

瞭解Catalyst 6500 S2T CEF專案

目錄

[簡介](#)

[必要條件](#)

[需求](#)

[採用元件](#)

[背景資訊](#)

[網路圖表](#)

[識別通過分散式轉發引擎的CEF條目](#)

[刪除CEF條目](#)

[新增CEF條目](#)

[新增和刪除VRF路由表的條目](#)

簡介

本檔案介紹搭載Sup2T監督器的Cisco Catalyst 6500如何線上卡硬體中為在Cisco IOS軟體上設定的 (Cisco快速轉送) CEF專案程式，以便實現封包轉送。

必要條件

需求

思科建議您瞭解以下主題：

- Cisco Express Forwarding(CEF)
- Cisco Catalyst 6500 系列交換器
- 思科分散式轉送卡(DFC)

採用元件

本檔案中的資訊是根據以下硬體和軟體版本：

- Cisco Catalyst 6500 WS-X6848-GE-TX (含DFC4) 線路卡。
- 採用IOS版本15.2.1SY5上監督器2T的Cisco Catalyst 6500

本文中的資訊是根據特定實驗室環境內的裝置所建立。文中使用到的所有裝置皆從已清除 (預設) 的組態來啟動。如果您的網路運作中，請確保您瞭解任何指令可能造成的影響。

背景資訊

大多數Cisco多層交換機都使用CEF作為第3層交換機制。網路工程師必須瞭解CEF的工作原理，以便每天對網路故障、資料包丟失或資料包延遲情況進行故障排除。

獨立模式下的Sup2T Supervisor或作為VSS當前被許多企業網路部署為核心交換機，實際上匯聚了所有其他路由或交換裝置。這也意味著轉發大多數域內和域間流量，以便成功將資料包傳送到其目標。要做到這一點，Sup2T必須具備通過靜態或路由協定動態獲知的正確路由資訊。

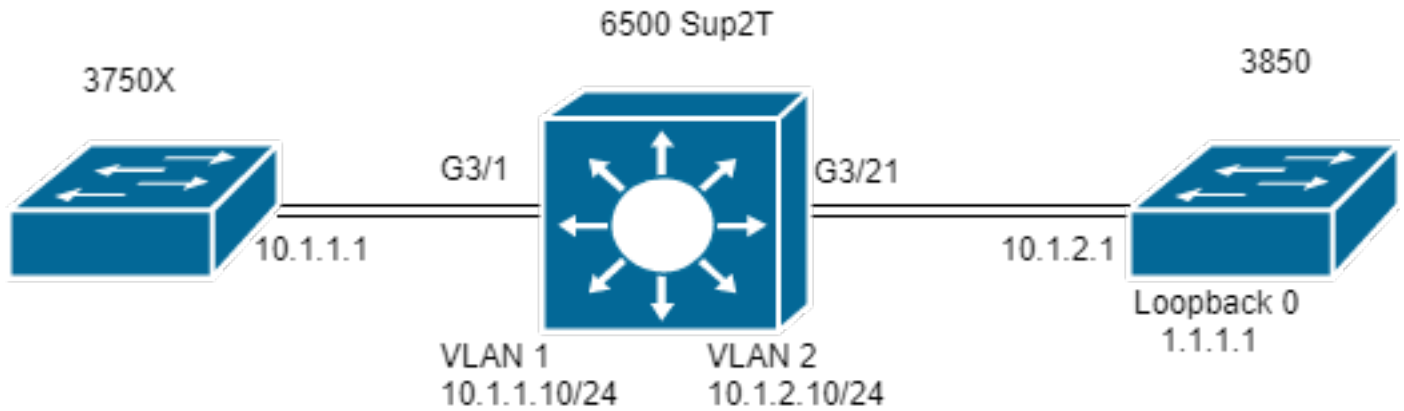
在模組化機箱中，除了管理引擎之外，可能還存在多個轉發引擎。為了提高資料包交換效能，某些線卡 (特別是新一代線卡，如C6800-32P10G) 已經包含了自己的轉發引擎，CEF條目的查詢在本地執行，並使資源以最佳方式分配到通過不同線卡入流的流量。這些線卡稱為分散式轉發卡(DFC)。

由於多種原因，跨所有轉發引擎共用的這些CEF條目可能無法在HW中分配，從軟體缺陷情況、資源耗盡到高CPU條件，並阻止交換機有足夠的時間更新所有條目，這可能導致一系列不希望發生的事件。

網路圖表

網路

:



```
Switch#show module 3
```

```
----- Mod Ports Card Type Model Serial No. ----
-----
10/100/1000mb Ethernet WS-X6848-GE-TX SAL2003X5AH -----
----- 3 48 CEF720 48 port
-----
3 Distributed Forwarding Card WS-F6K-DFC4-A SAL2003X5AH 1.4 Ok
```

識別通過分散式轉發引擎的CEF條目

在圖中，獨立6506交換器安裝有Supervisor 2T，以及線卡WS-6848-GE-TX（插槽3中有一個DFC）。主機3750X透過連線埠G3/1連線到線卡，會將流量傳送到3850的Loopback 0位址1.1.1.1。

為此，3750X具有到IP地址1.1.1.1的靜態路由，該路由通過下一跳10.1.1.10（即Sup2T交換機中VLAN 1的SVI）到達。Sup2T交換器需要透過下一個躍點10.1.2.1（連線到VLAN 2中Sup2T的3850介面），在IP 1.1.1.1/32的靜態路由專案中將此流量路由到3850交換器。

```
MXC.CALO.3750X#show ip route | inc 1.1.1.1
S 1.1.1.1 [1/0] via 10.1.1.10
```

```
MXC.CALO.Sup2T#show ip route | inc 1.1.1.1
S 1.1.1.1 [1/0] via 10.1.2.1
```

```
CALO.MXC.3850#show ip route | inc 1.1.1.1
C 1.1.1.1 is directly connected, Loopback1
```

請注意，為簡便起見，3750X和3850交換機均通過同一線卡連線到6500。這表示流量會在本地查詢並本地轉發。

封包透過Gi3/1進入Sup2T交換器，最終到達轉送引擎（因為這是DFC）。轉送引擎解析此封包中的目的地IP位址欄位，並在程式化CEF專案上尋找最佳相符專案（最長遮罩）。

由於這是DFC卡，這意味著它有自己的CEF條目並對其進行驗證，因此有必要使用**attach [dec]或attach switch [1-2] mod [dec]** for VSS命令連線到線卡。

現在，您應該在DFC提示符下，使用命令**show platform hardware cef**或**show platform hardware cef vpn 0**返回為常規路由表程式設計的所有CEF條目（VPN 0/無VRF）。

由於目標是字首1.1.1.1/32，因此您使用命令**show platform hardware cef vpn 0 lookup 1.1.1.1**。該命令返回字首1.1.1.1的最佳匹配項以及它用於實際轉發流量的匹配項：

```
MXC.CALO.Sup2T#attach 3
Trying Switch ...
```

Entering CONSOLE for Switch
Type "^C^C^C" to end this session

```
MXC.CALO.Sup2T-dfc3#show platform hardware cef vpn 0
Codes: decap - Decapsulation, + - Push Label
Index Prefix Adjacency
32 0.0.0.0/32 receive
33 255.255.255.255/32 receive
34 10.1.85.254/32 glean
35 10.1.85.5/32 receive
36 10.1.86.5/32 receive
[snip...]
```

```
MXC.CALO.Sup2T-dfc3#show platform hardware cef vpn 0 lookup 1.1.1.1
Codes: decap - Decapsulation, + - Push Label
Index Prefix Adjacency
262 1.1.1.1/32 V12 ,0c11.678b.f6f7
```

CEF條目就存在了，它是通過命令 `ip route 1.1.1.1 255.255.255.255 10.1.2.1` 在IOS軟體中程式設計的靜態條目而程式設計的。

您還可以通過命令 `show platform hardware cef 1.1.1.1 detail` (返回鄰接條目) 驗證此條目是否命中，並且與此條目一起轉發流量：

```
MXC.CALO.Sup2T-dfc3#show platform hardware cef 1.1.1.1 detail
Codes: M - mask entry, V - value entry, A - adjacency index, NR- no_route bit
LS - load sharing count, RI - router_ip bit, DF: default bit
CP - copy_to_cpu bit, AS: dest_AS_number, DGTv - dgt_valid bit
DGT: dgt/others value
```

```
Format:IPV4 (valid class vpn prefix)
M(262 ): 1 F 2FFF 255.255.255.255
V(262 ): 1 0 0 1.1.1.1
(A:114689, LS:0, NR:0, RI:0, DF:0 CP:0 DGTv:1, DGT:0)
```

最後，鄰接條目顯示如何重寫資料包，以及此鄰接條目是否重寫了流量：

```
MXC.CALO.Sup2T-dfc3#show platform hardware cef adjacencies entry 114689 detail
```

RIT fields: The entry has a Layer2 Format

```
-----
|decr_ttl = YES | pipe_ttl = 0 | utos = 0
|-----|-----|-----
|l2_fwd = 0 | rmac = 0 | ccc = L3_REWRITE
|-----|-----|-----
|rm_null_lbl = YES| rm_last_lbl = YES| pv = 0
|-----|-----|-----
|add_shim_hdr= NO | rec_findex = N/A | rec_shim_op = N/A
|-----|-----|-----
|rec_dti_type = N/A | rec_data = N/A
|-----|-----|-----
|modify_smac = YES| modify_dmac = YES| egress_mcast = NO
|-----|-----|-----
|ip_to_mac = NO
|-----|-----|-----
|dest_mac = 0c11.678b.f6f7 | src_mac = d8b1.902c.9680
|-----|-----|-----
|
```

```
Statistics: Packets = 642
Bytes = 75756 <<<<
```

`dest_mac`和`src_mac`是主要關注值，表示為此資料包寫入的新L2報頭。目的MAC地址`0c11.678b.f6f7`為`10.1.2.1`，即3850 (到達1.1.1.1的下一跳)：

```
MXC.CALO.Sup2T#show ip arp 10.1.2.1
Protocol Address Age (min) Hardware Addr Type Interface
Internet 10.1.2.1 30 0c11.678b.f6f7 ARPA Vlan2
```

此外，**Statistics**欄位顯示流量實際抵達此鄰接專案，並相應地重寫L2標頭。

刪除CEF條目

刪除CEF條目可以幫助我們刪除任何可能程式設計錯誤（例如，指向錯誤的鄰接條目）的條目，甚至是用於培訓目的的CEF條目。它還提供修改路由路徑的方法。

要刪除CEF條目，您需要瞭解CEF條目是按順序程式設計的，並且分配了硬體索引，例如：

```
MXC.CALO.Sup2T-dfc3#show platform hardware cef vpn 0
```

代碼：decap — 解除封裝，+ — 推送標籤

```
MXC.CALO.Sup2T-dfc3#show platform hardware cef vpn 0
...
Index Prefix Adjacency 259 10.1.2.255/32 receive 260 10.1.1.1/32 V11 ,a0ec.f930.3f40 261
10.1.2.1/32 V12 ,0c11.678b.f6f7 262 1.1.1.1/32 V12 ,0c11.678b.f6f7 <<<< Our CEF entry of
interest has a HW index of 262.
...
```

此硬體索引是刪除CEF條目的最重要元素，因為它被用作參考。但是，為了對其執行任何更改，必須將其轉換為軟體控制代碼。您可以使用**test platform hardware cef index-conv hw_to_sw [hw index]**

```
MXC.CALO.Sup2T-dfc3#test platform hardware cef index-conv hw_to_sw 262
hw index: 262 ----> sw handle: 101
```

現在您已瞭解軟體控制代碼，就可以使用**test platform hardware cef v4-delete [sw handle] mask [mask length] vpn [dec]**指令繼續刪除CEF專案了

```
MXC.CALO.s2TVSS-sw2-dfc3#test platform hardware cef v4-delete 101 mask 32 vpn 0
test_ipv4_delete: done.
```

附註： 遮罩長度值為32，因為這是一個主機特定路由(1.1.1.1/32)

現在，刪除我們的CEF條目：

```
MXC.CALO.Sup2T-dfc3#show platform hard cef vpn 0 1.1.1.1
Codes: decap - Decapsulation, + - Push Label
Index Prefix Adjacency
```

```
MXC.CALO.Sup2T-dfc3#show platform hard cef vpn 0
[snip...]
259 10.1.2.255/32 receive
260 10.1.1.1/32 V11 ,a0ec.f930.3f40
261 10.1.2.1/32 V12 ,0c11.678b.f6f7
288 224.0.0.0/24 receive <<<<<< Index 262 no longer exists in the CEF entries.
289 10.1.85.0/24 glean
```

請注意，**test platform hardware cef vpn 0**命令是在DFC提示符下執行的。如此一來，CEF條目已從DFC的CEF表中刪除，而不是從Supervisor中刪除，您必須非常小心從哪個轉發引擎中刪除這些條目。

流量發生變化時存在不可見性風險（在實驗室測試的情況下），這可能是由於另一個CEF條目被命中所致。請考慮始終匹配最精確的一個（最長掩碼）。在本實驗中，它命中：

```
MXC.CALO.Sup2T-dfc3#show plat hard cef vpn 0 lookup 1.1.1.1
```

```
Codes: decap - Decapsulation, + - Push Label
```

```
Index Prefix Adjacency
```

```
262048 0.0.0.0/0 glean
```

那麼，此條目實際如何處理資料包呢？：

```
MXC.CALO.Sup2T-dfc3#show platform hardware cef adjacencies entry 262048
```

```
RIT fields: The entry has a Recirc. Format
```

```
_____ |decr_ttl=NO | l2_fwd=NO | ccc = 6 |
add_shim_hdr = YES | _____ | _____ | _____ | _____ |rc_fidx=0 |
rc_shimop=1 | rc_dti_type=4 | rc_data = 0x10B
| _____ | _____ | _____ | _____ | _____ Statistics: Packets = 2163 Bytes =
255234
```

Taken from a CPU packet capture using Catlayst 6500 NETDR tool. For NETDR capture tool details refer to: [Catalyst 6500 Series Switches Netdr Tool for CPU-Bound Packet Captures](#)

----- dump of incoming inband packet -----

```
l2idb Pol, l3idb V11, routine inband_process_rx_packet, timestamp 01:00:17.841
dbus info: src_vlan 0x1(1), src_indx 0xB40(2880), len 0x82(130)
bpdu 0, index_dir 0, flood 0, dont_lrn 0, dest_indx 0x5FA4(24484), CoS 0
cap1 0, cap2 0
78020800 00018400 0B400100 82000000 1E000464 2E000004 00000010 5FA45BDD
destmac D8.B1.90.2C.96.80, srcmac A0.EC.F9.30.3F.40, shim ethertype CCF0
earl 8 shim header IS present:
version 0, control 64(0x40), lif 1(0x1), mark_enable 1,
feature_index 0, group_id 0(0x0), acos 0(0x0),
ttl 14, dti 4, dti_value 267(0x10B)
10000028 00038080 010B
ethertype 0800
protocol ip: version 0x04, hlen 0x05, tos 0x00, totlen 100, identifier 51573
df 0, mf 0, fo 0, ttl 255, src 10.1.1.1, dst 1.1.1.1
icmp type 8, code 0
```

----- dump of outgoing inband packet -----

```
l2idb NULL, l3idb V12, routine etsec_tx_pak, timestamp 01:03:56.989
dbus info: src_vlan 0x2(2), src_indx 0x380(896), len 0x82(130)
bpdu 0, index_dir 0, flood 0, dont_lrn 0, dest_indx 0x0(0), CoS 0
cap1 0, cap2 0
00020000 0002A800 03800000 82000000 00000000 00000000 00000000 00000000
destmac 0C.11.67.8B.F6.F7, srcmac D8.B1.90.2C.96.80, shim ethertype CCF0
earl 8 shim header IS present:
version 0, control 0(0x0), lif 16391(0x4007), mark_enable 0,
feature_index 0, group_id 0(0x0), acos 0(0x0),
ttl 15, dti 0, dti_value 540674(0x84002)
000800E0 0003C008 4002
ethertype 0800
protocol ip: version 0x04, hlen 0x05, tos 0x00, totlen 100, identifier 50407
df 0, mf 0, fo 0, ttl 254, src 10.1.1.1, dst 1.1.1.1
icmp type 8, code 0
```

現在，所有通過線卡3進入的目的地為1.1.1.1的流量都會使用填充碼報頭重新循環並傳送到CPU。有時，系統會顯示另一個具有drop鄰接的0.0.0.0/0條目，而不是此CEF條目，並且執行完全相同的操作。

附註：評估刪除哪些CEF條目。CPU使用率過高可能是由於此原因。通常配置預設路由0.0.0.0/0，並基於該路由轉發流量（並會導致資料包丟失）。


```
MXC.CALO.Sup2T(config)#arp 10.1.2.1 0c11.678b.f6f7 arpa <<< Static ARP configuration
```

```
MXC.CALO.Sup2T#show ip arp 10.1.2.1
Protocol Address Age (min) Hardware Addr Type Interface
Internet 10.1.2.1 - 0c11.678b.f6f7 ARPA <<< Now the static ARP entry is
complete
```

```
// Attaching to DFC3...
```

```
MXC.CALO.Sup2T-sw2-dfc3#show plat hard cef 10.1.2.1 detail
[snip...]
Format:IPV4 (valid class vpn prefix)
M(53 ): 1 F 2FFF 255.255.255.255
V(53 ): 1 0 0 10.1.2.1
(A:114689, LS:0, NR:0, RI:0, DF:0 CP:0 DGTv:1, DGT:0)
```

The ARP entry exist in CEF table for DFC3. Same Adjacency Index result as before...

現在您已瞭解這些鄰接條目的作用，您終於可以繼續新增CEF條目。在最後一節中，字首1.1.1.1/32的CEF條目通過test platform hardware cef v4-delete 命令刪除。現在，通過test platform hardware cef v4-insert [prefix] 命令將其新增回來 [掩碼長度] vpn [vpn編號]鄰接關係[鄰接索引]

若要驗證這一點，請使用命令test platform hardware cef v4-insert 1.1.1.1 32 vpn 0 adjacency 114689。該條目已重新新增到DFC CEF表中：

```
MXC.CALO.Sup2T-sw2-dfc3#test platform hardware cef v4-insert 1.1.1.1 32 vpn 0 adjacency 114689
test_ipv4_insert: done: sw_index = 42
```

```
MXC.CALO.Sup2T-sw2-dfc3#show plat hard cef vpn 0 1.1.1.1
Codes: decap - Decapsulation, + - Push Label
Index Prefix Adjacency
54 1.1.1.1/32 V12 ,0c11.678b.f6f7
```

Ping from the 3750X to Loopback 0 is successful and HW forwarded by 6500 DFC.

```
MXC.CALO.Sup2T-sw2-dfc3#show platform hard cef adj entry 114689
```

```
Index: 114689 -- Valid entry (valid = 1) --
```

```
RIT fields: The entry has a Layer2 Format
```

```
-----
|decr_ttl=YES | l2_fwd=NO | ccc = 4 | add_shim_hdr = NO
|_____||_____||_____||_____
-----
```

```
Statistics: Packets = 684
```

```
Bytes = 80712
```

```
// Logs in 3850
```

```
CALO.MXC.385024XU#show logging [snip...] *Jan 23 05:59:56.911: ICMP: echo reply sent, src
1.1.1.1, dst 10.1.1.1, topology BASE, dscp 0 topoid 0 *Jan 23 05:59:57.378: ICMP: echo reply
sent, src 1.1.1.1, dst 10.1.1.1, topology BASE, dscp 0 topoid 0 *Jan 23 05:59:57.390: ICMP: echo
reply sent, src 1.1.1.1, dst 10.1.1.1, topology BASE, dscp 0 topoid 0
```

新增和刪除VRF路由表的條目

在所有上述步驟中進行的配置中，show platform hardware cef命令中的vpn 0字串已被強制執行。即使看起來完全沒有必要，因為命令預設情況下會返回常規路由表或vpn 0的條目，但這是故意這樣做的，目的是請始終牢記，系統會通過您新增和刪除CEF條目1.1.1.1/32的文檔，在特定路由表例項(VRF)中新增或刪除條目。但是，某些字首很可能存在於不同的VRF中(即e.10.x.x.x)以及刪除、新增或修改錯誤的VRF的CEF條目會導致負面影響。

為VRF TEST_VRF刪除字首為1.1.1.1/32的CEF條目。有關新增CEF條目的詳細說明，請參閱本文檔的新增CEF條目部分。

為了新增VRF，請使用ip vrf forwarding [VRF-NAME]命令將6500交換機中的SVI更改為建議的VRF，最後在TEST_VRF表中新增相同的靜態路由：

