

BGP-4 的多协议扩展

RFC 4760

陶志豪

zhihao.tao@outlook.com

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

Network Working Group
Request for Comments: 4760
Obsoletes: 2858
Category: Standards Track

T. Bates
Cisco Systems
R. Chandra
Sona Systems
D. Katz
Y. Rekhter
Juniper Networks
January 2007

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

BGP-4 的多协议扩展

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（2007）。

概述

本文档定义了 BGP-4 的扩展，使其能够携带多种网络层协议的路由信息（如 IPv6，IPX，L3VPN 等）。扩展是向后兼容的 - 支持扩展的路由器可以与不支持扩展的路由器互操作。

1. 介绍

BGP-4[BGP-4]承载的三条 IPv4 特有的信息 (a) NEXT_HOP 属性 (表示为 IPv4 地址), (b) AGGREGATOR (包含 IPv4 地址), (c) NLRI (表示为 IPv4 地址前缀)。本文假设任何 BGP speakers (包括本文档中定义的支持多协议能力) 必须具有 IPv4 地址 (将在 AGGREGATOR 属性中使用)。因此, 使 BGP-4 支持多个网络层协议的路由, 只有两件事必须要添加到 BGP-4 是 (a) 将特定网络层协议与下一跳信息相关联的能力, (b) 将特定的网络层协议与 NLRI 相关联的能力。识别个别网络层协议相关联的下一跳信息和 NLRI 语义, 本文档使用地址族 (如[IANA-AF]) 和子地址族 (如本文件所述) 相结合的方式。

可以进一步观察下一跳信息 (由 NEXT_HOP 属性提供的信息), 仅与可达目的地公告结合使用是有意义的 (和必要) - 而与不可达目的地公告结合使用 (从服务中撤销的路由), 下一跳信息毫无意义。这表明公告的可达目的地应与公告的下一跳组合用于这些目的地址的, 而那些公告的可达目的地应与公告的不可达目的地分离使用。

提供向后兼容性, 以及将多协议能力简单地引入 BGP-4, 这样文档使用两个新属性, 即多协议可达性 NLRI (MP_REACH_NLRI) 和多协议不可达 NLRI (MP_UNREACH_NLRI)。MP_REACH_NLRI 用于携带一组可达目的地与下一跳信息一起用于转发到这些目的地。MP_UNREACH_NLRI 用于携带一组不可达目的地。这两个属性是可选的, 不可传递的。这样一个不支持多协议功能 BGP speaker 将会忽略这些属性携带的信息, 且不会传递给其他 BGP speaker。

2. 要求规格

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。

3. 多协议可达 NLRI - MP_REACH_NLRI (类型代码 14) :

这是一个可选的非传递属性, 可以用于以下用途:

- (a) 公告可达路由到一个 peer
- (b) 允许路由器公告路由器网络层地址, 通过在 NLRI 信息的 MP_REACH 属性中列出, 用于到达目的地址的下一跳。

属性编码如下所示:

+	-----	+
	Address Family Identifier (2 octets)	
+	-----	+
	Subsequent Address Family Identifier (1 octet)	
+	-----	+
	Length of Next Hop Network Address (1 octet)	
+	-----	+
	Network Address of Next Hop (variable)	
+	-----	+

Reserved (1 octet)	
+-----+	+-----+
Network Layer Reachability Information (variable)	
+-----+	+-----+

这些领域的用途和含义如下：

地址家族标识符（AFI）：

该字段与子地址族标识符字段相结合标识要到达的下一跳字段中携带的地址必须属于哪一组网络层协议，下一跳的地址被编码的方式，以及接下来网络层可达性信息的语义。如果下一跳被允许来自不止一个网络层协议，下一跳的编码必须提供一种确定其网络层协议的方法。

当前定义的地址族标识符字段值是在 IANA 的地址族号注册表[IANA-AF]中指定的。

子地址族标识符（SAFI）：

该字段与地址族标识符字段相结合标识要到达的下一跳字段中携带的地址必须属于哪组网络层协议，下一跳的地址被编码的方式，以及接下来网络层可达性信息的语义。如果下一跳被允许来自不止一个网络层协议，下一跳的编码必须提供一种确定其网络层协议的方法。

下一跳网络地址长度：

一个 1 字节的字段，其值表示“下一跳的网络地址”字段的长度，以八位字节为单位。

下一跳网络地址：

包含在目的地系统路径上下一个路由器的网络地址的可变长度字段。该网络层协议与网络地址下一跳相关联，由携带在属性中<AFI, SAFI>的组合识别。

保留：

必须将 1 个八位字节的字段设置为 0，并且收到后应该被忽略。

网络层可达性信息（NLRI）：

可变长度字段，列出这个属性中正在被公告的可行路由的 NLRI。NLRI 的语义由属性中携带的<AFI, SAFI>组确定。

当子地址族标识符字段设置为本文档中定义的值之一，每个 NLRI 都是按照文件“NLRI 编码”部分中描述的编码。

MP_REACH_NLRI 路径属性中携带的下一跳信息定义了路由器的网络层地址，其应该用作 UPDATE 消息的 MP_NLRI 属性中列出的目的地的下一跳。

下一跳信息的规则与在 NEXT_HOP BGP 属性中携带的信息（参见[BGP-4]第 5.1.3）的规则相同。

携带 MP_REACH_NLRI 的 UPDATE 消息也必须携带 ORIGIN 和 AS_PATH 属性（在 EBGp 和 IBGP 交换中）。此外，在 IBGP 交换这样的消息也必须携带 LOCAL_PREF 属性。

不包含 NLRI 的 UPDATE 消息，而不是 MP_REACH_NLRI 属性中的编码，不应携带 NEXT_HOP 属性。如果这样的消息包含 NEXT_HOP 属性，则 BGP speaker 接收消息应该忽略此属性。

UPDATE 消息不应包含相同的地址前缀（相同的<AFI, SAFI>）在以下多个字段中：WITHDRAWNROUTES 字段，网络可达性信息字段，MP_REACH_NLRI 字段和 MP_UNREACH_NLRI 字段。此形式的 UPDATE 消息的处理是未定义。

4. 多协议不可达 NLRI - MP_UNREACH_NLRI（类型代码 15）：

这是一个可选的非传递属性，可以用于从服务中撤出多条不可行路由的。

属性编码如下所示：

	Address Family Identifier (2 octets)	
	Subsequent Address Family Identifier (1 octet)	
	Withdrawn Routes (variable)	

这些领域的用途和含义如下：

地址家族标识符（AFI）：

该字段与子地址族标识符字段相结合标识要到达的下一跳字段中携带的地址必须属于哪一组网络层协议，下一跳的地址被编码的方式，以及接下来网络层可达性信息的语义。如果下一跳被允许来自不止一个网络层协议，下一跳的编码必须提供一种确定其网络层协议的方法。

当前定义的地址族标识符字段值是在 IANA 的地址族号注册表[IANA-AF]中指定的。

子地址族标识符（SAFI）：

该字段与地址族标识符字段相结合标识要到达的下一跳字段中携带的地址必须属于哪组网络层协议，下一跳的地址被编码的方式，以及接下来网络层可达性信息的语义。如果下一跳被允许来自不止一个网络层协议，下一跳的编码必须提供一种确定其网络层协议的方法。

撤销路由网络层可达性信息：

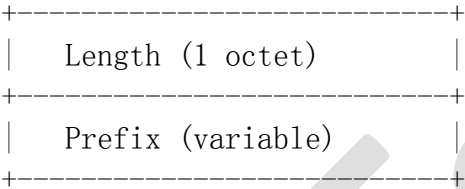
列出从服务中撤销的 NLRI 的可变长度字段。NLRI 的语义是由属性中携带的<AFI, SAFI>组合确定。

当子地址族标识符字段设置为本文档中定义的值之一，每个 NLRI 按照文件“NLRI 编码”部分中指定的编码。

包含 MP_UNREACH_NLRI 的 UPDATE 消息不需要携带任何其他路径属性。

5. NLRI 编码

网络层可达性信息被编码为一个或多个形式为<length, prefix>的 2 元组，其字段描述见下面：



这些领域的用途和含义如下：

- a) 长度：
长度字段指示地址前缀的长度（以位为单位）。长度为零表示前缀匹配所有（由地址族指定）地址（前缀本身，零字节）。
- b) 前缀：
前缀字段包含一个地址前缀，后跟足够尾随位，使字段的末尾落在八位字节的边界上。请注意，尾随位的值是无关紧要的。

6. 子地址族标识符

本文档为 MP_REACH_NLRI 和 MP_UNREACH_NLRI 属性中携带的子地址族标识符字段定义了以下值：

- 1 - 用于单播转发的网络层可达性信息
- 2 - 用于组播转发的网络层可达性信息

实现可以支持所有，一些或者没有本文档中定义的子地址族标识符值。

7. 错误处理

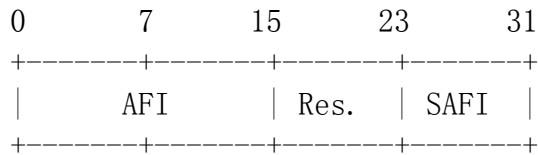
如果 BGP Speaker 从邻居收到一条包含 MP_REACH_NLRI 或 MP_UNREACH_NLRI 属性的 UPDATE 消息，如果 speaker 确定该属性不正确，speaker 必须删除从邻居接收的所有 BGP 路由，其 AFI/SAFI 与在不正确的 MP_REACH_NLRI 和 MP_UNREACH_NLRI 属性中携带的 AFI/SAFI 相同。在 BGP 会话的持续时间内接收到 UPDATE 消息，Speaker 应该忽略这个会话随后接收到的所有关于这个 AFI/SAFI 的路由。

此外，speaker 在收到 UPDATE 消息时可以终止 BGP 会话。Notification 消息代码/子代码指示“UPDATE 消息错误”/“可选属性错误”，会话应该被终止。

8. BGP 能力公告的使用

使用多协议扩展的 BGP speaker 应该使用能力公告程序[BGP-CAP]来确定, speaker 是否可以与特定的 peer 使用多协议扩展。

Capabilities 可选参数中的字段设置如下。能力代码字段设置为 1（表示多协议扩展功能）。能力长度字段设置为 4。能力值字段定义为：



该领域的用途和意义如下：

- AFI - 地址族标识符（16 位），与在多协议扩展的编码方式相同。
- Res. - 保留（8 位）字段。应由发件者设置为 0，并被接收器忽略。
- 请注意，对于不忽视该领域的接收者, 将字段值设置为 0 可能会产生问题。另外，如果曾经尝试重新定义这个域，定义是有问题的。
- SAFI - 子地址族标识符（8 位），与在多协议扩展的编码方式相同。

支持多个<AFI, SAFI>元组的 speaker 包括它们功能中的多个功能可选参数。

在一对 BGP speakers 之间的特定的<AFI, SAFI>，有路由信息的双向交换，每一个这样的 speaker 必须向另一方（通过能力公告机制）发布公告，其具有支持特定<AFI, SAFI>路由的能力。

9. IANA 注意事项

如本文档所述，MP_REACH_NLRI 和 MP_UNREACH_NLRI 属性包含子地址族标识符（SAFI）域。SAFI 名称空间在本文档中定义。IANA 注册并维护 SAFI 命名空间的值，如下所示：

- 本文档中分配了 SAFI 值 1 和 2。
- 保留 SAFI 值 3。它由 RFC 2858 分配使用，这从来没有得到充分的实现，所以不被本文档赞同。
- SAFI 值 5 至 63 由 IANA 按标准行动流程分配，[RFC2434]中定义，或早期 IANA 分配过程，在[RFC4020]中定义。
- SAFI 值 67 至 127 由 IANA 分配，使用 RFC 2434 中定义的“First Come First Served”策略。
- SAFI 值 0 和 255 被保留。
- SAFI 值 128 到 240 是之前的“私有使用”范围的一部分。在批准本文档时，路由区域主管向 IANA 提供未使用的值。这些未使用的值，即 130, 131, 135 到 139 和 141 至 240 被认为是保留的来避免冲突。

- SAFI 值 241 至 254 用于“私人使用”，该范围的值不由 IANA 分配。

10. 与 RFC 2858 比较

本文档使用下一跳信息与携带在 NEXT_HOP BGP 路径属性中的信息一致。

本文档删除了 SAFI 3 的定义，并废弃了 SAFI 3。

本文档改变了 SAFI 空间的分区。特别，在 RFC 2858 中，SAFI 值 128 到 240 是“私有使用”范围的一部分。本文档规定了此范围，分配目前正在使用的是由 IANA 认可，未使用的值，即 130, 131, 135 至 139 和 141 至 240，应视为保留。

此文档将“编号 SNPA s”字段重命名为“保留”，从 MP_REACH_NLRI 属性中删除其余的 SNPA 相关的信息。

11. 与 RFC 2283 比较

本文档限制 MP_REACH_NLRI 属性只能携带一个<AFI, SAFI, 下一跳信息...>单一实例。此文档限制 MP_UNREACH_NLRI 属性仅携带一个<AFI, SAFI, ...>单一实例。

本文件阐明不附带 NLRI 的 UPDATE 消息的处理，除了在 MP_REACH_NLRI 属性中的编码。

本文件澄清了 MP_REACH_NLRI 或 MP_UNREACH_NLRI 属性存在错误的处理。

本文档规定了 BGP 能力公告结合多协议扩展的使用。

最后，本文档包括“IANA 考虑”部分。