

## How to Select a Scheduling Policy

<https://campus.barracuda.com/doc/3539013/>

### In this article:

The Barracuda Load Balancer supports multiple scheduling methods to determine which Real Server that is associated with a Service gets the next new connection. On an ongoing basis each Real Server is assigned a weight, which indicates the proportion of the load that this Real Server will bear relative to other Real Servers. Weights are either calculated dynamically using Adaptive Scheduling, or they are pre-assigned. These Real Server weights are then used by the scheduling algorithm, which is either Weighted Round-Robin or Weighted Least Connections, to determine which Real Server gets the next connection.

### Adaptive Scheduling

The Adaptive Scheduling feature polls the Real Servers frequently and assigns weights to those Real Servers using the information gathered. The parameter polled may be:

- CPU Load, determined by an SNMP query. If you wish to use this and you have Real Servers running a version of Windows, refer to [How to Configure Adaptive Scheduling](#).
- Number of Windows Terminal Server sessions, determined by an SNMP query. In order to use this option, Real Servers must allow the Barracuda Load Balancer SNMP access to the community specified in the SNMP Community String box. This option is not available if the Service type is Layer 7 - RDP (see [Scheduling for a Service with type Layer 7 - RDP](#) on page 48).
- A URL provided by each Real Server which specifies a load value. If this option is selected, the Barracuda Load Balancer will poll the URL `http://[Real Server IP Address]/barracuda_load/` and expect the output to look like `LOAD=23` (showing the load as an integer between 0 and 100). Weights are assigned to each Real Server using the formula  $(100 - \text{LOAD})$ . For example, if the Load URL value is 23, the Real Server is assigned a weight of 77. In order for the URL query to work, you must create a load determination script and make the results available by running a web server on the Real Server that responds to the poll at the Real Server's IP address and port 80.

If, for example, all Real Servers have the same value for CPU load, then the Real Servers will be assigned the same weight. These weights will change as the value of the CPU Load for each Real Server varies.

Configure adaptive scheduling for a Service by editing it using the **BASIC > Services** page. On the Service Detail page, select the adaptive scheduling algorithm to use when making weight adjustments.

## Pre-Assigned Weight

---

As an alternative to adaptive scheduling, static weights for each Real Server can be used. If some of the Real Servers are faster or have more capacity than others, you can tell the Barracuda Load Balancer to direct more traffic to them by increasing their weight relative to the other Real Servers.

Configure the static weight for a Real Server by editing it on the **BASIC > Services** page. On the **Real Server Detail** page, enter a weight value to be compared against the weights of all other Real Servers for this Service. For example, a Real Server with a weight of 50 will get half the amount of traffic as a Real Server with a weight of 100, but will get twice that of a Real Server with a weight of 25.

If the Service is configured to use adaptive scheduling, these static weight values are ignored.

## Scheduling Policies

---

The Barracuda Load Balancer considers the weight values for the Real Servers and then applies a scheduling algorithm, either Weighted Round-Robin or Weighted Least Connections, to determine which Real Server gets the next connection.

In Weighted Round-Robin, Real Servers with higher weights get more connections than those with lower weights and Real Servers with equal weights get equal connections. The scheduling sequence is generated according to the Real Server weights. New connections are directed to the different Real Servers based on the scheduling sequence in a round-robin manner. The shortcoming with this method is that a majority of long-lived connections may go to the same Real Server.

In **Weighted Least Connections**, the Barracuda Load Balancer considers the number of live connections that each Real Server has, as well as the weight values. The Real Servers with higher weight values will receive a larger percentage of live connections at any one time. The Barracuda Load Balancer dynamically checks the number of live connections for each Real Server.

Weighted Least Connections is the recommended choice.

To configure whether Weighted Round-Robin or Weighted Least Connections will be used for a Service, edit the Service on the **BASIC > Services** page.

---

## Scheduling for a Service with type Layer 7 - RDP

---

If the Service type is Layer 7 - RDP, the Barracuda Load Balancer keeps track of the number of RDP sessions on each Real Server. This number is used in conjunction with Real Server weights when selecting which Real Server gets the next new session. The Real Server weights are determined by either one of these adaptive scheduling methods:

- Executing an SNMP GET for the CPU load on the Real Servers;
- Polling a URL provided by each Real Server which specifies a load value;
- or by retrieving pre-configured static weights (from the **Real Server Detail** page).

The number of active RDP sessions and the Real Server weights are used as input to the Weighted Round Robin or Weighted Least Connections algorithm.

On the **Service Detail** page the Terminal Sessions adaptive scheduling option is disabled for Layer 7 - RDP Services. Because the number of RDP sessions on each Real Server is maintained internally, there is no need for the adaptive scheduling algorithm to issue an SNMP query to get the number of active Windows Terminal Sessions.

## Viewing Current Connections

---

To see the number of current open connections/requests/sessions with each Service and each Real Server, navigate to the **BASIC > Server Health** page. The bars on the page display the approximate percentage of all traffic that is currently connected to each Service or Real Server.

Sometimes it may appear that a Real Server is handling more traffic than it should be based on its calculated weight. This is caused by persistence. If clients that were previously connected reconnect within a short period of time, they are directed to the same Real Server regardless of its current load.

© Barracuda Networks Inc., 2024 The information contained within this document is confidential and proprietary to Barracuda Networks Inc. No portion of this document may be copied, distributed, publicized or used for other than internal documentary purposes without the written consent of an official representative of Barracuda Networks Inc. All specifications are subject to change without notice. Barracuda Networks Inc. assumes no responsibility for any inaccuracies in this document. Barracuda Networks Inc. reserves the right to change, modify, transfer, or otherwise revise this publication without notice.