CHAPTER 45

Configuring Cache Services By Using WCCP

This chapter describes how to configure your Catalyst 3750 switch to redirect traffic to wide-area application engines (such as the Cisco Cache Engine 550) by using the Web Cache Communication Protocol (WCCP). This software release supports only WCCP version 2 (WCCPv2).

WCCP is a Cisco-developed content-routing technology that you can use to integrate wide-area application engines—referred to as *application engines*—into your network infrastructure. The application engines transparently store frequently accessed content and then fulfill successive requests for the same content, eliminating repetitive transmissions of identical content from servers. Application engines accelerate content delivery and ensure maximum scalability and availability of content. In a service-provider network, you can deploy the WCCP and application engine solution at the points of presence (POPs). In an enterprise network, you can deploy the WCCP and application engine solution at the regional site and the small branch office.

To use this feature, the stack master must be running the IP services image. Unless otherwise noted, the term *switch* refers to a standalone switch and to a switch stack.

Note For complete syntax and usage information for the commands used in this chapter, see the "WCCP Router Configuration Commands" section in the "*System Management Commands*" part of the *Cisco IOS Configuration Fundamentals Command Reference, Release 12.4* on Cisco.com.

This chapter consists of these sections:

- Understanding WCCP, page 45-1
- Configuring WCCP, page 45-5
- Monitoring and Maintaining WCCP, page 45-10

Understanding WCCP

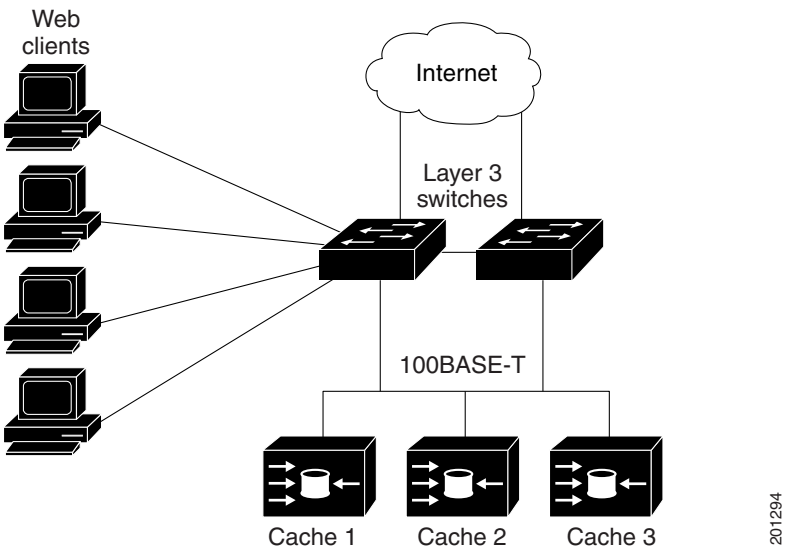
The WCCP and Cisco cache engines (or other application engines running WCCP) localize traffic patterns in the network, enabling content requests to be fulfilled locally.

WCCP enables supported Cisco routers and switches to transparently redirect content requests. With transparent redirection, users do not have to configure their browsers to use a web proxy. Instead, they can use the target URL to request content, and their requests are automatically redirected to an application engine. The word *transparent* means that the end user does not know that a requested file (such as a web page) came from the application engine instead of from the originally specified server.

When an application engine receives a request, it attempts to service it from its own local cache. If the requested information is not present, the application engine sends a separate request to the end server to retrieve the requested information. After receiving the requested information, the application engine forwards it to the requesting client and also caches it to fulfill future requests.

With WCCP, the application-engine cluster (a series of application engines) can service multiple routers or switches, as shown Figure 45-1.

Figure 45-1 Cisco Cache Engine and WCCP Network Configuration



WCCP Message Exchange

This sequence of events describes the WCCP message exchange:

1. The application engines send their IP addresses to the WCCP-enabled switch by using WCCP, signaling their presence through a *Here I am* message. The switch and application engines communicate to each other through a control channel based on UDP port 2048.
2. The WCCP-enabled switch uses the application engine IP information to create a cluster view (a list of application engines in the cluster). This view is sent through an *I see you* message to each application engine in the cluster, essentially making all the application engines aware of each other. A stable view is established after the membership of the cluster remains the same for a certain amount of time.
3. When a stable view is established, the application engine in the cluster with the lowest IP address is elected as the designated application engine.

WCCP Negotiation

In the exchange of WCCP protocol messages, the designated application engine and the WCCP-enabled switch negotiate these items:

- Forwarding method (the method by which the switch forwards packets to the application engine). The switch rewrites the Layer 2 header by replacing the packet destination MAC address with the target application engine MAC address. It then forwards the packet to the application engine. This forwarding method requires the target application engine to be directly connected to the switch at Layer 2.
- Assignment method (the method by which packets are distributed among the application engines in the cluster). The switch uses some bits of the destination IP address, the source IP address, the destination Layer 4 port, and the source Layer 4 port to determine which application engine receives the redirected packets.
- Packet-return method (the method by which packets are returned from the application engine to the switch for normal forwarding). These are the typical reasons why an application engine rejects packets and starts the packet-return feature:
  - The application engine is overloaded and has no room to service the packets.
  - The application engine receives an error message (such as a protocol or authentication error) from the server and uses the dynamic client bypass feature. The bypass enables clients to bypass the application engines and to connect directly to the server.

  The application engine returns a packet to the WCCP-enabled switch to forward to the server as if the application engine is not present. The application engine does not intercept the reconnection attempt. In this way, the application engine effectively cancels the redirection of a packet to the application engine and creates a bypass flow. If the return method is generic-route encapsulation (GRE), the switch receives the returned packet through a GRE tunnel that is configured in the application engine. The switch CPU uses Cisco Express Forwarding to send these packets to the target server. If the return method is Layer 2 rewrite, the packets are forwarded in hardware to the target server. When the server responds with the information, the switch uses normal Layer 3 forwarding to return the information to the requesting client.

MD5 Security

WCCP provides an optional security component in each protocol message to enable the switch to use MD5 authentication on messages between the switch and the application engine. Messages that do not authenticate by MD5 (when authentication of the switch is enabled) are discarded by the switch. The password string is combined with the MD5 value to create security for the connection between the switch and the application engine. You must configure the same password on each application engine.

Packet Redirection and Service Groups

You can configure WCCP to classify traffic for redirection, such as FTP, proxy-web-cache handling, and audio and video applications. This classification, known as a *service group*, is based on the protocol type (TCP or UDP) and the Layer 4 source destination port numbers. The service groups are identified either by well-known names such as web-cache, which means TCP port 80, or a service number, 0 to 99. Service groups are configured to map to a protocol and Layer 4 port numbers and are established and maintained independently. WCCP allows dynamic service groups, where the classification criteria are provided dynamically by a participating application engine.

You can configure up to 8 service groups on a switch or switch stack and up to 32 cache engines per service group. WCCP maintains the priority of the service group in the group definition. WCCP uses the priority to configure the service groups in the switch hardware. For example, if service group 1 has a priority of 100 and looks for destination port 80, and service group 2 has a priority of 50 and looks for source port 80, the incoming packet with source and destination port 80 is forwarded by using service group 1 because it has the higher priority.

WCCP supports a cluster of application engines for every service group. Redirected traffic can be sent to any one of the application engines. The switch supports the mask assignment method of load balancing the traffic among the application engines in the cluster for a service group.

After WCCP is configured on the switch, the switch forwards all service group packets received from clients to the application engines. However, these packets are not redirected:

- Packets originating from the application engine and targeted to the server.
- Packets originating from the application engine and targeted to the client.
- Packets returned or rejected by the application engine. These packets are sent to the server.

You can configure a single multicast address per service group for sending and receiving protocol messages. When there is a single multicast address, the application engine sends a notification to one address, which provides coverage for all routers in the service group, for example, 225.0.0.0. If you add and remove routers dynamically, using a single multicast address provides easier configuration because you do not need to specifically enter the addresses of all devices in the WCCP network.

You can use a router group list to validate the protocol packets received from the application engine. Packets matching the address in the group list are processed, packets not matching the group list address are dropped.

To disable caching for specific clients, servers, or client/server pairs, you can use a WCCP redirect access control list (ACL). Packets that do not match the redirect ACL bypass the cache and are forwarded normally.

Before WCCP packets are redirected, the switch examines ACLs associated with all inbound features configured on the interface and permits or denies packet forwarding based on how the packet matches the entries in the ACL.

Note Only **permit** ACL entries are supported in WCCP redirect lists.

When packets are redirected, the output ACLs associated with the redirected interface are applied to the packets. Any ACLs associated with the original port are not applied unless you specifically configure the required output ACLs on the redirected interfaces.

WCCP and Switch Stacks

WCCP support is the same for a switch stack as for a standalone switch. WCCP configuration information is propagated to all switches in the stack. All switches in the stack, including the stack master, process the information and program their hardware. For more information about switch stacks, see Chapter 5, "Managing Switch Stacks."

The stack master performs these WCCP functions:

- It receives protocol packets from any WCCP-enabled interface and sends them out any WCCP-enabled interface in the stack.
- It processes the WCCP configuration and propagates the information to all stack members.
- It distributes the WCCP information to any switch that joins the stack.
- It programs its hardware with the WCCP information it processes.

Stack members receive the WCCP information from the master switch and program their hardware.

Unsupported WCCP Features

These WCCP features are not supported in this software release:

- Packet redirection on an outbound interface that is configured by using the **ip wccp redirect out** interface configuration command. This command is not supported.
- The GRE forwarding method for packet redirection is not supported.
- The hash assignment method for load balancing is not supported.
- There is no SNMP support for WCCP.

Configuring WCCP

These sections describe how to configure WCCP on your switch:

- Default WCCP Configuration, page 45-5
- WCCP Configuration Guidelines, page 45-6
- Enabling the Cache Service, page 45-6 (required)

Default WCCP Configuration

Table 45-1 shows the default WCCP configuration.

Table 45-1 Default WCCP Configuration

Feature	Default Setting
WCCP enable state	WCCP services are disabled.
Protocol version	WCCPv2.
Redirecting traffic received on an interface	Disabled.

WCCP Configuration Guidelines

Before configuring WCCP on your switch, make sure to follow these configuration guidelines:

- The application engines and switches in the same service group must be in the same subnetwork directly connected to the switch that has WCCP enabled.
- Configure the switch interfaces that are connected to the clients, the application engines, and the server as Layer 3 interfaces (routed ports and switch virtual interfaces [SVIs]). For WCCP packet redirection to work, the servers, application engines, and clients must be on different subnets.
- Use only nonreserved multicast addresses when configuring a single multicast address for each application engine.
- WCCP entries and PBR entries use the same TCAM region. WCCP is supported only on the templates that support PBR: access, routing, and dual IPv4/v6 routing.
- When TCAM entries are not available to add WCCP entries, packets are not redirected and are forwarded by using the standard routing tables.
- The number of available policy-based routing (PBR) labels are reduced as more interfaces are enabled for WCCP ingress redirection. For every interface that supports service groups, one label is consumed. The WCCP labels are taken from the PBR labels. You need to monitor and manage the labels that are available between PBR and WCCP. When labels are not available, the switch cannot add service groups. However, if another interface has the same sequence of service groups, a new label is not needed, and the group can be added to the interface.
- The routing maximum transmission unit (MTU) size configured on the stack member switches should be larger than the client MTU size. The MAC-layer MTU size configured on ports connected to application engines should take into account the GRE tunnel header bytes.
- You cannot configure WCCP and VPN routing/forwarding (VRF) on the same switch interface.
- You cannot configure WCCP and PBR on the same switch interface.
- You cannot configure WCCP and a private VLAN (PVLAN) on the same switch interface.

Enabling the Cache Service

For WCCP packet redirection to operate, you must configure the switch interface connected to the client to redirect inbound packets.

This procedure shows how to configure these features on routed ports. To configure these features on SVIs, see the configuration examples that follow the procedure.

Note Before configuring WCCP commands, configure the SDM template, and reboot the switch. For more information, see Chapter 8, “Configuring SDM Templates.”

Beginning in privileged EXEC mode, follow these steps to enable the cache service, to set a multicast group address or group list, to configure routed interfaces, to redirect inbound packets received from a client to the application engine, enable an interface to listen for a multicast address, and to set a password. This procedure is required.

	Command	Purpose
Step 1	**configure terminal**	Enter global configuration mode.
Step 2	**ip wccp** {**web-cache** | *service-number*} [**group-address** *groupaddress*] [**group-list** *access-list*] [**redirect-list** *access-list*] [**password** *encryption-number password*]	Enable the cache service, and specify the service number which corresponds to a dynamic service that is defined by the application engine. By default, this feature is disabled. (Optional) For **group-address** *groupaddress*, specify the multicast group address used by the switches and the application engines to participate in the service group. (Optional) For **group-list** *access-list*, if a multicast group address is not used, specify a list of valid IP addresses that correspond to the application engines that are participating in the service group. (Optional) For **redirect-list** *access-list*, specify the redirect service for specific hosts or specific packets from hosts. (Optional) For **password** *encryption-number password*, specify an encryption number. The range is 0 to 7. Use 0 for not encrypted, and use 7 for proprietary. Specify a password name up to seven characters in length. The switch combines the password with the MD5 authentication value to create security for the connection between the switch and the application engine. By default, no password is configured, and no authentication is performed. You must configure the same password on each application engine. When authentication is enabled, the switch discards messages that are not authenticated.
Step 3	**interface** *interface-id*	Specify the interface connected to the application engine or the server, and enter interface configuration mode.
Step 4	**no switchport**	Enter Layer 3 mode.
Step 5	**ip address** *ip-address subnet-mask*	Configure the IP address and subnet mask.
Step 6	**no shutdown**	Enable the interface.
Step 7	**exit**	Return to global configuration mode. Repeat Steps 3 through 7 for each application engine and server.
Step 8	**interface** *interface-id*	Specify the interface connected to the client, and enter interface configuration mode.
Step 9	**no switchport**	Enter Layer 3 mode.
Step 10	**ip address** *ip-address subnet-mask*	Configure the IP address and subnet mask.
Step 11	**no shutdown**	Enable the interface.
Step 12	**ip wccp** {**web-cache** | *service-number*} **redirect in**	Redirect packets received from the client to the application engine. Enable this on the interface connected to the client.
Step 13	**ip wccp** {**web-cache** | *service-number*} **group-listen**	(Optional) When using a multicast group address, **group-listen** enables the interface to listen for the multicast address. Enable this on the interface connected to the application engine.

	Command	Purpose
Step 14	**exit**	Return to global configuration mode. Repeat Steps 8 through 13 for each client.
Step 15	**end**	Return to privileged EXEC mode.
Step 16	**show ip wccp web-cache** and **show running-config**	Verify your entries.
Step 17	**copy running-config startup-config**	(Optional) Save your entries in the configuration file.

To disable the cache service, use the **no ip wccp web-cache** global configuration command. To disable inbound packet redirection, use the **no ip wccp web-cache redirect in** interface configuration command. After completing this procedure, configure the application engines in the network.

This example shows how to configure routed interfaces and to enable the cache service with a multicast group address and a redirect access list. Gigabit Ethernet port 1 is connected to the application engine, is configured as a routed port with an IP address of 172.20.10.30, and is re-enabled. Gigabit Ethernet port 2 is connected through the Internet to the server, is configured as a routed port with an IP address of 175.20.20.10, and is re-enabled. Gigabit Ethernet ports 3 to 6 are connected to the clients and are configured as routed ports with IP addresses 175.20.30.20, 175.20.40.30, 175.20.50.40, and 175.20.60.50. The switch listens for multicast traffic and redirects packets received from the client interfaces to the application engine.

```
Switch# configure terminal
Switch(config)# ip wccp web-cache group-address 224.1.1.100 redirect list 12
Switch(config)# access-list 12 permit host 10.1.1.1
Switch(config)# interface gigabitethernet1/0/1
Switch(config-if)# no switchport
Switch(config-if)# ip address 172.20.10.30 255.255.255.0
Switch(config-if)# no shutdown
Switch(config-if)# ip wccp web-cache group-listen
Switch(config-if)# exit
Switch(config)# interface gigabitethernet1/0/2
Switch(config-if)# no switchport
Switch(config-if)# ip address 175.20.20.10 255.255.255.0
Switch(config-if)# no shutdown
Switch(config-if)# exit
Switch(config)# interface gigabitethernet1/0/3
Switch(config-if)# no switchport
Switch(config-if)# ip address 175.20.30.20 255.255.255.0
Switch(config-if)# no shutdown
Switch(config-if)# ip wccp web-cache redirect in
Switch(config-if)# exit
Switch(config)# interface gigabitethernet1/0/4
Switch(config-if)# no switchport
Switch(config-if)# ip address 175.20.40.30 255.255.255.0
Switch(config-if)# no shutdown
Switch(config-if)# ip wccp web-cache redirect in
Switch(config-if)# exit
Switch(config)# interface gigabitethernet1/0/5
Switch(config-if)# no switchport
Switch(config-if)# ip address 175.20.50.40 255.255.255.0
Switch(config-if)# no shutdown
Switch(config-if)# ip wccp web-cache redirect in
Switch(config-if)# exit
Switch(config)# interface gigabitethernet1/0/6
Switch(config-if)# no switchport
Switch(config-if)# ip address 175.20.60.50 255.255.255.0
```

```
Switch(config-if)# no shutdown
Switch(config-if)# ip wccp web-cache redirect in
Switch(config-if)# exit
```

This example shows how to configure SVIs and how to enable the cache service with a multicast group list. VLAN 299 is created and configured with an IP address of 175.20.20.10. Gigabit Ethernet port 1 is connected through the Internet to the server and is configured as an access port in VLAN 299. VLAN 300 is created and configured with an IP address of 172.20.10.30. Gigabit Ethernet port 2 is connected to the application engine and is configured as an access port in VLAN 300. VLAN 301 is created and configured with an IP address of 175.20.30.50. Fast Ethernet ports 3 to 6, which are connected to the clients, are configured as access ports in VLAN 301. The switch redirects packets received from the client interfaces to the application engine.

Note Only **permit** ACL entries are being used in the redirect-list; **deny** entries are unsupported.

```
Switch# configure terminal
Switch(config)# ip wccp web-cache 80 group-list 15
Switch(config)# access-list 15 permit host 171.69.198.102
Switch(config)# access-list 15 permit host 171.69.198.104
Switch(config)# access-list 15 permit host 171.69.198.106
Switch(config)# vlan 299
Switch(config-vlan)# exit
Switch(config)# interface vlan 299
Switch(config-if)# ip address 175.20.20.10 255.255.255.0
Switch(config-if)# exit
Switch(config)# interface gigabitethernet1/0/1
Switch(config-if)# switchport mode access
Switch(config-if)# switchport access vlan 299
Switch(config)# vlan 300
Switch(config-vlan)# exit
Switch(config)# interface vlan 300
Switch(config-if)# ip address 171.69.198.100 255.255.255.0
Switch(config-if)# exit
Switch(config)# interface gigabitethernet1/0/2
Switch(config-if)# switchport mode access
Switch(config-if)# switchport access vlan 300
Switch(config-if)# exit
Switch(config)# vlan 301
Switch(config-vlan)# exit
Switch(config)# interface vlan 301
Switch(config-if)# ip address 175.20.30.20 255.255.255.0
Switch(config-if)# ip wccp web-cache redirect in
Switch(config-if)# exit
Switch(config)# interface gigabitethernet1/0/3 - 6
Switch(config-if-range)# switchport mode access
Switch(config-if-range)# switchport access vlan 301
Switch(config-if-range)# exit
```

Monitoring and Maintaining WCCP

To monitor and maintain WCCP, use one or more of the privileged EXEC commands in Table 45-2:

Table 45-2 Commands for Monitoring and Maintaining WCCP

Command	Purpose
clear ip wccp web-cache	Removes statistics for the web-cache service.
show ip wccp web-cache	Displays global information related to WCCP.
show ip wccp web-cache detail	Displays information for the switch and all application engines in the WCCP cluster.
show ip interface	Displays status about any IP WCCP redirection commands that are configured on an interface; for example, `Web Cache Redirect is enabled / disabled.`
show ip wccp web-cache view	Displays which other members have or have not been detected.