of outstanding transactions is identified.

4. The identified server is assigned the next batch of transactions.

The batch size is a user-configured parameter. Changes in the batch size may impact CPU load and network throughput. As batch size increases, CPU load decreases, and network throughput increases. However, if a large batch size is used, all available server resources may not be fully utilized. As batch size decreases, CPU load increases and network throughput decreases.

Note   There is no set number for large or small batch sizes. A batch with more than 50 transactions is considered large and a batch with fewer than 25 transactions is considered small.

Note   If a server group contains ten or more servers, we recommend that you set a high batch size to reduce CPU load.

Transaction Load Balancing Across RADIUS Server Groups

You can configure load balancing either per-named RADIUS server group or for the global RADIUS server group. The load balancing server group must be referred to as "radius" in the authentication, authorization, and accounting (AAA) method lists. All public servers that are part of the RADIUS server group are then load balanced.

You can configure authentication and accounting to use the same RADIUS server or different servers. In some cases, the same server can be used for preauthentication, authentication, or accounting transactions for a session. The preferred server, which is an internal setting and is set as the default, informs AAA to use the same server for the start and stop record for a session regardless of the server cost. When using the preferred server setting, ensure that the server that is used for the initial transaction (for example, authentication), the preferred server, is part of any other server group that is used for a subsequent transaction (for example, accounting).

The preferred server is not used if one of the following criteria is true:

- The **load-balance method least-outstanding ignore-preferred-server** command is used.

- The preferred server is dead.

- The preferred server is in quarantine.

- The want server flag has been set, overriding the preferred server setting.

The want server flag, an internal setting, is used when the same server must be used for all stages of a multistage transaction regardless of the server cost. If the want server is not available, the transaction fails.

You can use the **load-balance method least-outstanding ignore-preferred-server** command if you have either of the following configurations:

- Dedicated authentication server and a separate dedicated accounting server

- Network where you can track all call record statistics and call record details, including start and stop records and records that are stored on separate servers

If you have a configuration where authentication servers are a superset of accounting servers, the preferred server is not used.

Note If a third-party RADIUS load balancer is used and RADIUS packets are routed based on the NAS source port, it is recommended to move to any other rule based on the following Attribute-Value Pairs (AVPs):

- If the load balancer uses NAS source port in the Access-Request to load balance, rules may not work as expected as the source port in NAS might change during transaction.

- If the load balancer compares AVPs between Access-Challenge and Access-Request to route packets, you will need to use the AVP value of t-State.

- If the load balancer compares AVPs in Access-Request from NAS, you will need to use one or a combination of the following AVPs:

  - t-State value

  - Calling-Station-ID and NAS IP or Identifier

Prerequisites for RADIUS Server Load Balancing

- Authentication, Authorization, and Accounting (AAA) must be configured on the RADIUS server.

- AAA RADIUS server groups must be configured.

- RADIUS must be configured for functions such as authentication, accounting, or static route download.

Restrictions for RADIUS Server Load Balancing

- Incoming RADIUS requests, such as Packet of Disconnect (POD) requests are not supported.

- Load balancing is not supported on proxy RADIUS servers and private server groups.

- Load balancing is not supported on Central Web Authentication (CWA).

Enabling Load Balancing for a Named RADIUS Server Group (CLI)

Procedure

	Command or Action	Purpose
Step 1	**configure terminal** **Example:** `Device# configure terminal`	Enters global configuration mode.

	Command or Action	Purpose
Step 2	**aaa group server radius** *group-name* **Example:** Device(config)# aaa group server radius rad-sg	Enters server group configuration mode.
Step 3	**server ip-address** [**auth-port port-number**] [**acct-port port-number**] **Example:** Device(config-sg-radius)# server 192.0.2.238 auth-port 2095 acct-port 2096	Configures the IP address of the RADIUS server for the group server.
Step 4	**load-balance method least-outstanding** [**batch-size** *number*] [**ignore-preferred-server**] **Example:** Device(config-sg-radius)# load-balance method least-outstanding batch-size 30	Enables the least-outstanding load balancing for a named server group.
Step 5	**end** **Example:** Device(config-sg)# end	Exits server group configuration mode and enters privileged EXEC mode.