

Cisco serie 12000 Internet Router Architecture: Packet Switching

Sommario

[Introduzione](#)

[Prerequisiti](#)

[Requisiti](#)

[Componenti usati](#)

[Convenzioni](#)

[Premesse](#)

[Packet Switching: Panoramica](#)

[Packet Switching: Schede di linea Engine 0 e Engine 1](#)

[Packet Switching: Schede di linea Engine 2](#)

[Packet Switching: Switching delle celle in un fabric](#)

[Packet Switching: Trasmissione di pacchetti](#)

[Riepilogo flusso di pacchetti](#)

[Informazioni correlate](#)

[Introduzione](#)

In questo documento vengono esaminati i principali elementi dell'architettura di Cisco serie 12000 Internet Router — pacchetti di switching. I pacchetti di switching sono radicalmente diversi da qualsiasi memoria condivisa o architettura Cisco basata su bus. Utilizzando un fabric crossbar, Cisco 12000 fornisce grandi quantità di larghezza di banda e scalabilità. Inoltre, lo switch 12000 utilizza code di output virtuali per eliminare il blocco Head of Line all'interno del fabric dello switch.

[Prerequisiti](#)

[Requisiti](#)

Nessun requisito specifico previsto per questo documento.

[Componenti usati](#)

Le informazioni fornite in questo documento si basano sui seguenti componenti hardware:

- Cisco serie 12000 Internet Router

Le informazioni discusse in questo documento fanno riferimento a dispositivi usati in uno specifico ambiente di emulazione. Su tutti i dispositivi menzionati nel documento la configurazione è stata ripristinata ai valori predefiniti. Se la rete è operativa, valutare attentamente eventuali conseguenze derivanti dall'uso dei comandi.

Convenzioni

Per ulteriori informazioni sulle convenzioni usate, consultare il documento [Cisco sulle convenzioni nei suggerimenti tecnici](#).

Premesse

(la decisione di commutazione su un Cisco 12000 viene presa dalle schede di linea (LC). Per alcuni LC, un ASIC (Application-Specific Integrated Circuit) dedicato commuta effettivamente i pacchetti. Distributed Cisco Express Forwarding (dCEF) è l'unico metodo di commutazione disponibile.

Nota: I motori 0, 1 e 2 non sono gli ultimi motori sviluppati da Cisco. Ci sono anche schede di linea Engine 3, 4 e 4+, con altri da seguire. Le schede di linea del motore 3 sono in grado di eseguire le feature di spigolo a velocità di linea. Più alto è il motore di layer 3, maggiore è il numero di pacchetti commutati nell'hardware. È possibile trovare alcune informazioni utili sulle diverse schede di linea disponibili per i Cisco serie 12000 Router e il motore su cui sono basati su [Cisco serie 12000 Internet Router: Domande frequenti](#).

Packet Switching: Panoramica

I pacchetti vengono sempre inoltrati dalla scheda di linea in entrata (LC). La LC in uscita esegue solo QoS (Quality of Service) in uscita dipendente dalla coda, ad esempio WRED (Weighted Random Early Detection) o CAR (Committed Access Rate). La maggior parte dei pacchetti viene commutata dal VLAN usando il protocollo distribuito Cisco Express Forwarding (dCEF). Solo i pacchetti di controllo (ad esempio gli aggiornamenti del routing) vengono inviati al Gigabit Route Processor (GRP) per l'elaborazione. Il percorso di commutazione del pacchetto dipende dal tipo di motore di commutazione usato sul LC.

Questo è quello che succede quando arriva un pacchetto:

1. Un pacchetto viene inserito nel PLIM (physical layer interface module). Qui possono accadere diverse cose: Un ricetrasmittitore trasforma i segnali ottici in segnali elettrici (la maggior parte delle schede di linea CSR dispone di connettori in fibra) Il frame L2 viene rimosso (SANE, ATM (Asynchronous Transfer Mode), Ethernet, HDLC (High-Level Data Link Control)/protocollo Point-to-Point - PPP) Le celle ATM vengono ricomposte i pacchetti che non superano il controllo di ridondanza ciclico (CRC) vengono ignorati
2. Quando il pacchetto viene ricevuto ed elaborato, si accede direttamente alla memoria in una piccola memoria (circa 2 x buffer MTU (Maximum Transmission Unit)) chiamata "First In, First Out (FIFO) burst memory". La quantità di questa memoria dipende dal tipo di LC (da 128 KB a 1 MB).
3. Una volta che il pacchetto è completamente in memoria FIFO, un circuito integrato specifico dell'applicazione (ASIC, Application-Specific Integrated Circuit) sul PLIM contatta il buffer Management ASIC (BMA, Buffer Management ASIC) e richiede un buffer per inserire il pacchetto. Al BMA viene comunicato quali sono le dimensioni del pacchetto e alloca un buffer di conseguenza. Se il BMA non riesce a ottenere un buffer della giusta dimensione, il pacchetto viene scartato e il contatore "ignore" viene incrementato sull'interfaccia in arrivo. Non esiste alcun meccanismo di fallback come con altre piattaforme. Mentre questo accade, il PLIM potrebbe ricevere un altro pacchetto nella memoria burst FIFO, motivo per cui ha

dimensioni pari a 2 xMTU.

4. Se nella coda corretta è disponibile un buffer libero, il pacchetto viene archiviato dalla BMA nell'elenco delle code libere delle dimensioni appropriate. Questo buffer viene inserito nella coda raw, che viene esaminata dall'ASIC Salsa o dalla CPU R5K. La CPU R5K determina la destinazione del pacchetto consultando la tabella dCEF locale nella memoria RAM dinamica (DRAM), quindi sposta il buffer dalla coda raw a una coda ToFabric corrispondente allo slot di destinazione. Se la destinazione non è nella tabella CEF, il pacchetto viene scartato. Se il pacchetto è un pacchetto di controllo (ad esempio, aggiornamenti di routing), viene accodato alla coda del GRP e verrà elaborato dal GRP. Sono disponibili 17 code ToFab (16 unicast e 1 multicast). È disponibile una coda toFab per scheda di linea (inclusa la RP). Queste code sono note come "code di output virtuali" e sono importanti per evitare il blocco dell'head-of-line.
5. La ToFab BMA suddivide il pacchetto in pezzi da 44 byte, che sono il payload di quelle che saranno note come "Cisco Cells". A queste celle viene assegnata un'intestazione a 8 byte e un'intestazione buffer a 4 byte dal BMA frFab (dimensioni totali dei dati finora = 56 byte), quindi vengono accodate nella coda ToFab appropriata (a quel punto, il contatore #Qelem nel pool da cui proviene il buffer si abbassa di un byte e il contatore della coda ToFab si alza di un byte). Il "decisore" dipende dal tipo di motori di commutazione: Sulle schede Engine 2+, viene utilizzato un ASIC speciale per migliorare il modo in cui vengono scambiati i pacchetti. I pacchetti normali (IP/Tag, nessuna opzione, checksum) vengono elaborati direttamente dal PSA (Packet Switching ASIC), quindi ignorano la combinazione di coda raw/CPU/Salsa e vengono accodati direttamente nella coda toFab. Solo i primi 64 byte del pacchetto vengono passati attraverso l'ASIC di switching del pacchetto. Se il pacchetto non può essere scambiato dal PSA, viene accodato all'RawQ per essere gestito dalla CPU del LC, come spiegato in precedenza. A questo punto, la decisione di commutazione è stata presa e il pacchetto è stato accodato alla coda di output ToFab corretta.
6. I DMA toFab BMA (Direct Memory Access) trasformano le celle del pacchetto in piccoli buffer FIFO nell'ASIC (Fabric Interface). Esistono 17 buffer FIFO (uno per coda ToFab). Quando la FIA ottiene una cella dalla BMA toFab, aggiunge un CRC a 8 byte (dimensioni totali della cella - 64 byte; payload di 44 byte, 8 byte di intestazione di cella, 4 byte di intestazione di buffer). La FIA dispone di ASIC (Serial Line Interface) che eseguono quindi la codifica 8B/10B sulla cella (come FDDI (Fiber Distributed Data Interface) 4B/5B) e si prepara a trasmetterla sulla struttura. Può sembrare un sovraccarico notevole (44 byte di dati vengono convertiti in 80 byte in tutta la struttura!), ma non si tratta di un problema in quanto la capacità della struttura è stata predisposta di conseguenza.
7. Ora che una FIA è pronta a trasmettere, la FIA richiede l'accesso al fabric dall'attuale sistema di pianificazione e orologio (CSC). La CSC lavora su un algoritmo di equità piuttosto complesso. L'idea è che nessun LC può monopolizzare la larghezza di banda in uscita di qualsiasi altra carta. Si noti che anche se un LC desidera trasmettere i dati da una delle proprie porte, deve comunque passare attraverso il fabric. Questa operazione è importante perché, in caso contrario, una porta su un LC potrebbe monopolizzare tutta la larghezza di banda per una determinata porta su quello stesso LC. Renderebbe inoltre più complicato il progetto di commutazione. La FIA invia le celle attraverso il fabric dello switch al LC in uscita (specificato dai dati nell'intestazione Cisco Cell inserita dal motore di switching). L'algoritmo di equità è anche progettato per una corrispondenza ottimale; se la carta 1 vuole trasmettere alla carta 2 e la carta 3 vuole trasmettere contemporaneamente alla carta 4, ciò accade in parallelo. Questa è la grande differenza tra un fabric switch e un'architettura bus. Può essere paragonata a uno switch Ethernet o a un hub; su uno switch, se la porta A desidera inviare

dati alla porta B e la porta C desidera comunicare con la porta D, questi due flussi si verificano indipendentemente l'uno dall'altro. Su un hub, sono presenti problemi half-duplex quali collisioni e algoritmi di backoff e retry.

8. Le celle Cisco che escono dal fabric passano attraverso l'elaborazione SLI per rimuovere la codifica 8B/10B. In caso di errori, questi verrebbero visualizzati nell'output del comando `show controller fi` come "parità di cella". Per ulteriori informazioni, vedere [Come leggere l'output del comando show controller tramite fax](#).
9. Queste Celle Cisco sono DMA in FIFO su frFab FIAs, e poi in un buffer su frFab BMA. Il BMA frFab è quello che in realtà riassume le celle in un pacchetto. Come fa il frFab BMA a sapere quale buffer mettere le celle prima di ricomporle? Questa è un'altra decisione presa dal motore di commutazione della scheda di linea in entrata; poiché tutte le code sull'intera scatola sono delle stesse dimensioni e nello stesso ordine, il motore di commutazione ha semplicemente il Tx LC messo il pacchetto nella stessa coda numerica da cui è entrato nel router. Le code SDRAM frFab BMA possono essere visualizzate con il comando `show controller frfab queue` sul cavo LC. Vedere [Come leggere l'output del comando show controller frfab | tofab queue comandi su un router Cisco serie 12000 Internet](#) per ulteriori informazioni. Questa è sostanzialmente la stessa idea dell'output BMA toFab. I pacchetti arrivano e vengono inseriti in pacchetti che vengono rimossi dalle rispettive code libere. Questi pacchetti vengono inseriti nella coda da-fabric, accodati nella coda di interfaccia (esiste una coda per porta fisica) o nel file rawQ per l'elaborazione dell'output. Non succede molto nel rawQ: Replica multicast per porta, MDRR (Modified Deficit Round Robin): la stessa idea di DWFQ (Distributed Weighted Fair Queuing) e CAR di output. Se la coda di trasmissione è piena, il pacchetto viene scartato e il contatore di rilascio dell'output viene incrementato.
10. Il BMA frFab attende che la parte TX del PLIM sia pronta a inviare un pacchetto. Il BMA frFab esegue la riscrittura MAC effettiva (basandosi, ricordate, sulle informazioni contenute nell'intestazione della cella Cisco) e le DMA trasmettono il pacchetto a un piccolo buffer (di nuovo, 2xMTU) nel circuito PLIM. Il PLIM esegue l'incapsulamento ATM SAR e SONET, se necessario, e trasmette il pacchetto.
11. Il traffico ATM viene ricomposto (dalla SAR), segmentato (dalla BMA tofab), riassembleato (dalla BMA fromfab) e segmentato di nuovo (dalla SAR fromfab). Questo succede molto rapidamente.

Si tratta del ciclo di vita di un pacchetto, dall'inizio alla fine. Se volete sapere come si sente un GSR alla fine della giornata, leggete questo articolo 500.000 volte!

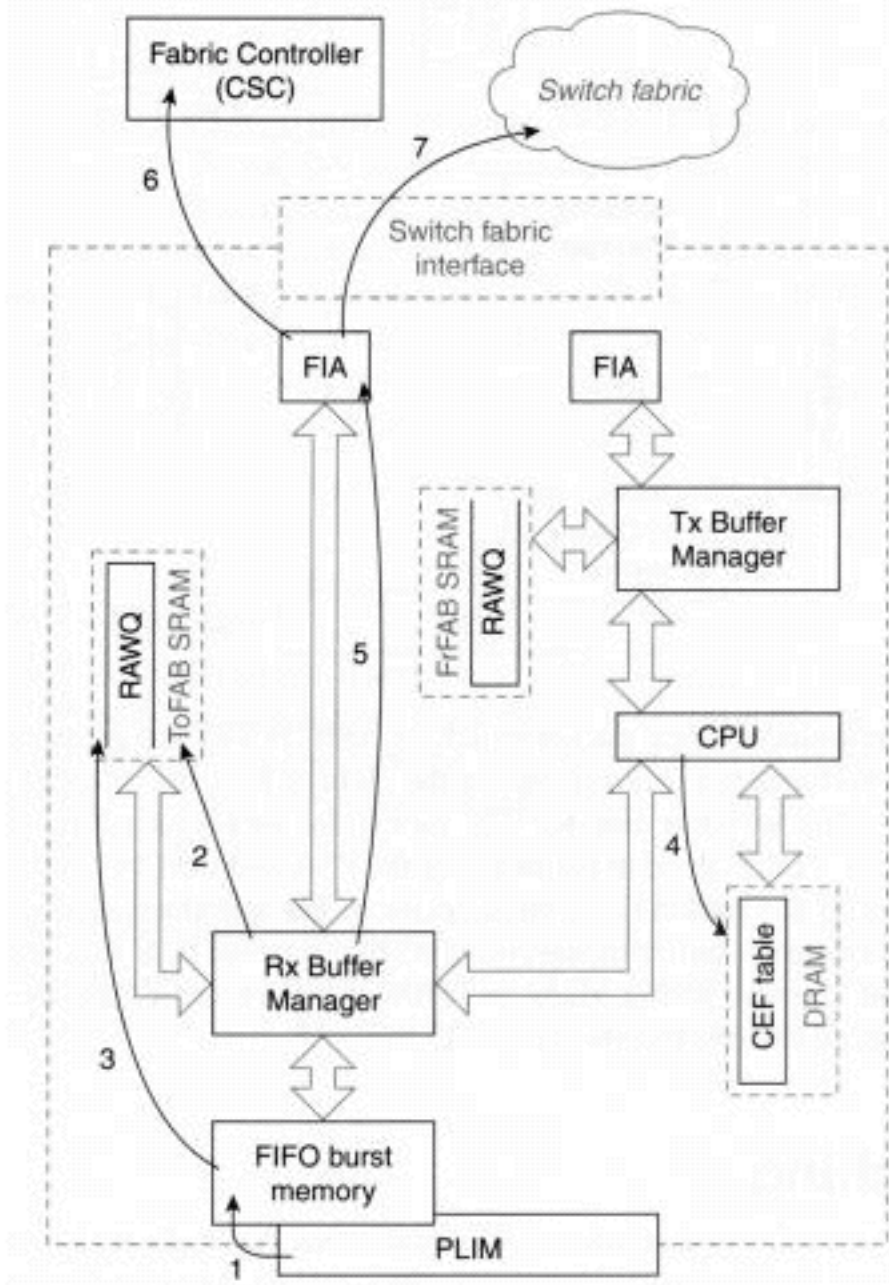
Il percorso di commutazione del pacchetto sul GSR dipende dal tipo di motore di inoltro sul LC. Ora esamineremo tutti i passaggi per il motore 0, il motore 1 e i due LC.

[Packet Switching: Schede di linea Engine 0 e Engine 1](#)

Le sezioni seguenti si basano sul libro *Inside Cisco IOS Software Architecture*, Cisco Press.

[Nella figura 1](#) seguente vengono illustrati i diversi passaggi durante la commutazione di un pacchetto per un LC del motore 0 o 1.

Figura 1: Percorso di switching del motore 0 e 1



Il percorso di commutazione per il motore 0 e il motore 1 LC è essenzialmente lo stesso, sebbene il motore di commutazione e il buffer manager del motore 1 LC siano stati migliorati per migliorare le prestazioni. Il percorso di commutazione è il seguente:

- **1.** Il processore di interfaccia (PLIM) rileva un pacchetto sul supporto di rete e inizia a copiarlo in una memoria FIFO chiamata **memoria burst** sul cavo LC. La quantità di memoria burst di cui dispone ciascuna interfaccia dipende dal tipo di LC; i tipici LC hanno da 128 KB a 1 MB di memoria burst.
- **Fase 2** - Il processore di interfaccia richiede un buffer di pacchetto dal BMA ricevente; il pool da cui viene richiesto il buffer dipende dalla lunghezza del pacchetto. Se non sono presenti buffer liberi, l'interfaccia viene scartata e il contatore "ignore" dell'interfaccia viene incrementato. Ad esempio, se un pacchetto da 64 byte arriva a un'interfaccia, BMA cerca di allocare un buffer di pacchetto da 80 byte. Se nel pool da 80 byte non esistono buffer liberi, i buffer non vengono allocati dal pool successivo disponibile.
- **3.** Quando un buffer libero viene allocato dal BMA, il pacchetto viene copiato nel buffer e accodato nella **coda raw** (RawQ) per essere elaborato dalla CPU. Un interrupt viene inviato alla CPU LC.

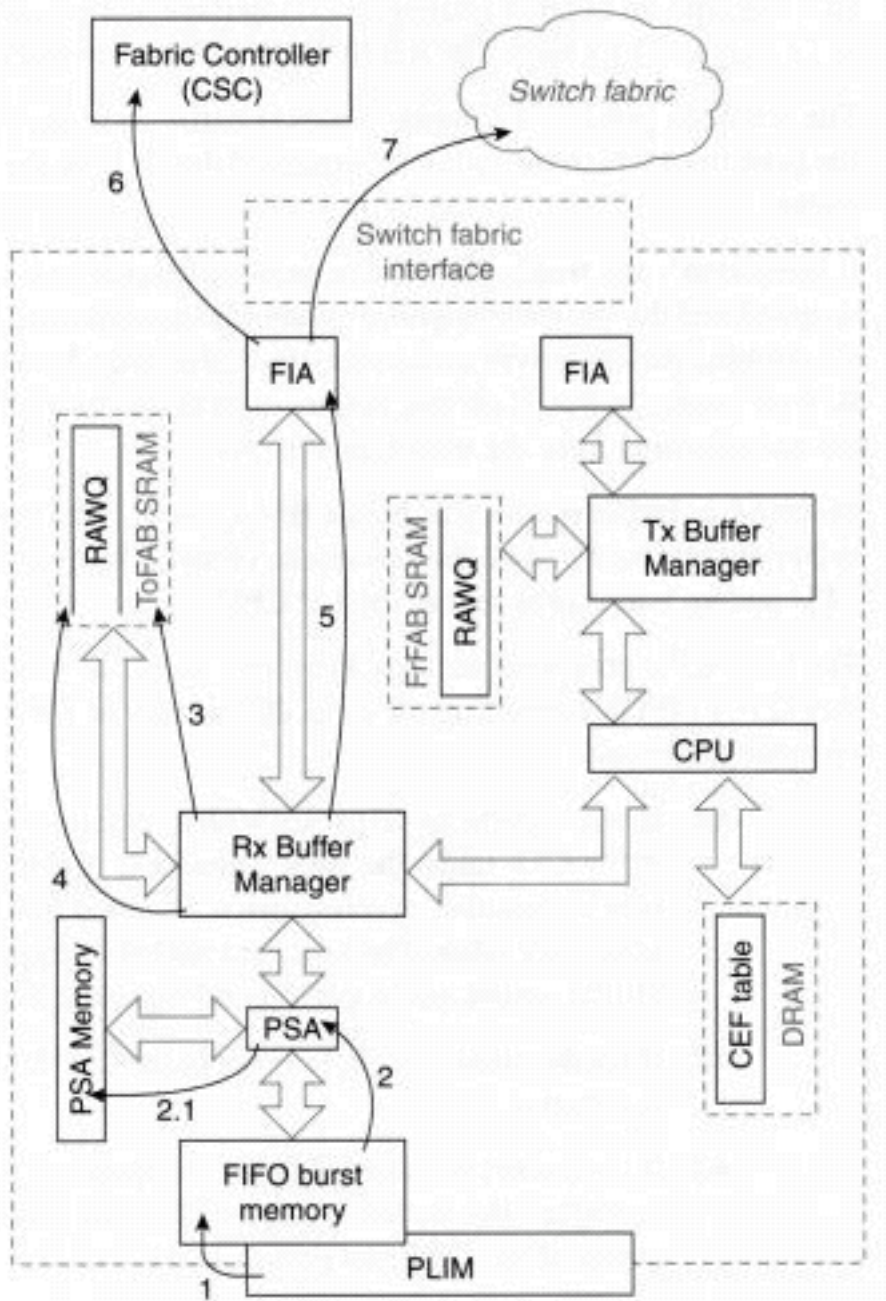
- **Fase 4** - La CPU del controller di dominio elabora ogni pacchetto nel RawQ non appena viene ricevuto (il RawQ è un FIFO), consultando la tabella Cisco Express Forwarding distribuita in DRAM per prendere una decisione di switching.**4.1** Se si tratta di un pacchetto IP unicast con un indirizzo di destinazione valido nella tabella CEF, l'intestazione del pacchetto viene riscritta con le nuove informazioni di incapsulamento ottenute dalla tabella adiacente al CEF. Il pacchetto commutato viene accodato nella coda di output virtuale corrispondente allo slot di destinazione.**4.2** Se l'indirizzo di destinazione non è nella tabella CEF, il pacchetto viene scartato.**4.3** Se il pacchetto è un pacchetto di controllo (ad esempio, un aggiornamento di routing), il pacchetto viene accodato nella coda di output virtuale del GRP ed elaborato dal GRP.
- **Fase 5** - Il ricevente BMA frammenta il pacchetto in celle da 64 byte e le consegna alla FIA per la trasmissione al LC in uscita.

Al termine della Fase 5, il pacchetto che è arrivato in una scheda LLC Engine 0/1 è stato commutato ed è pronto per essere trasportato attraverso il fabric dello switch come celle. Andare al punto 6 nella sezione [Packet Switching: Switching delle celle in Fabric](#).

[Packet Switching: Schede di linea Engine 2](#)

[La Figura 2](#) seguente mostra il percorso di switching del pacchetto quando i pacchetti arrivano in un LC del Engine 2, come descritto nel seguente elenco di passaggi.

Figura 2: Percorso di switching del motore 2



- 1. Il processore di interfaccia (PLIM) rileva un pacchetto sul supporto di rete e inizia a copiarlo in una memoria FIFO chiamata **memoria burst** sul cavo LC. La quantità di memoria burst di cui dispone ciascuna interfaccia dipende dal tipo di LC; i tipici LC hanno da 128 KB a 1 MB di memoria burst.
- 2. I primi 64 byte del pacchetto, chiamati intestazione, vengono passati attraverso il Packet Switching ASIC (PSA). 2.1 Il PSA commuta il pacchetto consultando la tabella CEF locale nella memoria PSA. Se il pacchetto non può essere commutato dal PSA, andare al punto 4; in caso contrario, andare al passo 3.
- **Passaggio 3** - Receive Buffer Manager (RBM) accetta l'intestazione dal PSA e la copia in un'intestazione di buffer libero. Se il pacchetto è più grande di 64 byte, anche la parte finale del pacchetto viene copiata nello stesso buffer libero nella memoria del pacchetto e viene accodata nella [coda di output virtuale](#) del LC in uscita. Andare al Passaggio 5.
- 4. Il pacchetto arriva in questa fase se il PSA non può selezionarlo. Questi pacchetti vengono inseriti nella **coda raw** (RawQ) e il percorso di commutazione è essenzialmente lo stesso del motore 1 e del motore 0 LC da questo punto (passaggio 4 nel caso del motore 0). Notare che i pacchetti commutati dal PSA non vengono mai inseriti nel RawQ e non viene inviato alcun

interrupt alla CPU.

- **Fase 5** - Il Fabric Interface Module (FIM) è responsabile della segmentazione dei pacchetti in [Cisco Cells](#) e dell'invio delle celle al Fabric Interface ASIC (FIA) per la trasmissione al LC in uscita.

Packet Switching: Switching delle celle in un fabric

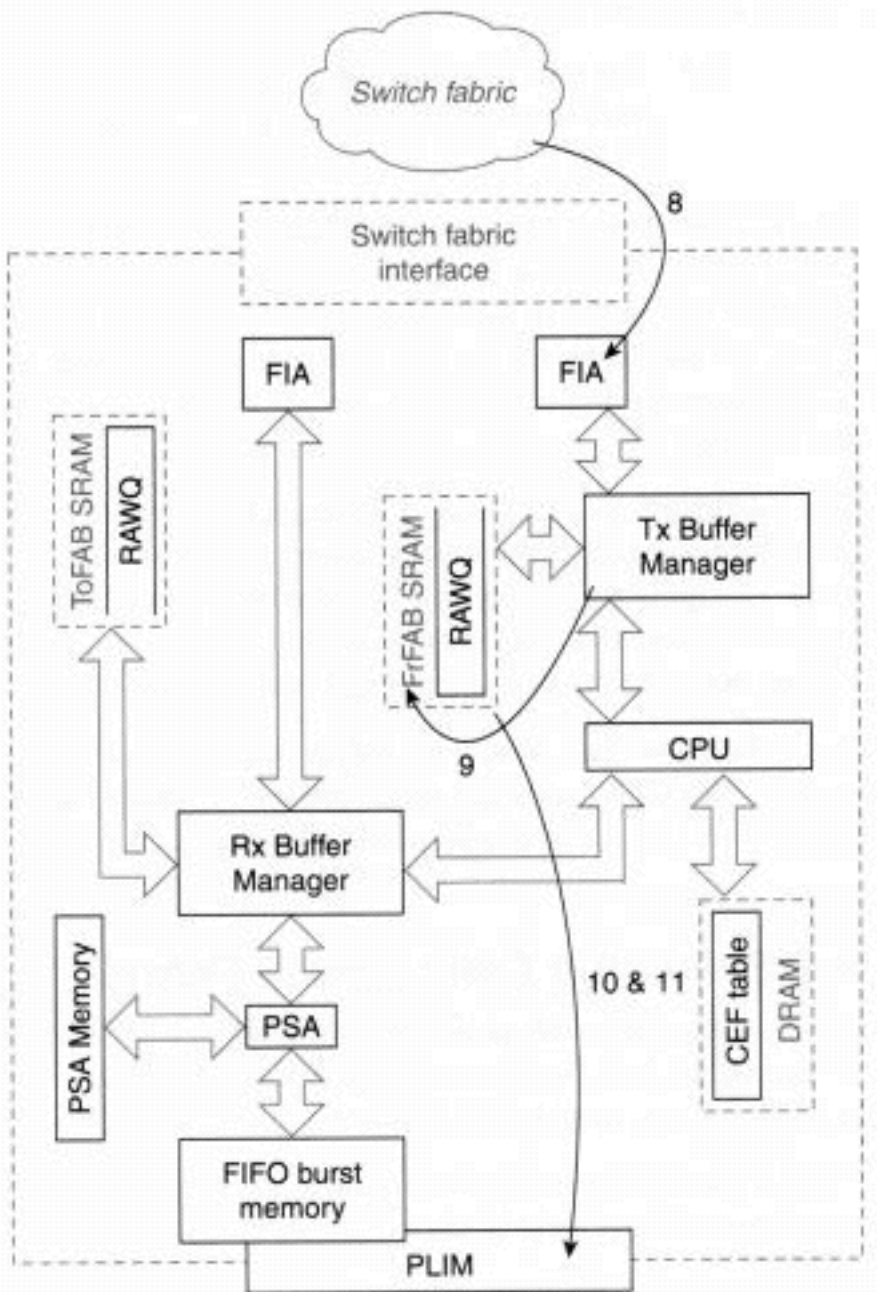
Si arriva in questa fase dopo che il motore di commutazione dei pacchetti ha cambiato i pacchetti. In questa fase, i pacchetti vengono segmentati in Cisco Cell e sono in attesa di essere trasmessi sul fabric di switching. I passaggi per questa fase sono i seguenti:

- **Fase 6** - La FIA invia una richiesta di sovvenzione al CSC, che pianifica il trasferimento di ciascuna cella nel fabric dello switch.
- **Passaggio 7** - Quando lo scheduler concede l'accesso al fabric dello switch, le celle vengono trasferite nello slot di destinazione. Si noti che le celle potrebbero non essere trasmesse tutte contemporaneamente; altre celle all'interno di altri pacchetti potrebbero essere interlacciate.

Packet Switching: Trasmissione di pacchetti

La figura 3 mostra l'ultima fase della commutazione di pacchetto. Le celle vengono ricomposte e il pacchetto viene trasmesso al supporto. Ciò avviene sulla scheda di linea in uscita.

Figura 3: Cisco 12000 Packet Switching: Fase di trasmissione



- **Fase 8** - Le celle scambiate attraverso il tessuto arrivano nella scheda di linea di destinazione attraverso la FIA.
- **Passaggio 9** - Il gestore del buffer di trasmissione alloca un buffer dalla memoria del pacchetto di trasmissione e ricompono il pacchetto in questo buffer.
- **Passaggio 10** - Quando il pacchetto viene ricostruito, il BMA di trasmissione accoda il pacchetto nella coda di trasmissione dell'interfaccia di destinazione sul LC. Se la coda di trasmissione dell'interfaccia è piena (il pacchetto non può essere accodato), il pacchetto viene scartato e il contatore di **rilascio della coda di output** viene incrementato. **Nota:** nella direzione di trasmissione, l'unico momento in cui i pacchetti vengono inseriti in RawQ è quando la CPU LC deve eseguire un'elaborazione prima della trasmissione. Ad esempio, la frammentazione IP, il multicast e l'output CAR.
- **Passaggio 11** - Il processore di interfaccia rileva un pacchetto in attesa di essere trasmesso, rimuove il buffer dalla memoria di trasmissione, lo copia nella memoria FIFO interna e trasmette il pacchetto sul supporto.

[Riepilogo flusso di pacchetti](#)

I pacchetti IP che attraversano il router 12000 vengono elaborati in tre fasi:

- Scheda di linea in ingresso in tre sezioni: PLIM (Physical Line Interface Module) in entrata: conversione da ottica a elettrica, elaborazione non-framing, HDLC e PPP di Synchronous Optical Network (SONET)/Synchronous Digital Hierarchy (SDH). Inoltro IP: decisione di inoltro basata sulla ricerca FIB e l'accodamento in una delle code unicast in entrata o multicast. Gestione delle code in ingresso e interfaccia fabric: elaborazione RED (Random Early Detection)/WRED (Weighted Random Early Detection) sulle code in entrata e rimozione dalla coda verso il fabric per ottimizzare l'utilizzo del fabric.
- Commutazione dei pacchetti IP attraverso la struttura 12000 dalla scheda in entrata alla scheda in uscita o in uscita (in caso di multicast).
- Scheda di linea in uscita in tre sezioni: Egress Fabric Interface - Riassemblaggio dei pacchetti IP da inviare e inserimento in coda nelle code di uscita; elaborazione di pacchetti multicast. Gestione delle code in uscita: elaborazione RED/WRED sulle code in entrata e rimozione dalla coda verso il PLIM in uscita per ottimizzare l'utilizzo della linea in uscita. Egress PLIM - elaborazione HDLC e PPP, frame SONET/SDH, conversione da elettricità a ottiche.

[Informazioni correlate](#)

- [Supporto tecnico – Cisco Systems](#)