

在 BGP-4 中携带标签信息

RFC 3107

陶志豪

zhihao.tao@outlook.com

<https://github.com/netwiki/share-doc>

Request for Comments: 3107
Category: Standards Track

Juniper Networks
E. Rosen
Cisco Systems, Inc.
May 2001

Thank Zhihao Tao for your hard work in Translation. The translator spent countless nights and weekends, using his hard work to make it convenient for everyone.

If you have any questions, please send a email to zhihao.tao@outlook.com

在 BGP-4 中携带标签信息

备忘录状态

本文档为互联网社区规定的互联网标准化协议，并请讨论和建议来改进。

请参考当前版本的“互联网官方协议标准”（STD 1）和该协议的状态。此备忘录的传播不受限制。

版权声明

版权所有（C）互联网协会（2001）。

概述

本文档规定了用于特定路由的标签映射信息的方式，他们搭载在同一边界网关协议（BGP）更新消息中，用于分发本身路由。当 BGP 用于分发特定路由时，也可以用于分发多协议标签交换（MPLS）映射到该路由标签。

1. 要求规范

关键词“必须”，“不得”，“所需”，“已”，“不”，“应该”，“不应该”，“推荐”，“可能”和“可选”在文档[RFC2119]中所述。

2. 概述

本文档规定了用于特定路由的标签映射信息的方式，他们搭载在同一边界网关协议（BGP）更新消息中，用于分发本身路由。当 BGP 用于分发特定路由时，也可以用于分发多协议标签交换（MPLS）映射到该路由标签。

这可以在以下情况下是有用的：

- 如果两紧邻的标签交换路由器（LSR）也是 BGP Peer，那么标签分配可以不需要任何其他标签分配协议。
- 假设一个网络分为两级 LSR：外部 LSR，接入其他网络的；内部 LSR，只是与外部 LSR 之间流量服务。假设外部 LSRs 同时也是 BGP speakers。如果 BGP speakers 在向其他 BGP 邻居发送的路由更新中顺便携带上为该路由分配的 MPLS 标签，并且内部路由器也支持 MPLS，那么他们可以不需要从 BGP speaker 接收任何 BGP 路由。

如果外部路由器 A 需要发送一个数据包到目的地 D，而 A 到 D 的 BGP 下一跳是外部路由器 B，而 B 已经将分配标签 L 到 D，然后 A 首先 PUSH L 到数据包的标签栈。然后查看 IGP 找到去往 B 的下一跳为 C。如果 C 已经分发 MPLS 标签 M 到 B，POSH M，然后发送包到 C。

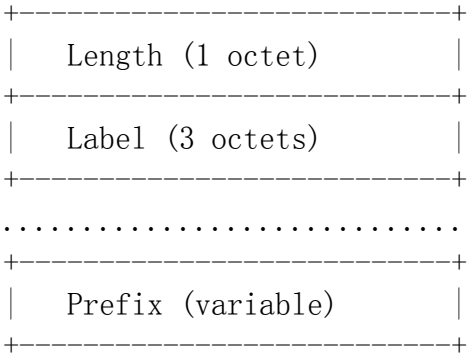
如果一组 BGP speakers 通过路由反射器[BGP-RR]交换路由，路由分发时捎带标签分配（通过路由反射器分配标签是非常好的行为）。这将非常明显的提升扩展性。如果路由反射器不在转发路径中，则它没有转发 MPLS 数据包的能力。

标签分配可以在 BGP 更新消息捎带（BGP-4 协议扩展属性[RFC 2283]）。标签为 NLRI 属性的一些字段，SAFI（“子地址族标识符”）字段表明包含一个标签。BGP speaker 可能不使用 BGP 发送标签到一个特定的 BGP 对等体，除非 peer 表示通过 BGP 能力通告，其能够处理 UPDATE 消息中特定 SAFI 信息。

3. 传递标签映射信息

标签映射信息是作为网络层可达性信息部分（NLRI）的多协议扩展属性。和之前一致 AFI 表示关联路由的地址族。SAFI 为 4 时表示 NLRI 中包含标签。

网络层可达性信息被编码为一个或多个元组的形式<长度、标签、前缀>，其字段描述如下：



这些字段的用法和含义如下：

- a) 长度：
长度字段表示地址前缀加上标签的位长度。
- b) 标签：
标签域携带一个或多个标签（对应于标签堆栈[MPLS-ENCAPS]）。每个标签编码为3个字节，其中高20位包含标签值，低阶位包含“栈底”（如[MPLS-ENCAPS]定义）。
- c) 前缀：
前缀字段包含地址前缀由足够的尾位的字节边界（8 字节宽）结束。注意，尾位的值是不相关的。

一个特定的路由（和它的地址前缀关联）指定的标签必须由 LSR 分配，其根据路由属性中下一跳的值辨识。

当一个 BGP speaker 重新分发路由时，分配给该路由的标签不能改变（除去省略），除非 speaker 改变路由下一跳的属性值。

BGP 路由器可以撤回先前的路由（以及该路由和标签之间的绑定），通过（a）公告新路由（和标签）具有和先前的路线相同的 NLRI，或（b）在 UPDATE 的撤销路由字段中列出先前的公告 NLRI 路由。携带在撤销路由中的标签信息（NLRI 部分功能）应设置 0x800000。（当终止 BGP 会话也会撤销之前公告的所有路由）。

4. 公告多路径到达目的地

BGP speaker 可以维护不止一条到指定的目的地路由（并向其同行发布公告），每一条路由都有自己的标签。

在 BGP speaker 公告多条到目的地路由，如果撤回一条路由，撤销时指定标签，则只有与相应标签的对应路由被撤销。

如果一条路由被撤销，在撤销时没有指定标签，只有相应没有标签的路由被撤销；有标签的路由被保留。

5. 公告能力

BGP speaker 使用多协议扩展携带标签映射信息，应该应用能力可选参数（[BGP-CAP]中定义），告知其同行这种能力。MP_EXT 能力代码，如[BGP-MP]的定义，是用来做公告（AFI，SAFI）对在一个特定的连接可用。

BGP speaker 不应该将此能力公告到另一个 BGP speaker，除非两个 speaker 之间存在标签交换路径（LSP）。

BGP speaker 能够处理多条到目的地的路由（正如上面描述的），应使用能力可选参数，告知其同行这种能力。此能力的值是 4。

6. BGP peer 不直接相邻时

考虑下面的 LSR 的拓扑结构：A—B—C—D。假设 D 给 A 分配一个标签 L。在这种拓扑结构中，A 不能简单地 PUSH 标签 L 到数据包的标签栈中，然后将产生的数据包发到 B，D 必须是唯一看到 L 在栈顶的 LSR。在数据包从 A 发到 B 之前，它必须 PUSH 上另一个标签，这是由 B 发配的。B 必须用另一个标签 SWAP 这个标签，这是由 C 分配的。换句话说，A 和 D 之间必须存在 LSP。如果不存在这样的 LSP，A 不能使用标签 L。A 不能使用的标签 L。这是事实，任何时候标签在非相邻的 LSRs 之间分发时，无论标签是否通过 BGP 或其他一些方式完成了分发。

本文件不指定任何程序步骤，确保在实际情况下，非相邻 LSRs 间的标签分配，只适用于适当的 MPLS 基础设施存在于网络或两个网络连接 LSRs 时。确保适当的基础设施存在是网络管理和运行的问题。

7. 安全考虑

当 LSR A 通过点到点接口直接连接到 LSR B 时，那么当 A 通过该接口接收数据包时，它知道他们来自 B。这使 A 很容易的丢弃，任何来自 B 的栈顶标签不在 A 分配给 B 的标签中的数据包。也就是说，A 可以很容易地确保 B 只能使用它有权使用的标签。可以使用这种技术防止“标签欺骗”，即 LSR 强加一个尚未正确分发给它的标签的情况。

本文中讨论的处理，通常标签分发对等体是分离，且不仅仅使用点到点链路，而是使用 MPLS 网络。这意味着当 LSR A 处理携带标签的数据包，它真的没有办法确定哪个 LSR B PUSH 栈顶标签。因此，它无法判断该标签是否是 B 有权使用的标签。其实，当路由反射器正在使用中，A 甚至可能不知道接收其标签映射的一组 LSR。所以前面防止标签欺骗的技巧不适用。

尽管使用其他技术来避免标签欺骗问题也是可能的。例如，假如从来不接受来自网络“外部”接口的携带标签的数据包，以及所有通过 IBGP 分发的 IBGP 分布式标签，而没有办法让不信任路由器将携带标签的数据包进入网络。普遍地，假定某个 IBGP 对等体（或某个路由反射器的 IBGP 对等体）不会尝试标签欺骗，因为它们都在控制单一管理。

实际上这种情况可以明显减弱。不需要拒绝接受来自外部接口所有携带标签的数据包。只需要确保任何在外部接口上接收的携带标签的数据包，有一个实际分配给该接口栈顶标签。

那么标签欺骗问题只会存在于在同一个接口信任和不信任的系统两者之间。一种避免这个问题方式是避免这种情况。

Zhihao Tao