

Data Forwarding 數據轉發機制與高可用性架構手冊

專題探討：封包交換演進、CEF 架構深度驗證、SSO & NSF 控制平面備援

在進階思科網路架構中，**數據轉發 (Data Forwarding)** 是決定網路吞吐效能與高可用性的基礎底層機制。本學習手冊基於思科官方 CCNP Encor 核心課程內容，深度整合課堂筆記與拓撲實驗，全面剖析傳統交換至思科快速轉發 (CEF) 的演進、底層硬體運作架構，以及高端設備特有的高可用性技術 (SSO & NSF)。

一、數據轉發模式的演進與瓶頸

Cisco 路由器隨著處理器與硬體架構的革新，發展出三代完全不同的數據轉發與交換模式：

1. 行程交換 (Process Switching)

這是最古老也最傳統的數據轉發模式。在此模式下，路由器的資料平面與控制平面未分離。

- **運作機制**：路由器接收到一個數據包時，會將其存入通用記憶體。隨後，CPU 被中斷訊號觸發，由 `ip_input` 行程逐包查閱控制平面的路由表 (Routing Table) 來找出目的地、下一跳和出站介面，並重新計算 Layer 2 標頭封裝，最後送入輸出介面。
- **效能瓶頸**：整個過程發生在 **CPU 層級**，每個數據包都會耗費 CPU 資源，導致大幅度的轉發延遲，容易使 CPU 負載過高。
- **配置指令**：在介面上輸入 `no ip route-cache` 即可強制啟用行程交換。

2. 快速交換 (Fast Switching)

快速交換是在 Process Switching 的基礎上發展而來的優化模式，其核心概念是引入了 **快速快取 (Fast Cache)**。

- **運作機制**：當路由到某個特定目的地的**第一個數據包**時，由於快取中無資料，仍必須由 CPU 進行行程交換查表轉發。但轉發的同時，CPU 會將查表得到的結果（目的地、下一跳、Layer 2 封裝細節）記錄到快速快取中。當後續的數據包到達時，直接比對快速快取的內容即可轉發，免除 CPU 再次查表。
- **效能瓶頸**：雖然改善了後續封包的效率，但每條流的第一個數據包仍有延遲。若網路上充滿突發的短流（大量不同的目的 IP），快取命中率將會非常低。此外，一旦路由表發生收斂變更，整個快速快取必須全部宣告失效並重新觸發查表。
- **配置指令**：介面下設定 `ip route-cache` 啟用快速交換。

3. 思科快速轉發 (Cisco Express Forwarding, CEF)

CEF 是目前思科設備的預設轉發方式，它顛覆了快速交換的被動式快取建立，採取了**主動拓撲驅動 (Topology-Driven)** 的設計架構。

- **運作機制**：在數據包尚未到達前，CEF 就會主動監聽控制平面的路由表 (RIB) 與 ARP 表的動態，並將其重組並同步為資料平面的專用轉發表。
- **硬體加速**：CEF 的轉發主要依靠專門的硬體晶片——**特殊應用積體電路 (ASICs)** 以及**網路處理單元 (NPUs)**。這使得封包能夠以硬體線速 (Wire-Speed) 進行極速轉發，不再受通用處理器的效能限制。

💡 關鍵核心觀念

行程交換 vs 快速交換 vs CEF

行程交換是「逐包由 CPU 查表」；快速交換是「首包 CPU 查表，後續包靠快取」；而 CEF 則是「拓撲變更時主動預建硬體轉發表，封包到達時直接硬體線速轉發，全程不驚動 CPU」。

二、CEF 的核心架構與元件分析

開啟 CEF 數據轉發時，所有的轉發決策只會在**資料平面 (CEF 模塊)** 進行。CEF 將控制平面的所有有用資訊分離並重構為兩個關鍵表格：

【控制平面 (Control Plane)】 → 監聽同步 → 【資料平面 (Data Plane) - CEF 模塊】
Routing Table (路由表 / RIB) → Forwarding Information Base (FIB 表)
ARP Table (ARP 映射表) → Adjacency Table (鄰接表)

1. 轉發資訊庫 (Forwarding Information Base, FIB)

FIB 是路由表 (Routing Table) 在資料平面上的一模一樣的鏡像與延伸。它包含了路由表中的所有網絡前綴，但經過了查詢演算法（如 IPv4 mtrie 8-8-8 樹狀結構）的優化。當控制平面的路由收斂完成，FIB 便同步建立，提供極速的目的地 IP 前綴查找，且徹底免除了傳統路由表的遞迴查詢 (Recursive Lookup)。

2. 鄰接表 (Adjacency Table)

鄰接表是控制平面 ARP 表的映射表。當路由器確認出站路徑後，必須封裝 Layer 2 標頭。鄰接表預先將下一跳 IP 所對應的 MAC 位址與輸出介面封裝資訊準備就緒。當 FIB 指向某個鄰接項目時，資料平面可立即進行硬體層級的 **Layer 2 標頭重寫 (L2 Rewrite)**，整個過程完全不使用 CPU 資源。

⚠ 異常處理：CEF 懲罰機制 (CEF Punt)

整個數據轉發流程皆在 CEF 硬體模組內完成，完全不會用到 CPU。然而，若有新到達的數據包在 CEF 表 (FIB) 中找不到對應前綴或下一跳鄰接項目尚未建立（未決鄰接 / Glean Adjacency），資料平面便會啟動懲罰機制 (Punt)，將該數據包轉發到 CPU 進行傳統的處理或觸發 ARP 請求。過多的 Punt 封包會導致 CPU 使用率飆高。

三、CEF 部署模式：集中式 vs 分散式

根據路由器的硬體分級與架構，CEF 支援兩種部署模式：

- **集中式轉發 (Centralized Forwarding)**：主要用於中低階設備。只有核心的路由處理器 (Route Processor, RP) 擁有轉發引擎。當線路卡 (Line Card) 收到封包後，需透過內部巴士或纖網送至 RP 完成 CEF 查表轉發。這屬於軟體 CEF 實作。
- **分散式轉發 (Distributed Forwarding - dCEF)**：用於模組化的高階高端設備。每一塊線路卡 (Line Card) 都配備了獨立的硬體轉發引擎與專屬的 CEF 晶片與記憶體。中央 RP 僅負責維護 RIB 與 ARP 表並同步給各個線路卡，各線路卡獨立做出轉發決定，完全不需要中央 RP 的介入，實現極致的轉發效能。

四、介面數據轉發狀態與表項驗證

在實際網路運維與除錯中，檢查設備在用哪個介面進行數據轉發，不是看路由表，而是直接查看最快最準的出站表項——CEF Table。以下是基於實驗拓撲 R1 與 R2 的指令驗證：

第一步 確認介面的 CEF 模式啟用狀態

使用 `show ip interface [介面名稱]` 指令檢視當前介面的二層與三層交換狀態：

```
R1# show ip interface ethernet 0/0
Ethernet0/0 is up, line protocol is up
Internet address is 1.1.1.1/24
...
IP fast switching is enabled
IP CEF switching is enabled          <-- 代表此 Port 正在使用 CEF 轉發模式
IP CEF switching turbo vector
IP route-cache flags are Fast, CEF
```

第二步 全面列出所有使用 CEF 的介面清單

若要確認全機有哪些介面納入 CEF 轉發範疇，可使用 `show cef interface` 指令，系統會將每一個處於 CEF 轉發模式的介面詳細排列出來。

第三步 精準查閱資料平面的出站介面與下一跳 (FIB 表驗證)

在拓撲中，當 R2 的 Loopback 0 配置了 `8.8.8.8/32`，我們可以在 R1 上查閱 FIB 項目，確認真實的出站路徑：

```
R1# show ip cef
Prefix                Next Hop                Interface
0.0.0.0/0             no route
0.0.0.0/8             drop
1.1.1.0/24            attached               Ethernet0/0
1.1.1.1/32            receive               Ethernet0/0
1.1.1.2/32            attached               Ethernet0/0
8.8.8.8/32            1.1.1.2               Ethernet0/0  <-- 最精準的出站路由與介面指示
127.0.0.0/8           drop
224.0.0.0/4           drop
255.255.255.255/32    receive
```

✓ 實驗室維護結論

不論路由表 (RIB) 如何顯示，底層硬體轉發的唯一絕對依據就是 `show ip cef` 的輸出結果。它是整個思科資料平面最即時、最具決定性的出站方向指標。

五、 高端設備高可用性專題：SSO & NSF 的完美協同

在配備雙控制平面（雙 Route Processors，簡稱雙 RPs）的 Cisco 高端電信級設備中，高可用性（High Availability）是不可或缺的指標。當主要控制處理器失效時，設備具備一套流暢的切換與轉發維護機制。

1. 核心元件定義

- **SSO (Stateful Switchover)**：狀態切換功能。它是一種控制平面冗餘特性，允許配備雙 RPs (RP1 ACTIVE / RP2 STANDBY) 的思科路由器，在 ACTIVE RP 正常運作時，將其路由器配置、控制平面的狀態資訊（如登入 Session、介面狀態等）即時同步到 STANDBY RP 上。
- **NSF (Nonstop Forwarding)**：不中斷轉發功能。這是一種資料平面維護特性，專門配合 SSO 解決控制平面切換期間的掉包問題。

2. SSO 獨立切換時的致命瓶頸

當 ACTIVE RP1 突然發生硬體故障時，SSO 機制會立即觸發切換，將備用的 STANDBY RP2 提升為 ACTIVE。此時：

```
ACTIVE RP1 故障 → SSO 觸發切換 → STANDBY RP2 接管控制權
→ 與鄰居的路由協議中斷 (Routing Protocol Adjacency Flap)
→ 舊的路由表 (RIB) 被清空 → 資料平面的 CEF 表項被隨之清除 (Purged)
→ ❌ 結果：在網路拓撲重新學習收斂完成前，數據包失去轉發依據，造成嚴重掉包！
```

3. SSO 協同 NSF 的完美解決方案

為了避免 SSO 切換時路由表清空導致的數據中斷，必須同時啟用 **NSF (Nonstop Forwarding)**：

- **運作機制**：當啟用 NSF 後，當控制平面發生 RP 切換且外面的路由協定短暫斷開時，NSF 會**通知並指示**資料平面的線路卡 (**Line Cards**) 暫時保留當前的 CEF 硬體表項一段時間（維持短暫的轉發狀態）。
- **完美結果**：在控制平面的控制權切換且路由協議重新收斂、relearn 的這段短暫時間內，資料平面的線路卡依然能夠**利用先前保留的 CEF 表項繼續正常轉發數據包**，直到控制平面完全恢復並送來最新的 FIB 更新。這達成了真正意義上的**控制平面斷線、資料平面永不中斷轉發**的電信級高可用性指標。