目 录

BGP

BGP 概述

BGP（Border Gateway Protocol，边界网关协议）是一种用于 AS（Autonomous System，自治系统）之间的动态路由协议。AS 是拥有同一选路策略，在同一技术管理部门下运行的一组路由器。

早期发布的三个版本分别是 BGP-1（RFC 1105）、BGP-2（RFC 1163）和 BGP-3（RFC 1267），当前使用的版本是 BGP-4（RFC 1771，已更新至 RFC 4271）。BGP-4 作为事实上的 Internet 外部路由协议标准，被广泛应用于 ISP（Internet Service Provider，因特网服务提供商）之间。

📖 **说明：**

下文中若不做特殊说明，所指的 BGP 均为 BGP-4。

BGP 特性描述如下：

- BGP 是一种外部网关协议（Exterior Gateway Protocol，EGP），与 OSPF、RIP 等内部网关协议（Interior Gateway Protocol，IGP）不同，其着眼点不在于发现和计算路由，而在于控制路由的传播和选择最佳路由。
- BGP 使用 TCP 作为其传输层协议（端口号 179），提高了协议的可靠性。
- BGP 支持 CIDR（Classless Inter-Domain Routing，无类别域间路由）。
- 路由更新时，BGP 只发送更新的路由，大大减少了 BGP 传播路由所占用的带宽，适用于在 Internet 上传播大量的路由信息。
- BGP 路由通过携带 AS 路径信息彻底解决路由环路问题。
- BGP 提供了丰富的路由策略，能够对路由实现灵活的过滤和选择。
- BGP 易于扩展，能够适应网络新的发展。

发送 BGP 消息的路由器称为 BGP 发言者（BGP Speaker），它接收或产生新的路由信息，并发布（Advertise）给其它 BGP 发言者。当 BGP 发言者收到来自其它自治系统的新路由时，如果该路由比当前已知路由更优、或者当前还没有该路由，它就把这条路由发布给自治系统内所有其它 BGP 发言者。

相互交换消息的 BGP 发言者之间互称对等体（Peer），若干相关的对等体可以构成对等体组（Peer group）。

BGP 在路由器上以下列两种方式运行：

- IBGP（Internal BGP）：当 BGP 运行于同一自治系统内部时，被称为 IBGP；
- EBGP（External BGP）：当 BGP 运行于不同自治系统之间时，称为 EBGP。

BGP 的消息类型

1. 消息头格式

BGP有 5 种消息类型：Open、Update、Notification、Keepalive和Route-refresh。这些消息有相同的报文头，其格式如 图 1所示。

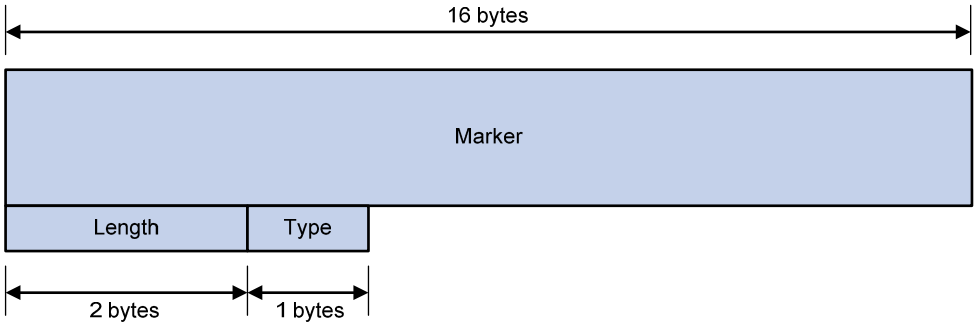

图1 BGP 消息的报文头格式

主要字段的解释如下：

- Marker：16 字节，用于标明 BGP 报文边界，所有比特均为“1”。
- Length：2 字节，BGP 消息总长度（包括报文头在内），以字节为单位。
- Type：1 字节，BGP 消息的类型。其取值从 1 到 5，分别表示 Open、Update、Notification、Keepalive 和 Route-refresh 消息。其中，前四种消息是在 RFC 1771 中定义，而 Type 为 5 的消息则是在 RFC 2918 中定义的。

2. Open

Open消息是TCP连接建立后发送的第一个消息，用于建立BGP对等体之间的连接关系。其消息格式如 图 2所示。

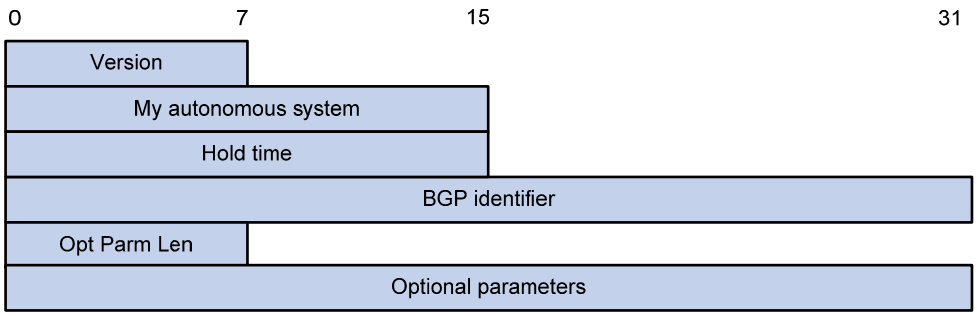

图2 BGP Open 消息格式

主要字段的解释如下：

- Version：BGP 的版本号。对于 BGP-4 来说，其值为 4。
- My autonomous system：本地 AS 号。通过比较两端的 AS 号可以确定是 EBGP 连接还是 IBGP 连接。
- Hold time：保持时间。在建立对等体关系时两端要协商 Hold Time，并保持一致。如果在这个时间内未收到对端发来的 Keepalive 消息或 Update 消息，则认为 BGP 连接中断。
- BGP identifier：BGP 标识符。以 IP 地址的形式表示，用来识别 BGP 路由器。
- Opt Parm Len（Optional Parameters Length）：可选参数的长度。如果为 0 则没有可选参数。
- Optional parameters：可选参数。用于多协议扩展（Multiprotocol Extensions）等功能。

3. Update

Update消息用于在对等体之间交换路由信息。它既可以发布可达路由信息，也可以撤销不可达路由信息。其消息格式如 图 3所示。

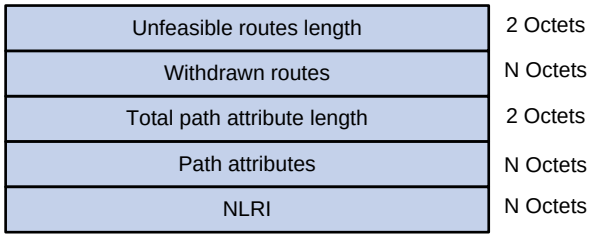

图3 BGP Update 消息格式

一条 Update 报文可以通告一类具有相同路径属性的可达路由，这些路由放在 NLRI（Network Layer Reachable Information，网络层可达信息）字段中，Path Attributes 字段携带了这些路由的属性，BGP 根据这些属性进行路由的选择；同时 Update 报文还可以携带多条不可达路由，被撤销的路由放在 Withdrawn Routes 字段中。

主要字段的解释如下：

- Unfeasible routes length：不可达路由字段的长度，以字节为单位。如果为 0 则说明没有 Withdrawn Routes 字段。
- Withdrawn routes：不可达路由的列表。
- Total path attribute length：路径属性字段的长度，以字节为单位。如果为 0 则说明没有 Path Attributes 字段。

- Path atributes：与 NLRI 相关的所有路径属性列表，每个路径属性由一个 TLV（Type-Length-Value）三元组构成。BGP 正是根据这些属性值来避免环路，进行选路，协议扩展等。
- NLRI（Network Layer Reachability Information）：可达路由的前缀和前缀长度二元组。

4. Notification

当BGP检测到错误状态时，就向对等体发出Notification消息，之后BGP连接会立即中断。其消息格式如 图 4所示。

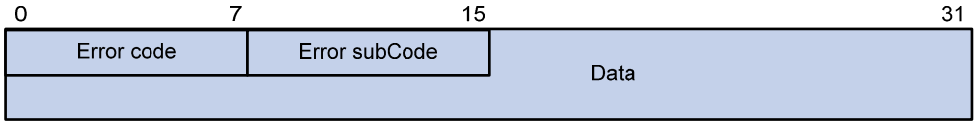

图4 BGP Notification 消息格式

主要字段的解释如下：

- Error code：差错码，指定错误类型。
- Error subcode：差错子码，错误类型的详细信息。
- Data：用于辅助发现错误的原因，它的内容依赖于具体的差错码和差错子码，记录的是出错部分的数据，长度不固定。

5. Keepalive

BGP 会周期性地向对等体发出 Keepalive 消息，用来保持连接的有效性。其消息格式中只包含报文头，没有附加其他任何字段。

6. Route-refresh

Route-refresh消息用来要求对等体重新发送指定地址族的路由信息。其消息格式如图 5所示。

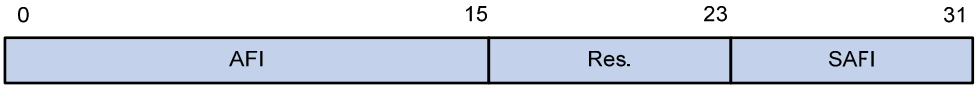

图5 BGP Route-refresh 消息格式

主要的字段解释如下：

- AFI：Address Family Identifier，地址族标识。
- Res.：保留，必须置 0。
- SAFI：Subsequent Address Family Identifier，子地址族标识。

BGP 的路由属性

1. 路由属性的分类

BGP 路由属性是一组参数，它对特定的路由进行了进一步的描述，使得 BGP 能够对路由进行过滤和选择。

事实上，所有的 BGP 路由属性都可以分为以下四类：

- 公认必须遵循（Well-known mandatory）：所有 BGP 路由器都必须能够识别这种属性，且必须存在于 Update 消息中。如果缺少这种属性，路由信息就会出错。
- 公认可选（Well-known discretionary）：所有 BGP 路由器都可以识别，但不要求必须存在于 Update 消息中，可以根据具体情况来选择。
- 可选过渡（Optional transitive）：在 AS 之间具有可传递性的属性。BGP 路由器可以不支持此属性，但它仍然会接收带有此属性的路由，并通告给其他对等体。
- 可选非过渡（Optional non-transitive）：如果 BGP 路由器不支持此属性，该属性被忽略，且不会通告给其他对等体。

BGP路由几种基本属性和对应的类别如 表 1所示。

表1 路由属性和类别

属性名称	类别
ORIGIN	公认必须遵循
AS_PATH	公认必须遵循
NEXT_HOP	公认必须遵循
LOCAL_PREF	公认可选
ATOMIC_AGGREGATE	公认可选
AGGREGATOR	可选过渡
COMMUNITY	可选过渡
MULTI_EXIT_DISC (MED)	可选非过渡
ORIGINATOR_ID	可选非过渡
CLUSTER_LIST	可选非过渡

2. 几种主要的路由属性

(1) 源（ORIGIN）属性

ORIGIN 属性定义路由信息的来源，标记一条路由是怎么成为 BGP 路由的。它有以下三种类型：

- IGP：优先级最高，说明路由产生于本 AS 内。
- EGP：优先级次之，说明路由通过 EGP 学到。
- incomplete：优先级最低，它并不是说明路由不可达，而是表示路由的来源无法确定。例如，引入的其它路由协议的路由信息。

(2) AS 路径（AS_PATH）属性

AS_PATH属性按一定次序记录了某条路由从本地到目的地址所要经过的所有AS号。当BGP将一条路由通告到其他AS时，便会把本地AS号添加在AS_PATH列表的最前面。收到此路由的BGP路由器根据AS_PATH属性就可以知道去目的地址所要经过的AS。离本地AS最近的相邻AS号排在前面，其他AS号按顺序依次排列。如 图6所示。

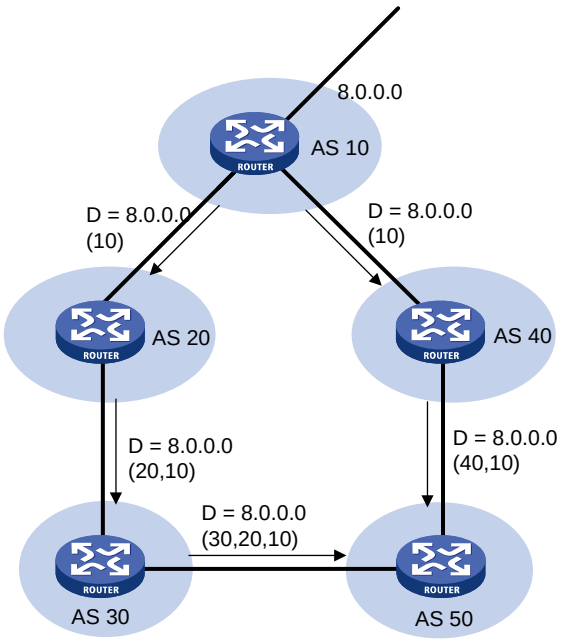

图6 AS_PATH 属性

通常情况下，BGP 不会接受 AS_PATH 中已包含本地 AS 号的路由，从而避免了形成路由环路的可能。

同时，AS_PATH属性也可用于路由的选择和过滤。在其他因素相同的情况下，BGP会优先选择路径较短的路由。比如在 图 6中，AS 50 中的BGP路由器会选择经过AS 40 的路径作为到目的地址 8.0.0.0 的最优路由。

在某些应用中，可以使用路由策略来人为地增加 AS 路径的长度，以便更为灵活地控制 BGP 路径的选择。

通过 AS 路径过滤列表，还可以针对 AS_PATH 属性中所包含的 AS 号来对路由进行过滤。

(3) 下一跳（NEXT_HOP）属性

BGP 的下一跳属性和 IGP 的有所不同，不一定就是邻居路由器的 IP 地址。

下一跳属性取值情况分为三种，如所示。

- BGP 发言者把自己产生的路由发给所有邻居时，将把该路由信息的下一跳属性设置为自己与对端连接的接口地址；
- BGP 发言者把接收到的路由发送给 EBGP 对等体时，将把该路由信息的下一跳属性设置为本地与对端连接的接口地址；
- BGP 发言者把从 EBGP 邻居得到的路由发给 IBGP 邻居时，并不改变该路由信息的下一跳属性。如果配置了负载分担，路由被发给 IBGP 邻居时则会修改下一跳属性。

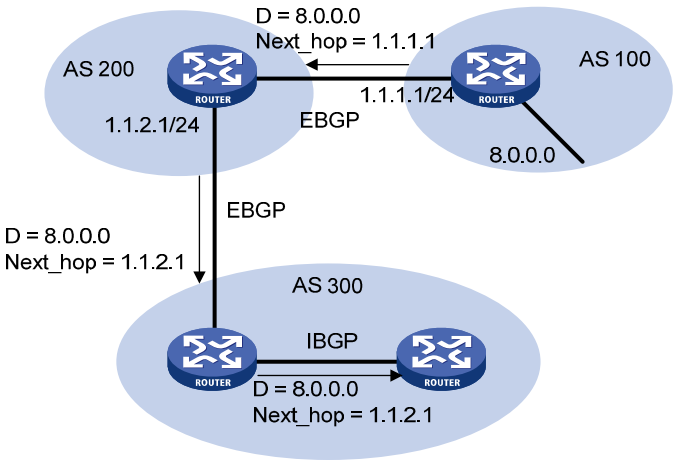

图7 下一跳属性

(4) MED（MULTI_EXIT_DISC）

MED 属性仅在相邻两个 AS 之间交换，收到此属性的 AS 一方不会再将其通告给任何其他第三方 AS。

MED属性相当于IGP使用的度量值（metrics），它用于判断流量进入AS时的最佳路由。当一个运行BGP的路由器通过不同的EBGP对等体得到目的地址相同但下一跳不同的多条路由时，在其它条件相同的情况下，将优先选择MED值较小者作为最佳路由。如 图 8所示，从AS 10 到AS 20 的流量将选择Router B作为入口。

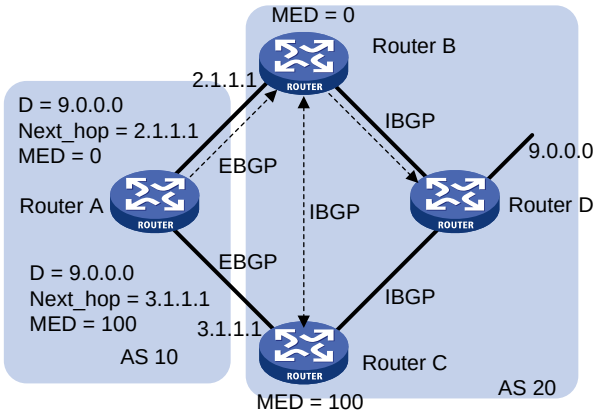

图8 MED 属性

通常情况下，BGP 只比较来自同一个 AS 的路由的 MED 属性值。

(5) 本地优先（LOCAL_PREF）属性

LOCAL_PREF 属性仅在 IBGP 对等体之间交换，不通告给其他 AS。它表明 BGP 路由器的优先级。

LOCAL_PREF属性用于判断流量离开AS时的最佳路由。当BGP的路由器通过不同的IBGP对等体得到目的地址相同但下一跳不同的多条路由时，将优先选择LOCAL_PREF属性值较高的路由。如 图 9所示，从AS 20 到AS 10 的流量将选择Router C作为出口。

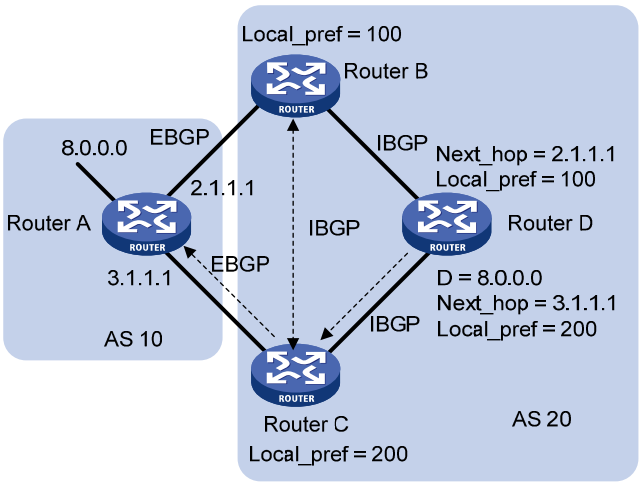

图9 LOCAL_PREF 属性

(6) 团体（COMMUNITY）属性

团体属性用来简化路由策略的应用和降低维护管理的难度。它是一组有相同特征的目的地址的集合，没有物理上的边界，与其所在的 AS 无关。公认的团体属性有：

- INTERNET：缺省情况下，所有的路由都属于 INTERNET 团体。具有此属性的路由可以被通告给所有的 BGP 对等体。

- NO_EXPORT：具有此属性的路由在收到后，不能被发布到本地 AS 之外。如果使用了联盟，则不能被发布到联盟之外，但可以发布给联盟中的其他子 AS。
- NO_ADVERTISE：具有此属性的路由被接收后，不能被通告给任何其他的 BGP 对等体。
- NO_EXPORT_SUBCONFED：具有此属性的路由被接收后，不能被发布到本地 AS 之外，也不能发布到联盟中的其他子 AS。

BGP 的选路规则

1. BGP 选择路由的策略

在目前的实现中，BGP 选择路由时采取如下策略：

- 首先丢弃下一跳（NEXT_HOP）不可达的路由；
- 优选 Preferred-value 值最大的路由；
- 优选本地优先级（LOCAL_PREF）最高的路由；
- 优选聚合路由；
- 优选 AS 路径（AS_PATH）最短的路由；
- 依次选择 ORIGIN 类型为 IGP、EGP、Incomplete 的路由；
- 优选 MED 值最低的路由；
- 依次选择从 EBGP、联盟、IBGP 学来的路由；
- 优选下一跳 Cost 值最低的路由；
- 优选 CLUSTER_LIST 长度最短的路由；
- 优选 ORIGINATOR_ID 最小的路由；
- 优选 Router ID 最小的路由器发布的路由。
- 优选地址最小的对等体发布的路由。

📖 说明：

- CLUSTER_ID 为路由反射器的集群 ID，CLUSTER_LIST 由 CLUSTER_ID 序列组成，反射器将自己的 CLUSTER_ID 加入 CLUSTER_LIST 中，若反射器收到路由中 CLUSTER_LIST 中包含有自己的 CLUSTER_ID，则丢弃该路由，从而避免群内环路的发生。
- 如果配置了负载分担，并且有多条到达同一目的地的路由，则根据配置的路由条数选择多条路由进行负载分担。

---

2. 应用 BGP 负载分担时的选路

在 BGP 中，由于协议本身的特殊性，它产生的路由的下一跳地址可能不是当前路由器直接相连的邻居。常见的一个原因是：IBGP 之间发布路由信息时不改变下一跳。这种情况下，为了能够将报文正确转发出去，路由器必须先找到一个直接可达的地址（查找 IGP 建立的路由表项），通过这个地址到达路由表中指示的下一跳。在上述过程中，去往直接可达地址的路由被称为依赖路由，BGP 路由依赖于这些路由指导报文转发。根据下一跳地址找到依赖路由的过程就是路由迭代（recursion）。

目前系统支持基于迭代的 BGP 负载分担，即如果依赖路由本身是负载分担的（假设有三个下一跳地址），则 BGP 也会生成相同数量的下一跳地址来指导报文转发。需要说明的是，基于迭代的 BGP 负载分担在系统上始终启用。

在实现方法上，BGP 的负载分担与 IGP 的负载分担有所不同：

- IGP 是通过协议定义的路由算法，对到达同一目的地址的不同路由，根据计算结果，将度量值（metric）相等的（如 RIP、OSPF）路由进行负载分担，选择的标准很明确（按 metric）。
- BGP 本身并没有路由计算的算法，它只是一个选路的路由协议，因此，不能根据一个明确的度量值决定是否对路由进行负载分担，但 BGP 有丰富的选路规则，可以在对路由进行一定的选择后，有条件地进行负载分担，也就是将负载分担加入到 BGP 的选路规则中去。

📖 **说明：**

- BGP 只对 AS_PATH 属性、ORIGIN 属性、LOCAL_PREF 和 MED 值完全相同的路由进行负载分担。
- BGP 负载分担特性适用于 EBGP、IBGP 以及联盟之间。
- 如果有多条到达同一目的地的路由，则根据配置的路由条数选择多条路由进行负载分担。

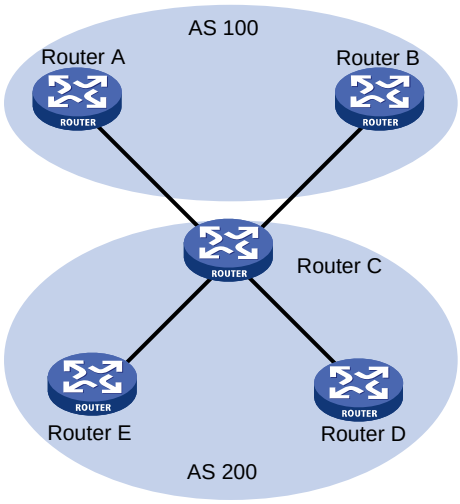

图10 BGP 负载分担示意图

在 图 10中，Router D和Router E是Router C的IBGP对等体。当Router A和Router B同时向Router C通告到达同一目的地的路由时，如果用户在Router C配置了负载分担（如balance 2），则当满足一定的选路规则后，并且两条路由具有相同的AS_PATH属性、ORIGIN属性、LOCAL_PREF和MED值时，Router C就把接收的两条路由同时加入到转发表中，实现BGP路由的负载分担。Router C只向Router D和Router E转发一次该路由，AS_PATH不变，但NEXT_HOP属性改变为Router C的地址，而不是原来的EBGP对等体地址。其它的BGP过渡属性将按最佳路由的属性传递。

3. BGP 发布路由的策略

在目前的实现中，BGP 发布路由时采用如下策略：

- 存在多条有效路由时，BGP 发言者只将最优路由发布给对等体；
- BGP 发言者只把自己使用的路由发布给对等体；
- BGP 发言者从 EBGP 获得的路由会向它所有 BGP 对等体发布（包括 EBGP 对等体和 IBGP 对等体）；
- BGP 发言者从 IBGP 获得的路由不向它的 IBGP 对等体发布；
- BGP 发言者从 IBGP 获得的路由发布给它的 EBGP 对等体（关闭 BGP 与 IGP 同步的情况下，IBGP 路由被直接发布；开启 BGP 与 IGP 同步的情况下，该 IBGP 路由只有在 IGP 也发布了这条路由时才会被同步并发布给 EBGP 对等体）；
- 连接一旦建立，BGP 发言者将把自己所有的 BGP 路由发布给新对等体。

IBGP 和 IGP 同步

同步是指 IBGP 和 IGP 之间的同步，其目的是为了避免出现误导外部 AS 路由器的现象发生。

如果一个AS中有非BGP路由器提供转发服务，经该AS转发的IP报文将可能因为目的地址不可达而被丢弃。如 图 11所示，Router E通过BGP从Router D可以学到Router A的一条路由 8.0.0.0/8，于是将到这个目的地址的报文转发给Router D，Router D查询路由表，发现下一跳是Router B。由于Router D从IGP学到了到Router B的路由，所以通过路由迭代，Router D将报文转发给Router C。但Router C并不知道去 8.0.0.0/8 的路由，于是将报文丢弃。

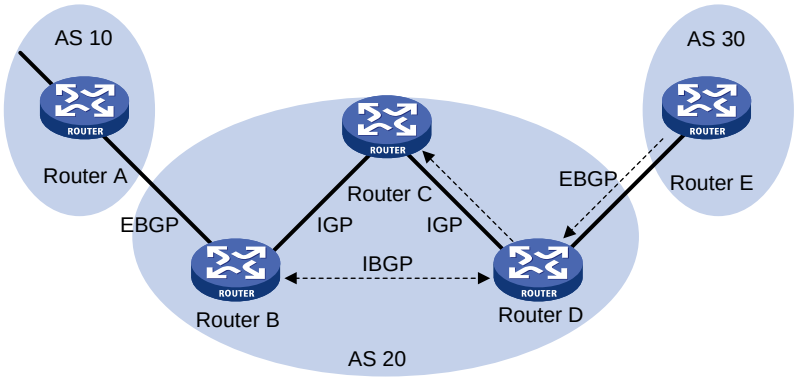

图11 IBGP 和 IGP 同步

如果设置了同步特性，在 IBGP 路由加入路由表并发布给 EBGP 对等体之前，会先检查 IGP 路由表。只有在 IGP 也知道这条 IBGP 路由时，它才会被发布给 EBGP 对等体。

在下面的情况中，可以关闭同步特性。

- 本AS不是过渡AS（图 11中的AS 20 就属于一个过渡AS）
- 本 AS 内所有路由器建立 IBGP 全连接

大规模 BGP 网络所遇到的问题

1. 路由聚合

在大规模的网络中，BGP 路由表十分庞大，使用路由聚合（Routes Aggregation）可以大大减小路由表的规模。

路由聚合实际上是将多条路由合并的过程。这样 BGP 在向对等体通告路由时，可以只通告聚合后的路由，而不是将所有的具体路由都通告出去。

目前系统支持自动聚合和手动聚合方式。使用后者还可以控制聚合路由的属性，以及决定是否发布具体路由。

2. BGP 路由衰减

BGP 路由衰减（Route Dampening）用来解决路由不稳定的问题。路由不稳定的主要表现形式是路由振荡（Route flaps），即路由表中的某条路由反复消失和重现。

发生路由振荡时，路由协议就会向邻居发布路由更新，收到更新报文的路由器需要重新计算路由并修改路由表。所以频繁的路由振荡会消耗大量的带宽资源和 CPU 资源，严重时会影响到网络的正常工作。

在多数情况下，BGP 协议都应用于复杂的网络环境中，路由变化十分频繁。为了防止持续的路由振荡带来的不利影响，BGP 使用衰减来抑制不稳定的路由。

BGP 衰减使用惩罚值来衡量一条路由的稳定性，惩罚值越高则说明路由越不稳定。路由每发生一次振荡（路由从激活状态变为未激活状态，称为一次路由振荡），BGP 便会给此路由增加一定的惩罚值（1000，此数值为系统固定，不可修改）。当惩罚值超过抑制阈值时，此路由被抑制，不加入到路由表中，也不再向其他 BGP 对等体发布更新报文。

被抑制的路由每经过一段时间，惩罚值便会减少一半，这个时间称为半衰期（Half-life）。当惩罚值降到再使用阈值时，此路由变为可用并被加入到路由表中，同时向其他 BGP 对等体发布更新报文。

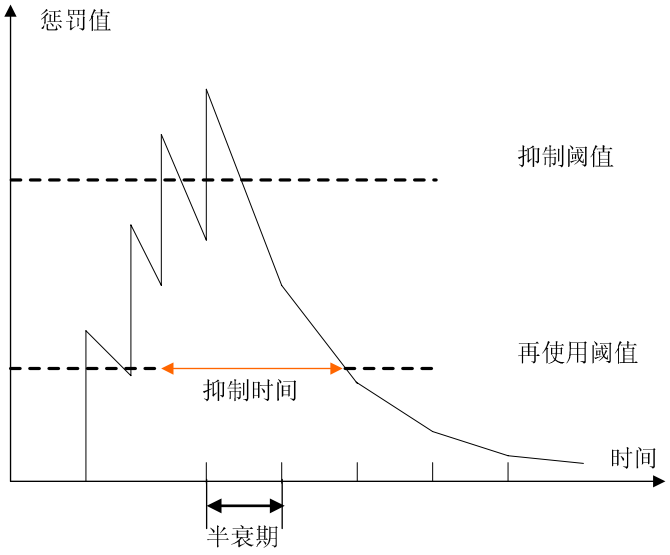

图12 BGP 衰减示意图

3. 对等体组

对等体组（Peer Group）是一些具有某些相同属性的对等体的集合。当一个对等体加入对等体组中时，此对等体将获得与所在对等体组相同的配置。当对等体组的配置改变时，组内成员的配置也相应改变。

在大型 BGP 网络中，对等体的数量会很多，其中很多对等体具有相同的策略，在配置时会重复使用一些命令，利用对等体组在很多情况下可以简化配置。

将对等体加入对等体组中，对等体与对等体组具有相同的路由更新策略，提高了路由发布效率。

4. 团体

对等体组可以使一组对等体共享相同的策略，而利用团体可以使多个 AS 中的一组 BGP 路由器共享相同的策略。团体是一个路由属性，在 BGP 对等体之间传播，它并不受到 AS 范围的限制。

BGP 路由器在将带有团体属性的路由发布给其它对等体之前，可以改变此路由原有的团体属性。

除了使用公认的团体属性外，用户还可以使用团体属性列表自定义扩展团体属性，以便更为灵活地控制路由策略。

5. 路由反射器

为保证 IBGP 对等体之间的连通性，需要在 IBGP 对等体之间建立全连接关系。假设在一个 AS 内部有 n 台路由器，那么应该建立的 IBGP 连接数就为 n(n-1)/2。当 IBGP 对等体数目很多时，对网络资源和 CPU 资源的消耗都很大。

利用路由反射可以解决这一问题。在一个 AS 内，其中一台路由器作为路由反射器 RR（Route Reflector），其它路由器作为客户机（Client）与路由反射器之间建立 IBGP 连接。路由反射器在客户机之间传递（反射）路由信息，而客户机之间不需要建立 BGP 连接。

既不是反射器也不是客户机的BGP路由器被称为非客户机（Non-Client）。非客户机与路由反射器之间，以及所有的非客户机之间仍然必须建立全连接关系。如 图 13 所示。

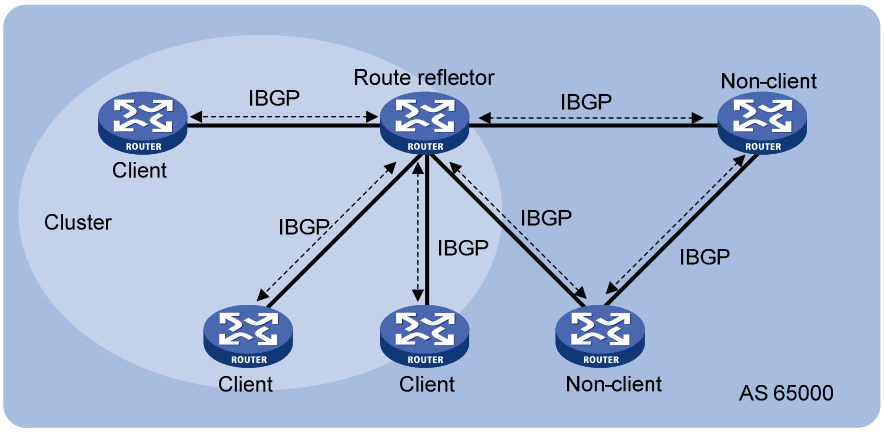

图13 路由反射器示意图

路由反射器和它的客户机组成了一个集群（Cluster）。某些情况下，为了增加网络的可靠性和防止单点故障，可以在一个集群中配置一个以上的路由反射器。这时，

位于相同集群中的每个路由反射器都要配置相同的Cluster_ID，以避免路由循环。如 图 14所示。

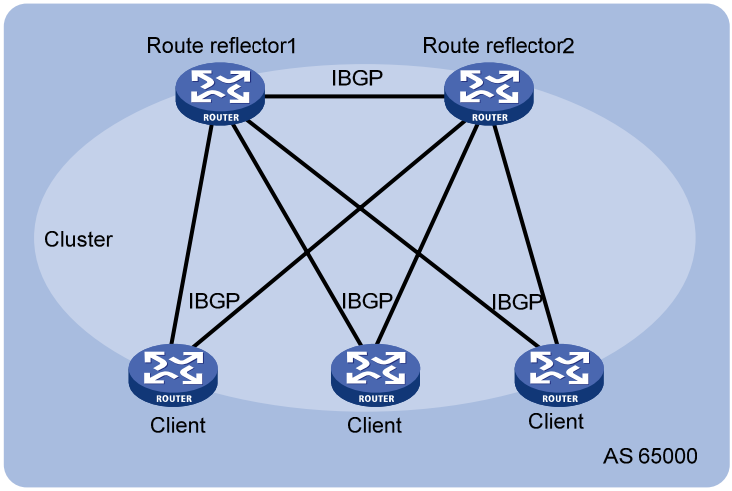

图14 多路由反射器

在某些网络中，路由反射器的客户机之间已经建立了全连接，它们可以直接交换路由信息，此时客户机到客户机之间的路由反射是没有必要的，而且还占用带宽资源。目前，系统支持配置相关命令来禁止在客户机之间反射路由。

📖 **说明：**

禁止客户机之间的路由反射后，客户机到非客户机之间的路由仍然可以被反射。

---

6. 联盟

联盟（Confederation）是处理AS内部的IBGP网络连接激增的另一种方法，它将一个自治系统划分为若干个子自治系统，每个子自治系统内部的IBGP对等体建立全连接关系，子自治系统之间建立联盟内部EBGP连接关系。如 图 15所示。

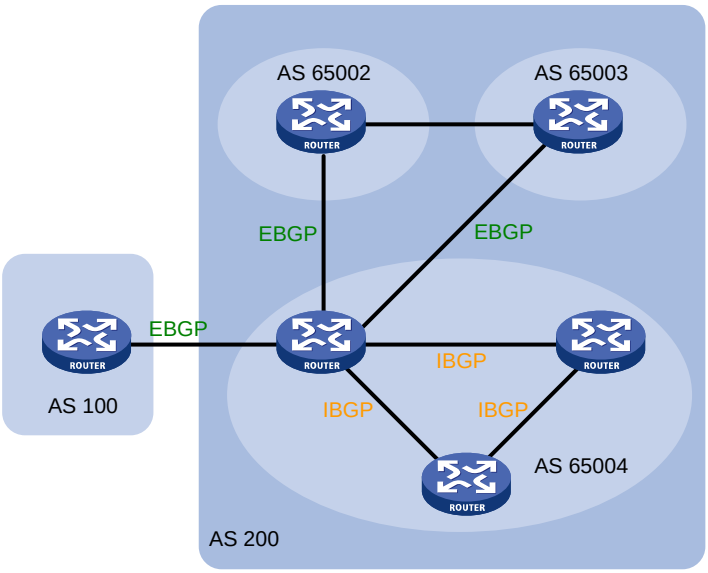

图15 联盟示意图

在不属于联盟的BGP发言者看来，属于同一个联盟的多个子自治系统是一个整体，外界不需要了解内部的子自治系统情况，联盟ID就是标识联盟这一整体的自治系统号，如 图 15中的AS 200 就是联盟ID。

联盟的缺陷是：从非联盟方案向联盟方案转变时，要求路由器重新进行配置，逻辑拓扑也要改变。

在大型 BGP 网络中，路由反射器和联盟可以被同时使用。

BGP GR

基于 BGP 的 GR Restarter 为了与 BGP 对等体建立一个 BGP 会话连接，首先要发送一个包含了 GR 能力的 OPEN 消息到对端，BGP 对等体收到该消息后，得知发送方已具有 GR 能力。这样，通过 OPEN 消息交互 GR 能力，GR Restarter 与其 BGP 对等体之间协商建立起 GR Session 连接。如果双方都没有交换 GR 能力的信息，建立起的会话也就不具备 GR 能力。

对于分布式设备，当进行主备倒换时，会话项将丢失，此时具备 GR 感知能力的 BGP 对等体会将所有与该 GR Restarter 有关的路由进行失效标记。但在 GR Time 内仍按照这些路由进行报文转发，这样确保了在从 BGP 对等体重新收集路由信息的过程中没有报文丢失。

对于分布式设备，主备倒换后，GR Restarter 会重新与 BGP 对等体建立 GR Session 连接，同时发送新的 GR 消息以宣告其重启完毕。此时两个 BGP 对等体间进行路由信息交换。交换完成后，GR Restarter 根据新的路由转发信息更新路由表和转发表，删除失效的路由，完成 BGP 协议收敛。

MP-BGP

1. MP-BGP 概述

传统的 BGP-4 只能管理 IPv4 单播路由信息，对于使用其它网络层协议（如 IPv6 等）的应用，在跨自治系统传播时就受到一定限制。

为了提供对多种网络层协议的支持，IETF 对 BGP-4 进行了扩展，形成 MP-BGP，目前的 MP-BGP 标准是 RFC 4760（Multiprotocol Extensions for BGP-4，BGP-4 的多协议扩展）。

支持 BGP 扩展的路由器与不支持 BGP 扩展的路由器可以互通。

2. MP-BGP 的扩展属性

BGP-4 使用的报文中，与 IPv4 地址格式相关的三条信息都由 Update 报文携带，这三条信息分别是：NLRI、路径属性中的 NEXT_HOP、路径属性中的 AGGREGATOR（该属性中包含形成聚合路由的 BGP 发言者的 IP 地址）。

为实现对多种网络层协议的支持，BGP-4 需要将网络层协议的信息反映到 NLRI 及 NEXT_HOP。MP-BGP 中引入了两个新的路径属性：

- MP_REACH_NLRI：Multiprotocol Reachable NLRI，多协议可达 NLRI。用于发布可达路由及下一跳信息。
- MP_UNREACH_NLRI：Multiprotocol Unreachable NLRI，多协议不可达 NLRI。用于撤销不可达路由。

这两种属性都是可选非过渡（Optional non-transitive）的，因此，不提供多协议能力的 BGP 发言者将忽略这两个属性的信息，不把它们传递给其它邻居。

3. 地址族

MP-BGP 采用地址族（Address Family）来区分不同的网络层协议，关于地址族的一些取值可以参考 RFC 1700（Assigned Numbers）。目前，系统实现了多种 MP-BGP 扩展应用，包括对 VPN 的扩展、对 IPv6 的扩展等。