

# Cisco 12000 Series Internet Router体系结构：数据包交换

## Contents

[Introduction](#)

[Prerequisites](#)

[Requirements](#)

[Components Used](#)

[Conventions](#)

[背景信息](#)

[数据包交换：概述](#)

[数据包交换：引擎0和引擎1线卡](#)

[数据包交换：引擎2线卡](#)

[数据包交换：跨结构交换信元](#)

[数据包交换：传送信息包](#)

[信息包流汇总](#)

[Related Information](#)

## [Introduction](#)

本文检查Cisco 12000 Series Internet Router的最重要的结构上元素--交换信息包。交换信息包是完全不同的与任何共享内存或基于总线的Cisco体系结构。通过使用Crossbar结构，Cisco 12000提供非常很多带宽和可扩展性。此外，12000使用虚拟输出输出队列排除在交换矩阵内的Head of Line封闭。

## [Prerequisites](#)

## [Requirements](#)

There are no specific requirements for this document.

## [Components Used](#)

本文的信息根据以下硬件：

- Cisco 12000 Series Internet Router

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, make sure that you understand the potential impact of any command.

## Conventions

有关文档规则的详细信息，请参阅 [Cisco 技术提示规则](#)。

## 背景信息

(在Cisco 12000的交换决定由线卡(LCs)完成。对于一些LCs，专用的Application-specific integrated circuit (ASIC)实际上转换信息包。Distributed Cisco Express Forwarding (DCEF)是可用唯一的交换方法。

**重新标明：**引擎0，1和2不是Cisco发展的最新的引擎。也有引擎3，4和4+线卡，与跟随的更多。引擎3线卡能够执行边缘功能在线路费率。越高第3层引擎，更多信息包在硬件方面得到交换。您能找到关于不同的线卡的一些有用的信息可用为他们在[Cisco 12000 Series Internet Router](#)的Cisco 12000 Series Router和引擎：[常见问题](#)。

## 数据包交换：概述

入口线卡总是转发信息包(LC)。出口LC只执行例如是依据队列的outbound服务质量(QoS) (加权随机早期检测(WRED)或承诺接入速率(CAR))。使用Distributed Cisco Express Forwarding (DCEF)，大多信息包由LC转换。仅控制数据包(例如路由更新)被发送到处理的千兆路由处理器 (GRP)。信息包交换的路径取决于在LC使用的交换引擎的种类。

这是发生了什么，当信息包来在：

1. 信息包进入物理层接口模块(PLIM)。多种事发生这里：收发器把光信号变成电子那些(多数CSR线卡有光纤连接器)L2去除构建帧(神志正常，异步传输模式(ATM)、以太网，高级数据链路控制(HDLC)/点对点协议-PPP)ATM信元被重新召集失败循环冗余校验(CRC)丢弃的信息包
2. 因为信息包收到并且被处理，它是直接存储器访问到称为“先入先出(FIFO)分段存储的”小的(大约2个x最大传输单元(MTU)缓冲区)内存。相当数量此内存取决于LC的种类(从128 KB到1 MB)。
3. 一旦信息包完全地在FIFO内存，在PLIM的application-specific integrated circuit (ASIC)与缓冲区管理ASIC (BMA)联系并且请求缓冲区放置信息包。BMA告诉什么大小信息包是，并且相应地分配缓冲区。如果BMA不能获得适当大小的缓冲区，信息包被丢弃，并且“ignore”计数器在流入的接口被增加。没有回退机制如同一些其他平台。当这继续时，PLIM可能收到在FIFO分段存储的另一个信息包，是它为什么在大小上是2xMTU。
4. 如果有空闲的缓存可用在正确的队列，BMA存储信息包在适当的大小自由队列列表。此缓冲区在自然状态的队列被放置，由辣调味汁ASIC或R5K CPU检查。R5K CPU通过参见其在动态RAM (DRAM)的本地dCEF表确定信息包的目的地，从自然状态的队列然后移动缓冲区向ToFabric队列与目的地地址槽相应。如果目的地不在CEF表里，信息包被丢弃。如果信息包是控制数据包(例如，路由更新)，排队对GRP的队列，并且由GRP处理。有17个Tofab队列(16单播，加上1组播)。有每线卡一个tofab队列(这包括RP)。这些队列叫作“虚拟输出输出队列”，并且是重要的，以便head-of-line封闭不发生。
5. Tofab BMA剪切信息包成44字节部分，是什么的有效载荷最终叫作“Cisco信元”。frfab BMA产生这些信元一个8字节报头和4字节缓冲报头(到目前为止总数据估量= 56个字节)，然后排队到适当的Tofab队列(到时，在缓冲区来自的池的#Qelem计数器由一个断开，并且Tofab队列计数器由一个上升)。“作决策者”取决于交换引擎的种类：在引擎2+卡，特殊ASIC用于改进信息包是交换式的方式。正常信息包(IP/Tag、没有选项，校验和)直接地由Packet Switching ASIC (PSA)处理，然后绕过原始的queue/CPU/Salsa组合和排队直接地在tofab队列上。信息包的仅

前64个字节通过分组交换ASIC。如果信息包不可能由PSA转换，信息包被排列对LC的CPU将处理的RawQ按照说明以前。这时，交换决定做了，并且信息包被排列了在适当的Tofab输出输出队列上。

6. tofab BMA DMAs (直接存储器访问)信息包的信元到在矩阵接口ASIC (FIA)的的小的FIFO缓冲区里。有17个FIFO缓冲区(—每个Tofab队列)。当FIA从tofab BMA时获得信元，添加一个8字节CRC (总信元大小- 64个字节;44字节有效载荷、8个字节信元头，4个字节缓冲报头)。FIA有串行线路然后进行8B/10B在信元的编码的接口(SLI) ASIC (类似光纤分布式数据接口(FDDI) 4B/5B)，并且准备在结构传输它。这也许似乎类似很多开销(44字节的数据获得把变成在结构间的80个字节!)，但是它不是问题，因为结构容量相应地设置了。
7. 既然FIA准备传输，FIA请求对结构的访问从当前活跃的卡调度程序和时钟(CSC)。CSC研究相当复杂公平算法。想法是LC没有允许垄断其他卡流出的带宽。注意，即使LC要从其自己的端口之一当中传输数据，必须仍然通过结构。这是重要的，因为，如果这没有发生，在LC的一个端口可能垄断一个特定端口的所有带宽在该同样LC。它也将做复杂化的交换设计。FIA发送在交换矩阵间的信元到他们流出的LC (指定由数据在交换引擎放置的那里Cisco信元头)。公平算法为最佳匹配也设计;如果card1要传输到卡2，并且卡3要同时传输到卡4，这平行发生。那是在交换矩阵和总线体系结构之间的大差值。认为它如类似于一台以太网交换机与集线器;在交换机，如果端口A要发送到端口B和端口C要与端口D谈，那两流独立彼此发生。在集线器上，有半双工问题例如冲突和备值并且再试算法。
8. 从结构出来的Cisco信元通过`SLI`处理去除8B/10B编码。如果那里这里任何错误，他们在show controller fia命令输出中将出现作为“信元奇偶校验”。请参阅[如何阅读输出的show controller fia命令](#)其他信息的。
9. 这些Cisco信元是DMA'd到在frfab FIAs的FIFO，然后到在frfab BMA的一个缓冲区。frfab BMA是实际上执行信元重新组装到信息包的那个。frfab BMA如何了解放置信元的什么缓冲区，在重新召集他们前？这是流入线卡交换引擎做出的另一决定;因为在整个机箱的所有队列是相同大小和按同一顺序，交换引擎在进入路由器的同一个编号队列安排Tx LC放置信息包。frfab BMA SDRAM队列可以用show controller frfab queue命令查看在LC。请参阅[如何阅读show controller frfab的输出|在Cisco 12000 Series Internet Router的tofab队列命令](#)关于详细资料。这基本上是想法和tofab BMA输出一样。信息包在从他们的各自自由队列离队的信息包进来和安置。这些信息包被放置到from-fabric队列，排队在接口队列(有每个物理端口一个队列)或输出处理的rawQ。并非在rawQ发生：每端口组播复制、改进的差额轮询(MDRR) -想法和分布式加权公平排队(DWFQ)一样和输出控制访问率。如果传输队列满，信息包被丢弃，并且输出丢弃计数器被增加。
10. frfab BMA等待，直到PLIM的TX部分准备发送信息包。frfab BMA执行实际MAC重写(基于，请切记，在Cisco信元头包含的信息)和DMAs信息包对在PLIM电路的一个小的(再，2xMTU)缓冲区。PLIM执行ATM SAR，并且SONET封装，只要适合的话，并且传输信息包。
11. ATM流量被重新召集(由SAR)，被分段(由tofab BMA)，被重新召集(由fromfab BMA)并且再被分段(由fromfab SAR)。这非常迅速发生。

那是信息包的生命周期，自始至终。如果要了解什么GSR感觉类似当晚，读此整个文章500,000次!

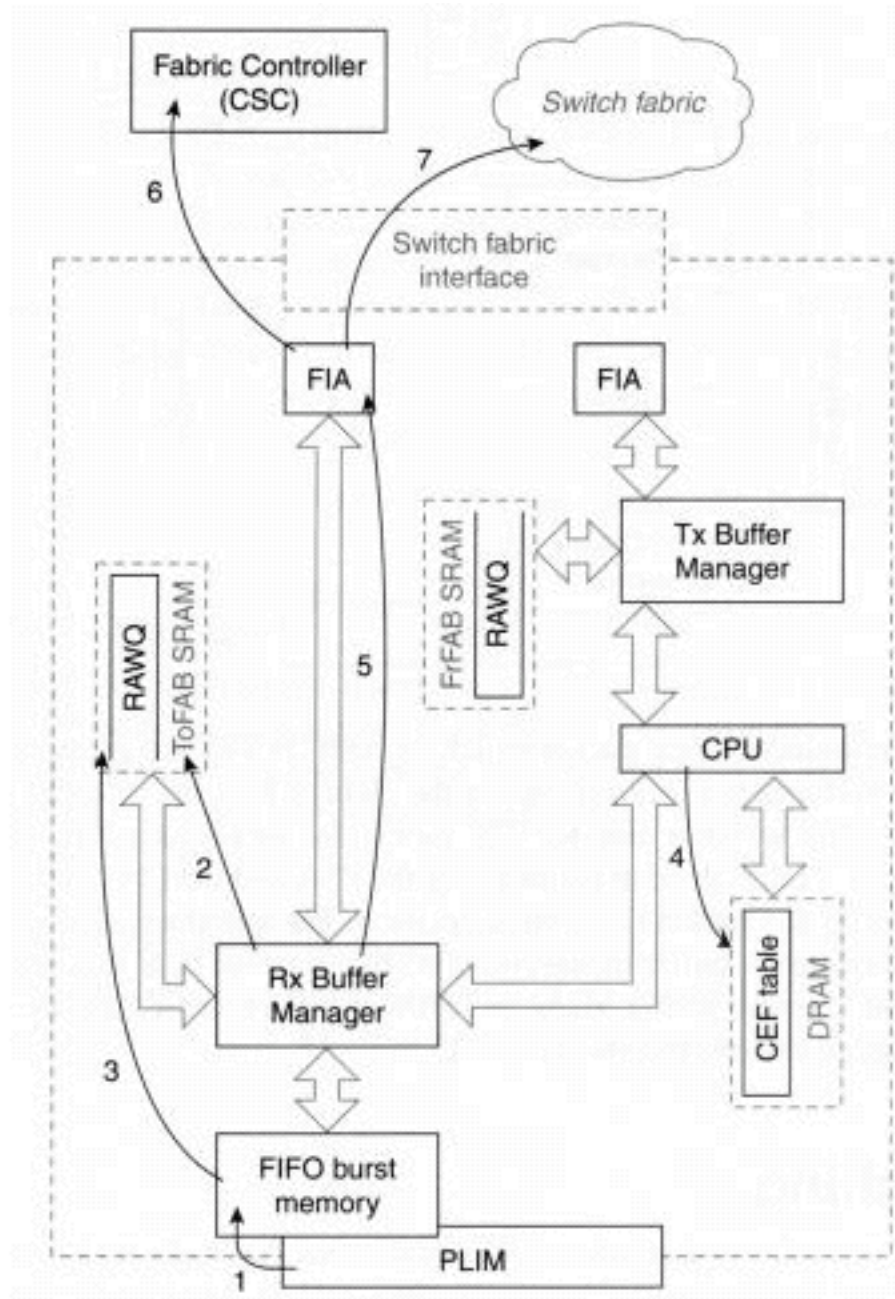
在GSR的信息包交换的路径取决于转发引擎的种类在LC的。现在我们将通过引擎0，引擎1和两个LCs的所有步骤。

## [数据包交换：引擎0和引擎1线卡](#)

下面的部分根据书在Cisco IOS软件体系结构里面，Cisco新闻。

在引擎0或引擎1 LC的，数据包交换期间下面的[图1](#)说明不同的步骤。

图 1：引擎0和引擎1交换路径



引擎0和引擎1 LC的交换路径根本是相同的，虽然引擎1 LC有一个改进的交换引擎和缓冲区管理程序更完善的性能的。交换路径如下：

- **Step1** -接口处理器(PLIM)发现在网络媒介的一个信息包并且开始复制它到称为在LC的分段存储的FIFO内存。每个接口有的相当数量分段存储取决于LC的种类;典型的LCs有128 KB对分段存储1 MB。
- **第2步**-接口处理器请求从接受BMA的一个信息包缓冲;缓冲区被请求的池取决于信息包的长度。如果没有任何空闲的缓存，接口被丢弃，并且接口的"ignore"计数器被增加。例如，如果64字节信息包到达接口，BMA设法分配80字节信息包缓冲。如果空闲的缓存不存在于80字节池，缓冲区从下个可用的池没有分配。
- **第3步**-当BMA时分配空闲的缓存，信息包在自然状态的队列(RawQ)被复制到缓冲区和被排列处理的由CPU。中断被发送到LC CPU。
- **第4步**- LC的CPU进程在RawQ的每个信息包，因为被接受(RawQ是FIFO)，参见在DRAM的本地分布式Cisco快速转发模式表做出交换决定。**4.1**如果这是一个单播IP信息包和一个有效目的地地址一起在CEF表里，信息包报头以从CEF邻接表得到的新的封装信息重写。交换数据包在虚拟输出输出队列被排列与目的地地址槽相应。**4.2**如果目的地地址不在CEF表里，信息包被丢

弃。4.3如果信息包是控制数据包(例如路由更新), 信息包在GRP的虚拟输出输出队列被排列并且由GRP处理。

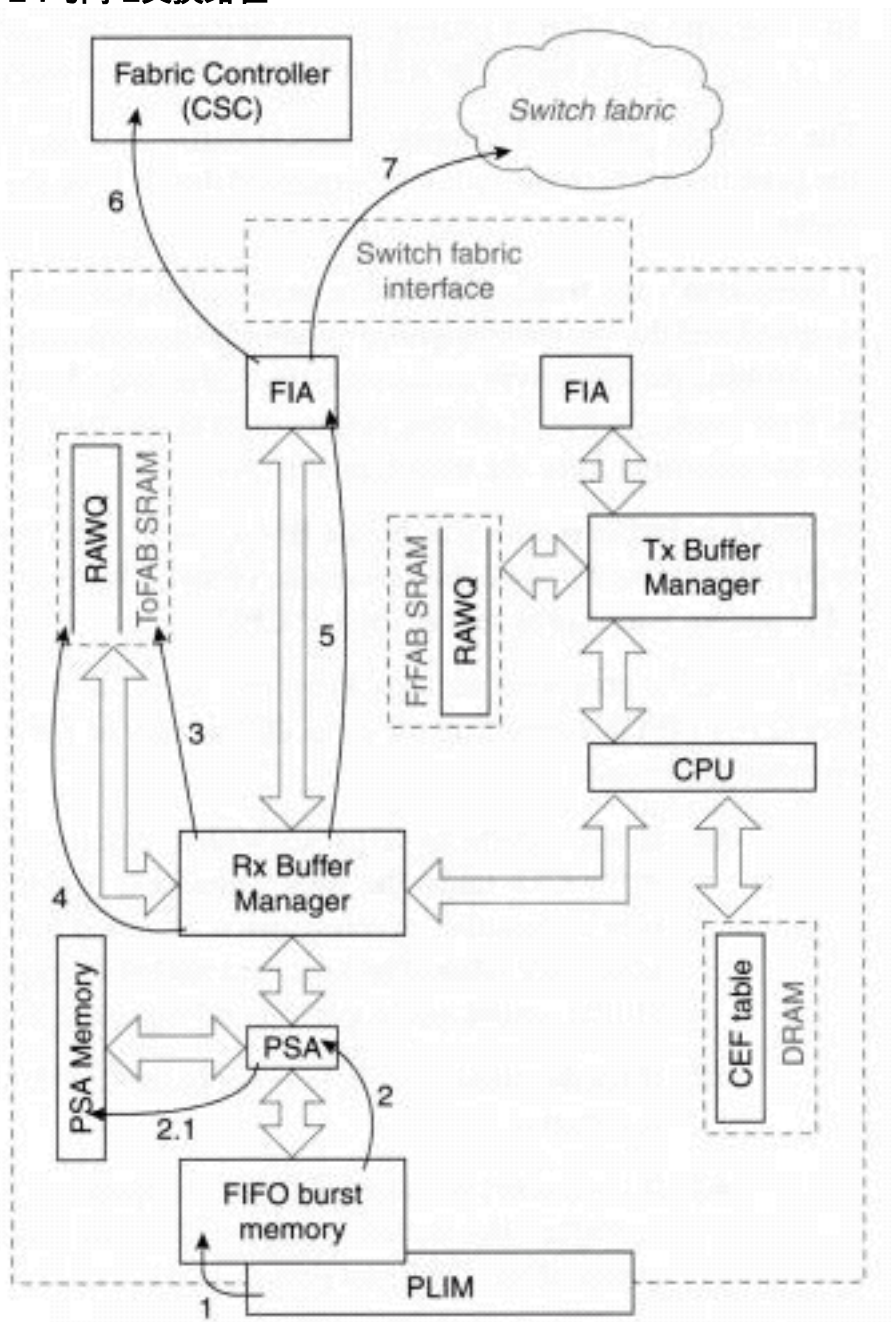
- **第5步**-接受BMA分段信息包到64字节信元, 并且递交这些对发射的FIA对outbound LC。

在第5步结束时, 到达到引擎0/1 LC的信息包被转换了并且准备在交换矩阵间被传输作为信元。进入在[部分Packet Switching](#)的步骤6 : [跨结构交换信元](#)。

## 数据包交换：引擎2线卡

下面的图2说明信息包交换的路径, 当信息包到达到引擎2 LC时, 正如步骤以下列表所描述。

图 2：引擎2交换路径



- **Step1** -接口处理器(PLIM)发现在网络媒介的一个信息包并且开始复制它到称为在LC的分段存储的FIFO内存。每个接口有的相当数量分段存储取决于LC的种类;典型的LCs有128 KB对分段存储1 MB。
- **第2步**-信息包的前64个字节, 称为报头, 通过Packet Switching ASIC (PSA)。2.1 PSA通过参

见在PSA内存的本地CEF表转换信息包。如果信息包不可能由PSA转换，请进入步骤4;否则，请继续对第3.步。

- **第3步**- Receive Buffer Manager (RBM)接受从PSA的报头并且复制它到一个自由缓冲报头。如果信息包大于64个字节，信息包的尾标在流出的LC[虚拟输出输出队列](#)也被复制到在信息包存储器的同一空闲的缓存和排队。进入步骤5。
- **第4步**-，如果不可能由PSA，交换信息包到达此步骤。这些信息包在[自然状态的队列\(RawQ\)](#)被放置，并且交换路径根本是相同的象为从此点(一旦引擎0)的第4步的引擎1和引擎0 LC。注意由PSA转换的信息包在RawQ和没有中断从未安置被发送到CPU。
- **第5步**- Fabric Interface Module (FIM)对分段信息包到[Cisco信元](#)和发送信元负责到发射的矩阵接口ASIC (FIA)到outbound LC。

## [数据包交换：跨结构交换信元](#)

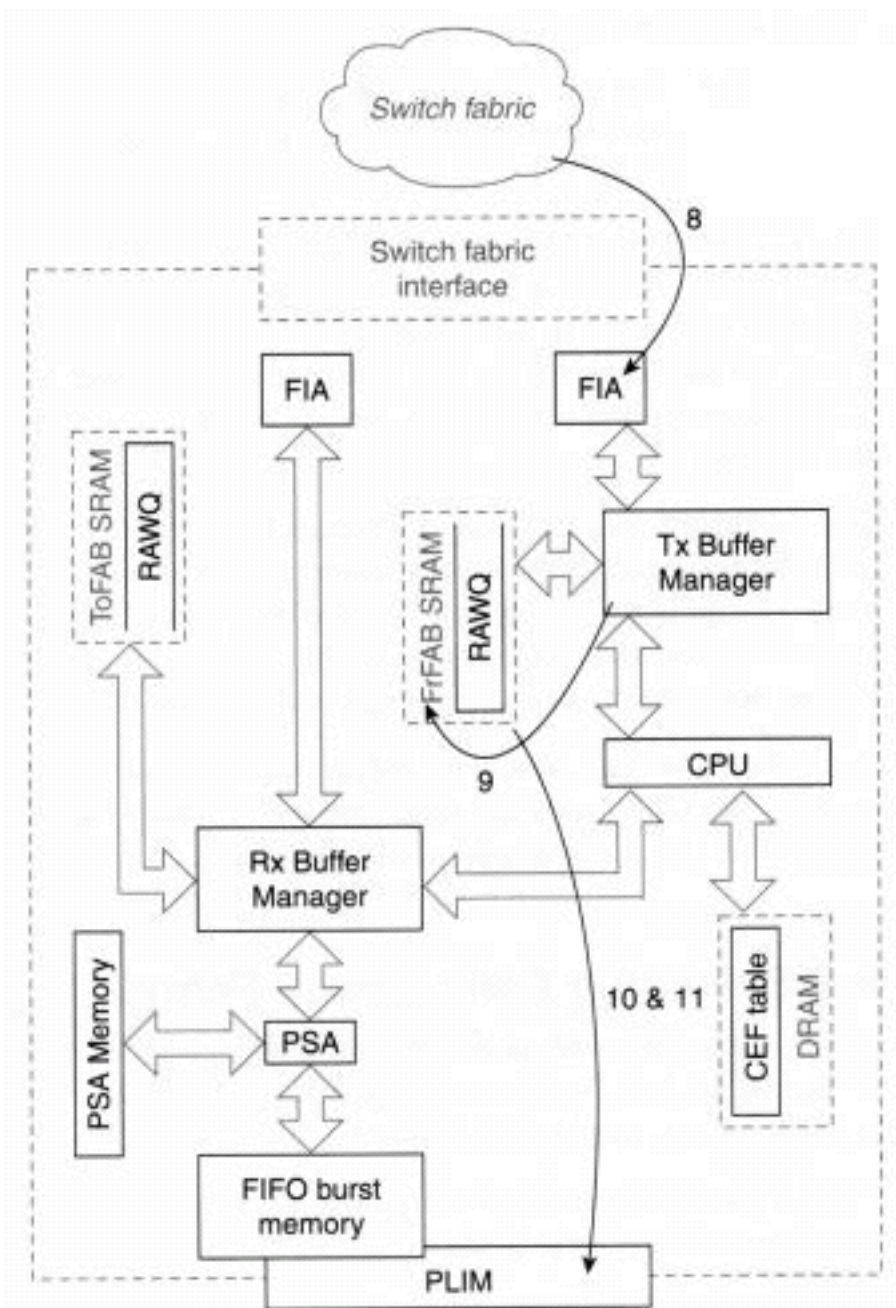
您在数据包交换引擎以后到达在此阶段转换信息包。在此阶段，信息包被分段到Cisco信元和等待在交换结构间传输。此阶段的步骤如下：

- **第6步**- FIA发送授予请求到CSC，安排在交换矩阵间的每个信元的转移。
- **第7步**-当调度程序准许对交换矩阵时的访问，信元调用到目的地地址槽。注意信元也许不同时被传输;也许插入在其他信息包内的其他信元。

## [数据包交换：传送信息包](#)

下面的图3显示数据包交换最后阶段。信元被重新召集，并且信息包传输在媒体上。这在outbound线卡发生。

图 3：Cisco 12000数据包交换：传输阶段



- **第8步**-在结构间被交换的信元到达目的地线卡通过FIA。
- **第9步**-传输缓冲区管理程序从传输信息包存储器分配一个缓冲区并且重新组装在此缓冲区的信息包。
- **第10步**-当信息包被重建时，传输BMA排列在目的地接口的传输队列上的信息包在LC。如果接口传输队列满(信息包不可能被排列)，信息包被丢弃，并且**输出队列丢弃**计数器被增加。  
**Note:** 在传送方向，当信息包在RawQ之时安置是，当LC CPU需要在发射前执行其中任一处理。示例包括IP分段、组播和输出控制访问率。
- **第11步**-接口处理器发现一个信息包等待传输，离队从传输内存的缓冲区，复制它到内部FIFO内存，并且传输在媒体的信息包。

## 信息包流汇总

横断12000的IP信息包在三个阶段内被处理：

- 在三个部分的入口线卡：入口PLIM (物理线路接口模块) -光学对电子转换，同步光网络 (SONET) /Synchronous数字体系(SDH) un-framing， HDLC和ppp处理。IP转发-根据FIB查找

和排队的转发决策到其中一个入口单播队列或组播队列。入口队列管理和矩阵接口-随机早期检测处理在入口队列和离队往结构的/Weighted随机早期检测(WRED)为了最大化结构利用率。

- 交换的IP信息包通过12000结构从进入卡到输出卡或输出卡(在组播的情况下)。
- 在三个部分的出口线卡：出口矩阵接口-重新组装IP信息包是发送和排队到出口队列;处理组播信息包。出口队列管理-在入口队列的RED/WRED处理和离队往出口PLIM最大化出口线路利用率。出口PLIM - HDLC和ppp处理，SONET/SDH构建帧，电子对光转换。

## [Related Information](#)

- [Technical Support - Cisco Systems](#)