PART 0

前序实验答疑

QUESTIONS & ANSWER

Lab5 S10

步骤10中，为什么示例图会出现RIP的路由条目？

```
     10.0.0.0/8 is variably subnetted, 7 subnets, 4 masks
C       10.0.0.0/24 is directly connected, FastEthernet0/0
O       10.0.1.0/24 [110/21] via 10.0.123.246, 01:13:36, FastEthernet0/1
R       10.0.20.0/30 [120/1] via 10.0.123.242, 00:00:14, Serial2/0
O       10.0.20.1/32 [110/12] via 10.0.123.246, 01:13:36, FastEthernet0/1
C       10.0.123.240/30 is directly connected, Serial2/0
C       10.0.123.244/30 is directly connected, FastEthernet0/1
O       10.0.123.248/29 [110/11] via 10.0.123.246, 01:13:42, FastEthernet0/1
```

```
R2(config)#interface loopback 0
R2(config-if)#ip address 10.0.20.1 255.255.255.252
```

Loopback接口通告方式：

- OSPF：无论设置的是什么子网掩码，均通告为/32
- RIP v2：以Loopback接口实际配置子网掩码为准

Lab5 S10

步骤10中，为什么示例图会出现RIP的路由条目？

- **实验报告示例(R2 Loopback 0配置为/30)：**
  OSPF通告为`10.0.20.1/32`，RIP通告为`10.0.20.1/30`
  R1上的两个协议分别产生了去往`10.0.20.1/32`和`10.0.20.1/30`的路由信息
  由于是不同网络的路由条目，都会被写入路由表
- **同学们的情况(R2 Loopback 0配置为/32)：**
  OSPF通告为`10.0.20.1/32`，RIP通告为`10.0.20.1/32`
  R1上的两个协议都产生了去往`10.0.20.1/32`的路由信息
  来自AD更小的OSPF的条目会被写入，而AD更大的RIP的则会被忽略

Lab5 S19

步骤19中，为什么我完全无法联通R9？我应该已经配置了Frame Relay映射呀

许多同学会错误地配置Port2:DLCI 102映射到Port11:DLCI 203

由于R5连接的是FRSW的Port1，此时对应R9 S2/0端口的是FRSW未连接设备的Port2

因此R9将无论如何都无法和R5正确通信

Port:DLCI	Port:DLCI
1:101	10:202
1:102	11:203

请同学们确保自己这里的配置中，源端口都是1

Lab5 S21

步骤21中，为什么我R7居然Ping通R5上对应R9的子接口了？

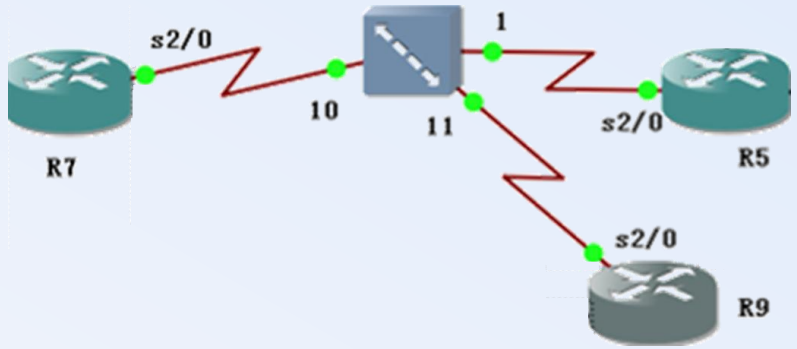
- Area 1、Area 2、Area 3使用10.X.0.0/16的网络地址进行分配，其中X为Area编号，例如Area 1的3个子网分别使用10.1.0.0/24、10.1.1.0/24、10.1.2.0/24等子网地址（同一个交换机上的多台路由器的接口属于同一个子网）

步骤22希望反映NBMA网络中，网络中多台设备虽然在同一逻辑网段，但不能像以太网广播/多播通信的特点，任意两台设备间通信都必须经过中心节点建立的点对点虚链路

FrameRelay会通过逆向ARP，使用DLCI与对端通信询问对端的IP，自动生成FR映射

相关接口在同一子网：R7发现相应接口与自己S2/0在同一子网，因此**直接二层转发**(本步骤中即按DLCI转发)；但R7没有另一子接口IP的FR映射，缺少对应DLCI导致无法转发

R5-7/R5-9不同子网：因为在不同子网选择**三层路由转发**，查路由表按查到OSPF通告的表项成功路由转发



Lab5 S21

[续上问] 路由器上两个接口不是不能在同一个子网吗?

不考虑FrameRelay等协议时，如果路由器上有两个相同子网的接口，在进行路由转发时我们会无法确定该数据包具体的出接口

但根据FrameRelayMap进行转发时，即使两个FrameRelay接口IP在相同子网，我们向该子网转发数据也只根据两个IP中哪一个有对应的DLCI映射进行判断，而我们要求这一映射应该是唯一的，从而消除了混淆

同学们应该也会注意到，当使用int [interface]配置子接口而不是int [interface] multipoint配置时，会报错警告不是相关的可用协议(如802.1Q等)

实验6

动态路由协议BGP配置

主讲：王信博

PART 01

BGP协议背景

BACKGROUND OF BGP PROTOCAL

OSPF协议为什么不够用?

OSPF 链路状态信息

OSPF中所有路由器都需要知道自己Area的全部链路状态信息和其他Area的链路总结信息

路由统计信息： 全世界

有关相关全局路由表项的统计信息

AS	前缀	路由
11万 IPv4: 77,965 IPv6: 36,499	**138万** IPv4: 1,121,434 IPv6: 254,933	**139万** IPv4: 1,130,559 IPv6: 261,130
RPKI 有效	RPKI 无效	RPKI 未知
82万 **(59%)** IPv4: 662,882 IPv6: 161,054	**1.5万** **(1.1%)** IPv4: 12,922 IPv6: 2,340	**55万** **(40%)** IPv4: 454,755 IPv6: 97,736

数据生成时间 2025年12月6日 UTC 14:00

Cloudflare Radar

过去 7 天 | 2025年12月6日 UTC 14:45

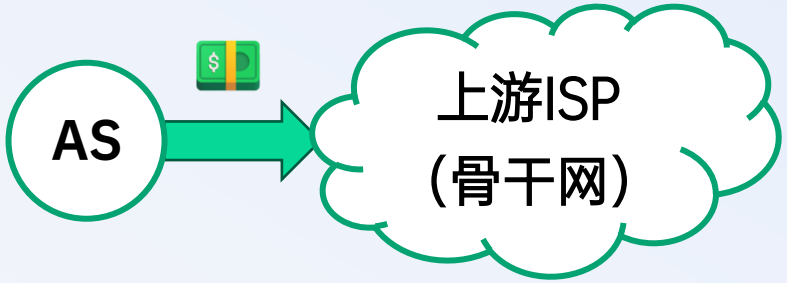
全球巨量的网络链路与设备会产生浩如烟海的链路状态信息，对于任何路由器而言都足以瞬间打爆其计算能力

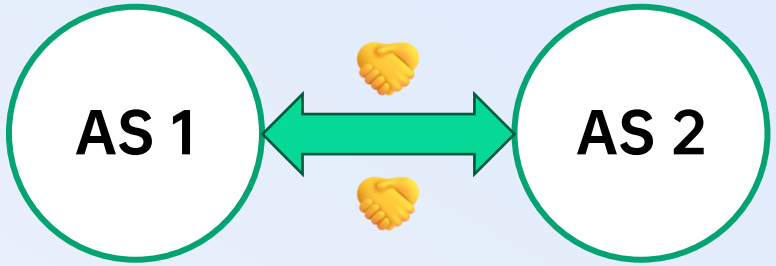
OSPF协议为什么不够用？

网络通信成本——此AS是我开，要想从此过，留下买路财

OSPF追求最短路径，但商业网络中最短路径、最低延迟≠最低成本、最省钱

网络服务商（ISP）的AS间通信有2种核心形式：



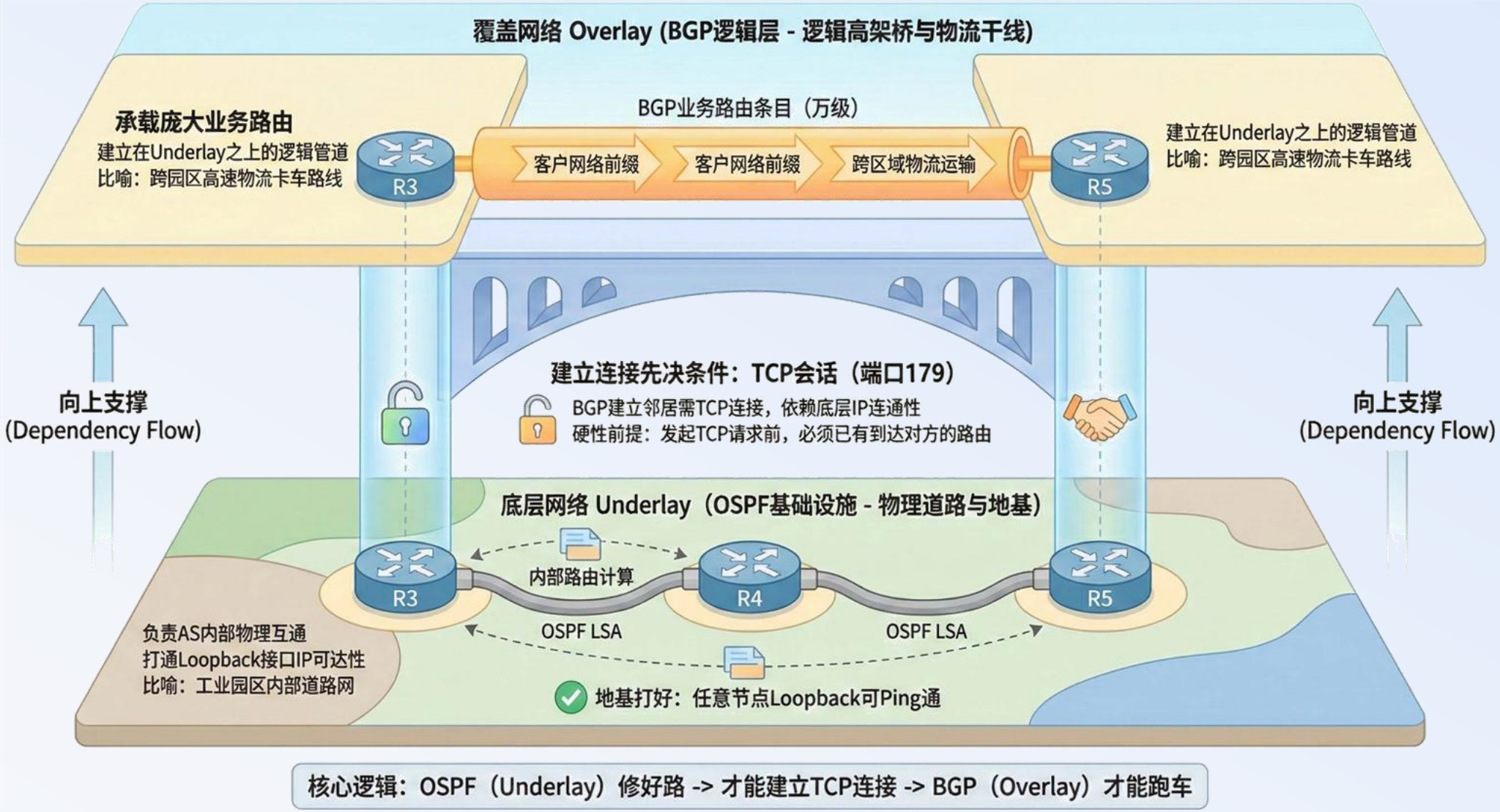


转接(Transit)：付费关系，小ISP向上游骨干网付费购买访问全球互联网的路权

对等(Peering)：免费关系，规模相当的ISP直接互联，免费交换彼此的用户路由，但不为对方转发去第三方的流量

如果为了路径最短而盲目选择路径，很可能因为AS间路由而赔得两手空空

BGP协议—底层网络上的Overlay



PART 02

实验原理

LAB PRINCIPLES

BGP运行模式

根据邻居所属的AS不同，有两种运行模式：eBGP和iBGP

特性	eBGP (External BGP)	iBGP (Internal BGP)
应用场景	不同自治系统之间	同一个自治系统内部
防环机制	AS_PATH：接收到的路由若包含自己的AS号则丢弃	**水平分割**：从iBGP邻居学到的路由，不传给其他iBGP邻居
TTL值	默认为1 (要求物理直连)	默认为255 (允许跨多跳)
路由修改	转发时会修改下一跳为自己	转发时**保留原始下一跳**不变

BGP路由通告

BGP在进行路由通告/更新时，具体传递了什么？

- **撤销路由信息**：通知对端原来通告的某个IP前缀不再可达，需要撤销相关路由信息
- **路径属性**：描述了通告路由的特性、策略和路径信息，公认必须遵循的属性有：
  - AS_PATH：有序AS编号列表，记录该路由信息从源头到本地BGP路由器经过的AS
  - Next_Hop：转发到目的网络时需要发送到的路由器IP，eBGP中通常是与目标路由器连接的接口IP，**iBGP中通常是eBGP学习到的IP，在内部保持不变(?)**
  - Origin：标记路由信息是如何注入到BGP系统的，优先级是IGP(i) - 通过`network`命令宣告 > EGP(e) – EGP注入 > Incomplete(?) - `redistribution`注入BGP
- **可达路由信息**：该Update信息中，路径属性所描述的目的地网络

为了减少开销并避免网络震荡，BGP会进行增量路由更新，且间隔时间很长，实验中需要更加耐心多等一会，或尝试清除BGP信息强制刷新

BGP路径选择

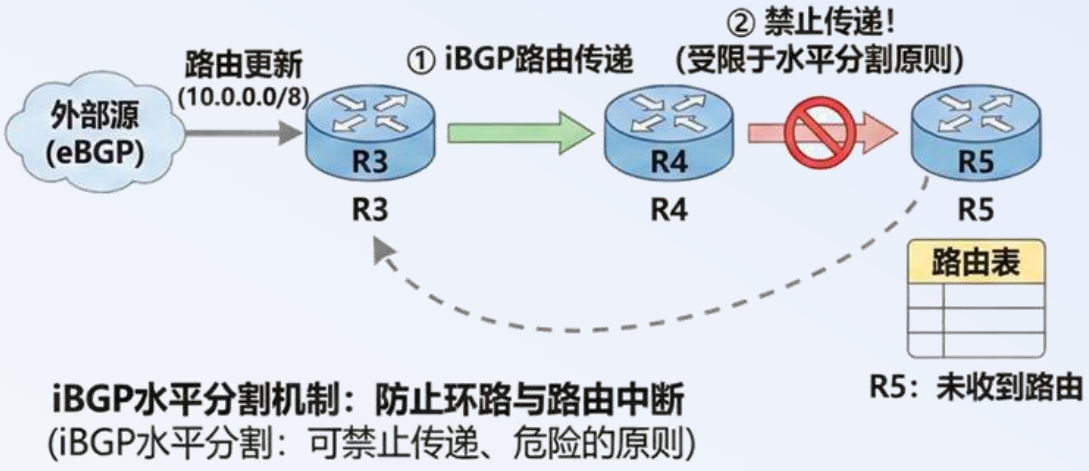
BGP以AS为核心，选取的路由路径也以经过的AS数量为最核心的因素

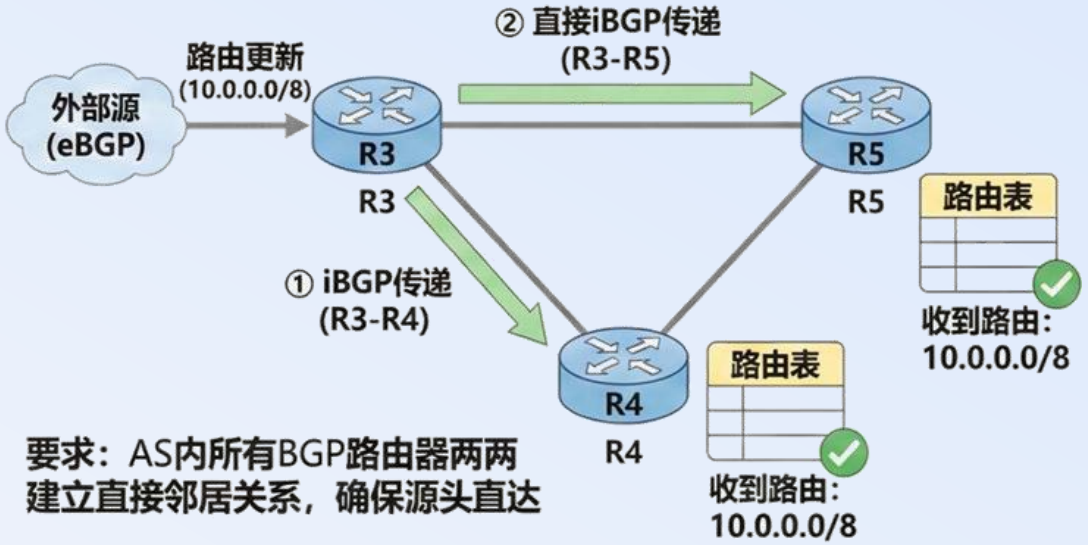
除去一些本地属性，可以看到对BGP而言最终的就是对成本影响最大的AS路径长度属性

优先级	选路属性	含义/商业逻辑
1	Weight	(Cisco私有) 本地权重，完全由管理员手工指定
2	Local Preference	本地优先级，决定流量离开本AS时走哪个出口
3	Local Origin	本地始发的路由优先
4	**AS_PATH Length**	**经过的AS数量越少越优**
5	Origin Code	IGP > EGP > Incomplete
6	MED	邻居建议的属性，决定流量进入本AS时走哪个入口

iBGP防环要求-S4

iBGP内路由在同一AS传递时AS_PATH不会变化，因此不能通过重复AS号来鉴别路由环路，为了避免环路，iBGP需要严格遵守**水平分割**原则，从一个iBGP邻居学到的路由信息禁止转发给其他iBGP邻居



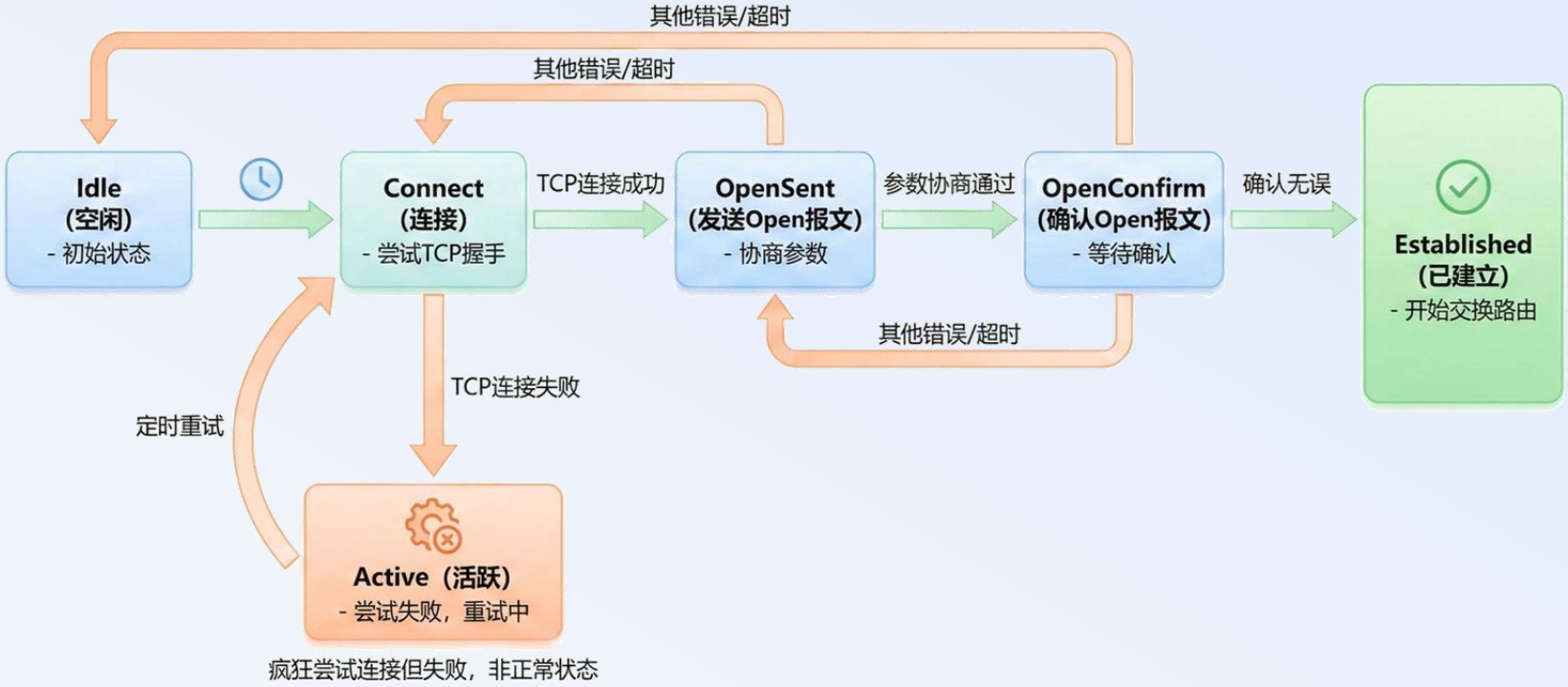


此时水平方向路由传递会中断，因此在AS内部，我们需要所有运行BGP的路由器都建立**全互联关系**，即两两建立邻居关系，以确保都能从源头收到路由更新

现实网络中两两组合开销巨大，维护困难，因此会引入路由反射器作为特权打破者，允许其将从iBGP邻居学到的路由转发给其他邻居，路由器只需要和路由反射器建立邻居关系

BGP下层协议与状态机-S5

BGP需要承载整个网络近140万条网络前缀，巨大的数据量下BGP如果像OSPF在网络层组播LSA，不仅会造成网络拥塞，分片/确认/重传和流控机制也需要自己实现——利用TCP

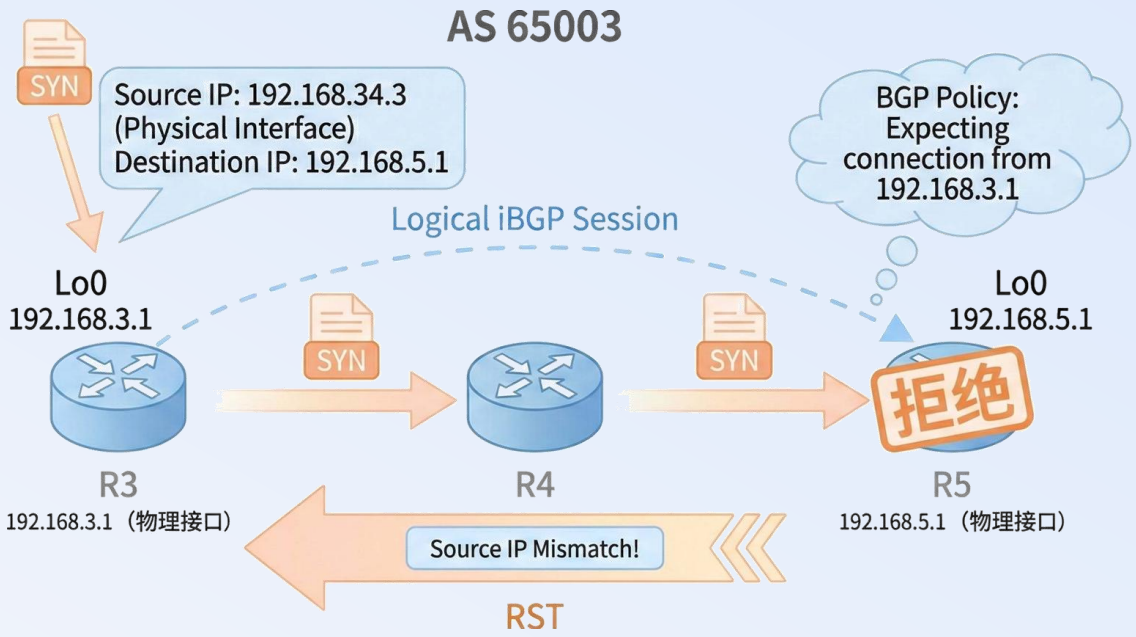


常见的误区是认为Active是个好状态，然而BGP中**Active表示TCP连接建立失败**，正在不断重试，如果邻居状态长期保持Active，可能说明网络不可达/配置错误

更新源与BGP的安全性检查-S6

BGP作为现代互联网的核心，任何错误的路由注入都可能瞬间造成网络瘫痪/流量劫持在全球传播，因此BGP的安全性与身份验证非常重要，它默认只会接受与配置的邻居地址匹配的TCP SYN包的连接，否则会直接RST

Cisco路由器**默认使用出接口IP**发起TCP连接，此时对端路由器收到的源IP将和我们配置的不同，导致邻居关系无法建立
因此需要使用`update-source`**手动指定**TCP三次握手的源地址使用我们所需接口的IP，以便通过BGP的检查



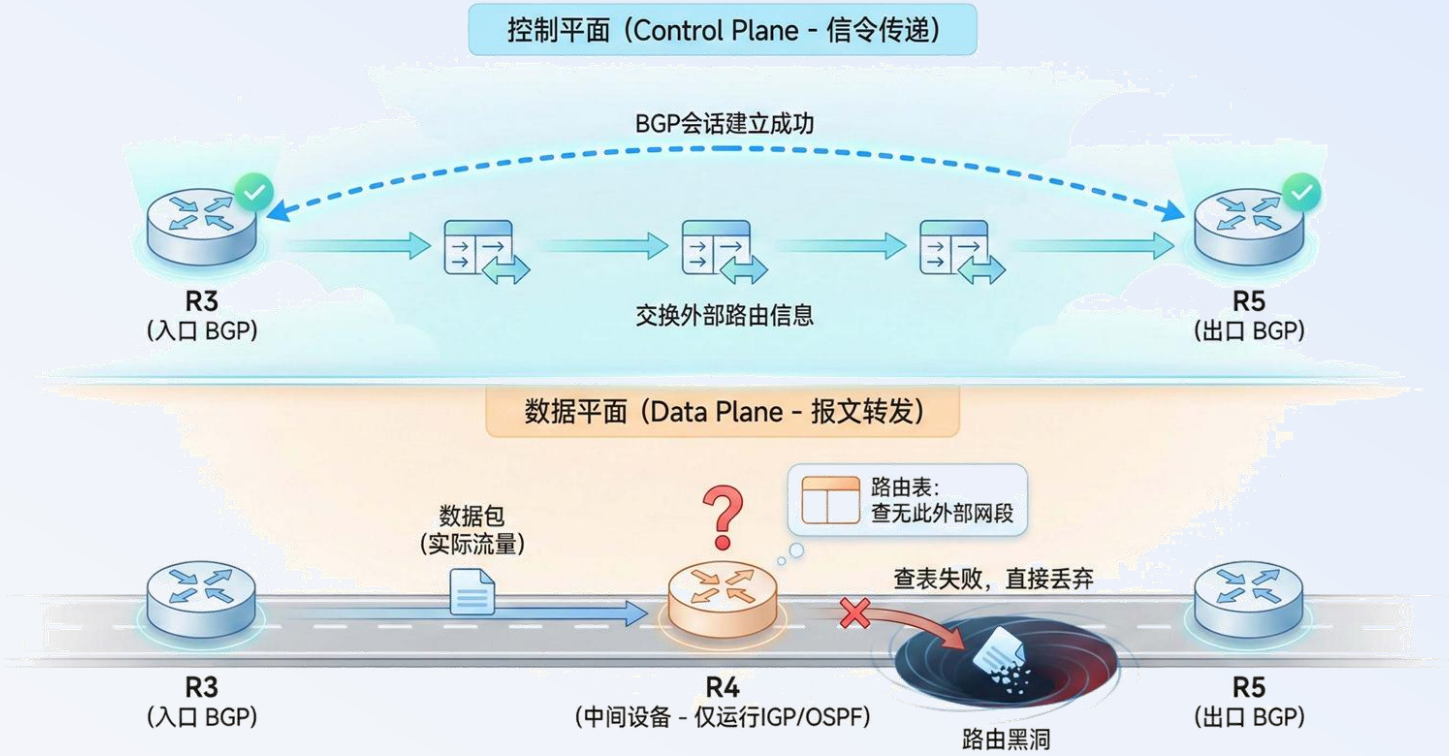
命令：`neighbor [IP] update-source [interface]`

例：`neighbor 192.168.5.1 update-source loopback 0`

路由黑洞-S16

我们在追踪R1到R2-R8间子网路由时，会发现路由在中途中断了，我们不是知道了正确的AS路径吗，为什么数据包会在中途迷路呢？这就是BGP协议中的"**路由黑洞**"

问题在于AS内部的普通路由器R4，它只运行了OSPF，因此**只有AS内部的路由**，不知道外部的BGP网段的情况，数据路过R4时，R4不知道如何转发，因此中断



BGP同步与重分发-S18

为了避免路由黑洞问题，早期BGP协议设计者指定了一个保守的同步规则：
BGP路由器向外(eBGP邻居)通告一条从内部(iBGP邻居)学到的路由前，必须检查自己的IGP路由表(如这里的OSPF)是否也有这条路由；没有则说明AS内没准备好转发这类流量，此时BGP会保持沉默，不通告该条路由，避免将外部流量吸引进路由黑洞中

然而同步规则只能避免路由黑洞，而不能消除路由黑洞，当我们必须经过该AS内部时，我们就需要通过重分发消除路由黑洞，即将BGP的路由注入内部的IGP(如OSPF)
但在当前互联网的规模下，启用同步意味着需要将全球BGP路由都注入到内部的OSPF协议中，这会瞬间撑爆IGP路由器的内存和CPU
现代Cisco路由器默认关闭同步/不启用重分发，而是借助MPLS(多协议标签交换)或GRE隧道，在R3-R5间建立隧道跨过R4，使得R4无需知道外部路由也可以承载相应数据

注意：重分发也需要双向配置（让AS内知道AS外BGP路由，也让AS外知道AS内的路由）

BGP路由过滤-S25

跨越不同ISP/企业网络边界时，连通性往往受合同、安全规则和流量成本制约，因此BGP不能只联通AS，也需要实现策略控制，允许管理员精确控制每一条路由的去留

本步骤中我们使用distribute-list控制路由过滤，限制前往PC3子网不能经过 AS 65008

命令：`access-list [id] deny [subnet] [wildcard-mask]`

`access-list [id] permit any`（ACL隐含Deny ALL，需显式允许不过滤的路由）

`neighbor [router-id] distribute-list [access-list-id] out`

通过以上命令，我们可以创建一个访问列表，禁止R8向R7传播相关子网的路由更新

实际大型网络中，网络工程师会用 Route-map(路由图) 来精细操作路由，例如：

- 修改 Local Preference 属性，强制让本公司的流量走带宽更便宜的ISP链路
- 修改 MED (Multi-Exit Discriminator) 属性，告诉邻居从这条路进入我的网络更近
- 利用 Community (团体) 属性，给路由打上标签，实现跨AS的批量策略控制

IPv6地址-S26

地址类型	前缀	对应IPv4	场景与规划建议
未指明	0...0/128	0...0/32	未配置IPv6主机的源地址，不作目的地址
环回	::1/128	127.x.x.x	与IPv4相同，但只有一个地址
多播	FF00::/8	224.0.0.0-239.255.255.255	提供广播/组播/ARP/邻居发现等功能
全球单播 GUA	其他所有	公网IP	全球唯一网络地址，运营商通常分配/48，企业内部再划分/64到各VLAN
唯一本地 ULA	FC00::/7	私网IP	公网不可路由，常用FD00::/8段规划内网
站点本地 Site-Local	FEC0::/10		类似IPv4私网地址，已废弃，由ULA取代
链路本地 Link-Local	FE80::/10	(类似)169.254.x.x	仅直连链路有效，接口强制自动生成存在路由邻居建立及作为网关下一跳的基础

BGP for IPv6-S26

随着网络技术的演进，我们面临着IPv6、组播、以及MPLS VPN等多种新型业务的需求
如果为每一种新业务都重新开发一个像BGP这样复杂的路由协议，网络设备的资源消耗和管理复杂度将是灾难性的

为了解决这个问题，IETF对BGP进行了扩展，诞生了MP-BGP（Multiprotocol BGP）
MP-BGP的核心思想是传输与业务分离，它引入了地址族（Address Family）的概念
路由器之间只需要建立一条TCP连接（BGP会话）就可以在这个**单一管道中同时传递**IPv4、IPv6、VPNv4等多种不同类型的路由信息，传递的这些不同路由信息共享同一套会话维护机制、防环机制和选路逻辑，但所承载的数据内容互不干扰
这种高度模块化的设计，使得BGP成为了当今网络世界中承载能力最强、适应性最广的万能协议

IPv6隧道-S33

为了连接被IPv4网络分隔的IPv6网络，我们可以利用**隧道技术**进行跨越