如何选择网络的最佳路由器交换路径

目录

简介

各种 Cisco 路由器和 Cisco IOS® 软件版本可用的交换路径太多。哪一个最适合您的网络，并且如何使它们全部工作？此白皮书企图说明了下列每一条交换路径，这样您便可以做出哪一条交换路径适合您的网络的最佳决策。

首先，检查转发进程本身。通过路由器转发数据包有三个步骤：

1. 确定数据包的目标地址是否可达。
2. 确定目的地的下一跳和到达下一跳的接口。
3. 重写数据包上的媒体访问控制 (MAC) 报头，使其成功到达下一跳。

上述每一个步骤对于数据包是否能够到达其目标地址而言都至关重要。

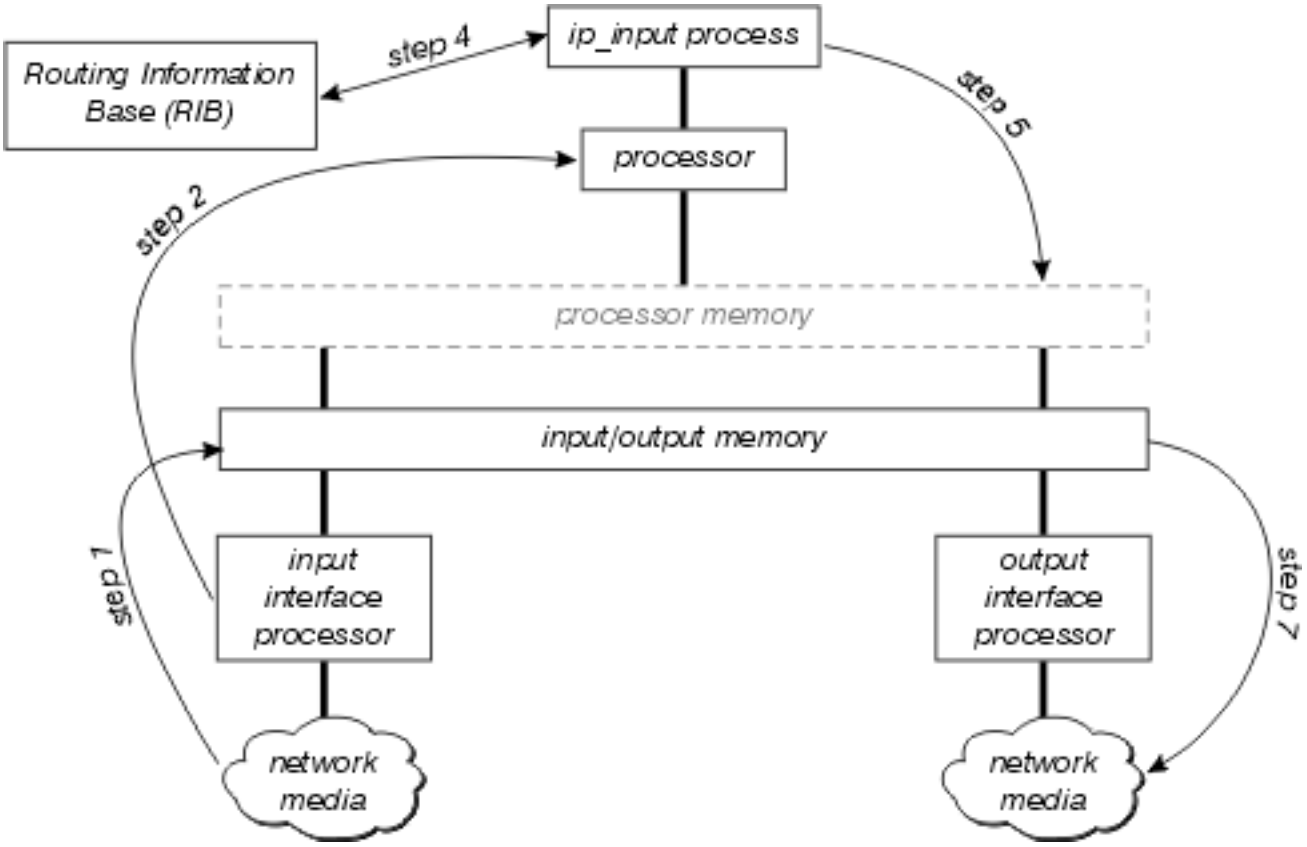
注意：在本文档中，IP交换路径用作示例；实际上此处提供的所有信息都适用于其他协议等同的交换路径（如果它们存在的话）。

过程交换

进程交换是交换路径中的最小共同特性；它可以在每种IOS版本和每种平台上提供，可用于正被交换的每种流量类型。进程交换由两个基本概念定义：

- 用于重写信息包上的MAC报头的转发决策和信息从路由表（从路由信息库或RIB）、地址解析协议(ARP)高速缓冲存储器或从包含MAC报头信息的某些其他列表（这些列表被映射到与路由器直接连接的每台主机的IP地址）中获取。
- 数据包由一个运行在 IOS 内的正常进程交换。换句话说，转发决策通过IOS调度程序做出，作为路由器上的其他程序的对等体运行，如路由协议。进行数据包进程交换时，不需要中断在路由器上正常运行的进程。

下图展示了进程交换路径。



可以更详细地了解此图：

1. 接口处理器首先检测出网络介质上有一个信息包，并把此信息包传输到路由器上的输入-输出内存。
2. 接口处理器生成接收中断。在此中断期间，中央处理器确定这是什么类型的信息包(假设它是IP信息包)，如果需要，则将它复制到处理器内存中(此决策独立于平台)。 最后，处理器把信息包放置在适当程序的输入队列中，并发布中断指令。
3. 当调度程序下一次运行时，它将记录ip_input输入队列中的信息包，并安排此程序运行。
4. 当ip_input运行时，查询RIB，确定下一跳和输出接口，然后查询ARP缓存，确定此下一跳的正确物理层地址。
5. **ip_input然后重写信息包MAC报头，然后将信息包放入正确出局接口的输出队列中。**
6. 信息包从出局接口的输出队列被复制到出局接口的传输队列；所有出站服务质量都发生在这两个队列之间。
7. 输出接口处理器在其传输队列中发现信息包，然后在网络媒介上传输信息包。

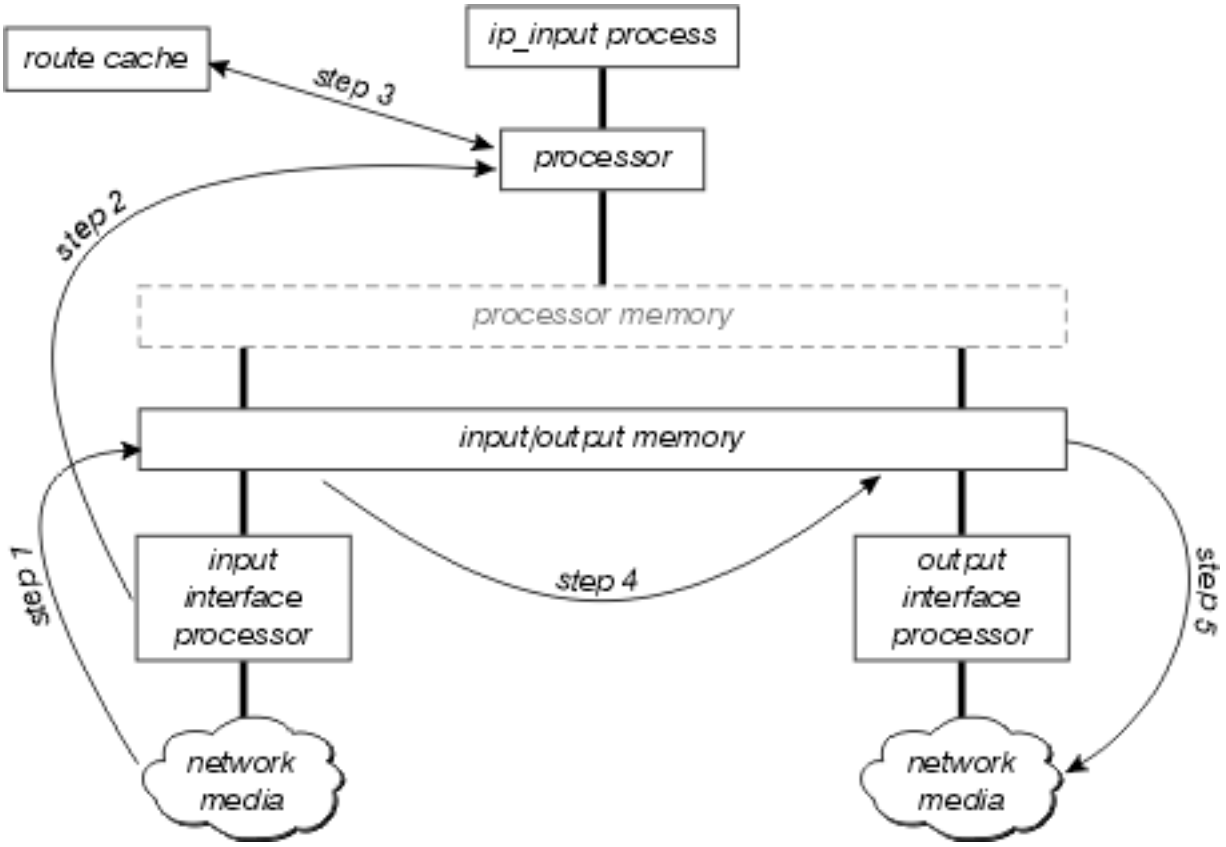
影响数据包交换的几乎所有的功能，如网络地址转换(NAT)和策略路由，会在进程交换路径中首次出现。经过证明和优化后，这些功能可能会也可能不会出现在中断上下文交换中。

中断 上下文 交换

中断上下文交换是 Cisco 路由器使用的第二种主要交换方法。中断上下文交换和进程交换之间的主要区别是：

- 交换数据包时需要中断处理器上当前运行的进程。信息包根据需要进行切换，而不只是安排好ip_input进程时才被转换。
- 处理器使用某种形式的路由缓存查找转换信息包需要的所有信息。

下图展示了中断上下文交换：



可以更详细地了解此图：

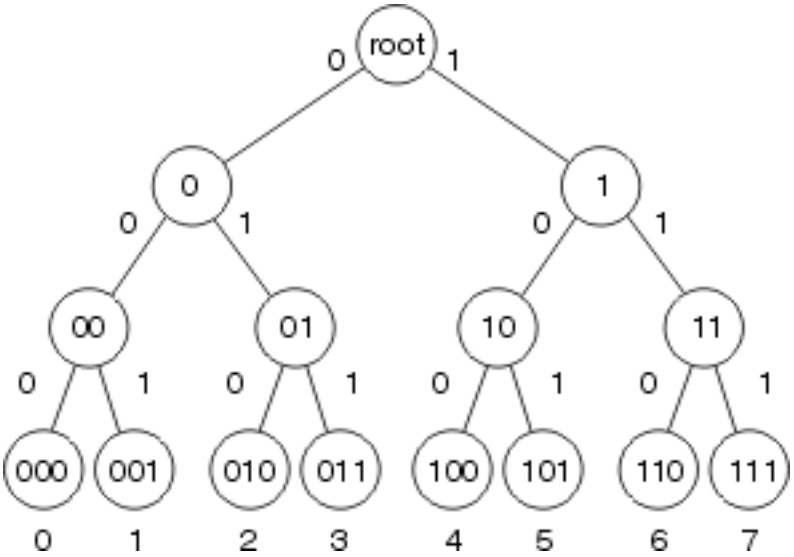
1. 接口处理器首先检测出网络介质上有一个信息包，并把此信息包传输到路由器上的输入-输出内存。
2. 接口处理器生成接收中断。在此中断期间，中央处理器确定这是什么类型的信息包(假设它是IP信息包)，然后开始交换该信息包。
3. 处理器在路由缓存进行搜索，确定数据包的目的地是否可及，输出接口应该是什么，通往该目的地的下一跳是什么，要成功到达下一条数据包的MAC报头应该是什么。处理器使用此信息重写数据包的 MAC 报头。
4. 信息包现在被复制到输出接口的传输或输出队列(根据各种要素)。 收到的中断信号现在返回，而中断发生前正在处理器上运行的程序则继续运行。
5. 输出接口处理器在其传输队列中发现信息包，然后在网络媒介上传输信息包。

阅读此说明后浮现在脑海的第一个问题是：“高速缓冲存储器中是什么？” 根据中断上下文交换的类型，有三种可能的答案：

- 快速交换
- 最佳交换
- Cisco 快速转发

快速交换

快速交换使用快速查找和参考的二分树，存储转发信息和MAC报头重写字符串。下图展示了一个二分树：



在快速交换中，可达性信息通过二分树上存在的节点表示为信息包的目的地。每个目的地的MAC报头和出局接口均作为树内节点的部分信息存储。二分树实际上可以有 32 个级别（上图中的树出于示例目的经过了很大程度的简化）。

要搜索某个二分树，您只需从所查找的（二进制）编号的左侧（带有最高有效位）开始，然后根据该编号在树中向右侧或左侧的分支进行搜索。例如，如果您要查找与此树中的数字4相关的信息，您将从右分支开始，因为第一个二进制数字是1。您将沿着树向下，比较（二进制）数字中的下一个数字，直到您到达结尾。

快速交换的特性

作为树节点的组成部分，快速交换具有二分树结构的结果和MAC报头重写信息的存储等几大特性。

- 由于路由表和快速缓存内容（例如，MAC 报头重写）之间不存在关联，生成缓存条目将涉及进程交换路径中必须完成的所有处理。所以，将在进程交换结束时生成快速缓存条目。
- 由于 ARP 缓存和快速缓存结构中的 MAC 报头（用于重写）之间不存在关联，因此当 ARP 列表更改时，快速缓存的某个部分必须失效（通过数据包进程交换重新创建）。
- 快速缓存只能为路由表内部的任何特定目的地建立某种深度的条目(某种前缀长度)。
- 没有办法从一个条目指向到快速缓存内的另一个条目(MAC报头和输出接口信息应该在节点之内)，因此快速缓存条目建立时，必须解决所有路由递归。换句话说，不能在快速缓存自身内部解析递归路由。

老化快速交换条目

为了让快速交换条目保持与路由表及 ARP 缓存的同步，并防止快速缓存中的未使用条目过度消耗路由器内存，1/20 的快速缓存会每分钟随机失效。如果路由器内存下降到非常低的水印下面，每分钟有1/5的 快速缓存条目失效。

快速交换前缀长度

如果每个目的地只能构建到一个前缀长度，那么快速交换可以构建的条目前缀长度是多大？在快速交换术语中，目的地是路由表或主要网络内部的单个可达目的地。生成给定缓存条目时用来决定前缀长度的规则是：

- 如果生成快速策略条目，总是缓存到 /32。

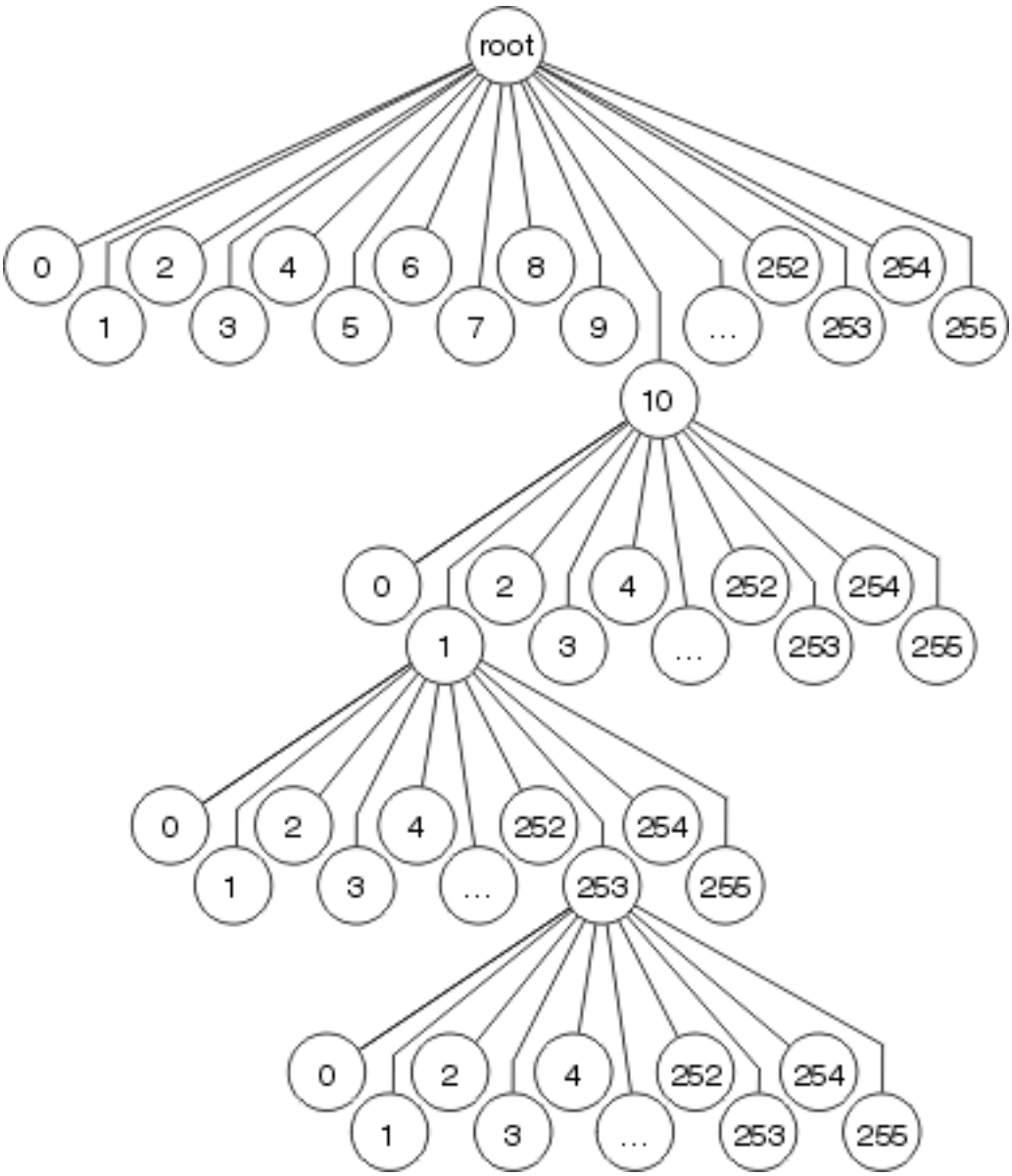
- 如果违反ATM虚拟电路上的多协议建立条目(MPOA VC) ，要一直缓存到/32。
- 如果网络未进行子网划分（是主网络条目）：如果直接连接，则使用 /32；否则将使用主网掩码。
- 如果是超网，则使用超网掩码。
- 如果网络已进行子网划分：如果直接连接，则使用 /32；如果有多个路径到此子网，则使用 /32；在其他所有情况下，使用此主网中最长的前缀长度。

负载共享

快速交换完全基于目标地址；负载共享按目标地址发生。如果有多条相等费用路径用于特定目的地网络，快速缓存有一个条目用于该网络内可及的每台主机，但所有数据流通过一条链路传输到特定主机。

最佳交换

最佳交换在256路多路树(256路mtree)中存储转发信息和MAC报头重写信息。 查寻前缀时，使用mtree减少必须采取的步骤的数量，如下图所示。



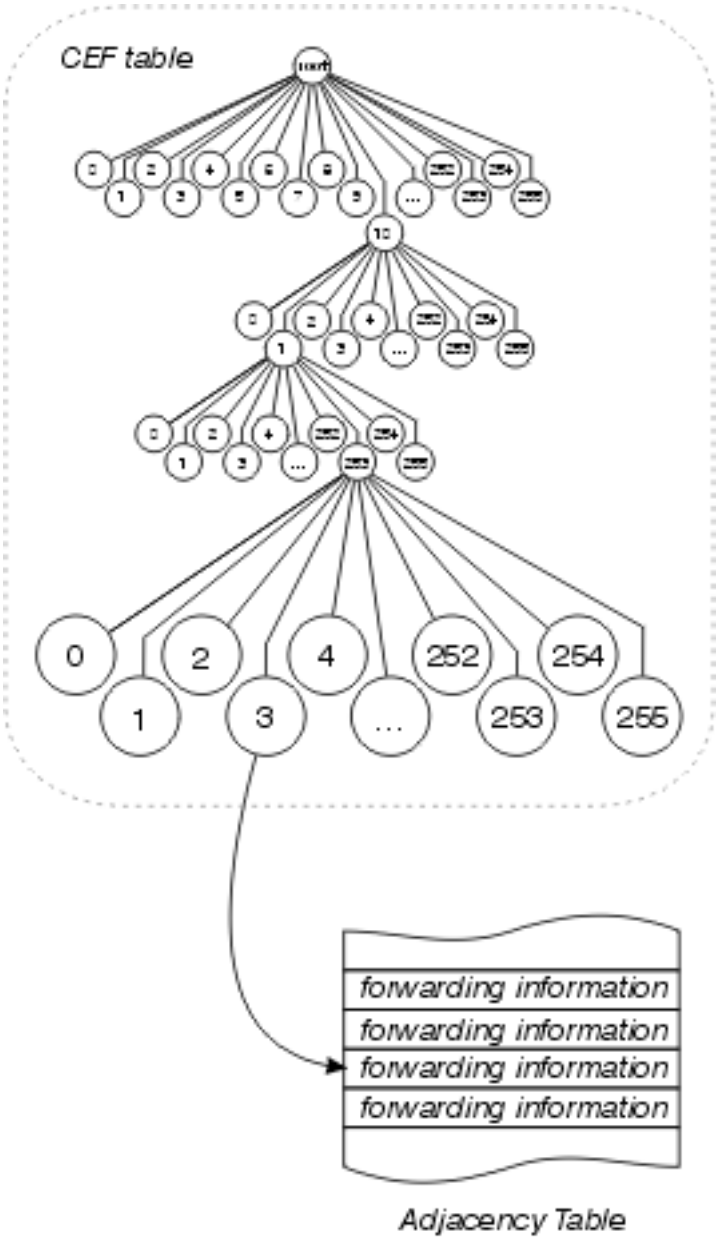
每个八位位组用于确定树的每个级别采取哪些256分组，这意味着找到任何目的地最多需要4次查找

。对于更短的前缀长度，可能只要求查找 1-3 次。MAC报头重写和输出接口信息存储为树节点的组成部分，因此缓存无效和过期仍然会在快速交换过程中发生。

最佳交换与快速交换一样，也确定每个缓存条目的前缀长度。

Cisco 快速转发

Cisco快速转发也使用256数据结构存储转发和MAC报头重写信息，但它不使用信息树。Cisco快速转发使用trie，意味着实际被搜索的信息不在数据结构中;相反地，数据保存在独立的数据结构中，并且trie仅仅是指向它。换句话说，与其在树内存储出局接口和MAC报头重写，Cisco快速转发不如在被称为邻接表的独立数据结构中存储此信息。



可达性信息(在Cisco快速转发表中)和转发信息的这种分离(在邻接表中)，具有某些好处：

- 邻接表可以分别从Cisco快速转发表构建，从而在不进程交换任何信息包的情况下能同时构建。
- 用于转发数据包的MAC报头重写不存储在缓存条目中，因此MAC报头重写字符串中的更改不要求缓存条目失效。
- 您可以直接指向转发信息，而不是递归下一跳，以解析递归路由。

根本上，消除了所有过期的高速缓冲存储器，并且高速缓冲存储器是根据包含在路由表和ARP缓存

的信息预先构建的。生成缓存条目时无需进行任何数据包的进程交换。

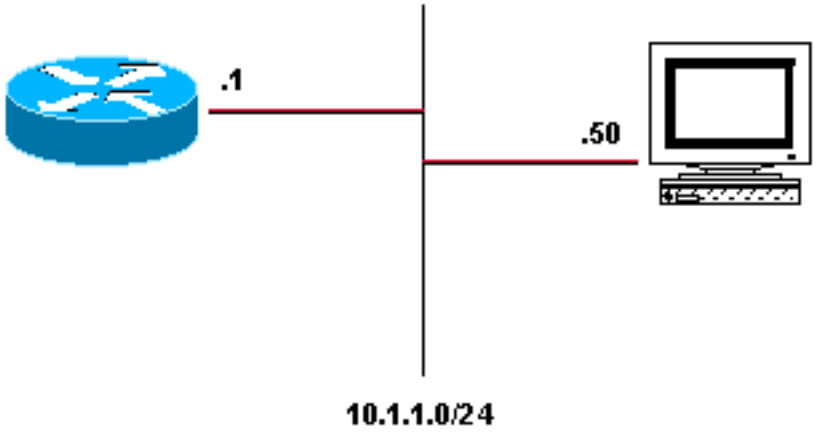
邻接表中的其他条目

邻接表可以包含除 MAC 报头重写字符串和出站接口信息之外的条目。在邻接表里可以安置的多种类型条目包括：

- **缓存 - 用于到达某个特定相邻主机或路由器的 MAC 报头重写字符串和出站接口。**
- **接收 - 路由器应该接收发往此 IP 地址的数据包。**这包括广播地址和配置在路由器上的地址。
- **丢弃 - 应该丢弃发往此 IP 地址的数据包。**这可以用于被访问控制列表拒绝或者被路由到NULL接口的数据流。
- **传送 - Cisco Express Forwarding 无法交换此数据包；**将其传递给下一个最佳交换方法（通常是快速交换），以便进行处理。
- **收集 - 直接连接下一跳，但是当前没有可用的 MAC 报头重写字符串。**

收集邻接

收集邻接条目表明，应该直接连接特定的下一跳，但没有 MAC 报头重写信息。如何生成和使用这些信息呢？默认情况下，运行Cisco快速转发并连接到广播网络的路由器(如下图所示）建立很多邻接表条目。



默认情况下生成的四个邻接表条目是：

```
10.1.1.0/24, version 17, attached, connected
0 packets, 0 bytes
  via Ethernet2/0, 0 dependencies
    valid glean adjacency
10.1.1.0/32, version 4, receive
10.1.1.1/32, version 3, receive
10.1.1.255/32, version 5, receive
```

请注意，有四个条目：三个接收和一个收集。每一个接收条目表示广播地址或路由器上配置的地址，而汇集条目则表示连接网络上的剩余地址空间。如果主机10.1.1.50收到信息包，那么路由器尝试交换它，并且查找此是否可以解决收集邻接。然后Cisco快速转发发出信号，10.1.1.50需要一个ARP缓存条目，ARP进程发送ARP信息包，并且相应的邻接表条目从新的ARP缓存信息中生成。此步骤完成后，邻接表有一个 10.1.1.50 条目。

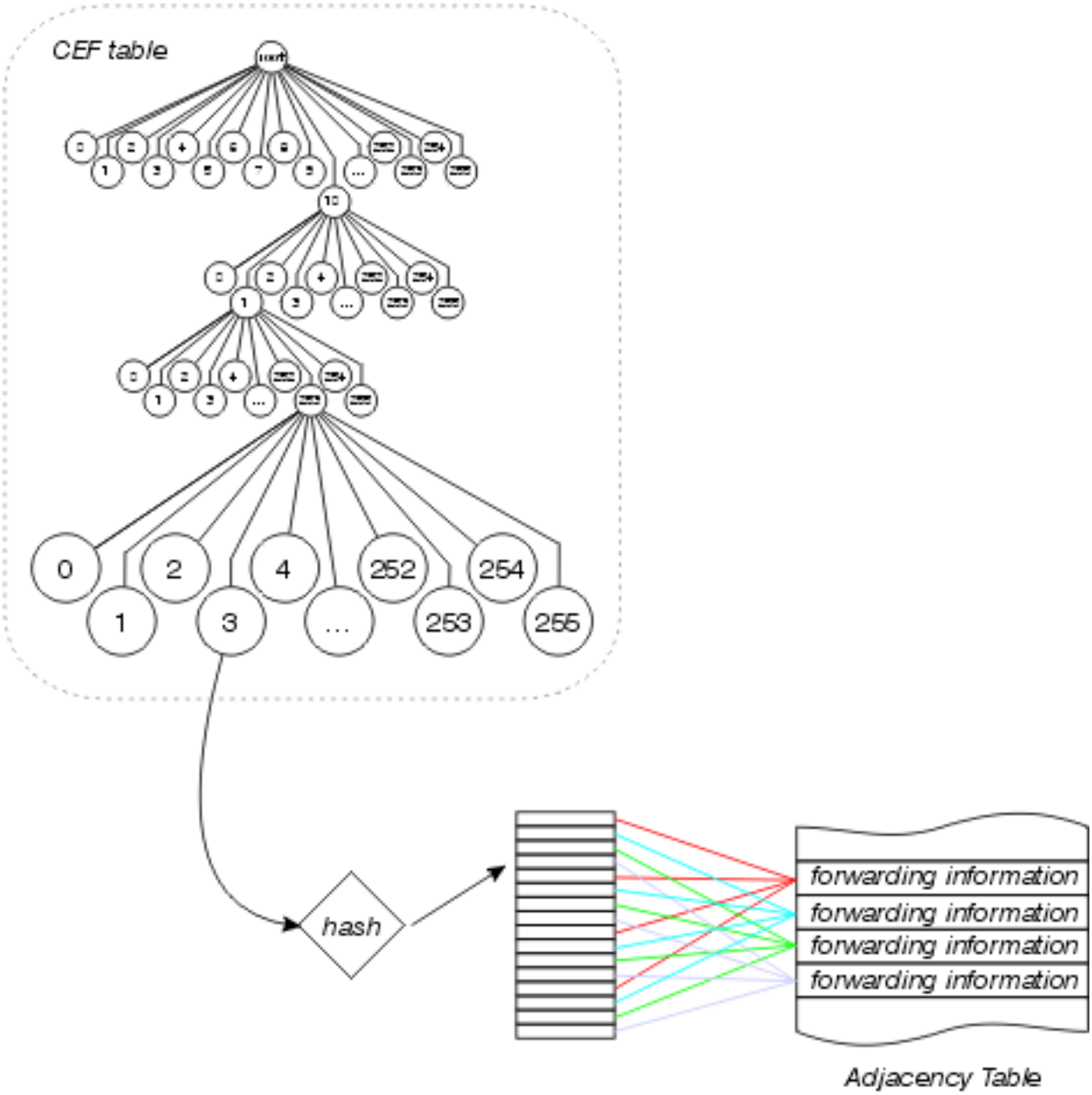
```
10.1.1.0/24, version 17, attached, connected
```

```
0 packets, 0 bytes
  via Ethernet2/0, 0 dependencies
    valid glean adjacency
10.1.1.0/32, version 4, receive
10.1.1.1/32, version 3, receive
10.1.1.50/32, version 12, cached adjacency 208.0.3.2
0 packets, 0 bytes
  via 208.0.3.2, Ethernet2/0, 1 dependency
    next hop 208.0.3.2, Ethernet2/0
    valid cached adjacency
10.1.1.255/32, version 5, receive
```

路由器下一个接收到的发往 10.1.1.50 的数据包将通过这个新的邻接进行交换。

负载共享

Cisco快速转发也利用Cisco快速转发表和邻接表之间的分离，提供比任何其它中断上下文交换模式更好的负载共享表。在 Cisco Express Forwarding 表和邻接表之间插入一个负载共享表，如下图所示：



Cisco快速转发表指向此加载共享表，其中包含指向可用并行路径的各种邻接表条目的指针。源地址和目的地址都通过Hash算法来确定每个信息包使用哪个负载分摊表条目。可配置每数据包负载共享

，在这种情况下，每个数据包各自使用不同的负载共享表条目。

每个加载共享表有16个条目，其中可用路径可以根据路由表中的流量共享计数器分开。如果路由表中的流量共享计数器的编号全部是1 (如同多条相等费用路径的情况)，每个可能的下一跳从负载共享表接收的指针数量相同。如果可用的路径数量不被平分成16 (因为有16个负载共享表条目)，那么一些路径将比其它路径有更多的条目。

从 Cisco IOS 软件版本 12.0 开始，负载共享表中的条目数量有所降低，这是为了确保每个路径都有适量的负载共享表条目。例如，如果路由表中有三个成本相等的路径，将只使用 15 个负载共享表条目。

哪条交换路径是最佳的？

只要可能，您希望您的路由器交换在中断上下文中进行，因为它至少比流程级交换更快速。Cisco Express Forwarding 交换确实比其他任何交换模式更快、更好。如果您使用的协议和IOS支持它，我们建议您使用“Cisco快速转发”。如果您用负载共享的一定流量的并行链路时特别如此。要确定支持 CEF 需要哪种 IOS，请访问 Cisco Feature Navigator（仅限注册用户）页面。

相关信息

- 如何验证 Cisco 快速转发交换
- 使用 Cisco 快速转发排除并行链路上的负载均衡故障
- 使用 CEF 执行负载均衡
- Cisco Express Forwarding 支持页
- IP 路由 支持页
- Cisco IOS 交换服务配置指南 12.1 版
- 技术支持 - Cisco Systems