

# 非等价负荷分担中CEF权重分布的理解

## 目录

[简介](#)

[先决条件](#)

[要求](#)

[使用的组件](#)

[UCMP概述](#)

[初始配置](#)

[度量权重/负载](#)

[UCMP方差确定](#)

[了解权重](#)

[确定权重值](#)

[权重](#)

[归一化权重](#)

[控制CEF重量/负荷比](#)

[示例 1：重量/负荷比为26/5](#)

[示例 2：重量/负荷比为30/1](#)

## 简介

本文档介绍了在IOS-XR中了解、配置和检验非等价多路径的各个方面。此外，我们还通过加权处理示例来展示到达目的地的路径度量如何影响链路上的负载。

## 先决条件

本文档没有必备条件。

## 要求

以下示例基于IOS-XR 6.4.1。

## 使用的组件

本文档中的信息都是基于特定实验室环境中的设备编写的。本文档中使用的所有设备最初均采用原始（默认）配置。如果您使用的是真实网络，请确保您已经了解所有命令的潜在影响。

## UCMP概述

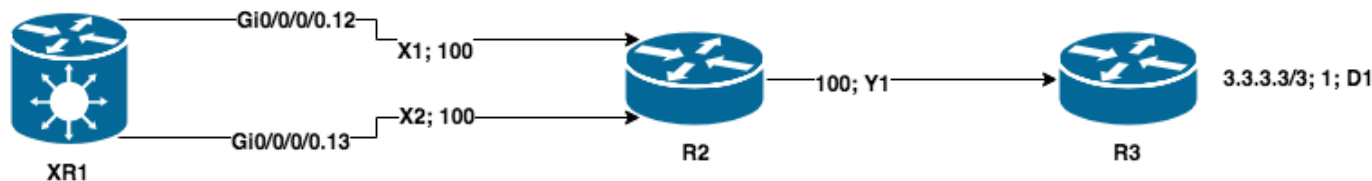
非等价多路径(UCMP)负载均衡能够以不同的成本在多条路径之间按比例均衡流量负载。通常，带宽较高的路径配置了较低的内部网关协议(IGP)度量，因此它们构成最短的IGP路径。

启用UCMP负载均衡后，协议可以使用更低带宽路径或更高成本的流量路径，并且可以将这些路径

安装到转发信息库(FIB)。这些协议仍在FIB中安装多条通往同一目的地的路径，但每条路径都将有一个与其关联的“负载度量/权重”。FIB使用此负载度量/权重来确定需要在较高带宽路径上发送的流量和需要在较低带宽路径上发送的流量。

传统上，EIGRP是唯一支持UCMP功能的IGP，但在IOS-XR UCMP中，所有IGP、静态路由和BGP均受支持。在本文档中，我们将以OSPF作为示例的基础来解释UCMP功能，但此处的信息也适用于IS-IS和其他支持UCMP的协议。

## 拓扑图



## 初始配置

```
XR1
!
hostname XR1
!
interface GigabitEthernet0/0/0/0.12
description TO R2
ipv4 address 12.0.0.1 255.255.255.0
encapsulation dot1q 12
!
interface GigabitEthernet0/0/0/0.13
description TO R2
ipv4 address 13.0.0.1 255.255.255.0
encapsulation dot1q 13
! router ospf 1 address-family ipv4 area 0 ! interface GigabitEthernet0/0/0/0.12 cost 100
!
interface GigabitEthernet0/0/0/0.13
cost 100
!
!
!
end
```

```
R2
!
hostname R2
!
interface Ethernet0/0.12
description TO XR1
encapsulation dot1Q 12
ip address 12.0.0.2 255.255.255.0
!
interface Ethernet0/0.13
description TO XR1
encapsulation dot1Q 13
ip address 13.0.0.2 255.255.255.0
!
interface Ethernet0/1
description TO R3
ip address 172.16.23.2 255.255.255.0
ip ospf cost 100
```

```

!
!
router ospf 1
 network 0.0.0.0 255.255.255.255 area 0
!
end

```

```

R3
!
hostname R3
!
interface Loopback0
 description FINAL_DESTINATION
 ip address 3.3.3.3 255.255.255.255
!
interface Ethernet0/0
 description TO R2
 ip address 172.16.23.3 255.255.255.0
!
router ospf 1
 network 0.0.0.0 255.255.255.255 area 0
!
end

```

## 度量权重/负载

在IOS-XR中，当我们安装到目的地的多条路径时，会为目的地分配一个权重值，该值指示特定链路的负载分布。此值与到目的地的路径度量成反比，开销越高，分配的权重越低。这允许CEF在路由到目的地时智能地执行链路负载共享。

安装ECMP路径时，为所有路径分配的权重值始终设置为0，这意味着流量负载均等共享。如果我们检查CEF，我们可以确认已为每个路径分配了0的权重。

```
RP/0/RP0/CPU0:XR1#show cef 3.3.3.3/32 detail
```

```

3.3.3.3/32, version 87, internal 0x1000001 0x0 (ptr 0xd689b50) [1], 0x0 (0xd820648), 0x0 (0x0)
Updated Nov 11 22:15:58.953
remote adjacency to GigabitEthernet0/0/0/0.12
Prefix Len 32, traffic index 0, precedence n/a, priority 1
 gateway array (0xd6b32f8) reference count 2, flags 0x0, source rib (7), 0 backups
   [3 type 3 flags 0x8401 (0xd759758) ext 0x0 (0x0)]
 LW-LDI[type=3, refc=1, ptr=0xd820648, sh-ldi=0xd759758]
 gateway array update type-time 1 Nov 11 22:15:58.953
 LDI Update time Nov 11 22:15:58.953
 LW-LDI-TS Nov 11 22:15:58.953
via 12.0.0.2/32, GigabitEthernet0/0/0/0.12, 4 dependencies, weight 0, class 0 [flags 0x0]
 path-idx 0 NHID 0x0 [0xe14b0a0 0x0]
 next hop 12.0.0.2/32
 remote adjacency
via 13.0.0.2/32, GigabitEthernet0/0/0/0.13, 4 dependencies, weight 0, class 0 [flags 0x0]
 path-idx 1 NHID 0x0 [0xe14b128 0x0]
 next hop 13.0.0.2/32
 remote adjacency

Load distribution: 0 1 (refcount 3)

Hash OK Interface Address
0 Y GigabitEthernet0/0/0/0.12 remote
1 Y GigabitEthernet0/0/0/0.13 remote

```

# UCMP方差确定

如果要启用UCMP，我们先从在XR1上以不同方式设置成本开始，为此，我们将设置如下成本：

```
router ospf 1
address-family ipv4
area 0
interface Loopback0
!
interface GigabitEthernet0/0/0/0.12
cost 50
!
interface GigabitEthernet0/0/0/0.13
cost 100
!
end
```

```
RP/0/RP0/CPU0:XR1#show route 3.3.3.3/32
```

```
Routing entry for 3.3.3.3/32
Known via "ospf 1", distance 110, metric 151, type intra area
Installed Nov 11 22:32:48.289 for 00:00:32
Routing Descriptor Blocks
12.0.0.2, from 3.3.3.3, via GigabitEthernet0/0/0/0.12
Route metric is 151
No advertising protos.
```

要考虑UCMP的其他路径，我们需要确定这些路径是否符合条件。IOS-XR对IS-IS和OSPF使用百分比标准，这基于ucmp variance <value>路由器进程命令。我们有两条路径：

路径度量1(pm1)= 151

路径度量2(pm2)= 201

无环路下一跳将根据UCMP  $\leq (\text{Variance} * \text{Primary path metric}) / 100$ 安装。

在本例中，要达到最坏路径度量(pm2)，主路径必须增长的量是151的134%，结果为202。这是我们需要配置的精确方差值，以使路径符合条件。

```
!
router ospf 1
ucmp variance 134
!

RP/0/RP0/CPU0:XR1#show route 3.3.3.3/32

Routing entry for 3.3.3.3/32
Known via "ospf 1", distance 110, metric 151, type intra area
Installed Nov 11 22:36:45.720 for 00:00:09
Routing Descriptor Blocks
12.0.0.2, from 3.3.3.3, via GigabitEthernet0/0/0/0.12
Route metric is 151, Wt is 4294967295
13.0.0.2, from 3.3.3.3, via GigabitEthernet0/0/0/0.13
Route metric is 151, Wt is 3226567396
No advertising protos.
```

[Spoiler](#)

**注意：**方差值对权重结果没有任何影响。在这种情况下，最小方差134或方差10000（最大值）将导致相同的权重结果，而成本值是影响所得权重的值，因为这些值彼此成反比。

**注意：**方差值对权重结果没有任何影响。在这种情况下，最小方差134或方差10000（最大值）将导致相同的权重结果，而成本值是影响所得权重的值，因为这些值彼此成反比。

## 了解权重

在IOS-XR中，我们有两种不同的权重，**权重**和**规范权重**。这些数据包的使用取决于特定平台支持多少个哈希桶，XRv9000分别支持32个哈希桶，ASR 9000和CRS-X分别支持64个哈希桶。这意味着，当路由器编程权重值时，权重不能超过特定平台的哈希桶限制。通过发出show cef <prefix> detail location <location>命令，我们可以观察规范化权重编程到什么。根据成本值集，我们有18, 13负载分配，这意味着已分配31个哈希桶(18+13)。

```
RP/0/RP0/CPU0:XR1#show cef 3.3.3.3/32 detail
```

```
3.3.3.3/32, version 23, internal 0x1000001 0x0 (ptr 0xd3ecb50) [1], 0x0 (0xd583610), 0x0 (0x0)
Updated Nov 11 22:36:45.723
remote adjacency to GigabitEthernet0/0/0/0.12
Prefix Len 32, traffic index 0, precedence n/a, priority 1
gateway array (0xd4163d8) reference count 1, flags 0x0, source rib (7), 0 backups
    [2 type 3 flags 0x8401 (0xd4bc7b8) ext 0x0 (0x0)]
LW-LDI[type=3, refc=1, ptr=0xd583610, sh-ldi=0xd4bc7b8]
gateway array update type-time 1 Nov 11 22:36:45.723
LDI Update time Nov 11 22:36:45.729
LW-LDI-TS Nov 11 22:36:45.729
  via 12.0.0.2/32, GigabitEthernet0/0/0/0.12, 6 dependencies, weight 4294967295, class 0 [flags 0x0]
    path-idx 0 NHID 0x0 [0xe14b1b0 0x0]
    next hop 12.0.0.2/32
    remote adjacency
  via 13.0.0.2/32, GigabitEthernet0/0/0/0.13, 6 dependencies, weight 3226567396, class 0 [flags 0x0]
    path-idx 1 NHID 0x0 [0xe14b128 0x0]
    next hop 13.0.0.2/32
    remote adjacency

Weight distribution:
slot 0, weight 4294967295, normalized_weight 18, class 0
slot 1, weight 3226567396, normalized_weight 13, class 0

Load distribution: 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 (refcount
2)
```

Hash	OK	Interface	Address
0	Y	GigabitEthernet0/0/0/0.12	remote
1	Y	GigabitEthernet0/0/0/0.12	remote
2	Y	GigabitEthernet0/0/0/0.12	remote
3	Y	GigabitEthernet0/0/0/0.12	remote
4	Y	GigabitEthernet0/0/0/0.12	remote
5	Y	GigabitEthernet0/0/0/0.12	remote
6	Y	GigabitEthernet0/0/0/0.12	remote
7	Y	GigabitEthernet0/0/0/0.12	remote
8	Y	GigabitEthernet0/0/0/0.12	remote
9	Y	GigabitEthernet0/0/0/0.12	remote
10	Y	GigabitEthernet0/0/0/0.12	remote
11	Y	GigabitEthernet0/0/0/0.12	remote

```

12 Y GigabitEthernet0/0/0/0.12 remote
13 Y GigabitEthernet0/0/0/0.12 remote
14 Y GigabitEthernet0/0/0/0.12 remote
15 Y GigabitEthernet0/0/0/0.12 remote
16 Y GigabitEthernet0/0/0/0.12 remote
17 Y GigabitEthernet0/0/0/0.12 remote
18 Y GigabitEthernet0/0/0/0.13 remote
19 Y GigabitEthernet0/0/0/0.13 remote
20 Y GigabitEthernet0/0/0/0.13 remote
21 Y GigabitEthernet0/0/0/0.13 remote
22 Y GigabitEthernet0/0/0/0.13 remote
23 Y GigabitEthernet0/0/0/0.13 remote
24 Y GigabitEthernet0/0/0/0.13 remote
25 Y GigabitEthernet0/0/0/0.13 remote
26 Y GigabitEthernet0/0/0/0.13 remote
27 Y GigabitEthernet0/0/0/0.13 remote
28 Y GigabitEthernet0/0/0/0.13 remote
29 Y GigabitEthernet0/0/0/0.13 remote
30 Y GigabitEthernet0/0/0/0.13 remote

```

我们可以观察到，规范化权重的总和表示平台分配的哈希桶数量，在这种情况下，根据此特定平台的限制，我们永远不能超过32个哈希桶。主路径(pm1)的权重始终设置为4294967295，即最大权重 $(2^{32}) - 1$ 。

## 确定权重值

### 权重

我们可以使用公式  $\text{权重} = \frac{\text{最佳成本}}{\text{最差成本}} * 4294967295$  轻松计算权重。例如，路径1和路径2的权重计算如下：

$\text{Weight\_path\_1} = \text{始终设置为} 4294967295$

$\text{Weight\_path\_2} = 151 / 201 * 4294967295 = 3226567470$

### [Spoiler](#)

**注意：**当计算这些值时，由于我们要进行浮点计算，因此在RIB和FIB中必须安装整数。

**注意：**当计算这些值时，由于我们要进行浮点计算，因此在RIB和FIB中必须安装整数。

### 归一化权重

如前所述，我们无法在CEF表中安装超过平台哈希桶数量的权重值，因为这需要在将权重编程到硬件之前对其进行规范化。平台根据公式  $\text{规范化权重} = (\text{路径权重} / \text{总权重}) * \text{最大桶大小}$  计算规范化权重。根据我们的示例，我们可以按如下方式计算：

$\text{normalized\_weight\_1} = (4294967295 * 32) / (3226567396 + 4294967295) = 18$

$\text{normalized\_weight\_2} = (3226567396 * 32) / (3226567396 + 4294967295) = 13$

### [Spoiler](#)

**注意：**当G.C.D等于1时，则使用上述方法，否则，如果G.C.D  $\neq 1$ ，归一化权值将得到的GCD按权值进行划分。

**注意：**当G.C.D等于1时，则使用上述方法，否则，如果G.C.D  $\neq 1$ ，归一化权重，按权重值对得到的GCD进行划分。

## 控制CEF重量/负荷比

在某些情况下，我们可能想确定需要配置什么特定路径度量值才能获得最终的权重/负载分布。我们可以通过更改链路的开销并基于达到或接近所需值之前确定适当的路径度量。请注意，我们可能需要的所有权重并非完全可能，但我们可以近似地计算所需的分布。

在继续之前，请考虑以下限制：

- a) 并非所有的权重/载荷分布都完全可能，但我们可以进行近似。
- b) 切勿超过哈希桶限制。 — 这意味着所有路径权重的总和不能超过哈希桶，如果发生这种情况，则必须规范化权重。这意味着，在添加所有权重时，我们不会超出哈希桶限制。
- c) ASR 9000和CRS-X有64个散列桶限制，XRv9000有32个散列桶限制。
- d) 当使用6.4.1之前时，权重分布是不同的，而权重最小的路径总是设置为权重1，而其它路径是该路径的倍数，这意味着它可以大于1。

### 示例 1：重量/负荷比为26/5

按照之前的相同拓扑，我们希望在两条链路之间分配26/5的权重。

- i) 最初，所有路径(100 + 100 + 1)的成本均等设置为201。
- ii) 如果将UCMP方差设置为最大值，则考虑所有下一跳。
- iii) 如果检查RIB，我们可以看到XR1执行ECMP的默认状态。

```
RP/0/RP0/CPU0:XR1#show cef 3.3.3.3/32 detail
```

```
3.3.3.3/32, version 27, internal 0x1000001 0x0 (ptr 0xd3ecb50) [1], 0x0 (0xd583610), 0x0 (0x0)
Updated Nov 11 23:08:25.290
remote adjacency to GigabitEthernet0/0/0/0.12
Prefix Len 32, traffic index 0, precedence n/a, priority 1
gateway array (0xd416218) reference count 2, flags 0x0, source rib (7), 0 backups
    [3 type 3 flags 0x8401 (0xd4bc6f8) ext 0x0 (0x0)]
LW-LDI[type=3, refc=1, ptr=0xd583610, sh-ldi=0xd4bc6f8]
gateway array update type-time 1 Nov 11 23:08:25.290
LDI Update time Nov 11 23:08:25.297
LW-LDI-TS Nov 11 23:08:25.297
  via 12.0.0.2/32, GigabitEthernet0/0/0/0.12, 4 dependencies, weight 4294967295, class 0 [flags
0x0]
    path-idx 0 NHID 0x0 [0xe14b1b0 0x0]
    next hop 12.0.0.2/32
    remote adjacency
  via 13.0.0.2/32, GigabitEthernet0/0/0/0.13, 4 dependencies, weight 4294967295, class 0 [flags
0x0]
    path-idx 1 NHID 0x0 [0xe14b128 0x0]
    next hop 13.0.0.2/32
    remote adjacency

Weight distribution:
slot 0, weight 4294967295, normalized_weight 1, class 0
slot 1, weight 4294967295, normalized_weight 1, class 0

Load distribution: 0 1 (refcount 3)
```

Hash	OK	Interface	Address
0	Y	GigabitEthernet0/0/0/0.12	remote
1	Y	GigabitEthernet0/0/0/0.13	remote

在本例中，我们将使用您需要以下权重的案例：

W1 = 26 ( 主要最佳成本 )

W2 = 5 ( 次要最佳成本 )

我们需要采用一条支路，对于此路径，开销应已知，在这种情况下，参考路径将是通过 Gi0/0/0/0.12的路径。支路路径将使用端到端开销预先计算，此路径所需的路径度量和权重为：

i)  $X1+Y1+D1 = 100 + 100 + 1 = 201$ 。( 注意拓扑中每个链路所连接的变量 )。

ii) 重量1 = 26

三) 重量2 = 5

iv.)  $pm1 = 201$  ( 主腿路径 ) ; 重量= 26

v.)  $pm2 = \text{未知}$  ( 辅助路径 ) ; 重量= 5

计算权重。

$pm2$ 的路径度量： $pm2 = (26/5) * 201 = 1045$

确定XR1上链路X2的开销。

$X2 = pm2 - (x2 + y1 + d1)$

$1045 - (100 + 100 + 1) = 844$

在X2链路上配置OSPF开销。

```
router ospf 1
  ucmp variance 10000
  area 0
  !
  interface GigabitEthernet0/0/0/0.13
    cost 844
```

验证权重/负荷分布，我们可以看到在CEF中所需的权重已按照我们在计算中预测的分配正确。

RP/0/RP0/CPU0:XR1#show cef 3.3.3.3/32 detail

```
3.3.3.3/32, version 37, internal 0x1000001 0x0 (ptr 0xd3ecce0) [1], 0x0 (0xd5835d8), 0x0 (0x0)
Updated Nov 11 23:17:47.945
remote adjacency to GigabitEthernet0/0/0/0.12
Prefix Len 32, traffic index 0, precedence n/a, priority 1
gateway array (0xd4163d8) reference count 1, flags 0x0, source rib (7), 0 backups
  [2 type 3 flags 0x8401 (0xd4bc7b8) ext 0x0 (0x0)]
LW-LDI[type=3, refc=1, ptr=0xd5835d8, sh-ldi=0xd4bc7b8]
gateway array update type-time 1 Nov 11 23:17:47.945
LDI Update time Nov 11 23:17:47.956
```







Hash	OK	Interface	Address
0	Y	GigabitEthernet0/0/0/0.12	remote
1	Y	GigabitEthernet0/0/0/0.12	remote
2	Y	GigabitEthernet0/0/0/0.12	remote
3	Y	GigabitEthernet0/0/0/0.12	remote
4	Y	GigabitEthernet0/0/0/0.12	remote
5	Y	GigabitEthernet0/0/0/0.12	remote
6	Y	GigabitEthernet0/0/0/0.12	remote
7	Y	GigabitEthernet0/0/0/0.12	remote
8	Y	GigabitEthernet0/0/0/0.12	remote
9	Y	GigabitEthernet0/0/0/0.12	remote
10	Y	GigabitEthernet0/0/0/0.12	remote
11	Y	GigabitEthernet0/0/0/0.12	remote
12	Y	GigabitEthernet0/0/0/0.12	remote
13	Y	GigabitEthernet0/0/0/0.12	remote
14	Y	GigabitEthernet0/0/0/0.12	remote
15	Y	GigabitEthernet0/0/0/0.12	remote
16	Y	GigabitEthernet0/0/0/0.12	remote
17	Y	GigabitEthernet0/0/0/0.12	remote
18	Y	GigabitEthernet0/0/0/0.12	remote
19	Y	GigabitEthernet0/0/0/0.12	remote
20	Y	GigabitEthernet0/0/0/0.12	remote
21	Y	GigabitEthernet0/0/0/0.12	remote
22	Y	GigabitEthernet0/0/0/0.12	remote
23	Y	GigabitEthernet0/0/0/0.12	remote
24	Y	GigabitEthernet0/0/0/0.12	remote
25	Y	GigabitEthernet0/0/0/0.12	remote
26	Y	GigabitEthernet0/0/0/0.12	remote
27	Y	GigabitEthernet0/0/0/0.12	remote
28	Y	GigabitEthernet0/0/0/0.12	remote
29	Y	GigabitEthernet0/0/0/0.12	remote
30	Y	GigabitEthernet0/0/0/0.13	remote