Configuring ECMP for Host Routes

This chapter describes how to configure the equal-cost multipathing (ECMP) protocol for host routes on the Cisco NX-OS switch.

This chapter includes the following sections:

- Information About ECMP for Host Routes, on page 1
- Guidelines for ECMP for Host Routes, on page 1
- Prerequisites for ECMP for Host Routes, on page 2
- Default Settings, on page 2
- Configuring ECMP for Host Routes, on page 2
- Configuring Weighted ECMP over BGP, on page 4
- Configuring Dynamic ECMP Group Resizing, on page 6
- Verifying the ECMP for Host Routes Configuration, on page 7
- Configuration Examples for ECMP for Host Routes, on page 7
- Additional References, on page 7

Information About ECMP for Host Routes

When you enable ECMP support for host routes, all unicast host routes are programmed into the longest-prefix match algorithm (LPM) table. ECMP for host routes is provided in the switch hardware. You configure this feature in the CLI using the **hardware profile unicast enable-host-ecmp** command.

Note Host entries are stored in the LPM routing table instead of the host table when ECMP is configured for IPv4 (/32) routes and IPv6 (/128) routes.

Guidelines for ECMP for Host Routes

ECMP for host routes has the following configuration guidelines and limitations:

- After enabling or disabling ECMP for host routes by using the [no] hardware profile unicast enable-host-ecmp command, ensure that you do the following:

- – Save the current configuration on the switch by using the copy running-config startup-config command.
- – Reload the switch by using the reload command so that the configuration can be applied.

Prerequisites for ECMP for Host Routes

ECMP for host routes has the following prerequisites:

- Before you use this command, we recommend that you disable Unicast Reverse Path Forwarding (URPF) globally on the switch using the **system urpf disable** command, and then save the configuration and reload the switch. Disabling URPF globally extends the LPM table size.

Default Settings

ECMP for host routes is disabled by default.

Configuring ECMP for Host Routes

This section includes the following topics:

Enabling the ECMP for Host Routes Feature

You can enable the ECMP for host routes feature.

Procedure

	Command or Action	Purpose

Step 1	**configure terminal**
	Example:
	`switch# configure terminal`
	`switch(config)#`
		Enters global configuration mode.

Step 2	(Optional) **system urpf disable** **Example:** `switch(config)# system urpf disable`	Disables URPF globally on the switch.
Step 3	**hardware profile unicast enable-host-ecmp** **Example:** `switch(config)# hardware profile unicast enable-host-ecmp`	Enables ECMP for host routes globally on the switch.

	Command or Action	Purpose
Step 4	**copy running-config startup-config** **Example:** `switch(config)# copy running-config startup-config`	Saves this configuration change.

Step 5	**reload**
	Example:
	`switch(config)# reload`
	`WARNING: This command will reboot the system`
	`Do you want to continue? (y/n) [n] y`
		Reloads the Cisco Nexus 3000 Series switches software.

Disabling the EMCP for Host Routes Feature

You can disable the ECMP for host routes feature.

Procedure

	Command or Action	Purpose

Step 1	**configure terminal**
	Example:
	`switch# configure terminal`
	`switch(config)#`
		Enters global configuration mode.

Step 2	**no hardware profile unicast enable-host-ecmp** **Example:** `switch(config)# no hardware profile unicast enable-host-ecmp`	Disables ECMP for host routes globally on the switch and removes all associated configuration.
Step 3	**copy running-config startup-config** **Example:** `switch(config)# copy running-config startup-config`	Saves this configuration change.

Step 4	**reload**
	Example:
	`switch(config)# reload`
	`WARNING: This command will reboot the system`
	`Do you want to continue? (y/n) [n] y`
		Reloads the Cisco Nexus 3000 Series switches software.

Configuring Weighted ECMP over BGP

ECMP is a mechanism that allows multiple routes to the same destination with different next-hops and it load-balances the routed traffic over those multiple next-hops.The basic ECMP works for most of the customers' requirements. The load entropy is the best way to maximize the link usage efficiency.

Often, the application distribution in the network can be unbalanced. The new clusters roll in at different over-subscription rates than the old clusters. The new clusters have powerful servers than the old clusters and they are capable of handling more load per CPU. As the network is not perfect, some control over routing behavior is needed. You can configure Weighted ECMP over BGP for balancing the load traffic and for administering control over the routing behavior.

The following use-cases are considered for this deployment:

- Unequal distribution of resources
- SDN or non-Homogeneous Fabric

Configuring Weighted ECMP over BGP

ECMP is a mechanism that allows multiple routes to the same destination with different next-hops and it load-balances the routed traffic over those multiple next-hops.The basic ECMP works for most of the customers' requirements. The load entropy is the best way to maximize the link usage efficiency.

Often, the application distribution in the network can be unbalanced. The new clusters roll in at different over-subscription rates than the old clusters. The new clusters have powerful servers than the old clusters and they are capable of handling more load per CPU. As the network is not perfect, some control over routing behavior is needed. You can configure Weighted ECMP over BGP for balancing the load traffic and for administering control over the routing behavior.

The following use-cases are considered for this deployment:

- Unequal distribution of resources
- SDN or non-Homogeneous Fabric

SDN/Non-Homogeneous Fabric

For SDN/non-homogeneous fabric, consider the following example:

- You have two instances of VIP1 (an any service).
- The server NIC IP address is advertised as the next hop (NH) for the VIP.
- The VIP1 is deployed evenly across the data center.
- Cluster 1 is oversubscribed 3:1 and Cluster 2 is oversubscribed 12:1.

As a result, the traditional ECMP delivers the traffic sub-optimally.

Solutions for the Use-Cases

The solution for the unequal distribution of the resources and sub-optimal traffic distribution use-cases is to configure Weighted ECMP over BGP. You can inject the VIP routes (from the host or the controller) and signal a weight for each instance. You can then aggregate the weights across the infrastructure and deliver the traffic in the direct proportion to the application deployment distribution.

Guidelines and Limitations for Configuring Weighted ECMP over BGP

See the following guidelines for configuring Weighted ECMP over BGP:

- Cisco Nexus 3100 platform switches support weighted ECMP only in non-resilient mode.
- BGP uses the Link Bandwidth EXTCOMM defined in the draft-ietf-idr-link-bandwidth-06.txt to implement the weighted ECMP feature.
- The weighted ECMP feature is supported for IPv4 address-family only.
- You can accept Link Bandwidth EXTCOMM from both iBGP and eBGP peers.
- Do not send the Link Bandwidth EXTCOMM in the BGP updates. The BGP controller/peer sends the link bandwidth to all the routers.
- For weights programming, the link bandwidth EXTCOMM has the link bandwidth encoded in bytes/second, as a four byte floating point integer, that is converted to kbits/second unsigned integer before downloading to RIB.
- The hardware ECMP width is fixed as 64 in size.

Displaying Link Bandwidth EXTCOMM fields

See the following example for the displaying the link bandwidth EXTCOMM fields:

```
Link Bandwidth EXTCOMM fields:
Link Bandwidth attribute – “40 04 00 64 47 80 00 00”
Where “40 04” specifies a Link Bandwidth EXTCOMM
Where “00 64” specifies a AS #.
Where “47 80 00 00” specifies the Link Bandwidth value as floating point integer.
The Link Bandwidth floating point value bits are encoded as follows:
#define IEEE_NUMBER_WIDTH 32 /* bits in number */
#define IEEE_EXP_WIDTH 8 /* bits in exponent */
#define IEEE_MANTISSA_WIDTH (IEEE_NUMBER_WIDTH - 1 - IEEE_EXP_WIDTH)
#define IEEE_SIGN_MASK 0x80000000
#define IEEE_EXPONENT_MASK 0x7F800000
#define IEEE_MANTISSA_MASK 0x007FFFFF
Link Bandwidth value programmed to RIB is calculated as follows:
uint32_t ieee_bw_32 = ntohl(GETLONG(&extcomm->value[4])); = 0x47800000
int64_t dmzlink_bw_64 = ptr_ieee_to_int64(&ieee_bw_32); = 65536
uint32_t value = 520 = (uint32_t)((dmzlink_bw_64 / 1000) * 8) = 520
```

Configuring Dynamic ECMP Group Resizing

Configuring the dynamic ECMP group resizing feature allows you to configure more number of ECMP groups on Cisco Nexus 3000 Series switches and Cisco Nexus 3100 platform switches. You can configure up to 1022 ECMP Groups. The ECMP groups sizes are not fixed in the hardware.

Note

- In Cisco Nexus 3000 Series switches, the dynamic ECMP group resizing feature is enabled by default. In Cisco Nexus 3100 platform switches, the dynamic ECMP group resizing feature is available in non-resilient mode.

The following new CLI command is introduced to configure more number of ECMP groups in Cisco Nexus 3100 platform switches:

```
(config)# no hardware profile ecmp resilient
Warning: The command will take effect after next reload.
(config)#
```

Verifying Dynamic ECMP Group Resizing

Use the following command to verify the configuration of the ECMP groups:

```
# show hardware profile status
slot 1
=======

Total LPM Entries = 7679.
Total Host Entries = 16384.
Reserved LPM Entries = 1024.
Max Host4/Host6 Limit Entries (shared)= 8192/4096*
Max Mcast Limit Entries = 4096.
Max Ucast IPv6 LPM Limit Entries = 2048.
Max Ucast IPv6 LPM_65_to_127 Limit Entries = 128.
Used LPM Entries (Total) = 4.
Used IPv4 LPM Entries = 1.
Used IPv6 LPM Entries = 2.
Used IPv6 LPM_65_to_127 Entries = 1.
Used Host Entries in LPM (Total) = 0.
Used Host4 Entries in LPM = 0.
Used Host6 Entries in LPM = 0.
Used Mcast Entries = 0.
Used Mcast OIFL Entries = 0.
Used Host Entries in Host (Total) = 0.
Used Host4 Entries in Host = 0.
Used Host6 Entries in Host = 0.
Max ECMP Table Entries = 1022.
Used ECMP Table Entries = 0.
Max ECMP Next Hop Table Entries = 16384.
Used ECMP Next Hop Table Entries = 0.
MFIB prefer-source-tree = Disabled/0/0.

*Unicast Host Table is in shared mode b/n v4 & v6...
#
```

Verifying the ECMP for Host Routes Configuration

To display the ECMP for host routes configuration information, perform one of the following tasks:

Command	Purpose
show hardware profile status	Displays the unicast and multicast routing entries in hardware tables.
show running-config	Displays the running system configuration.

Configuration Examples for ECMP for Host Routes

This example shows how to disable URPF and configure ECMP for host routes:

```
switch# configure terminal
switch(config)# system urpf disable
switch(config)# hardware profile unicast enable-host-ecmp
switch(config)# copy running-config startup-config
switch(config)# reload
```

This example show how to disable ECMP for host routes:

```
switch# configure terminal
switch(config)# no hardware profile unicast enable-host-ecmp
switch(config)# copy running-config startup-config
switch(config)# reload
```

Additional References

For additional information related to implementing ECMP for host routes, see the following sections:

Related Documents

Related Documents

Related Topic	Document Title
ECMP for host routes CLI commands	Cisco Nexus 3000 Series Command Reference