



链滴

HCIP-BGP

作者: [shuaiqijun](#)

原文链接: <https://ld246.com/article/1612175425875>

来源网站: [链滴](#)

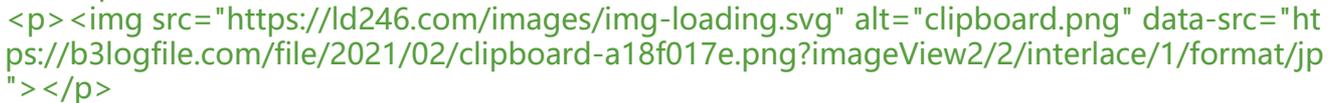
许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

1.概述

边界网关协议 (BGP) 是运行于 TCP 上的一种自治系统的路由协议。BGP 是唯一一个用来处理因特网大小的网络的协议，也是唯一能够妥善处理好不相关路由域间的多路连接的协议。BGP 构建在 EGP 的经验之上。BGP 系统的主要功能是和其他的 BGP 系统交换网络可达信息。网络可达信息包括出的自治系统 (AS) 的信息。这些信息有效地构造了 AS 互联的拓扑图并由此清除了路由环路，同时 AS 级别上可实施策略决策。

自治系统 AS (Autonomous System)

AS 是指在一个实体管辖下的拥有相同选路策略的 IP 网络。BGP 网络中的每个 AS 都被分配一个唯一的 AS 号，用于区分不同的 AS。AS 号分为 2 字节 AS 号和 4 字节 AS 号，其中 2 字节 AS 号的范围为 1 至 65535，其中 1-64511 为公有 AS 号，64512-65534 为私有 AS 号。4 字节 AS 号的范围为 1 至 4294967295。支持 4 字节 AS 号的设备能够与支持 2 字节 AS 号的设备兼容。这个编号由 IANA 配。





2.特点

BGP 属于外部或域间路由协议。BGP 的主要目标是为处于不同 AS 中的路由器之间进行路由通信提供保障。BGP 既不是纯粹的矢量距离协议，也不是纯粹的链路状态协议，通常被称为通路向量由协议。这是因为 BGP 在发布到一个目的网络的可达性的同时，包含了在 IP 分组到达目的网络过程所必须经过的 AS 的列表。通路向量信息是十分有用的，因为只要简单地查找一下 BGP 路由更新的 AS 编号就能有效地避免环路的出现。BGP 对网络拓扑结构没有限制，其特点包括：

(1) 实现自治系统间通信，传播网络的可达信息。BGP 是一个外部网关协议，允许一个 AS 与一个 AS 进行通信。BGP 允许一个 AS 向其他 AS 通告其内部的网络的可达性信息，或者是通过该 AS 可达的其他网络的路由信息。同时，AS 也能够从另一个 AS 中了解这些信息。与距离向量选路协议类似，BGP 为每个目的网络提供的是下一跳 (next-hop) 结点的信息。

(2) 多个 BGP 路由器之间的协调。如果在一个自治系统内部有多个路由器分别使用 BGP 与其自治系统中对等路由器进行通信，BGP 可以协调者一系列路由器，使这些路由器保持路由信息的一致。

(3) BGP 支持基于策略的选路 (policy-base routing)。一般的距离向量选路协议确切通告本选路中的路由。而 BGP 则可以实现由本地管理员选择的策略。BGP 路由器可以为域内和域间的网络可达性配置不同的策略。

(4) 可靠的传输。BGP 路由信息的传输采用了可靠地 TCP 协议。BGP 四层使用 tcp 179 号端

(5) 路径信息。在 BGP 通告目的网络的可达性信息时，处理指定目的网络的下一跳信息之外，告中还包括了通路向量 (path vector)，即去往该目的网络时需要经过的 AS 的列表，使接受者能够解去往目的网络的通路信息。

(6) 增量更新。BGP 不需要再所有路由更新报文中传送完整的路由数据库信息，只需要在启动交换一次完整信息。后续的路由更新报文只通告网络的变化信息。这种网络变化的信息称为增量 (delta)。

(7) BGP 支持无类型编制 (CIDR) 及 VLSM 方式。通告的所有网络都以网络前缀加子网掩码方式表示。

(8) 路由聚集。BGP 允许发送方把路由信息聚集在一起，用一个条目来表示多个相关的目的网，以节约网络带宽。

(9) BGP 还允许接收方对报文进行鉴别和认证，以验证发送方的身份。

(10) BGP 主要用于大型网络、大型集团、运营商、银行、国家电网、国家与国家之间的路由。

(11) BGP 可以跨路由器建立邻居 (因为 BGP 发送的报文都是单播)

(12) 一台路由器只能启用一个 BGP 进程

3.运行方式

如图 1 所示，BGP 按照运行方式分为 EBGP (External/Exterior BGP) 和 IBGP (Internal/Interior BGP)。

<p></p>

EBGP: 运行于不同 AS 之间的 BGP 称为 EBGP。为了防止 AS 间产生环路, 当 BGP 设备接收 EGP 对等体发送的路由时, 会将带有本地 AS 号的路由丢弃。

IBGP: 运行于同一 AS 内部的 BGP 称为 IBGP。为了防止 AS 内产生环路, BGP 设备不将从 IBGP 对等体学到的路由通告给其他 IBGP 对等体, 并与所有 IBGP 对等体建立全连接。为了解决 IBGP 对体的连接数量太多的问题, BGP 设计了路由反射器和 BGP 联盟。

<p>说明: </p>

<p>如果在 AS 内一台 BGP 设备收到 EBGP 邻居发送的路由后, 需要通过另一台 BGP 设备将该路由输给其他 AS, 此时推荐使用 IBGP。</p>

<p>BGP 报文交互中的角色</p>

<p>BGP 报文交互中分为 Speaker 和 Peer 两种角色。</p>

<p>Speaker: 发送 BGP 报文的设备称为 BGP 发言者 (Speaker), 它接收或产生新的报文信息, 发布 (Advertise) 给其它 BGP Speaker。</p>

<p>Peer: 相互交换报文的 Speaker 之间互称对等体 (Peer)。若干相关的对等体可以构成对等体 (Peer Group)。</p>

<p>BGP 的路由器号 (Router ID) </p>

<p>BGP 的 Router ID 是一个用于标识 BGP 设备的 32 位值, 通常是 IPv4 地址的形式, 在 BGP 会建立时发送的 Open 报文中携带。对等体之间建立 BGP 会话时, 每个 BGP 设备都必须有唯一的 Router ID, 否则对等体之间不能建立 BGP 连接。</p>

<p>BGP 的 Router ID 在 BGP 网络中必须是唯一的, 可以采用手工配置, 也可以让设备自动选取。省情况下, BGP 选择设备上的 Loopback 接口的 IPv4 地址作为 BGP 的 Router ID。如果设备上没配置 Loopback 接口, 系统会选择接口中最大的 IPv4 地址作为 BGP 的 Router ID。一旦选出 Router ID, 除非发生接口地址删除等事件, 否则即使配置了更大的地址, 也保持原来的 Router ID。</p>

<p>4.BGP 报文</p>

<p>BGP 对等体的建立、更新和删除等交互过程主要有 5 种报文、6 种状态机和 5 个原则。</p>

<p>BGP 对等体间通过以下 5 种报文进行交互, 其中 Keepalive 报文为周期性发送, 其余报文为触式发送: </p>

<p>Open 报文: 用于建立 BGP 对等体连接。</p>

<p>Update 报文: 用于在对等体之间交换路由信息。可以发布也可以撤销路由信息。</p>

<p>Notification 报文: 用于中断 BGP 连接。</p>

<p>Keepalive 报文: 用于保持 BGP 连接。</p>

<p>Route-refresh 报文: 用于在改变路由策略后请求对等体重新发送路由信息。只有支持路由刷新 (Route-refresh) 能力的 BGP 设备会发送和响应此报文。</p>

<p>5.BGP 状态机</p>

<p>BGP 对等体的交互过程中存在 6 种状态机: 空闲 (Idle)、连接 (Connect)、活跃 (Active)、Open 报文已发送 (OpenSent)、Open 报文已确认 (OpenConfirm) 和连接已建立 (Established)。在 BGP 对等体建立的过程中, 通常可见的 3 个状态是: Idle、Active 和 Established。</p>

<p>图 2 BGP 对等体交互过程</p>

<p></p>

Idle 状态是 BGP 初始状态。在 Idle 状态下, BGP 拒绝邻居发送的连接请求。只有在收到本设备 Start 事件后, BGP 才开始尝试和其它 BGP 对等体进行 TCP 连接, 并转至 Connect 状态。

Start 事件是由一个操作者配置一个 BGP 过程, 或者重置一个已经存在的过程或者路由器软件重置 BGP 过程引起的。

任何状态中收到 Notification 报文或 TCP 拆链通知等 Error 事件后, BGP 都会转至 Idle 状态。

- 在 Connect 状态下, BGP 启动连接重传定时器 (Connect Retry) , 等待 TCP 完成连接。

- 如果 TCP 连接成功, 那么 BGP 向对等体发送 Open 报文, 并转至 OpenSent 状态。
- 如果 TCP 连接失败, 那么 BGP 转至 Active 状态。
- 如果连接重传定时器超时, BGP 仍没有收到 BGP 对等体的响应, 那么 BGP 继续尝试和其它 BGP 对等体进行 TCP 连接, 停留在 Connect 状态。

- 在 Active 状态下, BGP 总是在试图建立 TCP 连接。

- 如果 TCP 连接成功, 那么 BGP 向对等体发送 Open 报文, 关闭连接重传定时器, 并转至 OpenSent 状态。
- 如果 TCP 连接失败, 那么 BGP 停留在 Active 状态。
- 如果连接重传定时器超时, BGP 仍没有收到 BGP 对等体的响应, 那么 BGP 转至 Connect 状态

- 在 OpenSent 状态下, BGP 等待对等体的 Open 报文, 并对收到的 Open 报文中的 AS 号、版本号、认证码等进行检查。

- 如果收到的 Open 报文正确, 那么 BGP 发送 Keepalive 报文, 并转至 OpenConfirm 状态。
- 如果发现收到的 Open 报文有错误, 那么 BGP 发送 Notification 报文给对等体, 并转至 Idle 状态。

- 在 OpenConfirm 状态下, BGP 等待 Keepalive 或 Notification 报文。如果收到 Keepalive 报文, 则转至 Established 状态, 如果收到 Notification 报文, 则转至 Idle 状态。
- 在 Established 状态下, BGP 可以和对等体交换 Update、Keepalive、Route-refresh 报文和 Notification 报文。

- 如果收到正确的 Update 或 Keepalive 报文, 那么 BGP 就认为对端处于正常运行状态, 将保持 BGP 连接。
- 如果收到错误的 Update 或 Keepalive 报文, 那么 BGP 发送 Notification 报文通知对端, 并转至 Idle 状态。
- Route-refresh 报文不会改变 BGP 状态。
- 如果收到 Notification 报文, 那么 BGP 转至 Idle 状态。
- 如果收到 TCP 拆链通知, 那么 BGP 断开连接, 转至 Idle 状态。

<p>BGP 对等体之间的交互原则</p>

<p>BGP 设备将最优路由加入 BGP 路由表, 形成 BGP 路由。BGP 设备与对等体建立邻居关系后, 取以下交互原则: </p>

- 从 IBGP 对等体获得的 BGP 路由, BGP 设备只发布给它的 EBGP 对等体。
- 从 EBGP 对等体获得的 BGP 路由, BGP 设备发布给它所有 EBGP 和 IBGP 对等体。
- 当存在多条到达同一目的地址的有效路由时, BGP 设备只将最优路由发布给对等体。
- 路由更新时, BGP 设备只发送更新的 BGP 路由。
- 所有对等体发送的路由, BGP 设备都会接收。

<p>| 参数 | 缺省值 |

| BGP | 未使能 |</p>

<table>

<thead>

<tr>

<th>Keepalive 消息发送间隔</th>

<th>60 秒</th>

<th></th>

</tr>

</thead>

</table>

<p>6.BGP 与 IGP 交互</p>

<p>BGP 与 IGP 在设备中使用不同的路由表，为了实现不同 AS 间相互通讯，BGP 需要与 IGP 进行互，即 BGP 路由表和 IGP 路由表相互引入。</p>

<p>BGP 引入 IGP 路由</p>

<p>BGP 协议本身不发现路由，因此需要将其他路由引入到 BGP 路由表，实现 AS 间的路由互通。一个 AS 需要将路由发布给其他 AS 时，AS 边缘路由器会在 BGP 路由表中引入 IGP 的路由。为了更好的规划网络，BGP 在引入 IGP 的路由时，可以使用路由策略进行路由过滤和路由属性设置，也可以置 MED 值指导 EBGP 对等体判断流量进入 AS 时选路。</p>

<p>BGP 引入路由时支持 Import 和 Network 两种方式：</p>

Import 方式是按协议类型，将 RIP、OSPF、ISIS 等协议的路由引入到 BGP 路由表中。为了保证引入的 IGP 路由的有效性，Import 方式还可以引入静态路由和直连路由。

Network 方式是逐条将 IP 路由表中已经存在的路由引入到 BGP 路由表中，比 Import 方式更精。

<p>IGP 引入 BGP 路由</p>

<p>当一个 AS 需要引入其他 AS 的路由时，AS 边缘路由器会在 IGP 路由表中引入 BGP 的路由。为避免大量 BGP 路由对 AS 内设备造成影响，当 IGP 引入 BGP 路由时，可以使用路由策略，进行路由过滤和路由属性设置。</p>

<p>IGP：路由发现和计算</p>

<p>BGP：路由的控制和优先</p>

<p>7.BGP 路由的生成方式</p>

<p>1.network //只能 network ip 路由表里的路由，路由和掩码需要一致</p>

<p></p>

<p>2.import</p>

<p>[R5-bgp]import-route ospf 1</p>

<p>[R3-bgp]import-route ospf 1</p>

<p>3.路由汇总</p>

<p>路由黑洞：能够学习到路由，但是无法访问。</p>

<p>8.BGP 示例 1</p>

<p></p>

<p>[R1]interface g0/0/0</p>

<p>[R1-GigabitEthernet0/0/0]ip add 10.1.12.1 24</p>

<p>[R1]interface loopback 0</p>

<p>[R1-LoopBack0]ip add 1.1.1.1 24</p>

<p>[R1-LoopBack0]interface loopback 1</p>

<p>[R1-LoopBack1]ip add 1.1.2.1 24</p>

<p>[R1]router id 1.1.1.1</p>

```
<p>[R1]ospf 1</p>
<p>[R1-ospf-1]area 5</p>
<p>[R1-ospf-1-area-0.0.0.0]network 10.1.12.0 0.0.0.255</p>
<p>[R1-ospf-1-area-0.0.0.0]network 1.1.1.1 0.0.0.0</p>
<p>[R1-ospf-1-area-0.0.0.0]network 1.1.2.1 0.0.0.0</p>
<p>[R1]bgp 100 //R1 AS 号</p>
<p>[R1-bgp]peer 2.2.2.2 as-number 100 //Loopback 接口属于逻辑接口，与物理接口相比，不
链路影响，减少 BGP 振荡。邻居的 AS 号为 100</p>
<p>[R1-bgp]peer 2.2.2.2 connect-interface loopback 0 //指定更新源 loopback0 为 R1 环
接口地址 1.1.1.1</p>
<p>[R2]interface g0/0/0</p>
<p>[R2-GigabitEthernet0/0/0]ip add 10.1.12.2 24</p>
<p>[R2-GigabitEthernet0/0/0]interface loopback 0</p>
<p>[R2-LoopBack0]ip add 2.2.2.2 24</p>
<p>[R2]interface g0/0/1</p>
<p>[R2-GigabitEthernet0/0/1]ip add 10.1.23.2 24</p>
<p>[R2-GigabitEthernet0/0/1]interface g0/0/2</p>
<p>[R2-GigabitEthernet0/0/2]ip add 10.1.25.2 24</p>
<p>[R2]router id 2.2.2.2</p>
<p>[R2]ospf 1</p>
<p>[R2-ospf-1]area 0</p>
<p>[R2-ospf-1-area-0.0.0.0]network 10.1.23.0 0.0.0.255</p>
<p>[R2-ospf-1-area-0.0.0.0]network 10.1.25.0 0.0.0.255</p>
<p>[R2-ospf-1-area-0.0.0.0]area 5</p>
<p>[R2-ospf-1-area-0.0.0.5]network 10.1.12.0 0.0.0.255</p>
<p>[R2-ospf-1-area-0.0.0.5]network 2.2.2.2 0.0.0.0</p>
<p>[R2-ospf-1-area-0.0.0.5]q</p>
<p>[R2-ospf-1]silent-interface g0/0/1 //将 g0/0/1 和 g0/0/2 配置为静默接口</p>
<p>[R2-ospf-1]silent-interface g0/0/2</p>
<p>[R2]bgp 100</p>
<p>[R2-bgp]peer 1.1.1.1 as-number 100</p>
<p>[R2-bgp]peer 1.1.1.1 connect-interface loopback 0</p>
<p>[R2-bgp]peer 10.1.23.3 as-number 200</p>
<p>[R2-bgp]peer 10.1.25.5 as-number 200</p>
<p></p>
<p>//reset bgp all 重置 BGP 邻居关系 生产环境不能使用</p>
<p></p>
<p>[R3]interface g0/0/0</p>
<p>[R3-GigabitEthernet0/0/0]ip add 10.1.23.3 24</p>
<p>[R3-GigabitEthernet0/0/0]interface g0/0/1</p>
<p>[R3-GigabitEthernet0/0/1]ip add 10.1.34.3 24</p>
<p>[R3-GigabitEthernet0/0/1]interface loopback 0</p>
<p>[R3-LoopBack0]ip add 3.3.3.3 24</p>
<p>[R3-LoopBack0]q</p>
<p>[R3]router id 3.3.3.3</p>
<p>[R3]ospf 1</p>
<p>[R3-ospf-1]area 0</p>
<p>[R3-ospf-1-area-0.0.0.0]network 3.3.3.3 0.0.0.0</p>
<p>[R3-ospf-1-area-0.0.0.0]network 10.1.23.0 0.0.0.255</p>
<p>[R3-ospf-1-area-0.0.0.0]network 10.1.34.0 0.0.0.255</p>
```

```
<p>[R3-ospf-1-area-0.0.0.0]q</p>
<p>[R3-ospf-1]silent-interface g0/0/0</p>
<p>[R3]bgp 200</p>
<p>[R3-bgp]peer 5.5.5.5 as-number 200</p>
<p>[R3-bgp]peer 5.5.5.5 connect-interface LoopBack 0</p>
<p>[R3-bgp]peer 4.4.4.4 as-number 200</p>
<p>[R3-bgp]peer 4.4.4.4 connect-interface LoopBack 0</p>
<p>[R3-bgp]peer 10.1.23.2 as-number 100</p>
<p>[R4]interface g0/0/0</p>
<p>[R4-GigabitEthernet0/0/0]ip add 10.1.34.4 24</p>
<p>[R4-GigabitEthernet0/0/0]interface g0/0/1</p>
<p>[R4-GigabitEthernet0/0/1]ip add 10.1.45.4 24</p>
<p>[R4-GigabitEthernet0/0/1]interface loopback 0</p>
<p>[R4-LoopBack0]ip add 4.4.4.4 24</p>
<p>[R4-LoopBack0]interface loopback 1</p>
<p>[R4-LoopBack1]ip add 4.4.5.4 24</p>
<p>[R4-LoopBack1]q</p>
<p>[R4]router id 4.4.4.4</p>
<p>[R4]ospf 1</p>
<p>[R4-ospf-1]area 0</p>
<p>[R4-ospf-1-area-0.0.0.0]network 4.4.4.4 0.0.0.0</p>
<p>[R4-ospf-1-area-0.0.0.0]network 4.4.5.4 0.0.0.0</p>
<p>[R4-ospf-1-area-0.0.0.0]network 10.1.34.0 0.0.0.255</p>
<p>[R4-ospf-1-area-0.0.0.0]network 10.1.45.0 0.0.0.255</p>
<p>[R4]bgp 200</p>
<p>[R4-bgp]peer 3.3.3.3 as-number 200</p>
<p>[R4-bgp]peer 3.3.3.3 connect-interface LoopBack 0</p>
<p>[R4-bgp]peer 5.5.5.5 as-number 200</p>
<p>[R4-bgp]peer 5.5.5.5 connect-interface LoopBack 1</p>
<p>[R5]interface g0/0/0</p>
<p>[R5-GigabitEthernet0/0/0]ip add 10.1.25.5 24</p>
<p>[R5-GigabitEthernet0/0/0]interface g0/0/1</p>
<p>[R5-GigabitEthernet0/0/1]ip add 10.1.45.5 24</p>
<p>[R5-GigabitEthernet0/0/1]interface loopback 0</p>
<p>[R5-LoopBack0]ip add 5.5.5.5 24</p>
<p>[R5-LoopBack0]q</p>
<p>[R5]router id 5.5.5.5</p>
<p>[R5]ospf 1</p>
<p>[R5-ospf-1]area 0</p>
<p>[R5-ospf-1-area-0.0.0.0]network 5.5.5.5 0.0.0.0</p>
<p>[R5-ospf-1-area-0.0.0.0]network 10.1.25.0 0.0.0.255</p>
<p>[R5-ospf-1-area-0.0.0.0]network 10.1.45.0 0.0.0.255</p>
<p>[R5-ospf-1-area-0.0.0.0]q</p>
<p>[R5-ospf-1]silent-interface g0/0/0</p>
<p>[R5]bgp 200</p>
<p>[R5-bgp]peer 10.1.25.2 as-number 100</p>
<p>[R5-bgp]peer 3.3.3.3 as-number 200</p>
<p>[R5-bgp]peer 3.3.3.3 connect-interface LoopBack 0</p>
<p>[R5-bgp]peer 4.4.5.4 as-number 200</p>
<p>[R5-bgp]peer 4.4.5.4 connect-interface LoopBack 0</p>
<p></p>
```