



石家庄铁道大学
SHIJIAZHUANG TIEDAO UNIVERSITY

计算机网络

第 13 讲 网络层



上讲内容回顾

- ◆ ICMP 报文的种类
- ◆ ICMP 的应用举例
- ◆ 有关路由选择协议的几个基本概念
- ◆ 内部网关协议 RIP



本讲内容

- ◆ 内部网关协议 OSPF
- ◆ 外部网关协议 BGP
- ◆ 路由器的构成
- ◆ IP 多播的基本概念
- ◆ 在局域网上进行硬件多播
- ◆ 因特网组管理协议 IGMP 和多播路由选择协议

4.5.3 内部网关协议 OSPF

(Open Shortest Path First)

1. OSPF 协议的基本特点

- “开放”表明 OSPF 协议不是受某一家厂商控制，而是公开发表的。
- “最短路径优先”是因为使用了 Dijkstra 提出的最短路径算法 SPF
- OSPF 只是一个协议的名字，它并不表示其他的路由选择协议不是“最短路径优先”。
- 是分布式的链路状态协议。

三个要点

- 向本自治系统中所有路由器发送信息，这里使用的方法是洪泛法。
- 发送的信息就是与本路由器相邻的所有路由器的链路状态，但这只是路由器所知道的部分信息。
 - “链路状态”就是说明本路由器都和哪些路由器相邻，以及该链路的“度量” (metric)。
- 只有当链路状态发生变化时，路由器才用洪泛法向所有路由器发送此信息。

链路状态数据库

(link-state database)

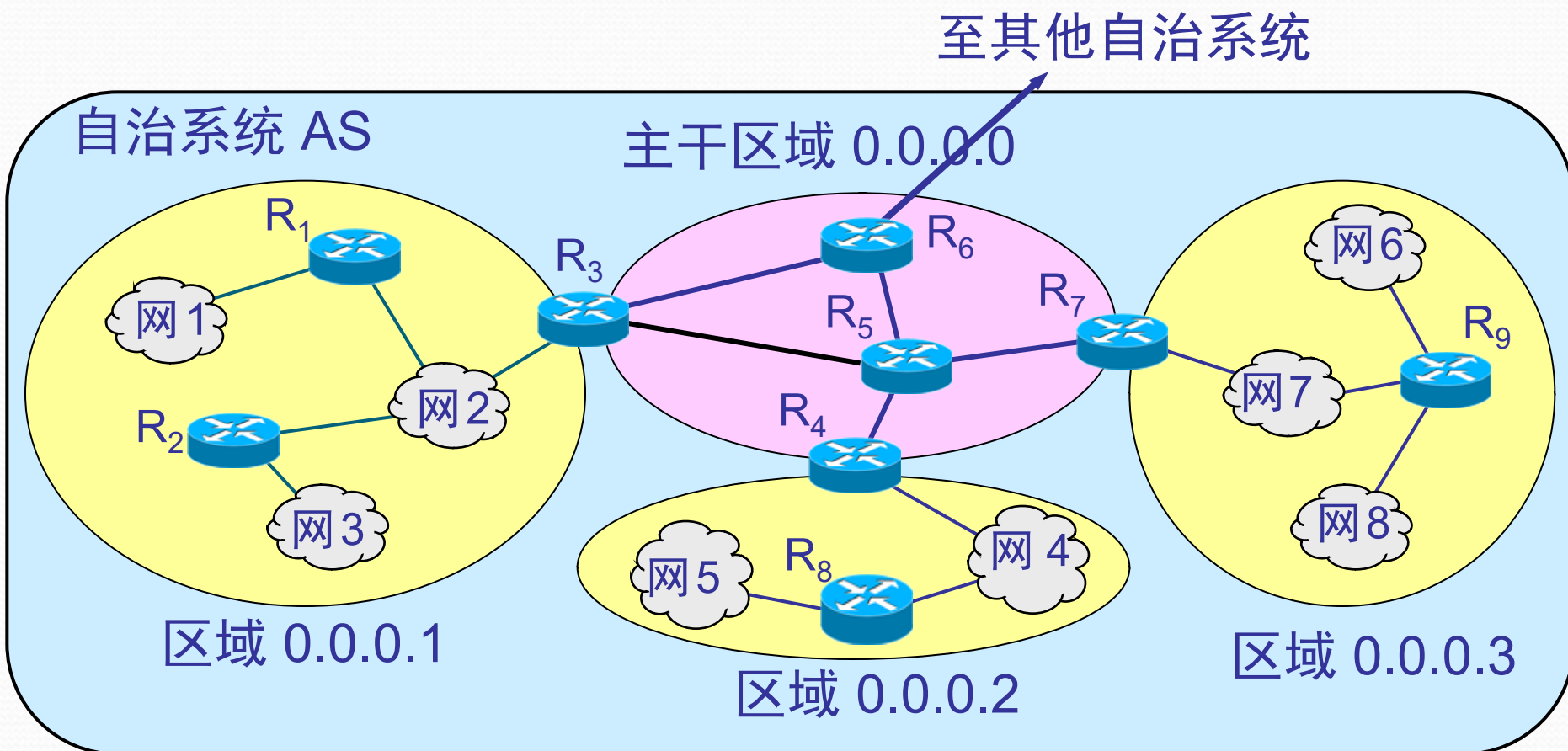
- 由于各路由器之间频繁地交换链路状态信息，因此所有的路由器最终都能建立一个链路状态数据库。
- 这个数据库实际上就是全网的拓扑结构图，它在全网范围内是一致的（这称为链路状态数据库的同步）。
- OSPF 的链路状态数据库能较快地进行更新，使各个路由器能及时更新其路由表。OSPF 的更新过程收敛得快是其重要优点。

→ 全网 → 内网各 (自治系统)

OSPF 的区域(area)

- 为了使 OSPF 能够用于规模很大的网络，OSPF 将一个自治系统再划分为若干个更小的范围，叫作区域。
- 每一个区域都有一个 32 位的区域标识符（用点分十进制表示）。
- 区域也不能太大，在一个区域内的路由器最好不超过 200 个。

OSPF 划分为两种不同的区域

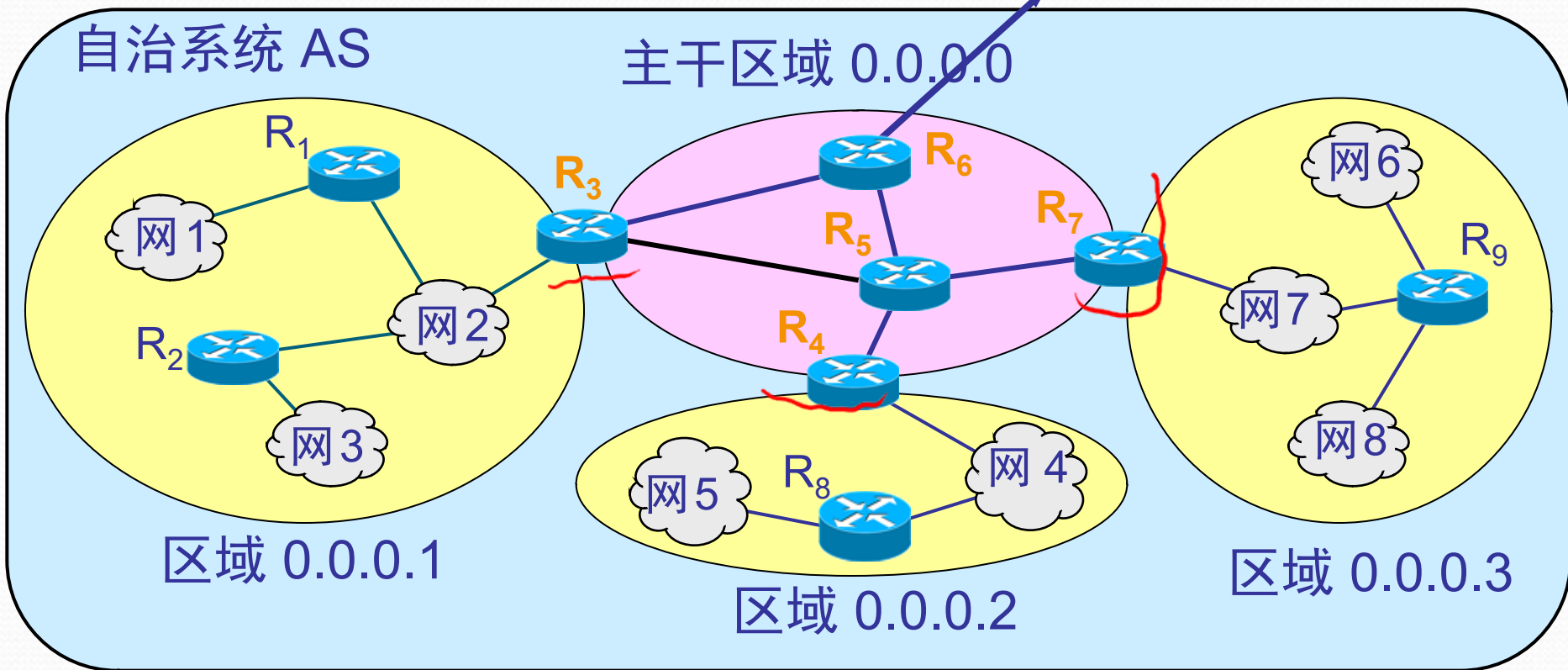


划分区域

- 划分区域的好处就是将利用洪泛法交换链路状态信息的范围局限于每一个区域而不是整个的自治系统，这就减少了整个网络上的通信量。
- 在一个区域内部的路由器只知道本区域的完整网络拓扑，而不知道其他区域的网络拓扑的情况。
- OSPF 使用层次结构的区域划分。在上层的区域叫作主干区域 (backbone area)。主干区域的标识符规定为0.0.0.0。主干区域的作用是用来连通其他在下层的区域。

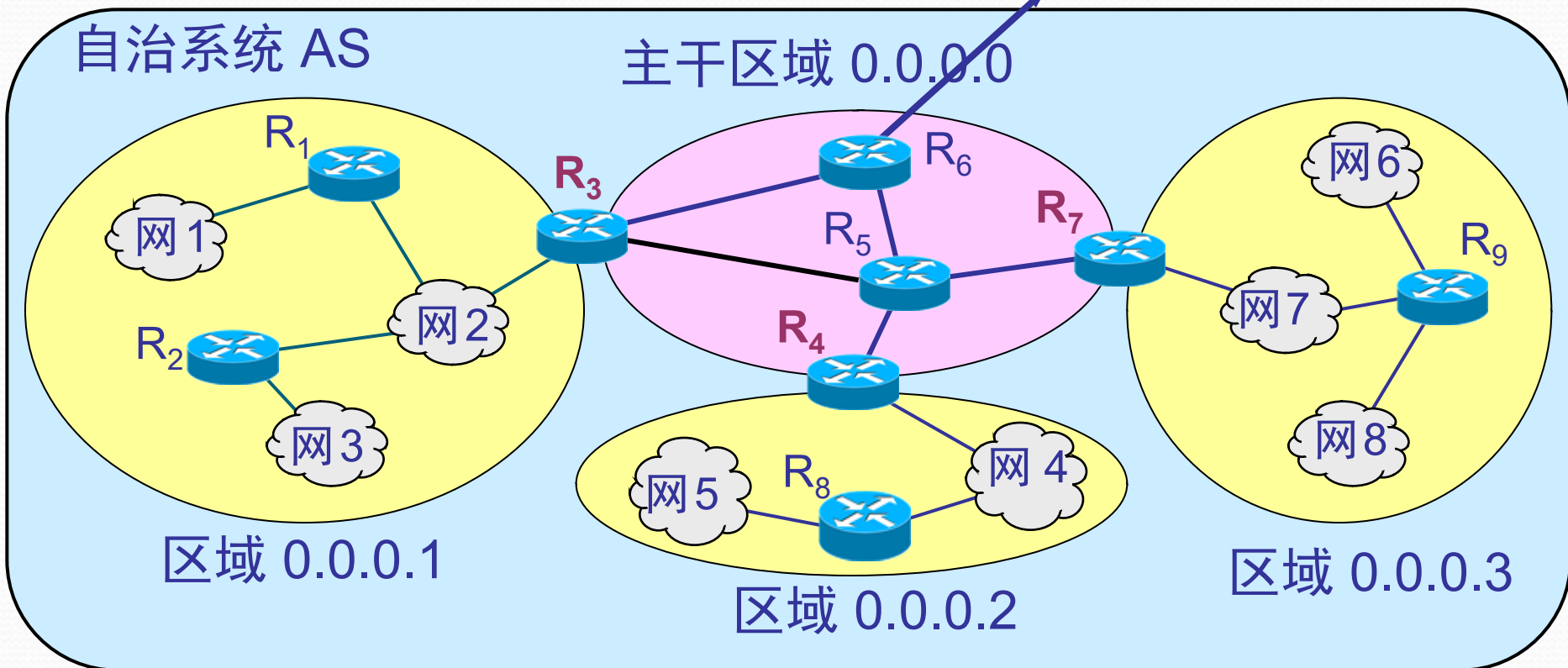
主干路由器

至其他自治系统



区域边界路由器

至其他自治系统



OSPF 直接用 IP 数据报传送

- OSPF 不用 UDP 而是直接用 IP 数据报传送。
- OSPF 构成的数据报很短。这样做可减少路由信息的通信量。
- 数据报很短的另一好处是可以不必将长的数据报分片传送。分片传送的数据报只要丢失一个，就无法组装成原来的数据报，而整个数据报就必须重传。

OSPF 的其他特点

- OSPF 对不同的链路可根据 IP 分组的不同服务类型 TOS 而设置成不同的代价。因此，OSPF 对于不同类型的业务可计算出不同的路由。
- 如果到同一个目的网络有多条相同代价的路径，那么可以将通信量分配给这几条路径。这叫作多路径间的负载平衡。
- 所有在 OSPF 路由器之间交换的分组都具有鉴别的功能。
- 支持可变长度的子网划分和无分类编址 CIDR。
- 每一个链路状态都带上一个 32 位的序号，序号越大状态就越新。

仿真 链接 防火墙 - gang of death - echo

配置好了，但是收包 对√ 数据

OSPF 分组

位 0 8 16 31

V2, V3

版 本	类 型✓	分 组 长 度✓
路 由 器 标 识 符✓		
区 域 标 识 符✓		
检 验 和✓	鉴 别 类 型✓	
鉴 别		
鉴 别		

24 字节

OSPF 分组首部

类型 1 至类型 5 的 OSPF 分组



IP数据报首部

OSPF 分组

IP 数据报

2. OSPF 的五种分组类型

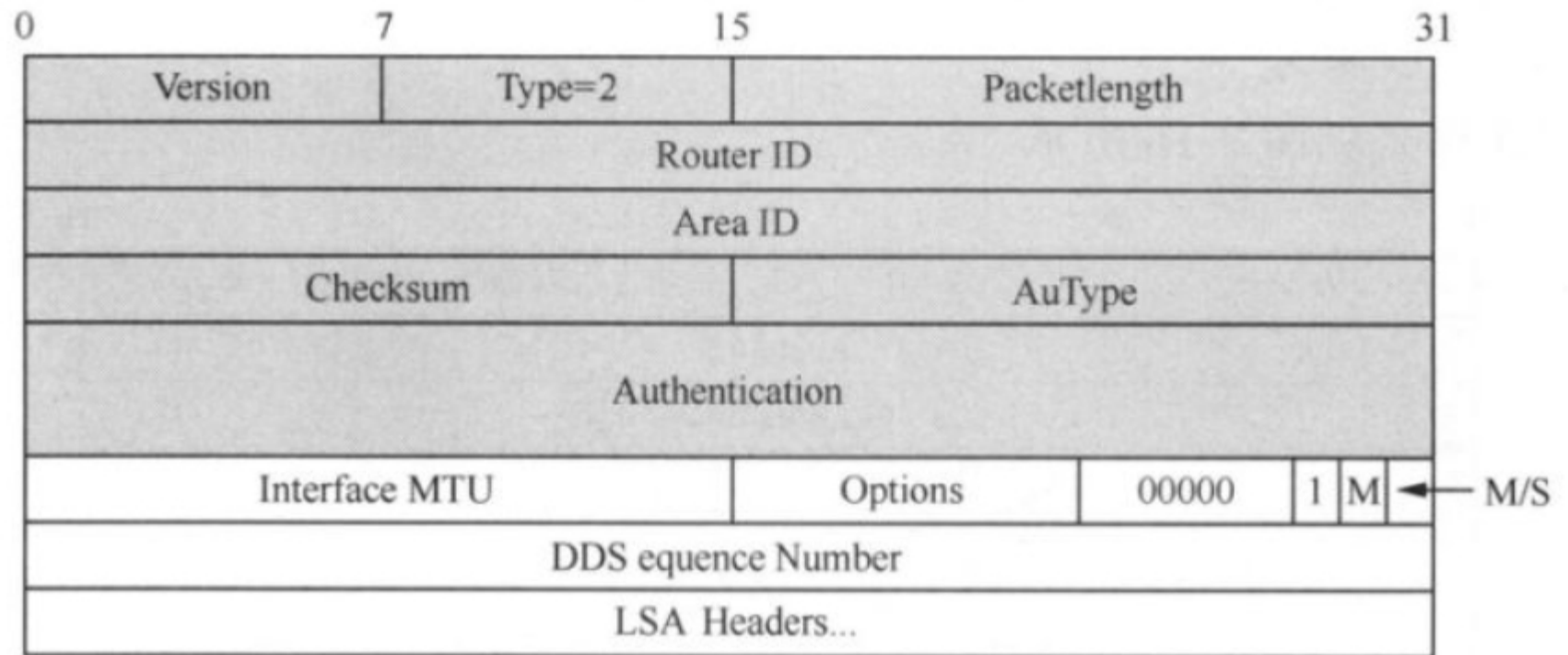
- 类型1, 问候(Hello)分组。
- 类型2, 数据库描述(Database Description)分组。
- 类型3, 链路状态请求(Link State Request)分组。
- 类型4, 链路状态更新(Link State Update)分组,
用洪泛法对全网更新链路状态。
- 类型5, 链路状态确认(Link State Acknowledgment)
分组。

链路状态数据库 25TB
LSDB

新

Link-State Advertisement

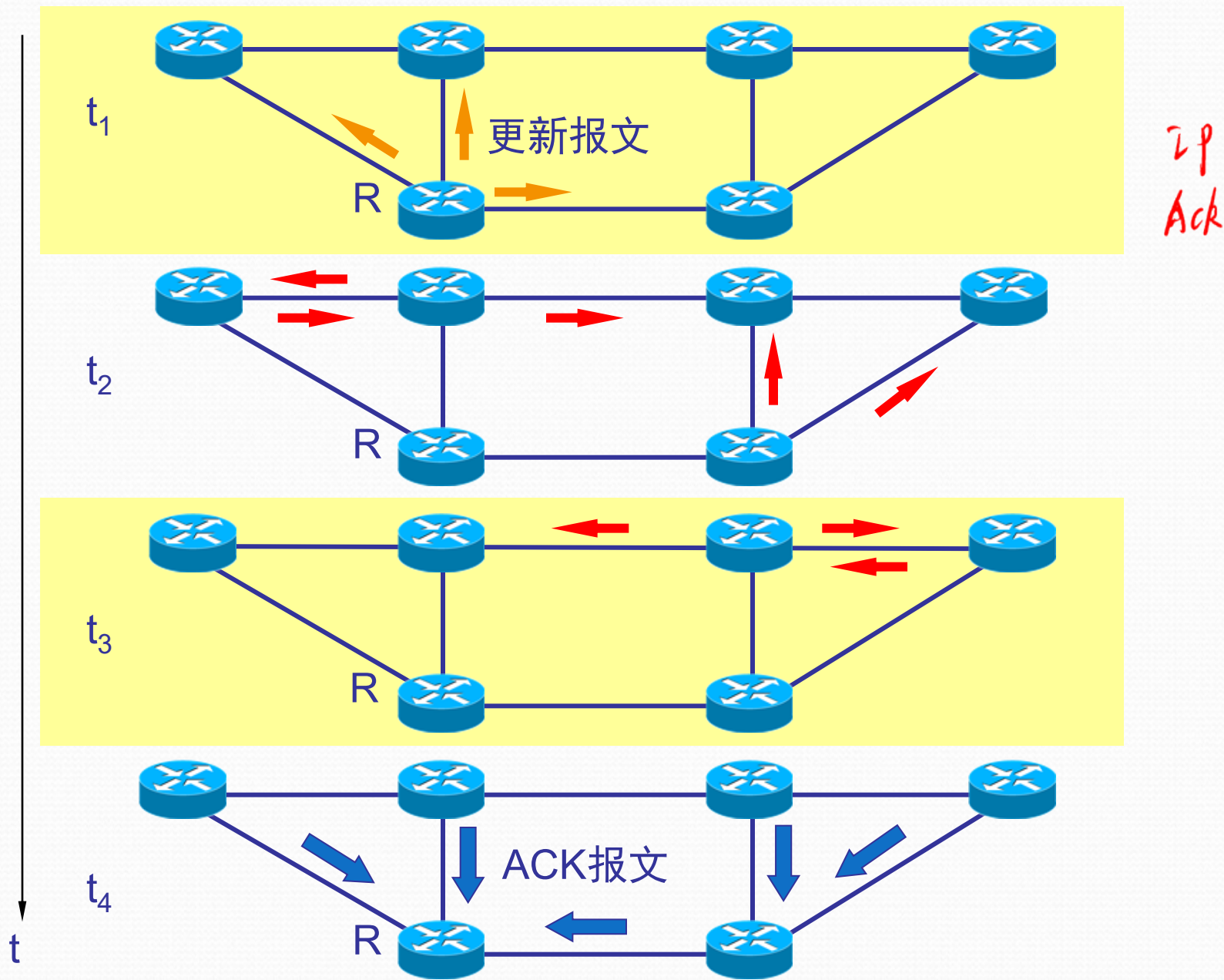
0	7	15	31
Version		Type=1	Packetlength
Router ID			
Area ID			
Checksum		AuType	
Authentication			
Network Mask			
HelloInterval		Options	Rtr Pri
Router DeadInterval			
Designated Router			
Backup Designated Router			
Neighbor			
...			



OSPF的基本操作



OSPF 使用的是可靠的洪泛法



OSPF 的其他特点

- OSPF 还规定每隔一段时间，如 30 分钟，要刷新一次数据库中的链路状态。
- 由于一个路由器的链路状态只涉及到与相邻路由器的连通状态，因而与整个互联网的规模并无直接关系。因此当互联网规模很大时，OSPF 协议要比距离向量协议 RIP 好得多。
- OSPF 没有“坏消息传播得慢”的问题，据统计，其响应网络变化的时间小于 100 ms。

指定的路由器

(designated router)

- 多点接入的局域网采用了指定的路由器的方法，使广播的信息量大大减少。
- 指定的路由器代表该局域网上的所有的链路向连接到该网络上的各路由器发送状态信息。

RIPng v2v6
OSPFv3

4.5.4 外部网关协议 BGP

ZGP — 中的

- BGP 是不同自治系统的路由器之间交换路由信息的协议。
- BGP 较新版本是 2006 年 1 月发表的 BGP-4（BGP 第 4 个版本），即 RFC 4271 ~ 4278。
- 可以将 BGP-4 简写为 BGP。

BGP 使用的环境却不同

- 因特网的规模太大，使得自治系统之间路由选择非常困难。对于自治系统之间的路由选择，要寻找最佳路由是很不现实的。
 - 当一条路径通过几个不同 AS 时，要想对这样的路径计算出有意义的代价是不太可能的。
 - 比较合理的做法是在 AS 之间交换“可达性”信息。
- 自治系统之间的路由选择必须考虑有关策略。
- 因此，边界网关协议 BGP 只能是力求寻找一条能够到达目的网络且**比较好的路由**（不能兜圈子），而**并非要寻找一条最佳路由**。

BGP 发言人

(BGP speaker)

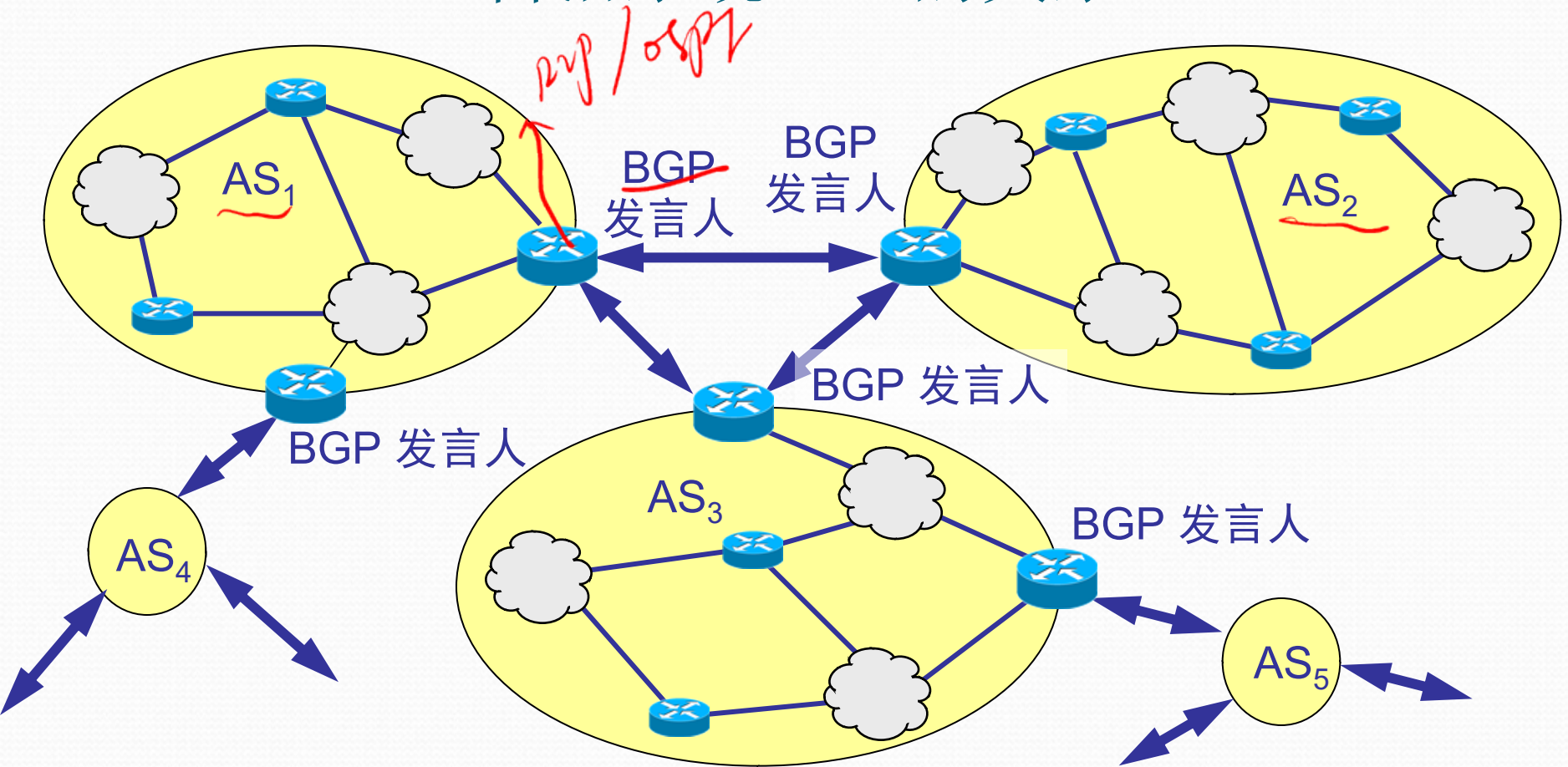
- 每一个自治系统的管理员要选择至少一个路由器作为该自治系统的“**BGP 发言人**”。
- 一般说来，两个 BGP 发言人都是通过一个共享网络连接在一起的，而 BGP 发言人往往就是 BGP 边界路由器，但也可以不是 BGP 边界路由器。

rip - udp, ospf - ip
BGP - tcp

BGP 交换路由信息

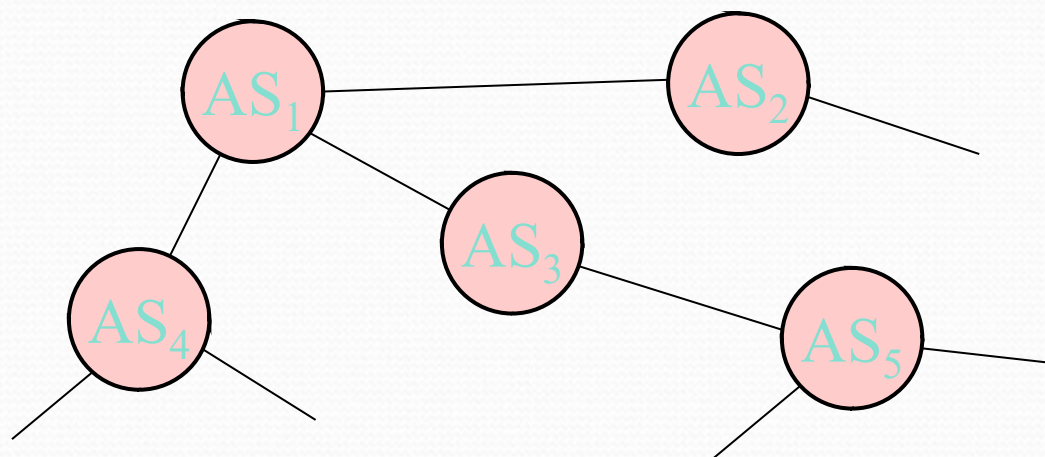
- 一个 BGP 发言人与其他自治系统中的 BGP 发言人要交换路由信息，就要先建立 TCP 连接，然后在此连接上交换 BGP 报文以建立 BGP 会话(session)，利用 BGP 会话交换路由信息。
- 使用 TCP 连接能提供可靠的服务，也简化了路由选择协议。
- 使用 TCP 连接交换路由信息的两个 BGP 发言人，彼此成为对方的邻站或对等站。

BGP 发言人和 自治系统 AS 的关系



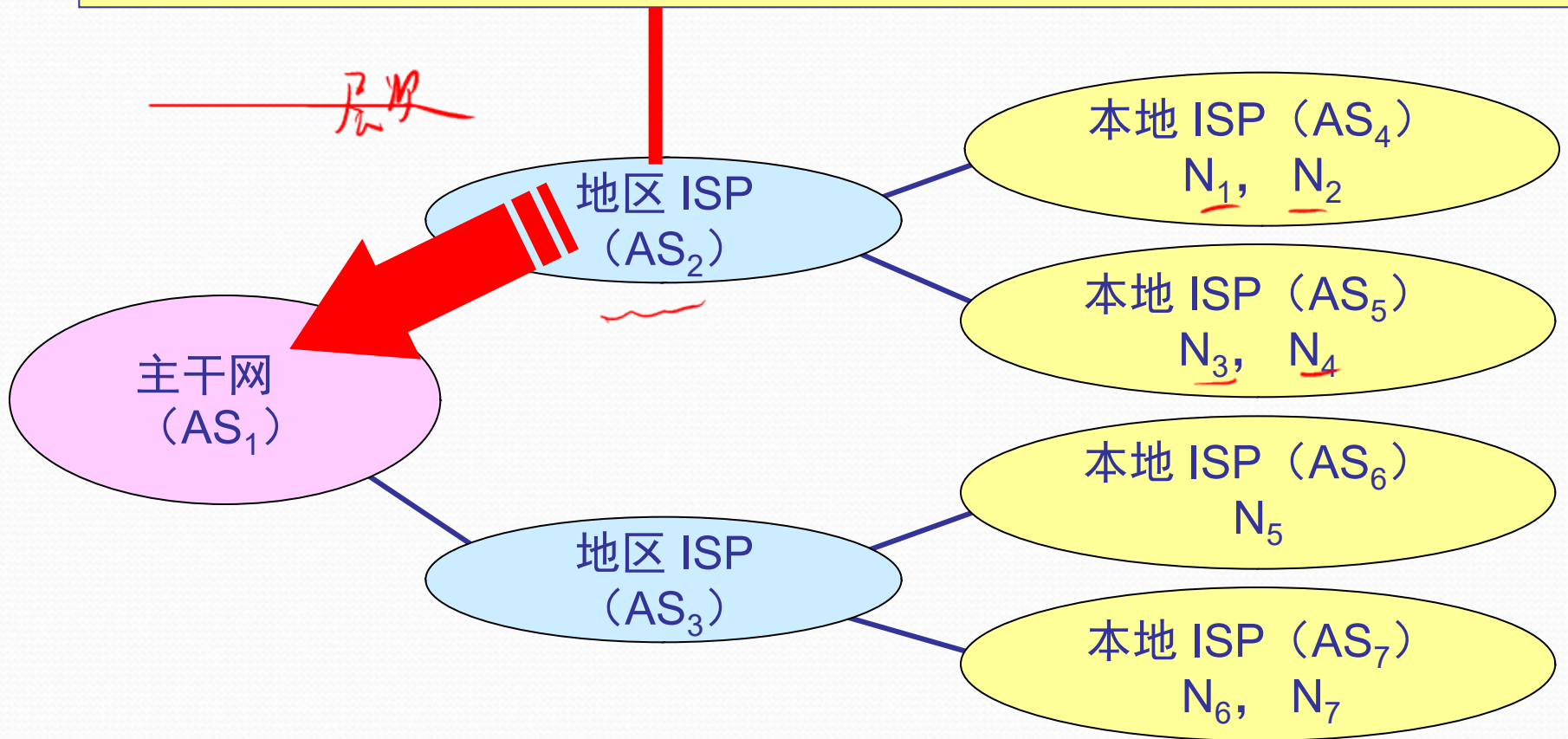
AS 的连通图举例

- BGP 所交换的网络可达性的信息就是要到达某个网络所要经过的一系列 AS。
- 当 BGP 发言人互相交换了网络可达性的信息后，各 BGP 发言人就根据所采用的策略从收到的路由信息中找出到达各 AS 的较好路由。



BGP 发言人交换路径向量

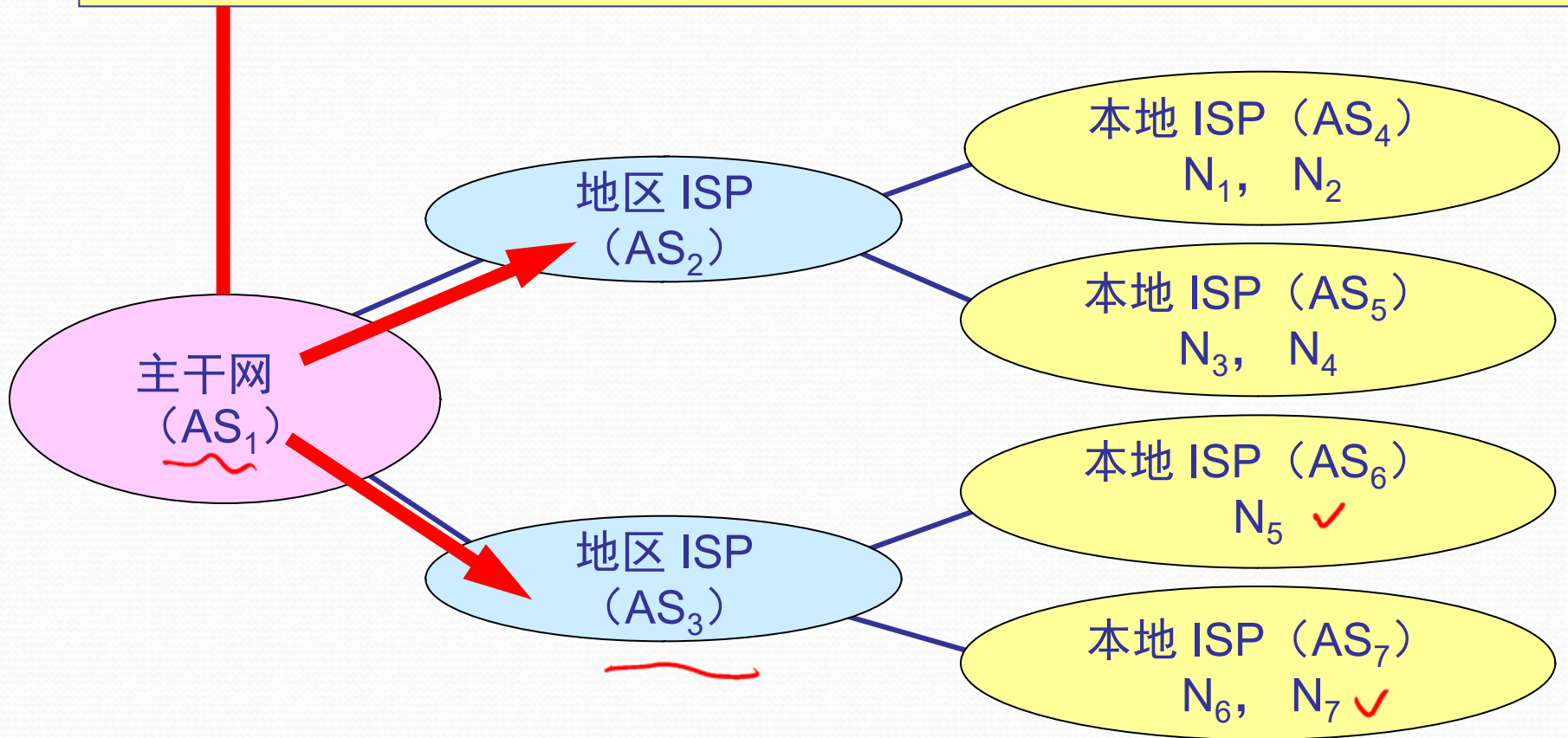
自治系统 AS_2 的 BGP 发言人通知主干网的 BGP 发言人：“要到达网络 N_1 , N_2 , N_3 和 N_4 可经过 AS_2 。”



BGP 发言人交换路径向量

EGP 的扩展

主干网还可发出通知：“要到达网络 N_5 , N_6 和 N_7 可沿路径 (AS_1, AS_3) 。”



BGP 协议的特点

- BGP 协议交换路由信息的结点数量级是自治系统数的量级，这要比这些自治系统中的网络数少很多。
- 每一个自治系统中 BGP 发言人（或边界路由器）的数目是很少的。这样就使得自治系统之间的路由选择不致过分复杂。

2/3/② BGP 协议的特点

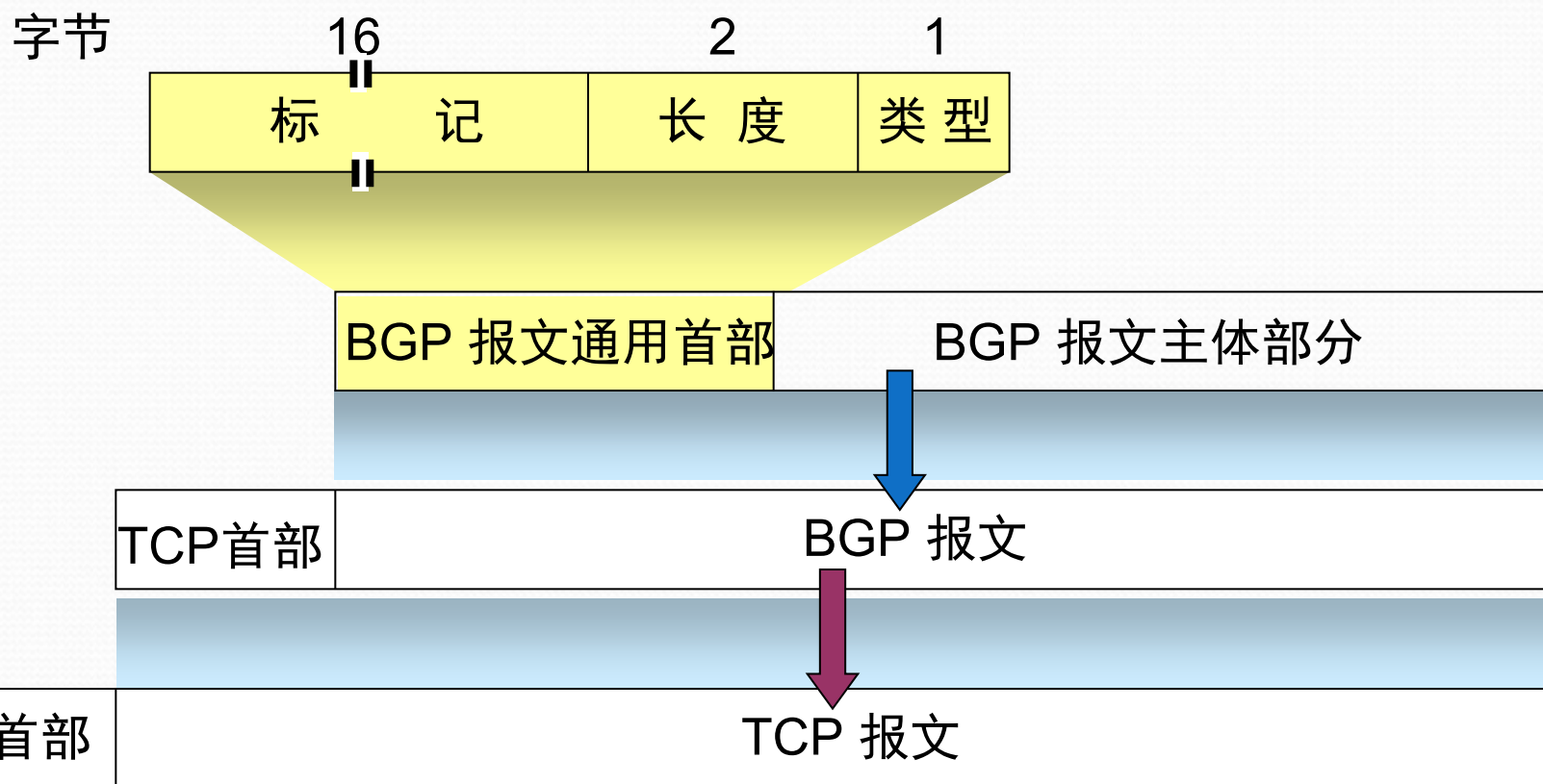
TCP - ?
ospf - ip
Rip - udp

- BGP 支持 CIDR，因此 BGP 的路由表也就应当包括目的网络前缀、下一跳路由器，以及到达该目的网络所要经过的各个自治系统序列。
- 在 BGP 刚刚运行时，BGP 的邻站是交换整个的 BGP 路由表。但以后只需要在发生变化时 更新有变化的部分。这样做对节省网络带宽和减少路由器的处理开销方面都有好处。

BGP-4 共使用四种报文

- (1) 打开(**OPEN**)报文，用来与相邻的另一个BGP发言人建立关系。
 - (2) 更新(**UPDATE**)报文，用来发送某一路由的信息，以及列出要撤消的多条路由。*更新的发送*
 - (3) 保活(**KEEPALIVE**)报文，用来确认打开报文和周期性地证实邻站关系。
 - (4) 通知(**NOTIFICATION**)报文，用来发送检测到的差错。
- 在 RFC 2918 中增加了 ROUTE-REFRESH 报文，用来请求对等端重新通告。

BGP 报文具有通用的首部



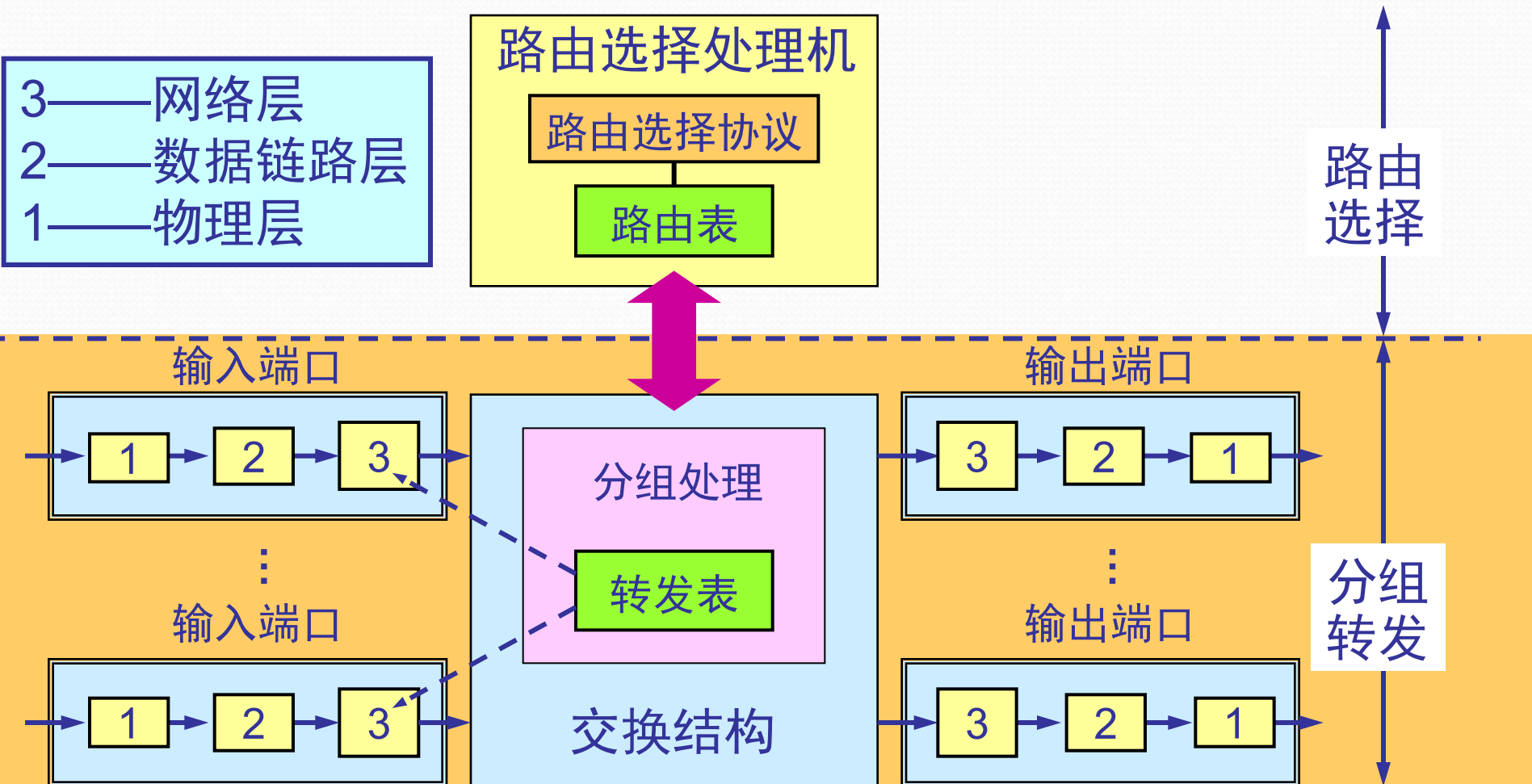
4.5.6

路由器在网际互连中的作用

1. 路由器的结构

- 路由器是一种具有多个输入端口和多个输出端口的专用计算机，其任务是转发分组。也就是说，将路由器某个输入端口收到的分组，按照分组要去的目的地（即目的网络），把该分组从路由器的某个合适的输出端口转发给下一跳路由器。
- 下一跳路由器也按照这种方法处理分组，直到该分组到达终点为止。

典型的路由器的结构



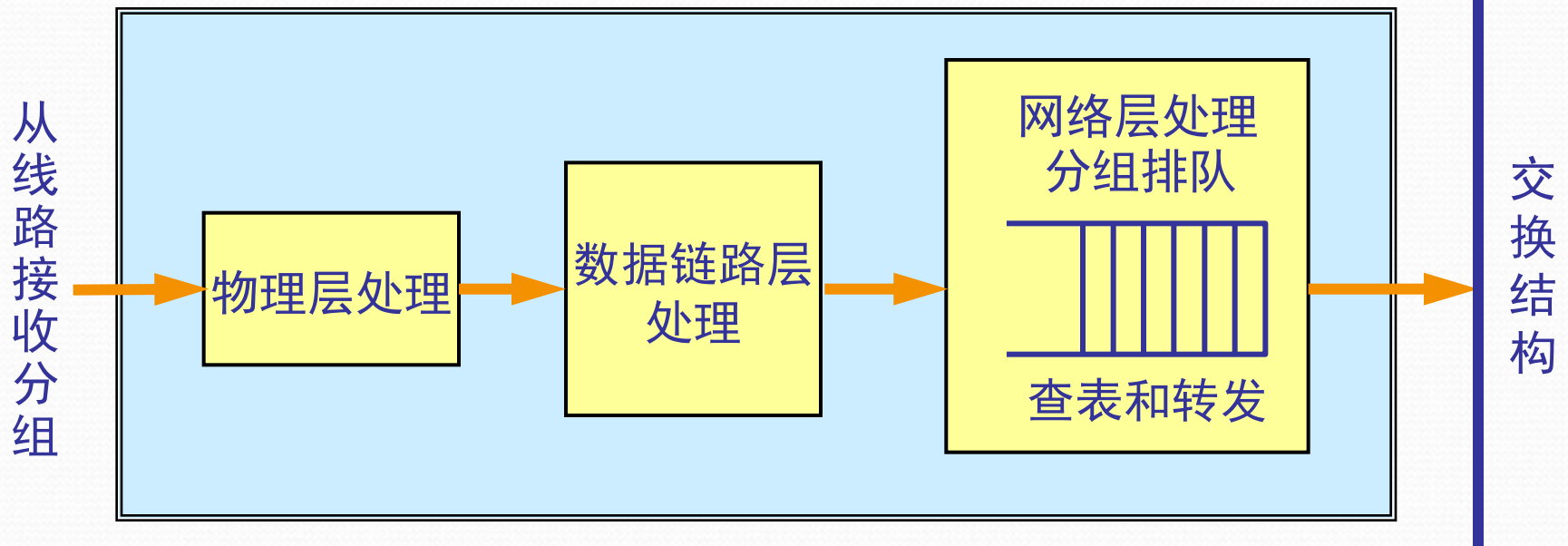
“转发”和“路由选择”的区别

- “**转发**”(forwarding)就是路由器根据转发表将用户的 IP 数据报从合适的端口转发出去。
- “**路由选择**”(routing)则是按照分布式算法，根据从各相邻路由器得到的关于网络拓扑的变化情况，动态地改变所选择的路由。
- 路由表是根据路由选择算法得出的。而转发表是从路由表得出的。
- 在讨论路由选择的原理时，往往不去区分转发表和路由表的区别，

输入端口对线路上收到的分组的处理

- 数据链路层剥去帧首部和尾部后，将分组送到网络层的队列中排队等待处理。这会产生一定的时延。

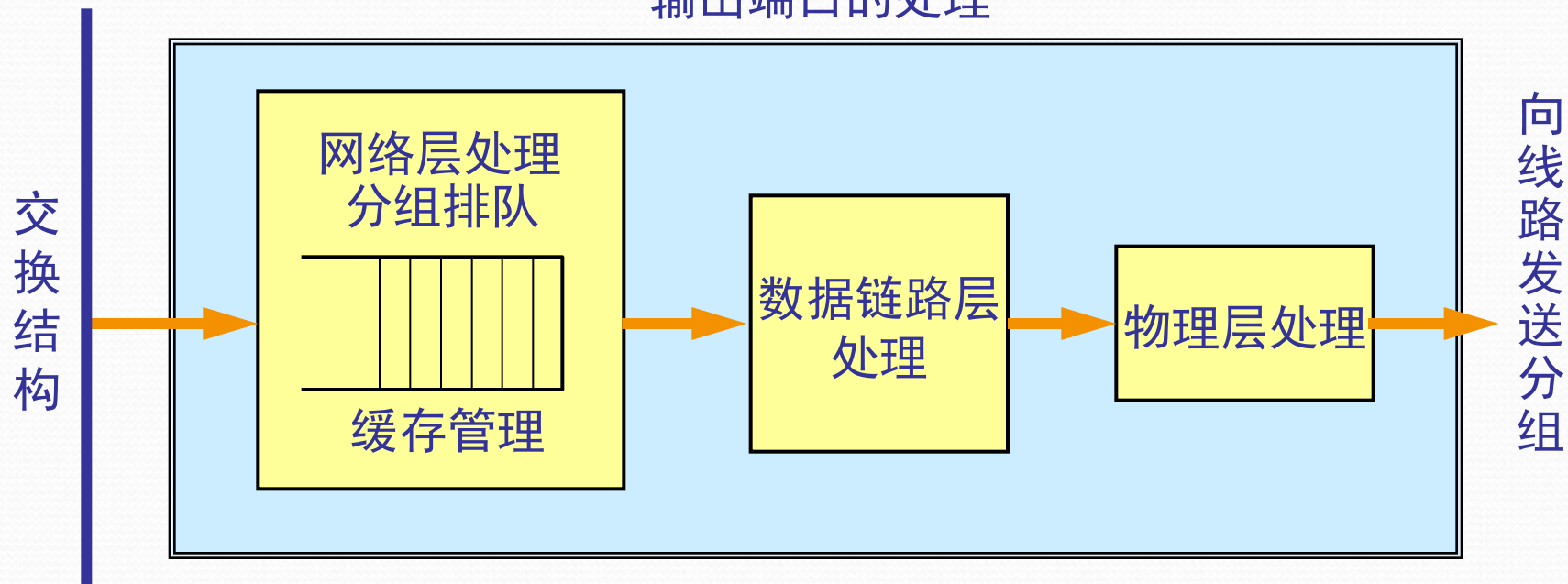
输入端口的处理



输出端口将交换结构传送来的分组发送到线路

- 当交换结构传送过来的分组先进行缓存。数据链路层处理模块将分组加上链路层的首部和尾部，交给物理层后发送到外部线路。

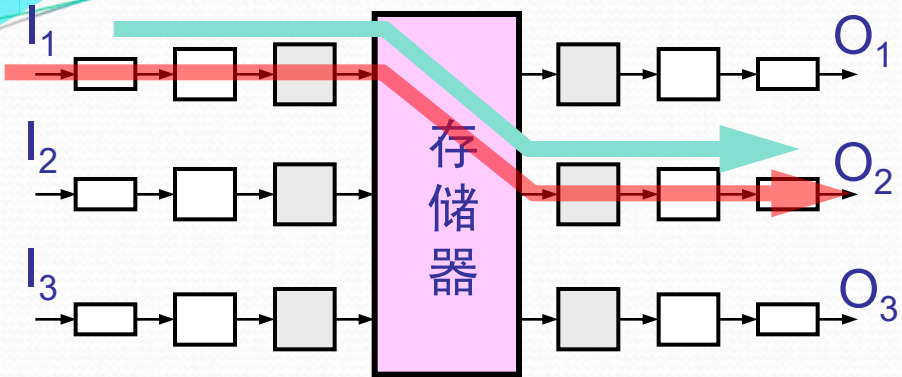
输出端口的处理



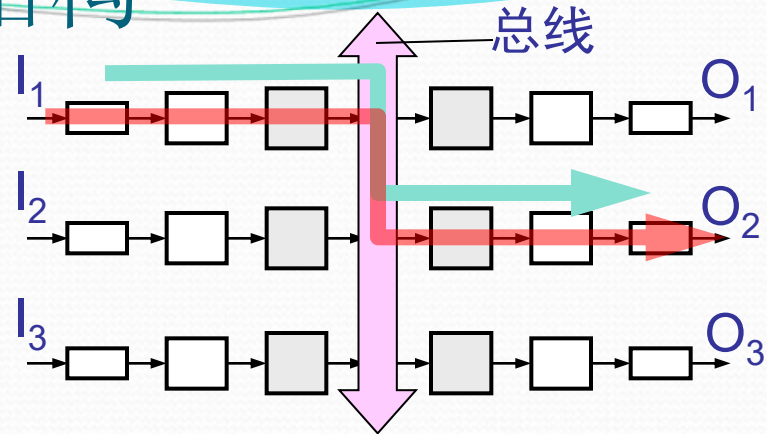
分组丢弃

- 若路由器处理分组的速率赶不上分组进入队列的速率，则队列的存储空间最终必定减少到零，这就使后面再进入队列的分组由于没有存储空间而只能被丢弃。
- 路由器中的输入或输出队列产生溢出是造成分组丢失的重要原因。

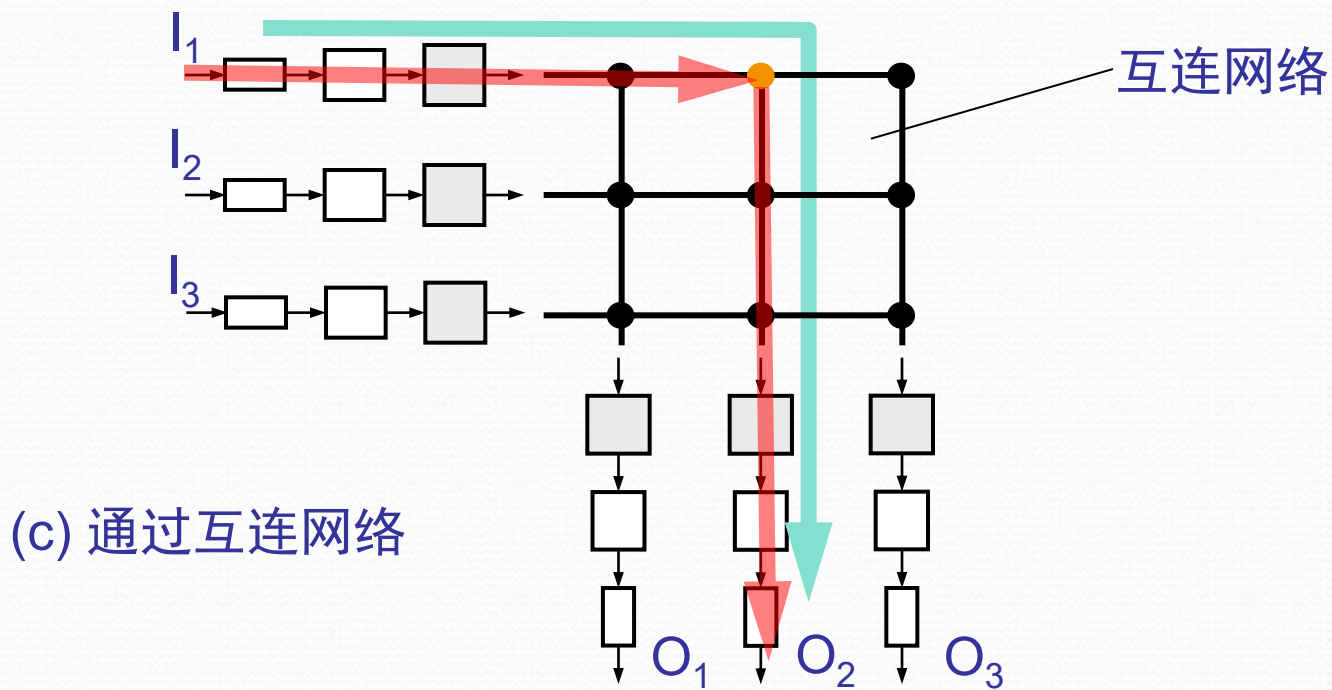
2 交换结构



(a) 通过存储器



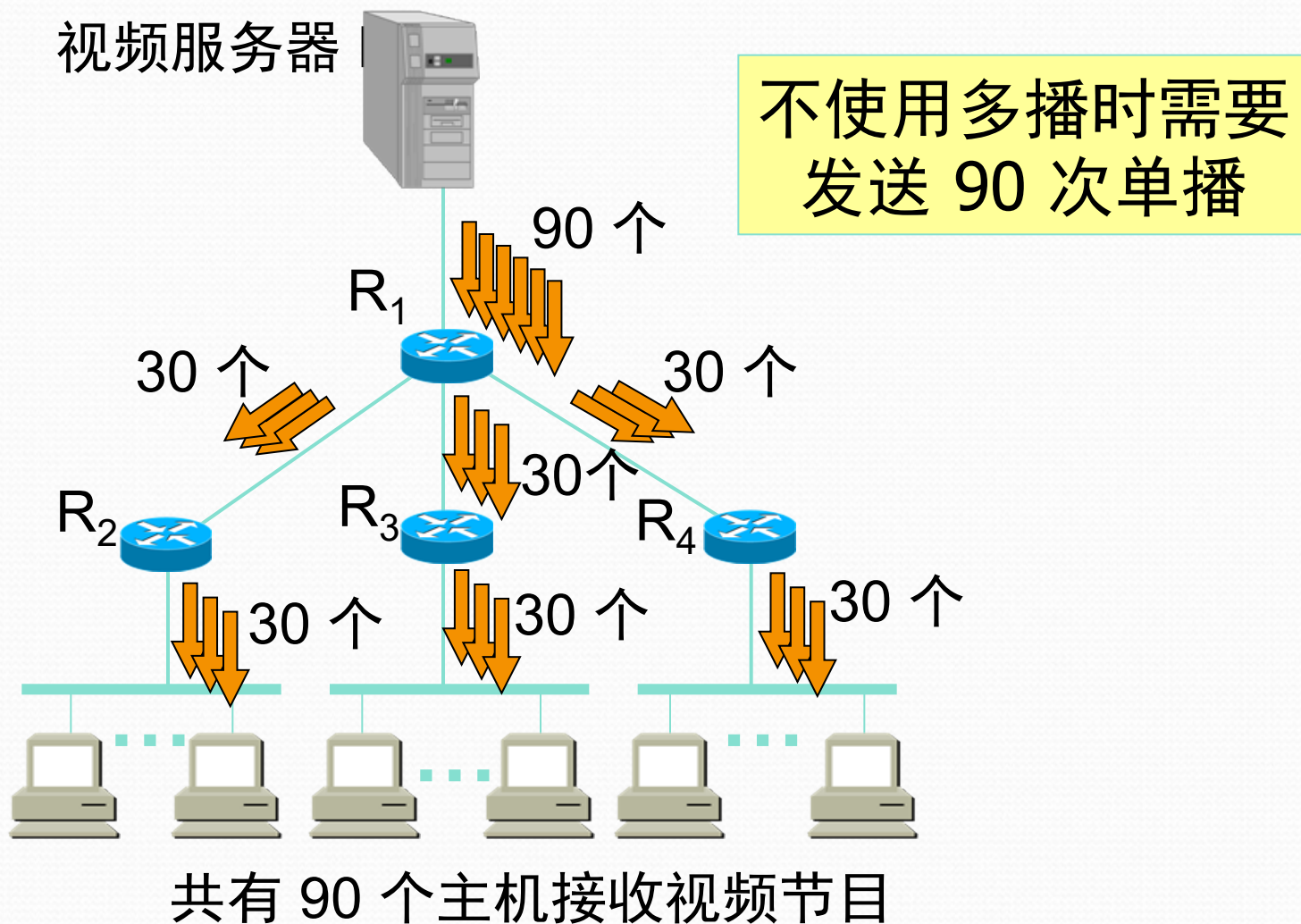
(b) 通过总线



(c) 通过互连网络

4.6 IP 多播

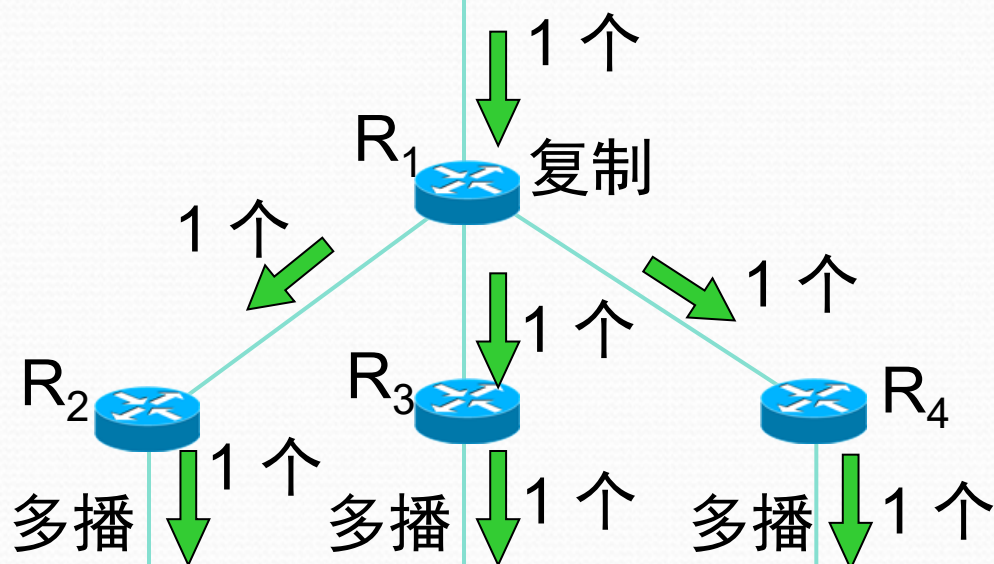
4.6.1 IP 多播的基本概念



多播可明显地减少 网络中资源的消耗

视频服务器 I

发送 1 次多播



多播组成员
共有 90 个

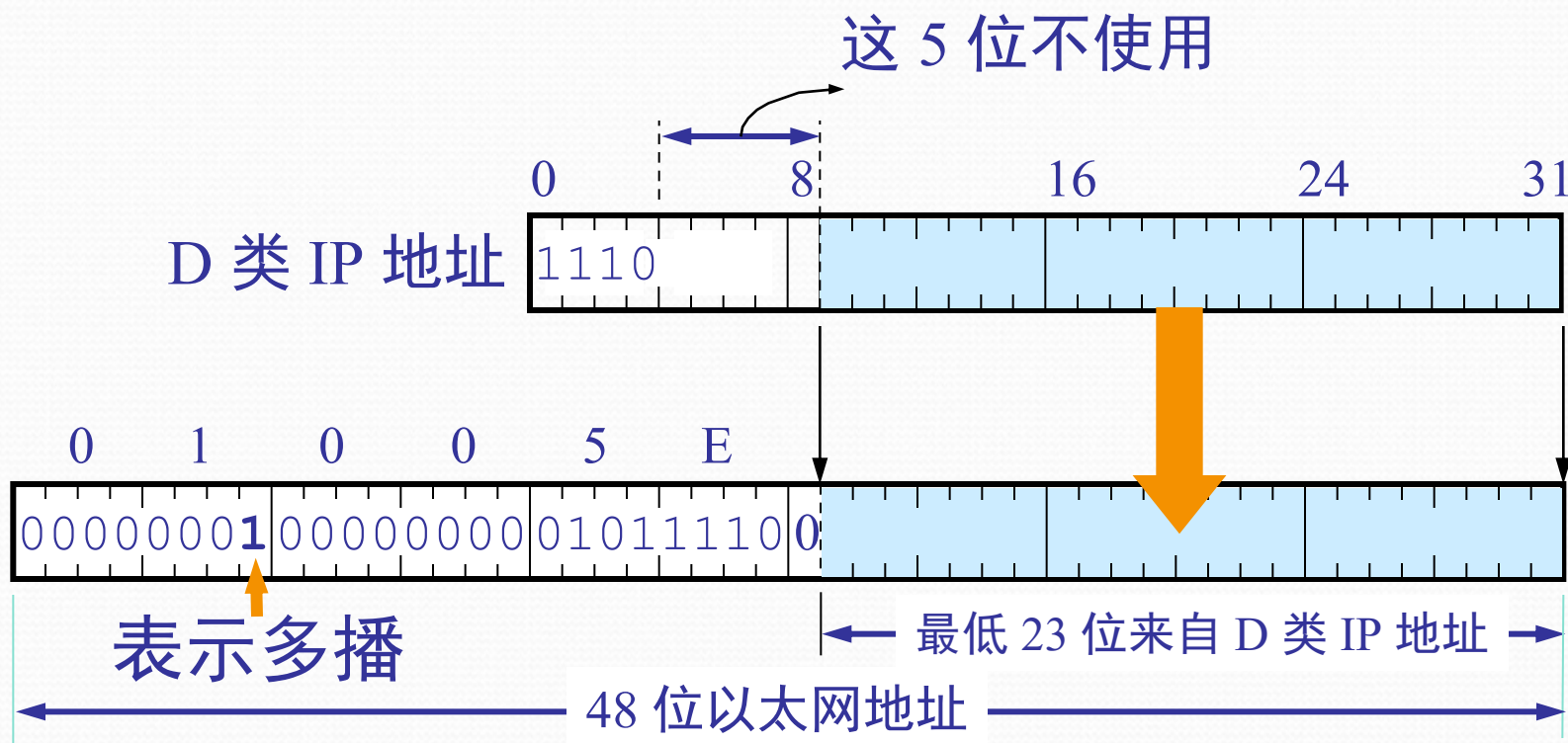
IP 多播的一些特点

- (1) 多播使用组地址——IP 使用 D 类地址支持多播。多播地址只能用于目的地址，而不能用于源地址。
- (2) 永久组地址——由因特网号码指派管理局 IANA 负责指派。
- (3) 动态的组成员
- (4) 使用硬件进行多播

4.6.2 在局域网上进行硬件多播

- 因特网号码指派管理局 IANA 拥有的以太网地址块的高 24 位为 00-00-5E。
- 因此 TCP/IP 协议使用的以太网多播地址块的范围是：从 00-00-5E-00-00-00
到 00-00-5E-FF-FF-FF
- D 类 IP 地址可供分配的有 28 位，在这 28 位中的前 5 位不能用来构成以太网硬件地址。

D 类 IP 地址 与以太网多播地址的映射关系



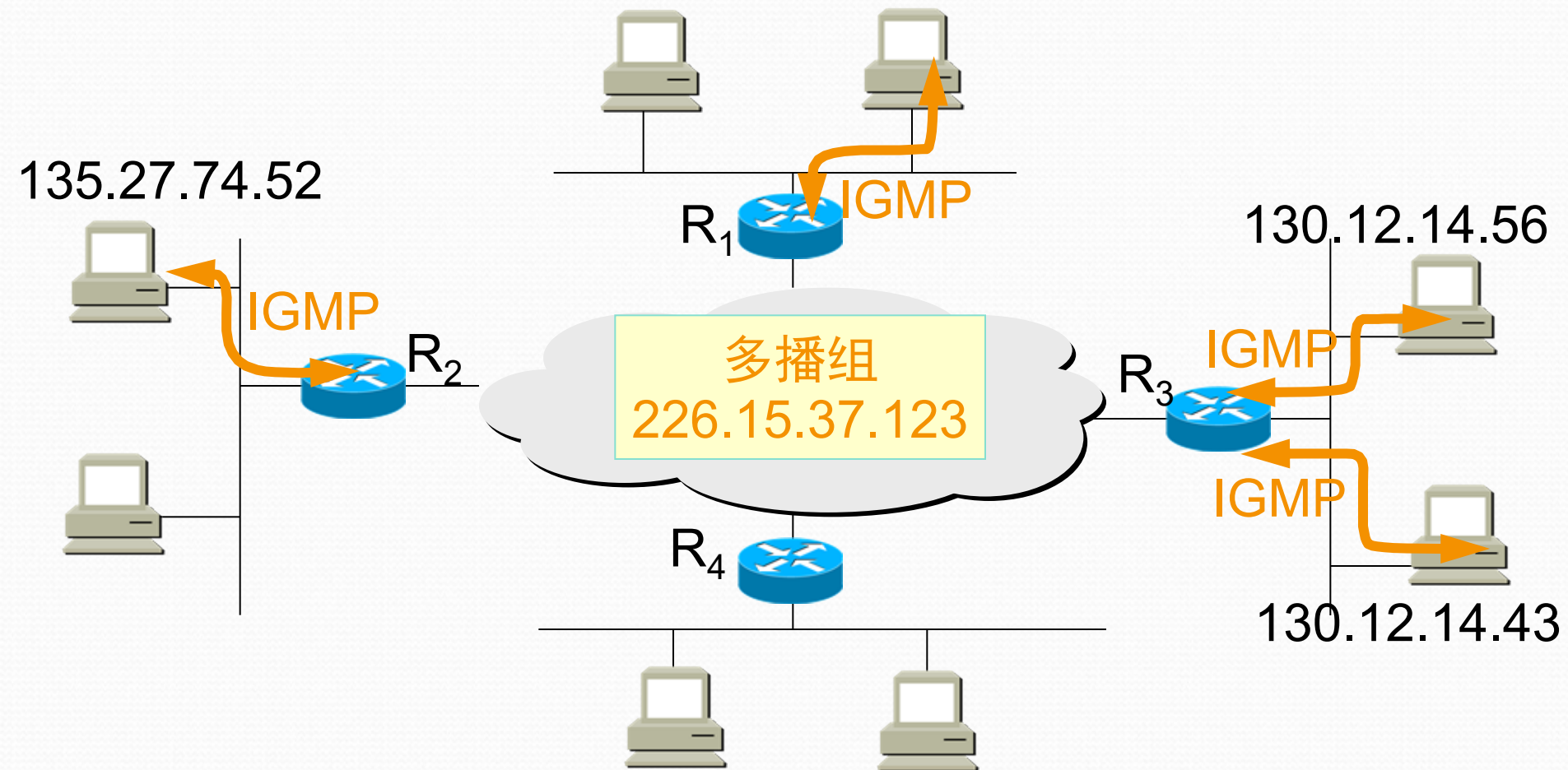
4.6.3 网际组管理协议 IGMP 和多播路由选择协议

1. IP多播需要两种协议

- 为了使路由器知道多播组成员的信息，需要利用网际组管理协议 IGMP (Internet Group Management Protocol)。
- 连接在局域网上的多播路由器还必须和因特网上的其他多播路由器协同工作，以便把多播数据报用最小代价传送给所有的组成员。这就需要使用多播路由选择协议。

IGMP 使多播路由器 知道多播组成员信息

128.56.24.34



IGMP 的本地使用范围

- IGMP 并非在因特网范围内对所有多播组成员进行管理的协议。
- IGMP 不知道 IP 多播组包含的成员数，也不知道这些成员都分布在哪些网络上。
- IGMP 协议是让连接在本地局域网上的多播路由器知道本局域网上是否有主机（严格讲，是主机上的某个进程）参加或退出了某个多播组。

多播路由选择协议

比单播路由选择协议复杂得多

- 多播转发必须**动态地**适应多播组成员的变化（这时网络拓扑并未发生变化）。请注意，单播路由选择通常是在网络拓扑发生变化时才需要更新路由。
- 多播路由器在转发多播数据报时，不能仅仅根据多播数据报中的目的地址，而是还要考虑这个多播数据报从什么地方来和要到什么地方去。
- 多播数据报可以由没有加入多播组的主机发出，也可以通过没有组成员接入的网络。

2. 网际组管理协议 IGMP

- 1989 年公布的 RFC 1112（IGMPv1）早已成为了因特网的标准协议。
- 1997 年公布的 RFC 2236（IGMPv2，建议标准）对 IGMPv1 进行了更新。
- 2002 年 10 月公布了 RFC 3376（IGMPv3，建议标准），宣布 RFC 2236（IGMPv2）是陈旧的。

IGMP 是整个网际协议 IP 的一个组成部分

- 和 ICMP 相似，IGMP 使用 IP 数据报传递其报文（即 IGMP 报文加上 IP 首部构成 IP 数据报），但它也向 IP 提供服务。
- 因此，我们不把 IGMP 看成是一个单独的协议，而是属于整个网际协议 IP 的一个组成部分。

IGMP 可分为两个阶段

- 第一阶段：当某个主机加入新的多播组时，该主机应向多播组的多播地址发送IGMP 报文，声明自己要成为该组的成员。本地的多播路由器收到 IGMP 报文后，将组成员关系转发给因特网上的其他多播路由器。

IGMP 可分为两个阶段

- 第二阶段：因为组成员关系是动态的，因此本地多播路由器要周期性地探测本地局域网上的主机，以便知道这些主机是否还继续是组的成员。
- 只要对某个组有一个主机响应，那么多播路由器就认为这个组是活跃的。
- 但一个组在经过几次的探测后仍然没有一个主机响应，则不再将该组的成员关系转发给其他的多播路由器。

IGMP 采用的一些具体措施

- 在主机和多播路由器之间的所有通信都是使用 IP 多播。
- 多播路由器在探询组成员关系时，只需要对所有的组发送一个请求信息的询问报文，而不需要对每一个组发送一个询问报文。默认的询问速率是每 125 秒发送一次。
- 当同一个网络上连接有几个多播路由器时，它们能够迅速和有效地选择其中的一个来探询主机的成员关系。

IGMP 采用的一些具体措施（续）

- 在 IGMP 的询问报文中有一个数值 N ，它指明一个最长响应时间（默认值为 10 秒）。当收到询问时，主机在 0 到 N 之间随机选择发送响应所需经过的时延。对应于最小时延的响应最先发送。
- 同一个组内的每一个主机都要监听响应，只要有本组的其他主机先发送了响应，自己就可以不再发送响应了。

3. 多播路由选择

- 多播路由选择协议尚未标准化。
- 一个多播组中的成员是动态变化的，随时会有主机加入或离开这个多播组。
- 多播路由选择实际上就是要找出以源主机为根结点的多播转发树。
- 在多播转发树上的路由器不会收到重复的多播数据报。
- 对不同的多播组对应于不同的多播转发树。同一个多播组，对不同的源点也会有不同的多播转发树。

转发多播数据报使用的方法

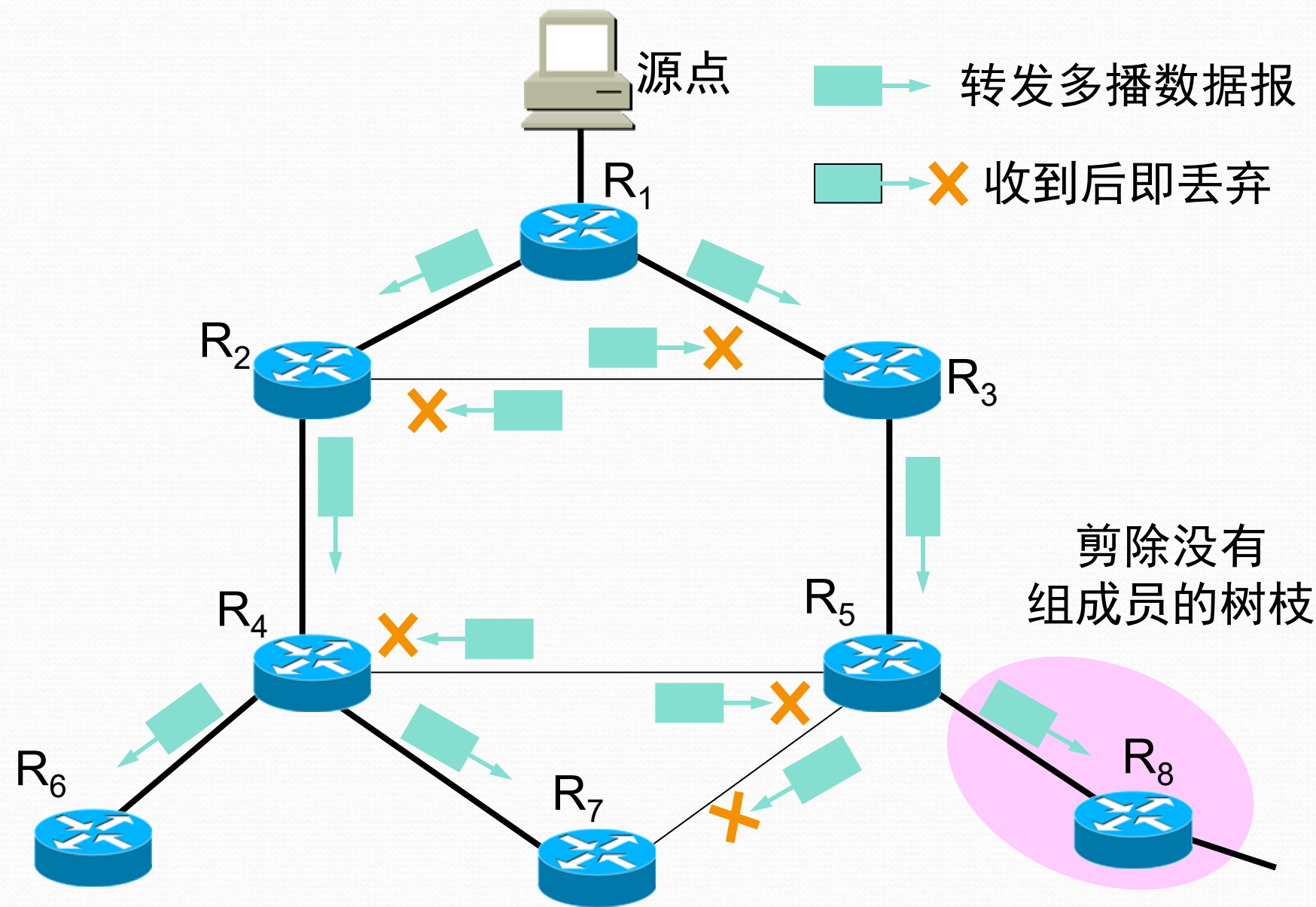
(1) 洪泛与剪除

- 这种方法适合于较小的多播组，而所有的组成员接入的局域网也是相邻接的。
- 一开始，路由器转发多播数据报使用洪泛的方法（这就是广播）。为了避免兜圈子，采用了叫做**反向路径广播 RPB** (Reverse Path Broadcasting) 的策略。

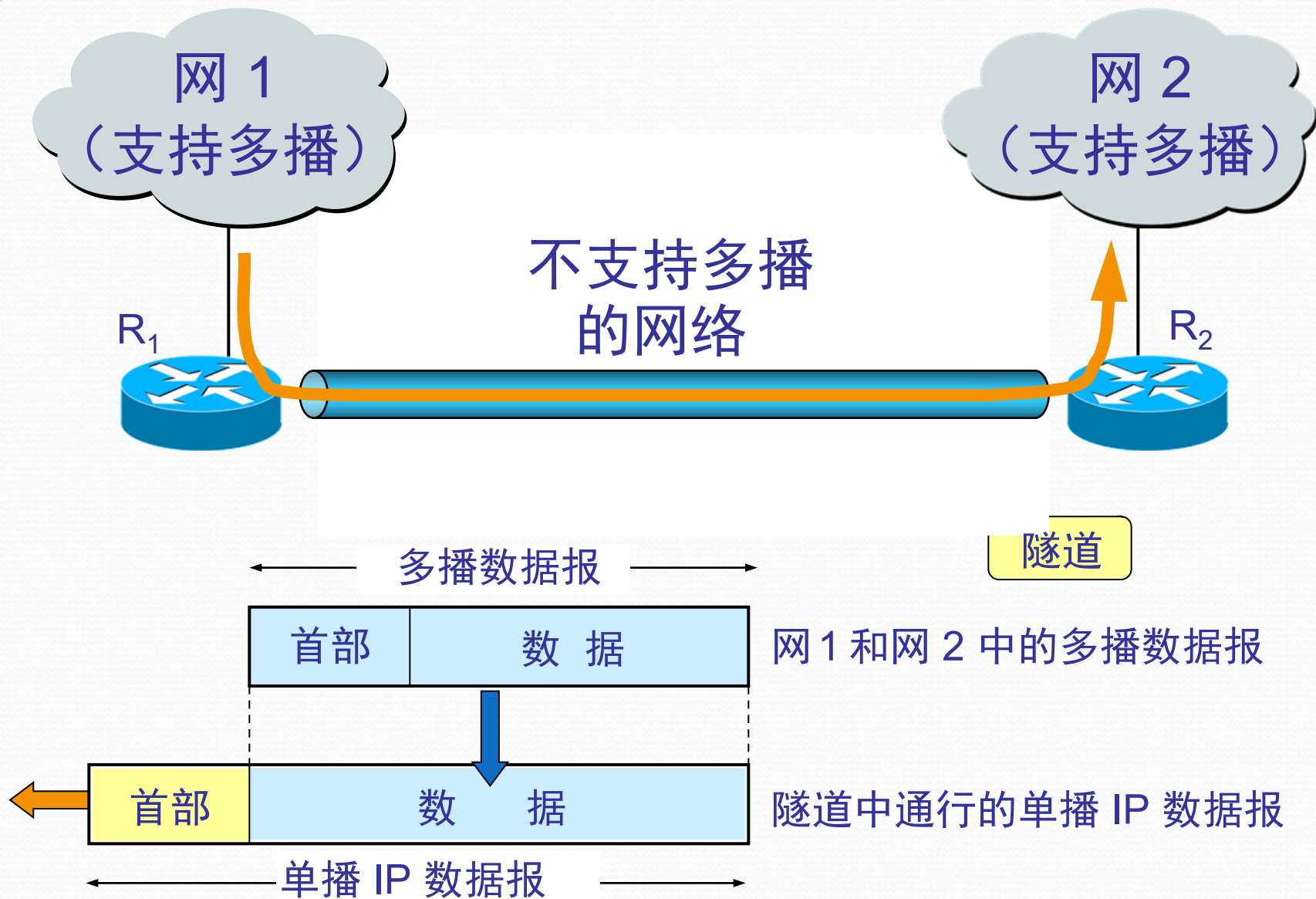
RPB 的要点

- 路由器收到多播数据报时，先检查是否从源点经最短路径传送来的。
- 若是，就向所有其他方向转发刚才收到的多播数据报（但进入的方向除外），否则就丢弃而不转发。
- 如果存在几条同样长度的最短路径），那么只能选择一条最短路径，选择的准则就是看这几条最短路径中的相邻路由器谁的 IP 地址最小。

反向路径广播 RPB 和剪除



(2) 隧道技术(tunneling)



(3) 基于核心的发现技术

- 这种方法对于多播组的大小在较大范围内变化时都适合。
- 这种方法是对每一个多播组 G 指定一个核心(core)路由器，给出它的 IP 单播地址。
- 核心路由器按照前面讲过的方法创建出对应于多播组 G 的转发树。

几种多播路由选择协议

- 距离向量多播路由选择协议 DVMRP (Distance Vector Multicast Routing Protocol)
- 基于核心的转发树 CBT (Core Based Tree)
- 开放最短通路优先的多播扩展 MOSPF (Multicast Extensions to OSPF)
- 协议无关多播-稀疏方式 PIM-SM (Protocol Independent Multicast-Sparse Mode)
- 协议无关多播-密集方式 PIM-DM (Protocol Independent Multicast-Dense Mode)

本讲总结

内部网关协议 OSPF

路由器的构成

IP 多播



作业

- 4-39, 4-40, 4-43