在MPLS网络中跟踪路由

目录

简介

本文档介绍多协议标签交换(MPLS)网络中的互联网控制消息协议(ICMP)跟踪路由行为，以及与LSP跟踪的快速比较。

背景信息

在IP环境中，任何节点在收到数据包时，如果生存时间(TTL)过期，它都会生成“TTL Exceeded” ICMP错误消息（类型=11，代码=0），并将其发送到数据包源地址。利用此概念，通过从1开始顺序发送TTL为UDP数据包，跟踪从源到目的地的IP路径。可以注意，此功能的基本要求是：

- 数据包的源地址可从传输节点到达
- ICMP不沿路径过滤

在MPLS环境中，传输提供商LSR可能并不总是具有到源地址的可达性，并且需要对MPLS域中的ICMP处理进行一些增强。

MPLS网络中的ICMP跟踪路由

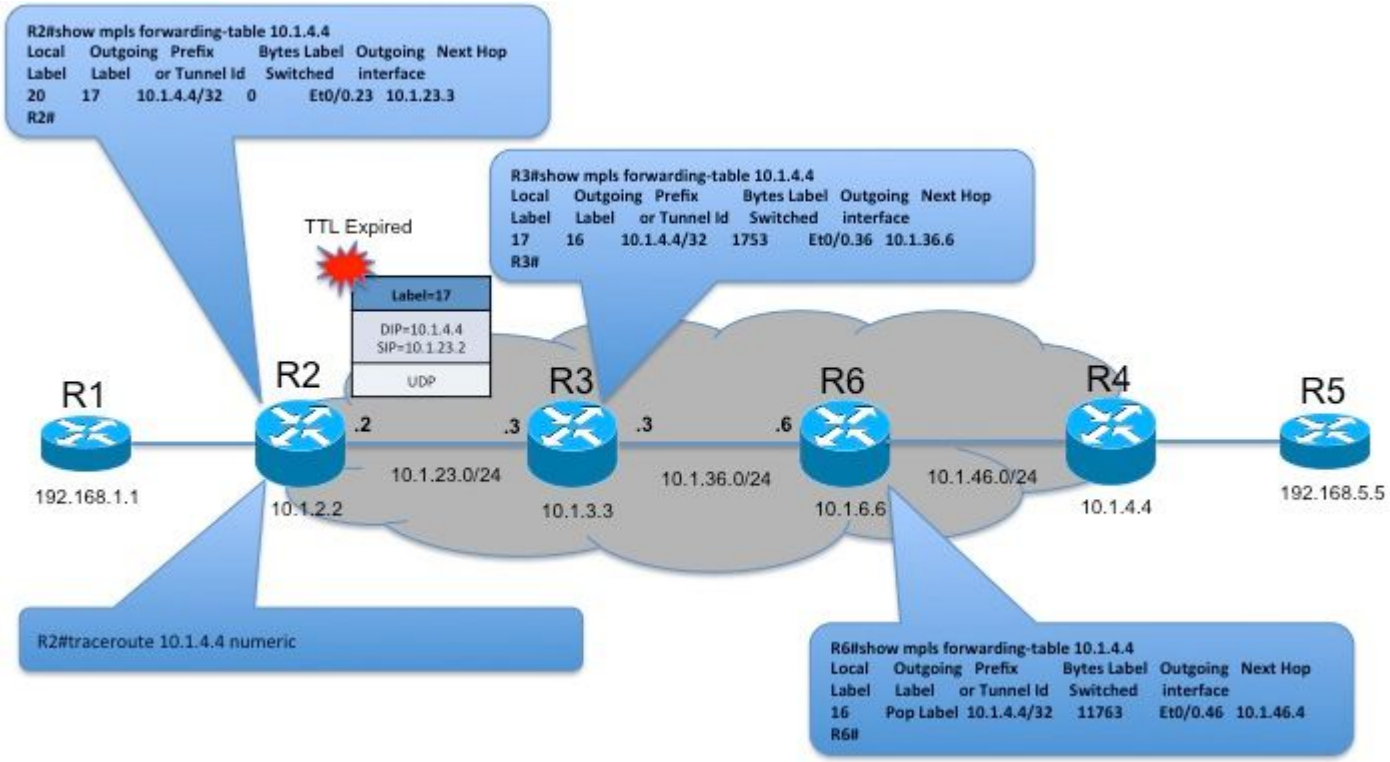
任何LSR在接收顶部标签上TTL=1的数据包时的默认行为遵循丢弃数据包并触发ICMP错误消息的传统IP行为。为了将ICMP消息路由到源，LSR将执行以下操作：

- 缓冲来自传入数据包的标签堆栈（以TTL=1接收的数据包）
- 生成ICMP错误消息，源作为其自己的地址，目的地作为来自接收数据包的源地址。
- 将标签堆栈底部（在步骤1之前缓冲）的所有标签附加为TTL=255（顶部标签除外）。
- 从缓冲的标签堆栈获取顶部标签并执行本地LFIB查找，以获取要交换的标签和关联的下一跳。
- 将新标签附加到TTL=255的堆栈顶部，然后通过发送。

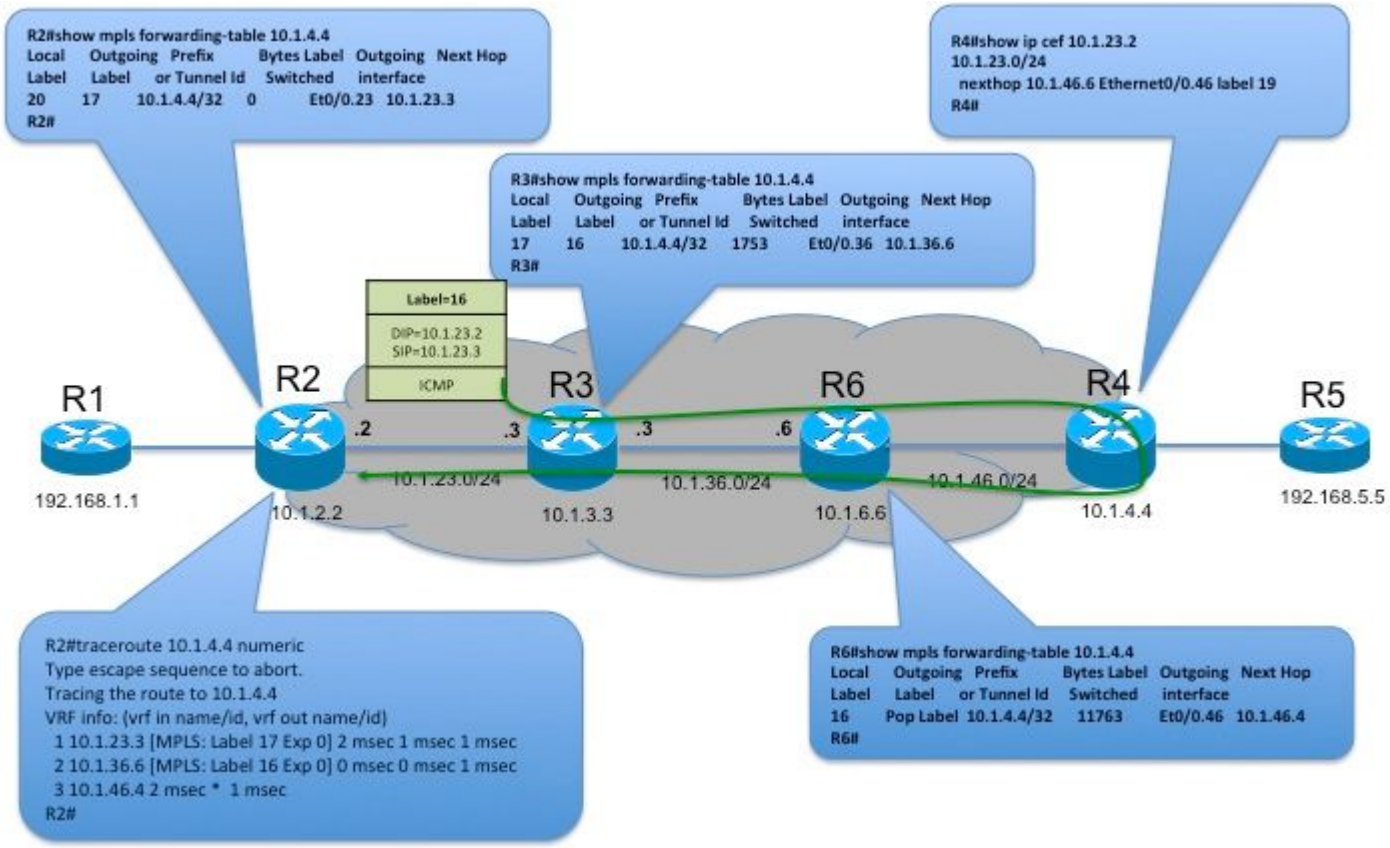
使用此方法，ICMP错误消息会从中转LSR遍历到出口LER，然后返回到入口LER到实际源。

从PE触发到远程PE的ICMP跟踪

以下是一个简单示例，解释当ICMP跟踪从PE触发到同一MPLS域内的远程PE时的行为：



在此拓扑中，当从R2触发ICMP跟踪路由到10.1.4.4时，第一个数据包的TTL为1。R3在接收数据包时会将TTL递减到0并触发ICMP生成机制。



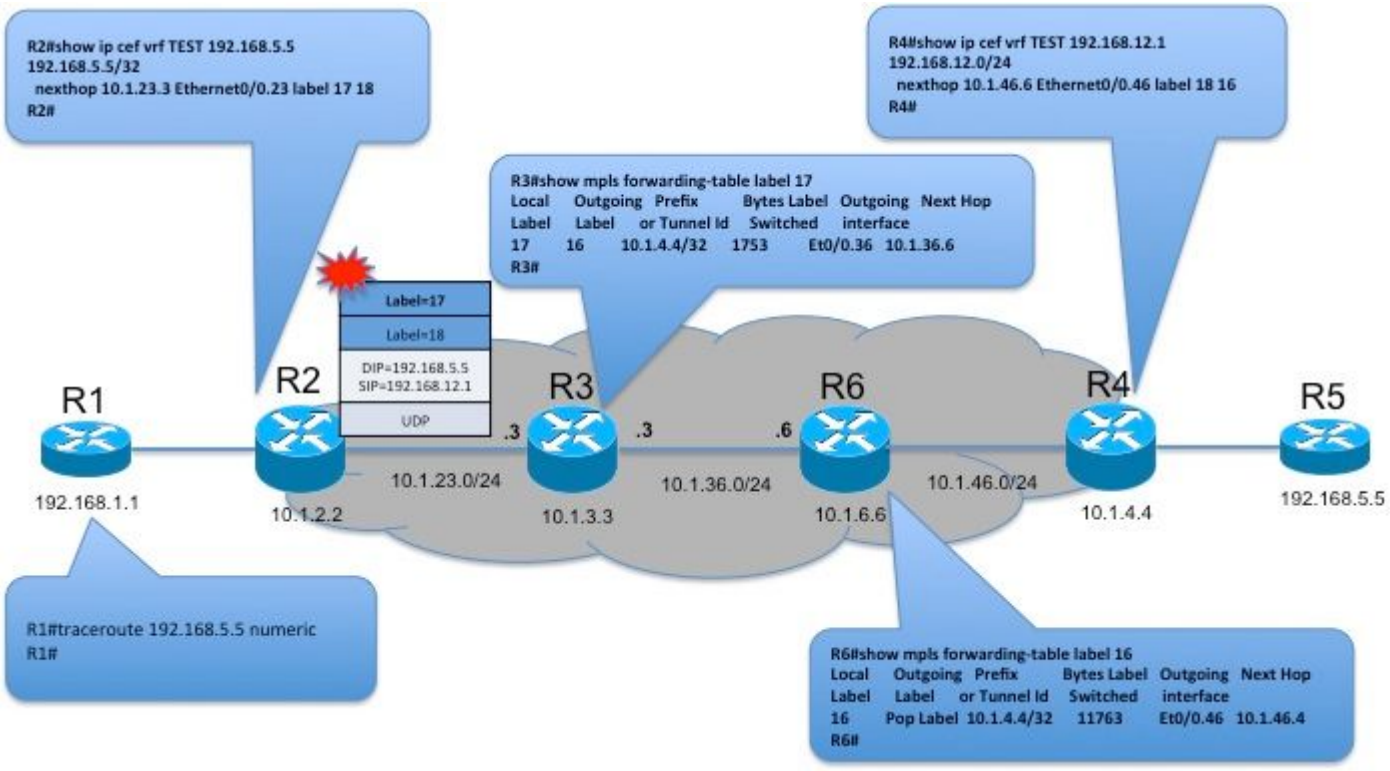
R3将缓冲标签堆栈并生成ICMP错误消息，并在ICMP负载中将来自缓冲区的传入标签堆栈包括在缓冲区中。它还使用来自已标记数据包的传入接口的源地址填充IP报头，目的地址作为已标记数据包的源地址。TTL设置为255。现在它从缓冲区推送标签堆栈，并查询LFIB表以在顶部标签上执行转发操作。在此拓扑中，收到的标签堆栈是17。在LFIB表中执行查找时，标签17与标签16交换，并转发

到下一跳R6。R6将弹出顶部标签，转发到R4,R4将IP将数据包转发回R2。

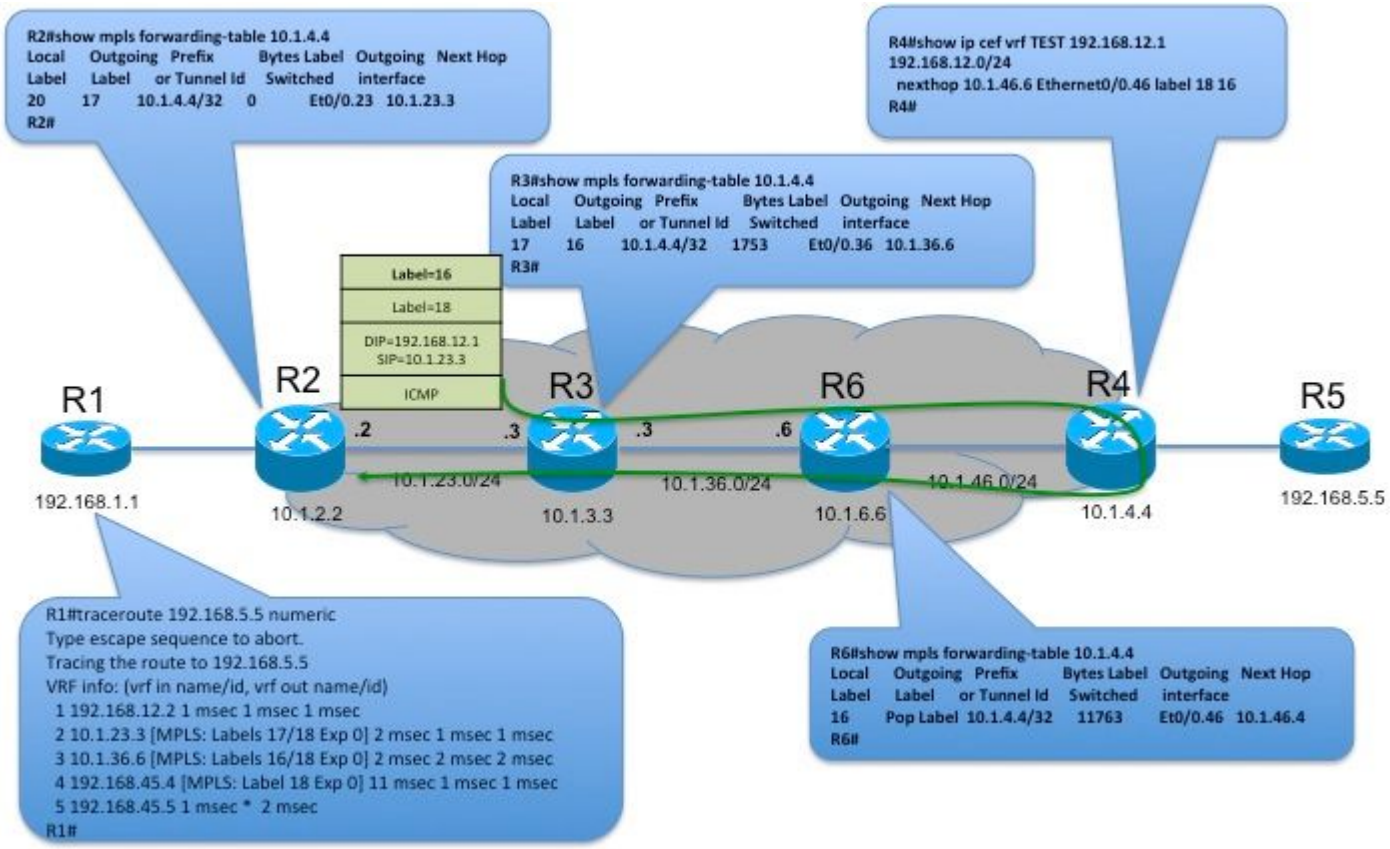
如R2的traceroute输出中所述，传入标签将按路径上的每一跳列出。

从CE触发到远程CE的ICMP跟踪

以下是一个简单示例，解释当ICMP跟踪通过MPLS域从CE触发到远程CE时的行为：



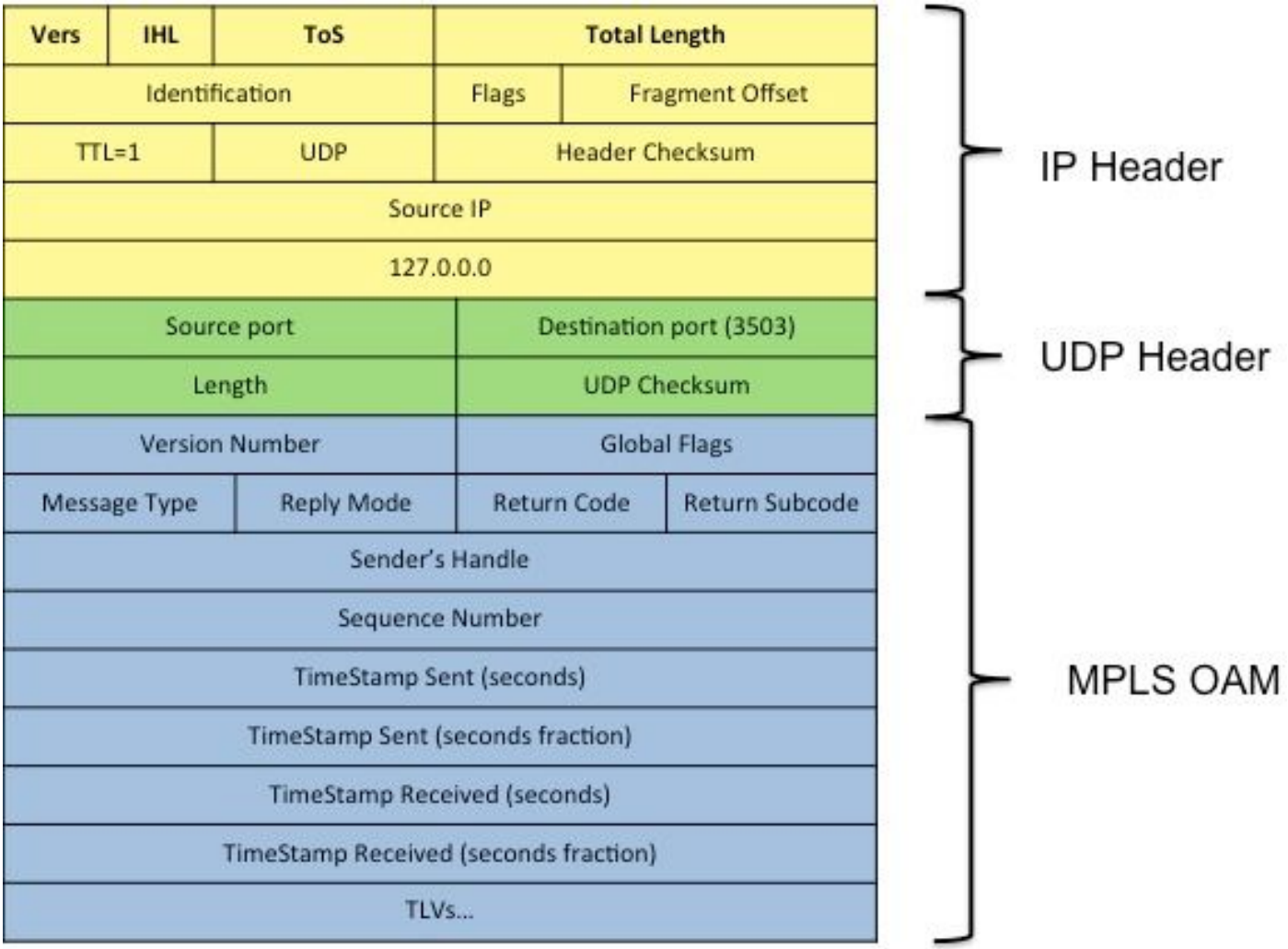
在此拓扑中，当从R1(CE)触发ICMP跟踪路由到192.168.5.5（远程CE）时，第一个数据包的TTL为1。这是正常的IP数据包，因此R2遵循生成ICMP并直接发送到R1的传统行为。第二个数据包的TTL=2在R3过期。



R3将缓冲标签堆栈并生成ICMP错误消息，并在ICMP负载中将来自缓冲区的传入标签堆栈包括在缓冲区中。它还使用来自已标记数据包的传入接口的源地址填充IP报头，目的地址作为已标记数据包的源地址。TTL设置为255。现在它从缓冲区推送标签堆栈，并查询LFIB表以在顶部标签上执行转发操作。在上述拓扑中，收到的标签堆栈是{17, 18}。在LFIB表中为顶部标签执行查找时，17将与标签16交换，并转发到下一跳R6。R6将弹出顶部标签并转发到R4。R4将使用VRF标签识别VRF并将数据包转发回R1。

如R1的traceroute输出中所述，传入标签堆栈按路径上的每一跳列出。

MPLS网络中的MPLS LSP跟踪路由

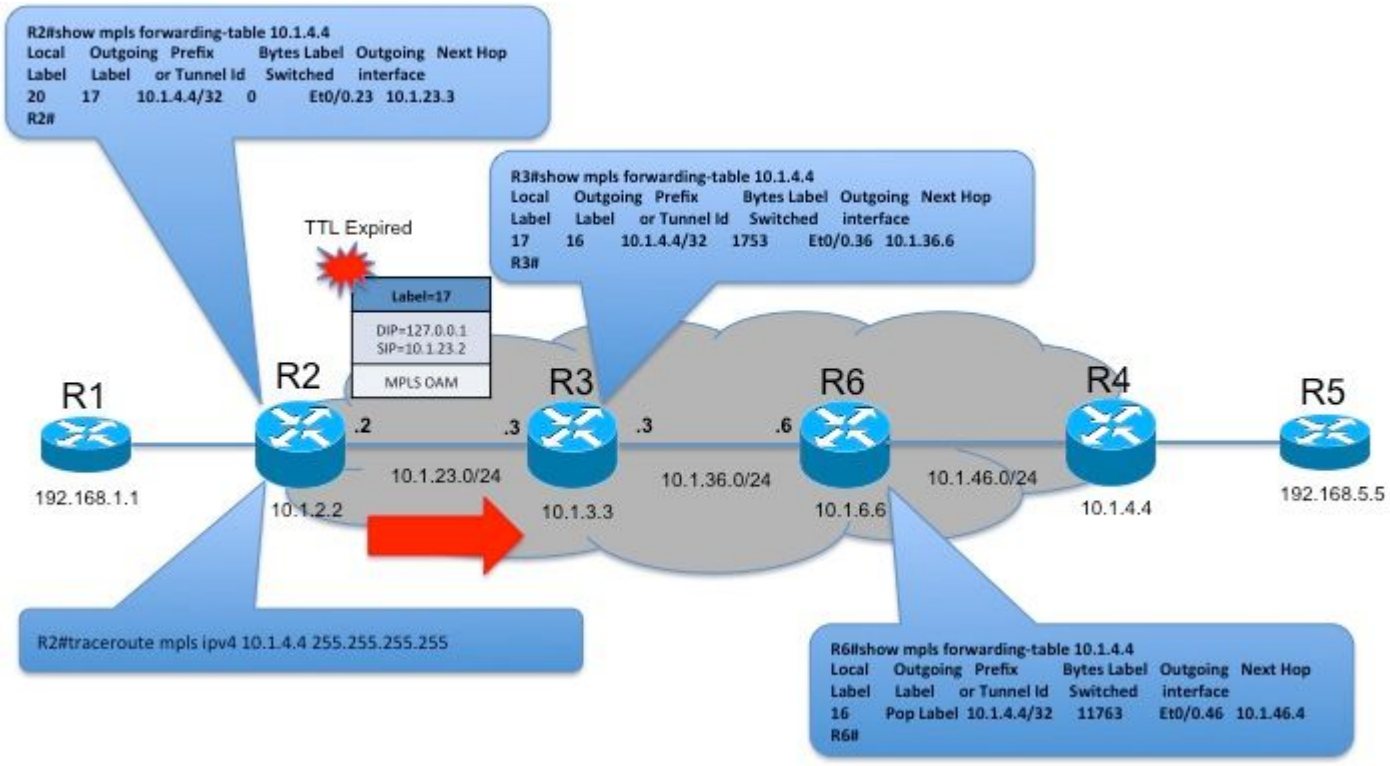


与基于ICMP的traceroute不同，LSP traceroute使用RFC4379中定义的机器。它使用IP/UDP封装，请求的目的地址设置为环回地址(127.0.0.0/8范围)。 预期LSP Ping将在同一MPLS域内触发，因此应答将直接发送给发起方。

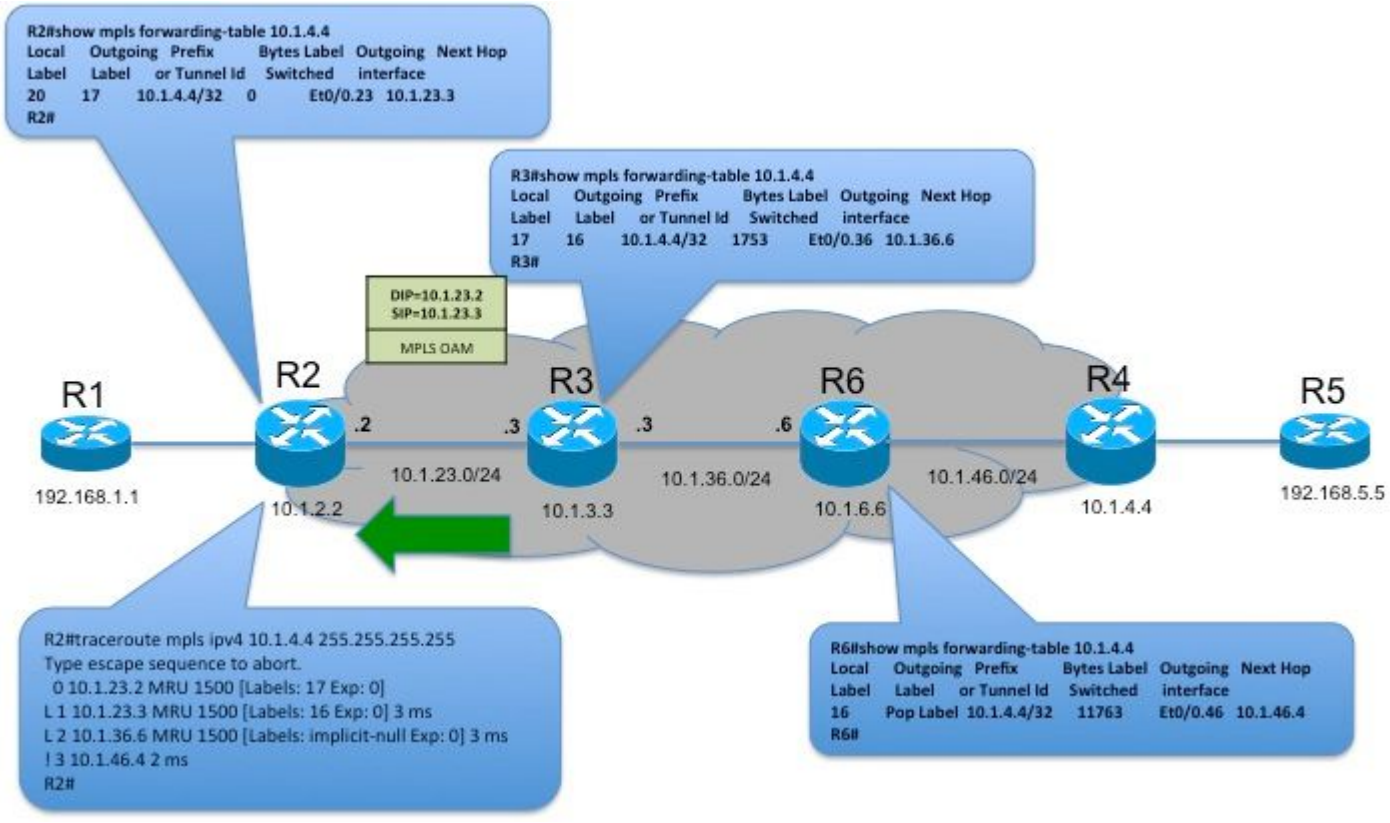
当从任何LSR触发LSP跟踪路由(“traceroute mpls ipv4 <FEC>”)时，有关要验证的FEC的详细信息将作为MPLS回应请求中的“目标FEC堆栈”包含在TLV中。此消息将从1开始按顺序在标签堆栈上以TTL发送。接收数据包时的任何传输LSR以及如果TTL过期，将处理IP数据包，因为目的地址是环回地址。并发送到CPU以进行MPLS OAM处理。

响应方可选择从收到的MPLS回应请求的标签堆栈获取标签，并从目标FEC堆栈TLV获取FEC详细信息，以根据本地控制平面信息验证标签堆栈，从而执行FEC验证。在跟踪情况下，响应器将在TLV中包括下行信息，如传出标签和下行邻居地址等，作为下行映射(DSMAP)TLV。（DDMAP将弃用DSMAP，因为它比DSMAP更灵活）。

从PE触发到远程PE的LSP跟踪



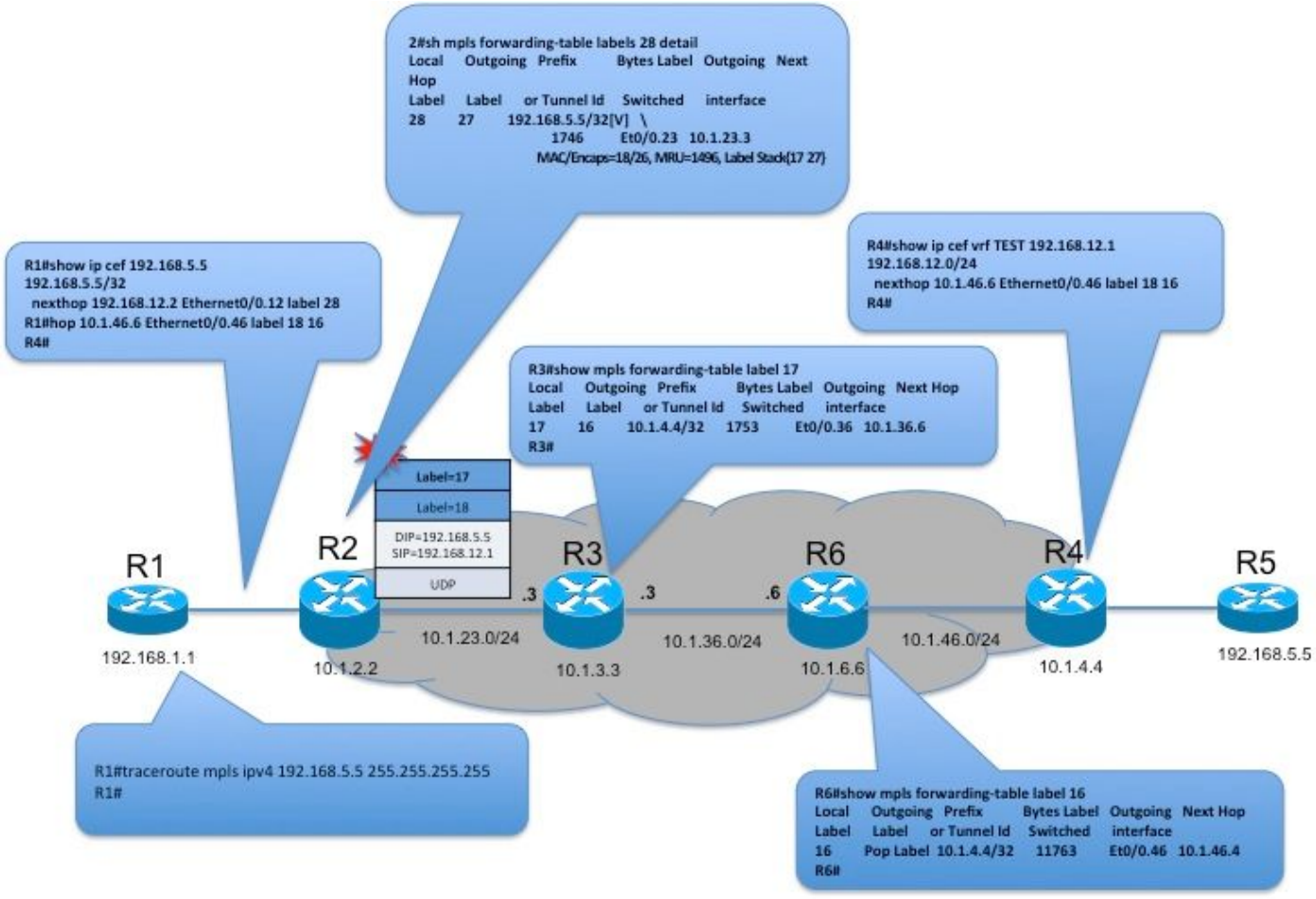
在此拓扑中，从R2触发LSP跟踪以验证LSP到前缀10.1.4.4/32。标签上的TTL将从1开始设置。R3在收到LSP后，将发送到CPU进行OAM处理。



R3将使用带有传出标签16的DSMAP TLV的MPLS回应应答和下游邻居详细信息等附加信息回复R2。与ICMP消息不同，MPLS回应应答将从响应方R3直接转发到发起方R2。

如R2的LSP traceroute输出中所述，传出标签堆栈将按路径中的每一跳列出。这与基于ICMP的traceroute不同，在后者中，输出中列出的标签将是传入标签堆栈。

从CE触发到远程CE的LSP跟踪

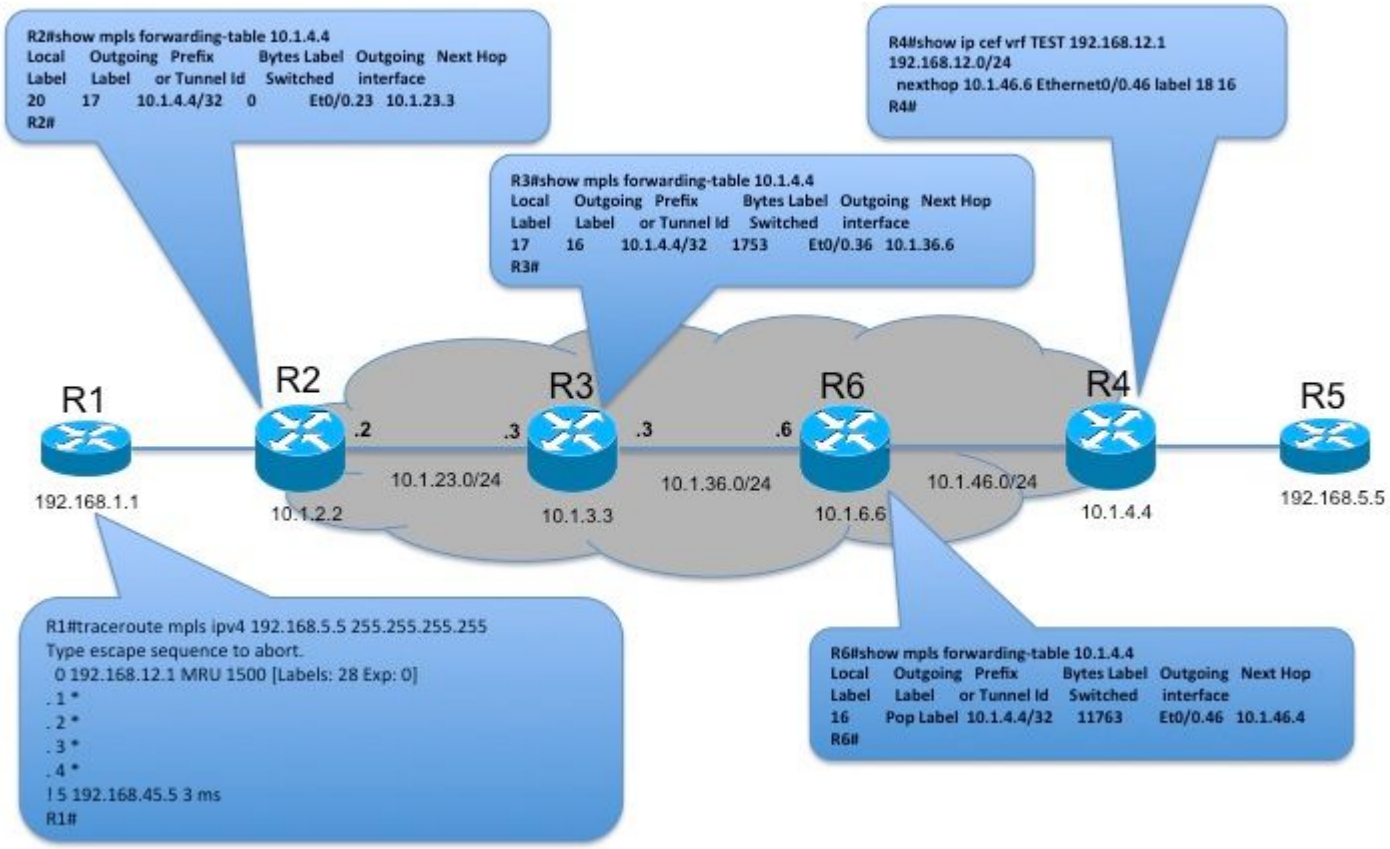


这适用于PE-CE之间启用MPLS的CSC类场景。在通过运营商MPLS域执行从CE到远程CE的LSP跟踪时，存在以下2个挑战：

- LSP回应应答将直接发送给发起方。因此，响应方必须具有到发起方的可达性。在上述拓扑中，R3可能没有到达R1的可达性，因为它位于VRF中。
- 对于标签堆栈中的每个标签，目标FEC堆栈中应包含相关的FEC详细信息以进行验证。发起方包含的FEC为1，而PE将推送2个标签。在上述拓扑中，R1发送FEC={192.168.5.5/32}的MPLS回应请求，并在堆栈中包含标签28。由于R2将标签28与{17, 27}交换，因此R3将在堆栈中接收带2标签的请求，而在TLV混淆FEC验证中接收1个FEC。

RFC6424定义了“FEC堆栈更改TLV”的概念，以解决问题2。此TLV将作为PUSH/POP包含在相关FEC的应答中，发起方可以在后续回应请求中包含该TLV。

draft-ietf-mpls-lsp-ping-relay-reply定义了在TLV中传输中继节点地址堆栈的概念，响应方可使用该堆栈来中继响应，即使它无法到达发起方。



Cisco IOS®当前不支持这2个问题，因此从CE到远程CE的LSP跟踪将仅列出入口PE和远程CE。这仅仅是为了完整性。

相关信息

- RFC 3032
- RFC 4379
- RFC 6424