

Architektur der Cisco Internet Router der Serie 1200: Paket-Switching

Inhalt

[Einführung](#)

[Voraussetzungen](#)

[Anforderungen](#)

[Verwendete Komponenten](#)

[Konventionen](#)

[Hintergrundinformationen](#)

[Paket-Switching: Übersicht](#)

[Paket-Switching: Line Cards für Engine 0 und Engine 1](#)

[Paket-Switching: Line Cards der Engine 2](#)

[Paket-Switching: Switching-Zellen über die Fabric hinweg](#)

[Paket-Switching: Pakete übertragen](#)

[Paketfluss-Zusammenfassung](#)

[Zugehörige Informationen](#)

Einführung

In diesem Dokument werden die wichtigsten Architekturelemente der Cisco Internet Router der Serie 12000 - Switching-Pakete - untersucht. Switching-Pakete unterscheiden sich grundlegend von allen gemeinsam genutzten Speicher- oder Bus-basierten Cisco Architekturen. Durch die Verwendung einer Kreuzschienen-Fabric bietet der Cisco 1200 eine sehr hohe Bandbreite und Skalierbarkeit. Darüber hinaus verwendet der 1200 virtuelle Ausgabewarteschlangen, um die Sperre des Line-Leiters innerhalb der Switch-Fabric zu verhindern.

Voraussetzungen

Anforderungen

Es gibt keine spezifischen Anforderungen für dieses Dokument.

Verwendete Komponenten

Die Informationen in diesem Dokument basieren auf der folgenden Hardware:

- Cisco Internet Router der Serie 1200

Die Informationen in diesem Dokument beziehen sich auf Geräte in einer speziell eingerichteten Testumgebung. Alle Geräte, die in diesem Dokument benutzt wurden, begannen mit einer gelöschten (Nichterfüllungs) Konfiguration. Wenn Ihr Netz Live ist, überprüfen Sie, ob Sie die

mögliche Auswirkung jedes möglichen Befehls verstehen.

Konventionen

Weitere Informationen zu Dokumentkonventionen finden Sie unter [Cisco Technical Tips Conventions](#) (Technische Tipps von Cisco zu Konventionen).

Hintergrundinformationen

(Die Switching-Entscheidung für einen Cisco 12000 wird von den Line Cards (LCs) getroffen. Bei einigen LCs schaltet ein dedizierter Application-Specific Integrated Circuit (ASIC) die Pakete tatsächlich um. Distributed Cisco Express Forwarding (dCEF) ist die einzige verfügbare Switching-Methode.

Anmerkung: Die Engines 0, 1 und 2 sind nicht die neuesten von Cisco entwickelten Engines. Es gibt auch Line Cards der Engine 3, 4 und 4+, weitere folgen. Line Cards der Engine 3 können Edge-Funktionen mit Leitungsgeschwindigkeit ausführen. Je höher die Layer-3-Engine, desto mehr Pakete werden in der Hardware geswitcht. Hier finden Sie einige nützliche Informationen zu den verschiedenen Linecards, die für den Cisco Router der Serie 12000 verfügbar sind, sowie zu dem Modul, auf dem diese Linecards basieren, auf dem [Cisco Internet Router der Serie 12000: Häufig gestellte Fragen](#).

Paket-Switching: Übersicht

Pakete werden immer von der Eingangs-Linecard (LC) weitergeleitet. Der ausgehende LC führt nur ausgehende Quality of Service (QoS) durch, die von der Warteschlange abhängig ist (z. B. Weighted Random Early Detection (WRED) oder Committed Access Rate (CAR)). Die meisten Pakete werden über verteilte Cisco Express Forwarding (dCEF) vom LC umgeleitet. Nur die Steuerungspakete (z. B. Routing-Updates) werden zur Verarbeitung an den Gigabit Route Processor (GRP) gesendet. Der Paketvermittlungspfad hängt vom Typ der auf dem LC verwendeten Switching Engines ab.

Dies geschieht, wenn ein Paket eingeht:

1. Ein Paket wird in das physische Layer Interface Module (PLIM) eingespeist. Verschiedene Dinge passieren hier: Ein Transceiver wandelt optische Signale in elektrische Signale um (die meisten CSR Line Cards haben Glasfaserstecker). L2-Framing wird entfernt (SANE, Asynchronous Transfer Mode (ATM), Ethernet, High-Level Data Link Control (HDLC)/Point-to-Point Protocol - PPP) ATM-Zellen werden reassembliert Pakete, die die CRC-Prüfung nicht bestehen, werden verworfen
2. Während das Paket empfangen und verarbeitet wird, wird es als Direct Memory Access in einem kleinen (etwa zweifachen Maximum Transmission Unit (MTU)-Puffer) Speicher bezeichnet, der als "First In, First Out (FIFO) Burst Memory" bezeichnet wird. Die Größe dieses Speichers hängt vom Typ des LC ab (von 128 KB bis 1 MB).
3. Wenn sich das Paket vollständig im FIFO-Speicher befindet, kontaktiert ein ASIC (Application-Specific Integrated Circuit) auf dem PLIM den Buffer Management ASIC (BMA) und fordert einen Puffer zum Einlegen des Pakets an. Der BMA wird mitgeteilt, wie groß das Paket ist, und es wird ein Puffer entsprechend zugewiesen. Wenn die BMA keinen Puffer der richtigen Größe erhält, wird das Paket verworfen und der Zähler "Ignore" wird auf der

eingehenden Schnittstelle erhöht. Es gibt keinen Fallback-Mechanismus wie bei anderen Plattformen. In diesem Fall erhält das PLIM möglicherweise ein weiteres Paket im FIFO-Burst-Speicher, weshalb es eine Größe von 2 x MTU aufweist.

4. Wenn in der rechten Warteschlange ein freier Puffer verfügbar ist, wird das Paket von der BMA in der freien Warteschlangenliste der entsprechenden Größe gespeichert. Dieser Puffer wird auf die Raw Queue gelegt, die von der Salsa ASIC oder der R5K CPU untersucht wird. Die R5K-CPU bestimmt das Ziel des Pakets, indem sie die lokale dCEF-Tabelle im Dynamic RAM (DRAM) abrufen und dann den Puffer aus der Raw Queue in eine ToFabric-Warteschlange verschiebt, die dem Zielsteckplatz entspricht. Wenn sich das Ziel nicht in der CEF-Tabelle befindet, wird das Paket verworfen. Wenn es sich bei dem Paket um ein Steuerungspaket handelt (z. B. Routing-Updates), wird es in die Warteschlange des GRP gestellt und vom GRP verarbeitet. Es gibt 17 ToFab-Warteschlangen (16 Unicast plus 1 Multicast). Pro Linecard gibt es eine ToFab-Warteschlange (einschließlich RP). Diese Warteschlangen werden als "virtuelle Ausgabewarteschlangen" bezeichnet und sind wichtig, damit eine Head-of-Line-Blockierung nicht erfolgt.
5. Die ToFab-BMA zerlegt das Paket in 44-Byte-Stücke, die Nutzlast für das, was später als "Cisco Cells" bezeichnet wird. Diese Zellen erhalten vom frFab-BMA einen 8-Byte-Header und einen 4-Byte-Pufferheader (Gesamtdatengröße bis jetzt = 56 Byte) und werden dann in die richtige ToFab-Warteschlange aufgenommen (an diesem Punkt wird der #Qelem-Zähler im Pool, von dem der Puffer stammt, um eins heruntergegangen und der ToFab-Zähler um eins erhöht). Der "Entscheidungsträger" hängt von der Art der Switching-Motoren ab: Auf Engine 2+-Karten wird ein spezieller ASIC verwendet, um das Switching von Paketen zu verbessern. Normale Pakete (IP/Tag, keine Optionen, Prüfsumme) werden direkt über den Packet Switching ASIC (PSA) verarbeitet, dann umgehen Sie die Rohwarteschlange/CPU/Salsa-Kombination und werden direkt in die Fab-Warteschlange aufgenommen. Nur die ersten 64 Byte des Pakets werden über den Packet Switching ASIC weitergeleitet. Wenn das Paket nicht vom PSA gewechselt werden kann, wird das Paket in den RawQ eingestellt, damit es wie zuvor erläutert von der CPU des LC verarbeitet wird. Zu diesem Zeitpunkt wurde die Switching-Entscheidung getroffen, und das Paket wurde in die richtige ToFab-Ausgabewarteschlange gestellt.
6. Die toFab-BMA-DMA (Direct Memory Access), die Zellen des Pakets in kleine FIFO-Puffer im Fabric Interface ASIC (FIA). Es gibt 17 FIFO-Puffer (einer pro ToFab-Warteschlange). Wenn die FIA eine Zelle aus der zuFab-BMA abrufen, fügt sie eine 8-Byte-CRC hinzu (Gesamtzellengröße: 64 Byte; 44 Byte Nutzlast, 8 Byte Zellenheader, 4 Byte Puffer Header). Die FIA verfügt über SLI-ASICs (Serial Line Interface), die anschließend die 8B/10B-Codierung in der Zelle ausführen (z. B. die Fiber Distributed Data Interface (FDDI) 4B/5B) und sich auf die Übertragung über die Fabric vorbereiten. Dies mag sehr viel Overhead bedeuten (44 Byte Daten werden in 80 Byte in der gesamten Fabric umgewandelt!), ist jedoch kein Problem, da die Fabric-Kapazität entsprechend bereitgestellt wurde.
7. Nachdem eine FIA nun für die Übertragung bereit ist, fordert die FIA den Zugriff auf die Fabric über den derzeit aktiven Card Scheduler und die Clock (CSC) an. Der CSC arbeitet mit einem ziemlich komplexen Fairness-Algorithmus. Die Idee ist, dass kein LC die ausgehende Bandbreite einer anderen Karte monopolisieren darf. Beachten Sie, dass selbst wenn ein LC Daten über einen seiner eigenen Ports übertragen möchte, diese dennoch die Fabric durchlaufen muss. Dies ist wichtig, da ein Port auf einem LC, falls dies nicht der Fall ist, die gesamte Bandbreite für einen bestimmten Port auf demselben LC monopolisieren könnte. Das Switching-Design wird zudem komplizierter. Die FIA sendet Zellen über die Switch-Fabric an ihren ausgehenden LC (spezifiziert durch Daten im Cisco Cell-Header, der

von der Switching-Engine dort gespeichert wird). Der Fairness-Algorithmus ist auch für eine optimale Abstimmung ausgelegt. Wenn die Karte 1 gleichzeitig auf Karte 2 übertragen und die Karte 3 gleichzeitig auf Karte 4 übertragen möchte, geschieht dies parallel. Das ist der große Unterschied zwischen einer Switch-Fabric und einer Bus-Architektur. Stellen Sie sich dies als Analogie zu einem Ethernet-Switch gegenüber einem Hub vor. Wenn Port A an Port B gesendet werden soll und Port C an Port D kommunizieren möchte, erfolgen diese beiden Datenflüsse unabhängig voneinander. Auf einem Hub gibt es Halbduplex-Probleme wie Kollisionen, Backoff- und Wiederholungs-Algorithmen.

8. Die Cisco Zellen, die aus der Fabric kommen, werden durch SLI-Verarbeitung verarbeitet, um die 8B/10B-Codierung zu entfernen. Wenn hier Fehler auftreten, werden sie in der Ausgabe des Befehls `show controller fia` als "cell parity" (Zellparität) angezeigt. Weitere Informationen finden Sie unter [How To Read the Output of the show controller fia Command](#) ([So lesen Sie die Ausgabe](#) des Befehls "`show controller`").
9. Diese Cisco Zellen sind DMAs in FIFOs auf den frFab FIAs und dann in einen Puffer auf dem frFab BMA. Der frFab-BMA ist derjenige, der tatsächlich Zellen in ein Paket reassembliert. Woher weiß die frFab BMA, in welchen Puffer die Zellen eingelassen werden, bevor sie wieder zusammengebaut werden? Dies ist eine weitere Entscheidung der Incoming Line Card Switching Engine. Da alle Warteschlangen im gesamten Feld dieselbe Größe haben und in derselben Reihenfolge angeordnet sind, hat die Switching-Engine lediglich die Tx-LC-Einheit, die das Paket in die gleiche Nummernwarteschlange stellt, von der aus es den Router eingegeben hat. Die frFab-BMA-SDRAM-Warteschlangen können mit dem Befehl `show controller fab queue` im LC angezeigt werden. Siehe [How To Read the Output of the show controller frab | Tab Queue Commands auf einem Cisco Internet Router der Serie 1200](#) für Details. Dies ist im Grunde dieselbe Idee wie die toFab-BMA-Ausgabe. Pakete werden in Paketen eingesendet, die aus den entsprechenden freien Warteschlangen entfernt werden. Diese Pakete werden in die Fabric-Warteschlange gestellt und in die Warteschlange für die Schnittstelle (es gibt eine Warteschlange pro physischem Port) oder in die Warteschlange für die Ausgangsverarbeitung gestellt. Im rawQ passiert nicht viel: Multicast-Replizierung pro Port, Modified Deficit Round Robin (MDRR) - dieselbe Idee wie Distributed Weighted Fair Queuing (DWFQ) und Ausgabe-CAR. Wenn die Übermittlungswarteschlange voll ist, wird das Paket verworfen und der Zähler für die Ausgabe-Drop erhöht.
10. Die frFab-BMA wartet, bis der TX-Teil des PLIM bereit ist, ein Paket zu senden. Die frFab-BMA führt die eigentliche MAC-Umschreibung durch (basierend, erinnern Sie sich, auf den Informationen im Cisco Cell-Header), und DMAs übertragen das Paket auf einen kleinen (wiederum 2xMTU) Puffer im PLIM-Schaltkreis. Das PLIM kapselt gegebenenfalls ATM SAR und SONET und überträgt das Paket.
11. Der ATM-Datenverkehr wird wieder aufgebaut (durch die SAR), segmentiert (durch die ToFab-BMA), reassembliert (durch das Formular-BMA) und wieder segmentiert (durch die Form-SAR). Das passiert sehr schnell.

Dies ist der Lebenszyklus eines Pakets, von Anfang bis Ende. Wenn Sie wissen wollen, wie sich eine GSR am Ende des Tages anfühlt, lesen Sie dieses ganze Whitepaper 500.000 Mal!

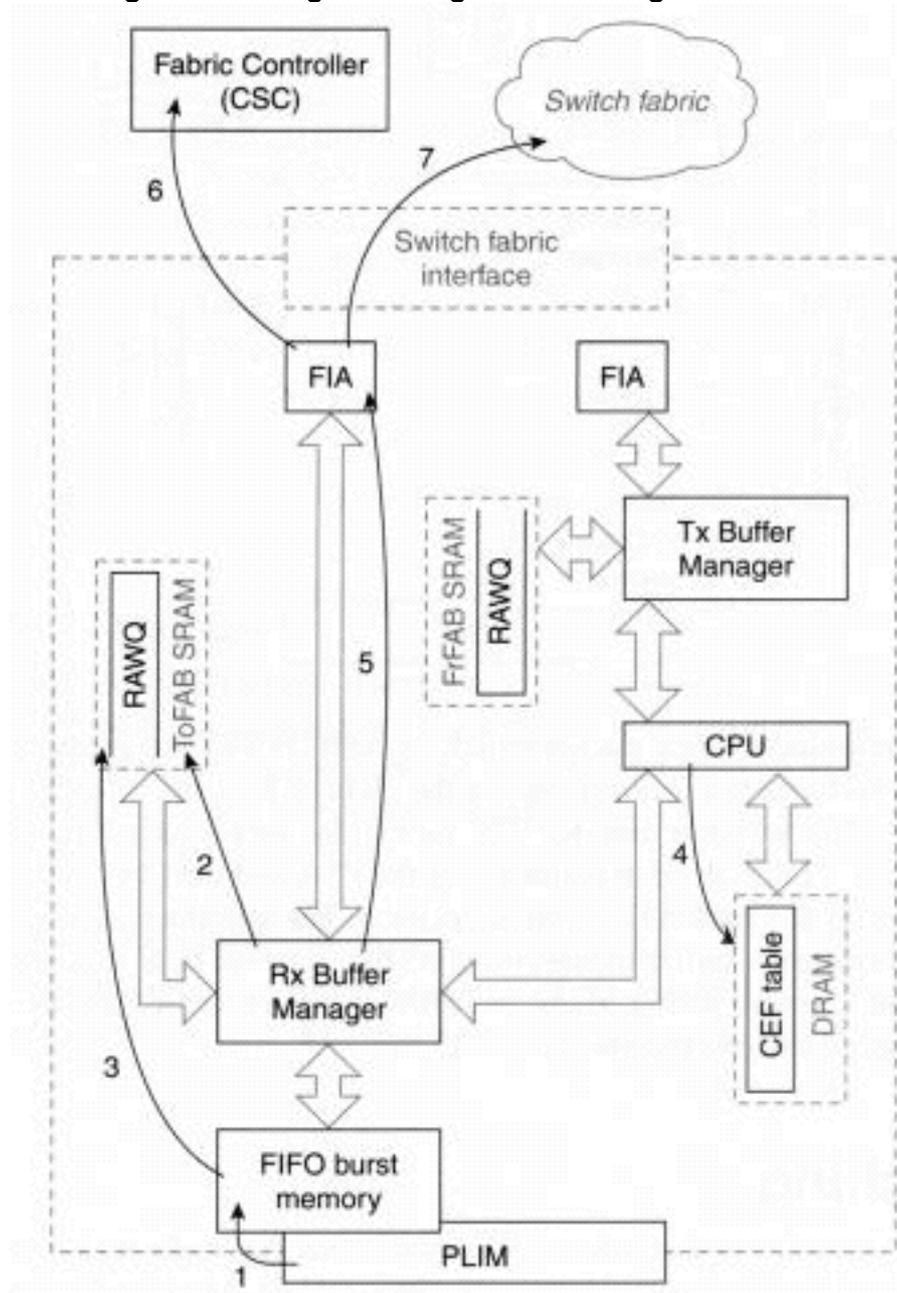
Der Paketvermittlungspfad im GSR hängt vom Weiterleitungstyp des LC ab. Jetzt werden wir alle Schritte für Engine 0, Engine 1 und die beiden LCs durchgehen.

[Paket-Switching: Line Cards für Engine 0 und Engine 1](#)

Die folgenden Abschnitte basieren auf dem Buch *Inside Cisco IOS Software Architecture*, Cisco Press.

[Abbildung 1](#) unten zeigt die verschiedenen Schritte beim Paket-Switching für Engine 0- oder Engine 1-LC.

Abbildung 1: Switching-Pfad Engine 0 und Engine 1



Der Switching-Pfad für die LC-Module Engine 0 und Engine 1 ist im Wesentlichen derselbe, obwohl die Engine 1 LC über eine erweiterte Switching-Engine und einen Puffer-Manager verfügt, um die Leistung zu erhöhen. Der Switching-Pfad ist wie folgt:

- **Schritt 1** - Der Schnittstellenprozessor (PLIM) erkennt ein Paket auf den Netzwerkmedien und beginnt damit, es in einen FIFO-Speicher zu kopieren, den so genannten **Burst-Speicher** auf dem LC. Die Menge an Burst-Speicher, die jede Schnittstelle hat, hängt vom Typ des LC ab. Typische LCs verfügen über einen Burst-Speicher von 128 KB bis 1 MB.
- **Schritt 2** - Der Schnittstellenprozessor fordert von der empfangenden BMA einen Paketpuffer an. Der Pool, aus dem der Puffer angefordert wird, hängt von der Länge des Pakets ab. Wenn

keine freien Puffer vorhanden sind, wird die Schnittstelle verworfen und der Zähler "ignore" der Schnittstelle erhöht. Wenn beispielsweise ein 64-Byte-Paket an eine Schnittstelle ankommt, versucht die BMA, einen 80-Byte-Paketpuffer zuzuweisen. Wenn im 80-Byte-Pool keine kostenlosen Puffer vorhanden sind, werden keine Puffer für den nächsten verfügbaren Pool zugewiesen.

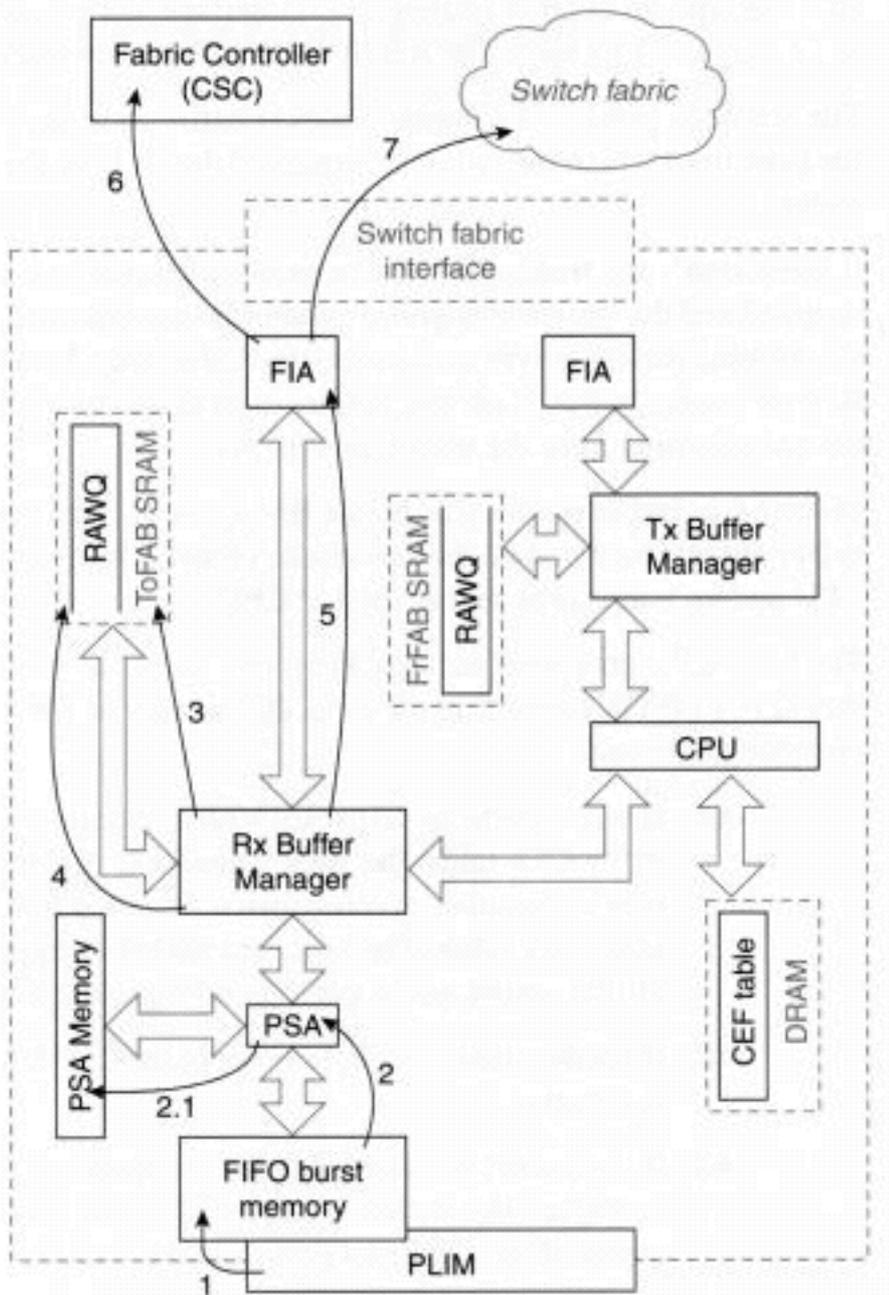
- **Schritt 3** - Wenn der BMA einen freien Puffer zuweist, wird das Paket in den Puffer kopiert und in die **Rohwarteschlange** (RawQ) eingestellt, um von der CPU verarbeitet zu werden. Ein Interrupt wird an die LC-CPU gesendet.
- **Schritt 4** - Die CPU des LC verarbeitet jedes Paket im RawQ, wie es empfangen wird (der RawQ ist ein FIFO). Die lokale verteilte Cisco Express Forwarding-Tabelle im DRAM wird zur Switching-Entscheidung herangezogen.**4.1** Wenn es sich um ein Unicast-IP-Paket mit einer gültigen Zieladresse in der CEF-Tabelle handelt, wird der Paket-Header mit den neuen Kapselungsinformationen aus der CEF-Adjacency-Tabelle neu geschrieben. Das geschichtete Paket wird in die virtuelle Ausgabewarteschlange gestellt, die dem Zielsteckplatz entspricht.**4.2** Wenn sich die Zieladresse nicht in der CEF-Tabelle befindet, wird das Paket verworfen.**4.3** Wenn es sich bei dem Paket um ein Steuerungskpaket (z. B. ein Routing-Update) handelt, wird das Paket in die virtuelle Ausgabewarteschlange des GRP gestellt und vom GRP verarbeitet.
- **Schritt 5** - Der empfangende BMA fragmentiert das Paket in Zellen mit 64 Byte und übergibt diese an die FIA, um es an den ausgehenden LC zu übertragen.

Am Ende von Schritt 5 wurde das Paket, das an eine LC-Engine 0/1 angeschlossen wurde, umgeschaltet und kann als Zellen über die Switch-Fabric übertragen werden. Fahren Sie im Abschnitt [Packet Switching](#) mit Schritt 6 fort: [Umschalten von Zellen über Fabric](#).

[Paket-Switching: Line Cards der Engine 2](#)

[Abbildung 2](#) unten zeigt den Pfad zum Paket-Switching, wenn die Pakete an einem Engine-2-LC eintreffen, wie in der folgenden Schrittliste beschrieben.

Abbildung 2: Engine 2-Switching-Pfad



- **Schritt 1** - Der Schnittstellenprozessor (PLIM) erkennt ein Paket auf den Netzwerkmedien und beginnt damit, es in einen FIFO-Speicher zu kopieren, den so genannten **Burst-Speicher** auf dem LC. Die Menge an Burst-Speicher, die jede Schnittstelle hat, hängt vom Typ des LC ab. Typische LCs verfügen über einen Burst-Speicher von 128 KB bis 1 MB.
- **Schritt 2** - Die ersten 64 Byte des Pakets, der so genannte Header, werden über den Packet Switching ASIC (PSA) weitergeleitet. **2.1** Das PSA schaltet das Paket um, indem es die lokale CEF-Tabelle im PSA-Speicher abfragt. Wenn das Paket nicht vom PSA gewechselt werden kann, fahren Sie mit Schritt 4 fort. Fahren Sie andernfalls mit Schritt 3 fort.
- **Schritt 3** - Der Receive Buffer Manager (RBM) akzeptiert den Header vom PSA und kopiert ihn in einen freien Puffer-Header. Wenn das Paket größer als 64 Byte ist, wird der Tail des Pakets auch im Paketspeicher in denselben freien Puffer kopiert und in die Warteschlange für die virtuelle Ausgangsausgabe des ausgehenden LC gestellt. Fahren Sie mit Schritt 5 fort.
- **Schritt 4** - Das Paket kommt an diesem Schritt an, wenn es nicht vom PSA gewechselt werden kann. Diese Pakete werden in die **Rohwarteschlange** (RawQ) gestellt, und der Switching-Pfad ist im Wesentlichen derselbe wie für Engine 1 und Engine 0 LC von diesem Punkt an (Schritt 4 bei Engine 0). Beachten Sie, dass die Pakete, die vom PSA geschickt

werden, niemals im RawQ platziert werden und dass keine Interrupt an die CPU gesendet wird.

- **Schritt 5** - Das Fabric Interface Module (FIM) ist für die Segmentierung der Pakete in [Cisco Zellen](#) und das Senden der Zellen an den Fabric Interface ASIC (FIA) für die Übertragung an den ausgehenden LC zuständig.

Paket-Switching: Switching-Zellen über die Fabric hinweg

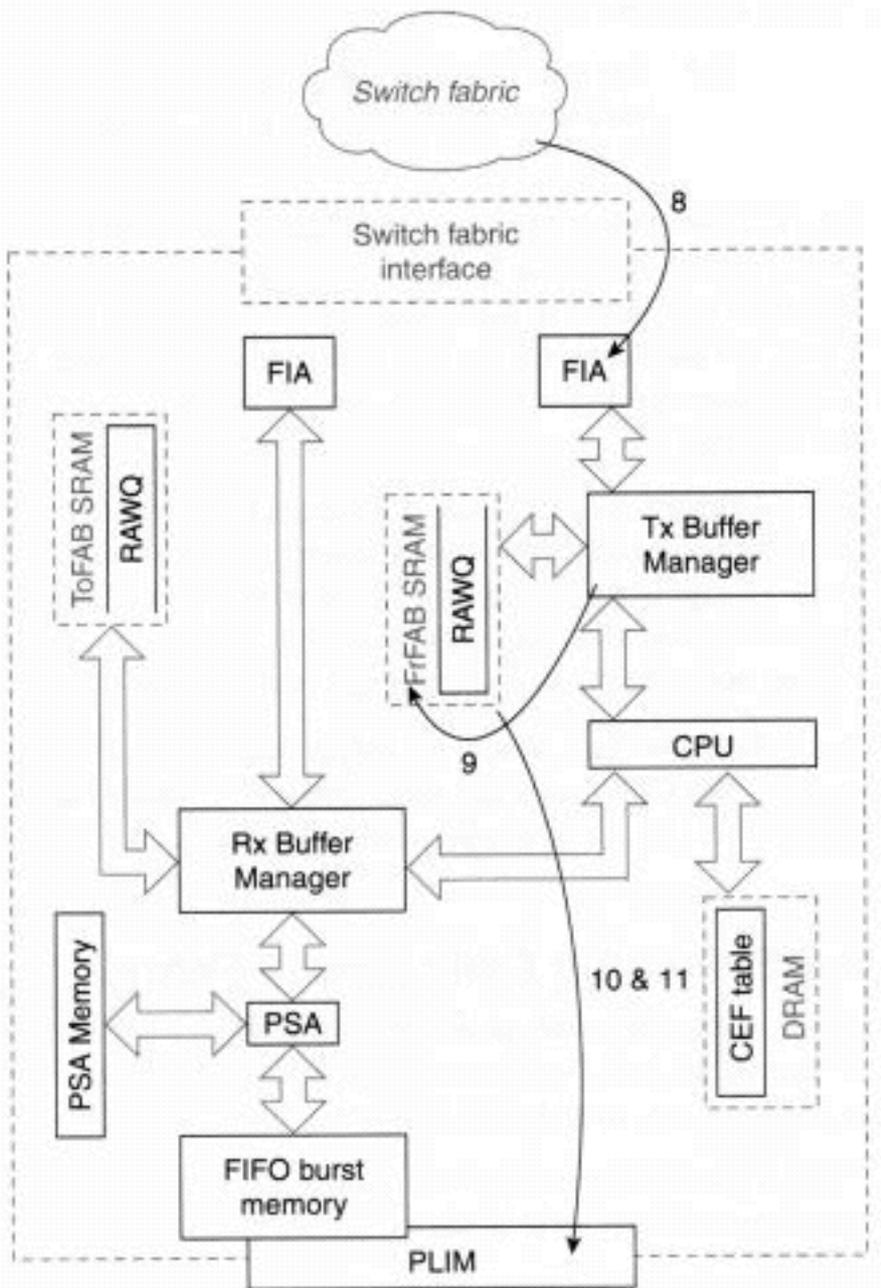
Sie gelangen zu dieser Phase, nachdem die Paket-Switching-Engine die Pakete umschaltet. In dieser Phase werden die Pakete in Cisco Zellen segmentiert und warten auf die Übertragung über die Switching-Fabric. Die Schritte für diese Phase sind wie folgt:

- **Schritt 6** - Die FIA sendet eine Finanzhilfeanfrage an den CSC, der die Übertragung jeder Zelle über die Switch-Fabric plant.
- **Schritt 7** - Wenn der Scheduler Zugriff auf die Switch-Fabric gewährt, werden die Zellen in den Zielsteckplatz übertragen. Beachten Sie, dass die Zellen möglicherweise nicht alle gleichzeitig übertragen werden; Andere Zellen in anderen Paketen können untereinander verschachtelt sein.

Paket-Switching: Pakete übertragen

Abbildung 3 zeigt die letzte Phase des Paket-Switching. Die Zellen werden neu zusammengesetzt, und das Paket wird auf das Medium übertragen. Dies erfolgt auf der ausgehenden Linecard.

Abbildung 3: Cisco 12000 Packet Switching: Übertragungsphase



- **Schritt 8:** Die über die Fabric geschwitchten Zellen gelangen über die FIA in die ZielLine Card.
- **Schritt 9 -** Der Sende-Puffer-Manager weist einen Puffer aus dem Speicher des Übertragungspakets zu und reassembliert das Paket in diesem Puffer.
- **Schritt 10 -** Wenn das Paket neu erstellt wird, ordnet die übertragende BMA das Paket in die Übermittlungswarteschlange der Zielschnittstelle im LC ein. Wenn die Übermittlungswarteschlange der Schnittstelle voll ist (das Paket kann nicht in die Warteschlange gestellt werden), wird das Paket verworfen und der **Drop-Zähler** der **Ausgabewarteschlange** erhöht.**Hinweis:** In der Übertragungsrichtung werden im RawQ nur Pakete platziert, bei denen die LC-CPU vor der Übertragung eine Verarbeitung durchführen muss. Beispiele sind IP-Fragmentierung, Multicast und Ausgabe-CAR.
- **Schritt 11 -** Der Schnittstellenprozessor erkennt ein Paket, das auf die Übertragung wartet, entfernt den Puffer aus dem übertragenden Speicher, kopiert es in den internen FIFO-Speicher und überträgt das Paket auf den Medien.

Paketfluss-Zusammenfassung

IP-Pakete, die die 12000 passieren, werden in drei Phasen verarbeitet:

- Eingangs-Linecard in drei Abschnitte: Eingangs-PLIM (Physical Line Interface Module) - Unframing, HDLC- und PPP-Verarbeitung (Optical to Electrical Converting, SONET)/Synchronous Digital Hierarchy (SDH) IP Forwarding - Weiterleitungsentscheidung basierend auf FIB-Suche und Warteschlangenverwaltung in einer der Eingangs-Unicast-Warteschlangen oder Multicast-Warteschlangen. Management von Eingangswarteschlangen und Fabric-Schnittstelle: Verarbeitung von Eingangswarteschlangen und Herabstufungen (Random Early Detection, RED)/Weighted Random Early Detection (WRED), um die Fabric-Nutzung zu maximieren.
- Switching von IP-Paketen über die 12.000 Fabric von der Eingangs- zur Ausgangs-Karte oder den Ausgangs-Karten (bei Multicast).
- Ausgangs-Linecard in drei Abschnitte: Egress Fabric Interface - Zusammenstellung der zu sendenden IP-Pakete und Einreihung in die Warteschlange für Ausgangs-Datenpuffer Verarbeitung von Multicast-Paketen. Management der Ausgangswarteschlange - ROT/WRED-Verarbeitung in den Eingangs-Warteschlangen und Entfernen der Warteschlange zum Ausgangs-PLIM, um die Ausgangs-Leitungsauslastung zu maximieren. Ausgangs-PLIM - HDLC- und PPP-Verarbeitung, SONET/SDH-Framing, Umwandlung von elektrisch zu optisch.

Zugehörige Informationen

- [Technischer Support – Cisco Systems](#)