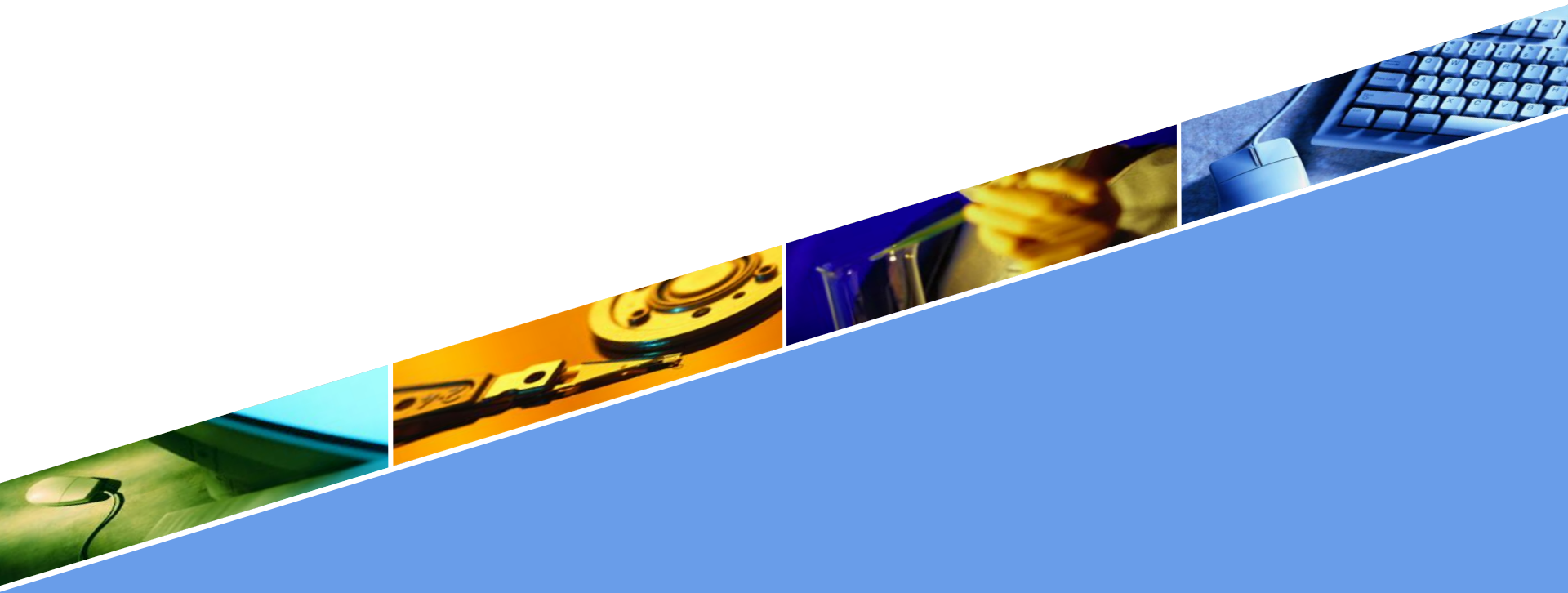
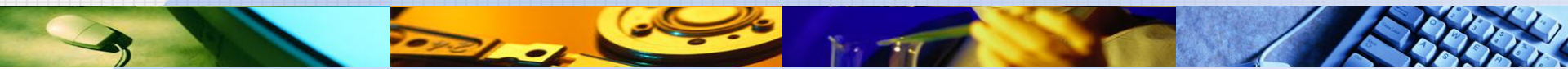
路由器和路由协议



主题 1

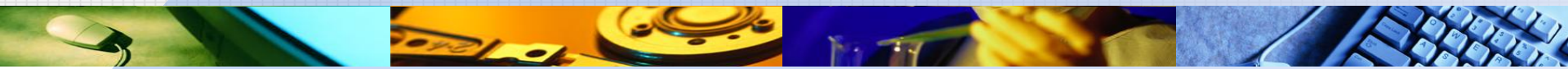


1. 路由器的结构和工作原理
2. 路由选择协议概述
3. RIP协议
4. OSPF协议
5. BGP协议

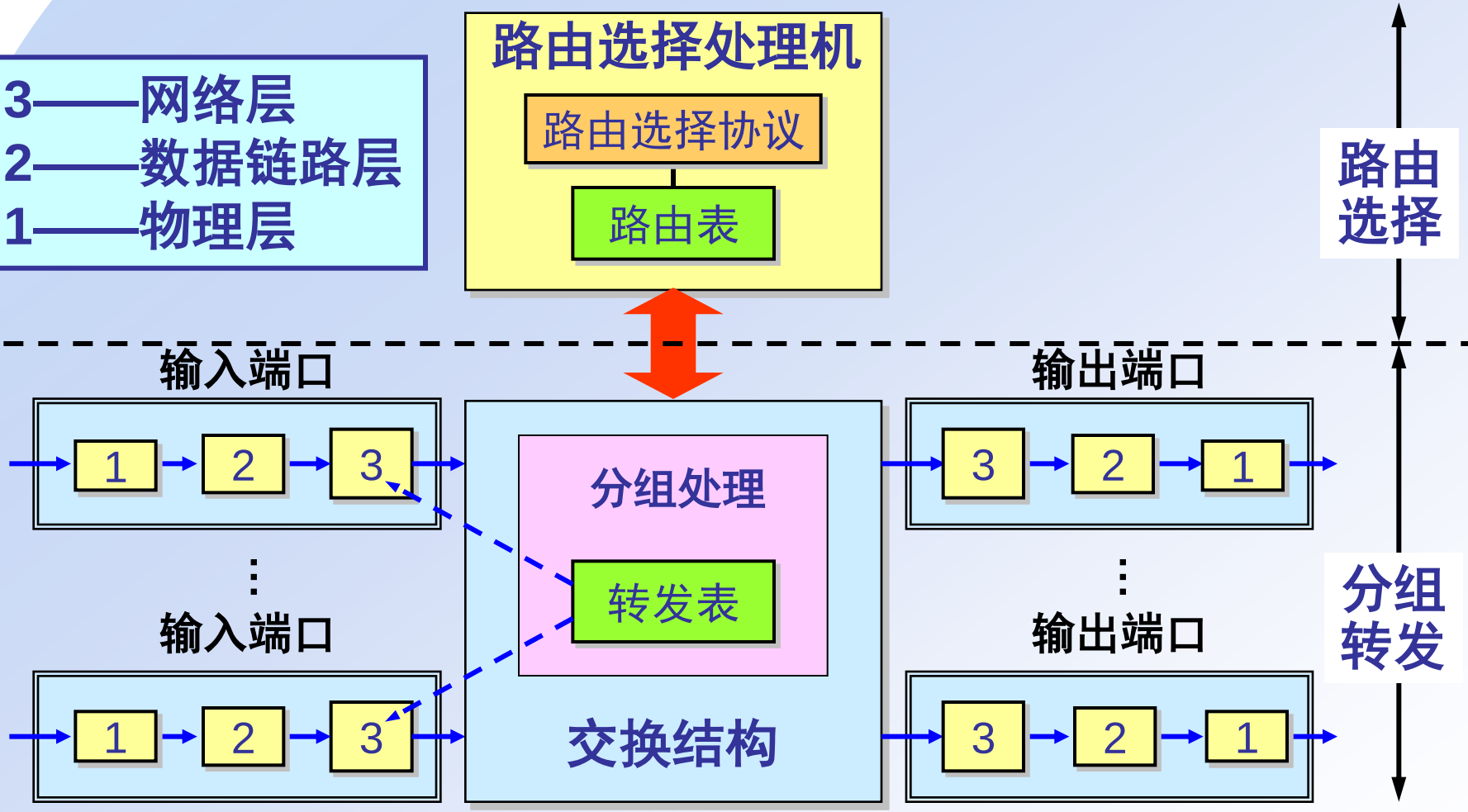
路由器的任务

- 路由器是IP网络互连设备。
- 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，其任务是分组转发和路由选择转发分组。
- 路由器执行网络互连功能，任务有两方面：
  - **分组转发**
  - **路由选择**

路由器的结构



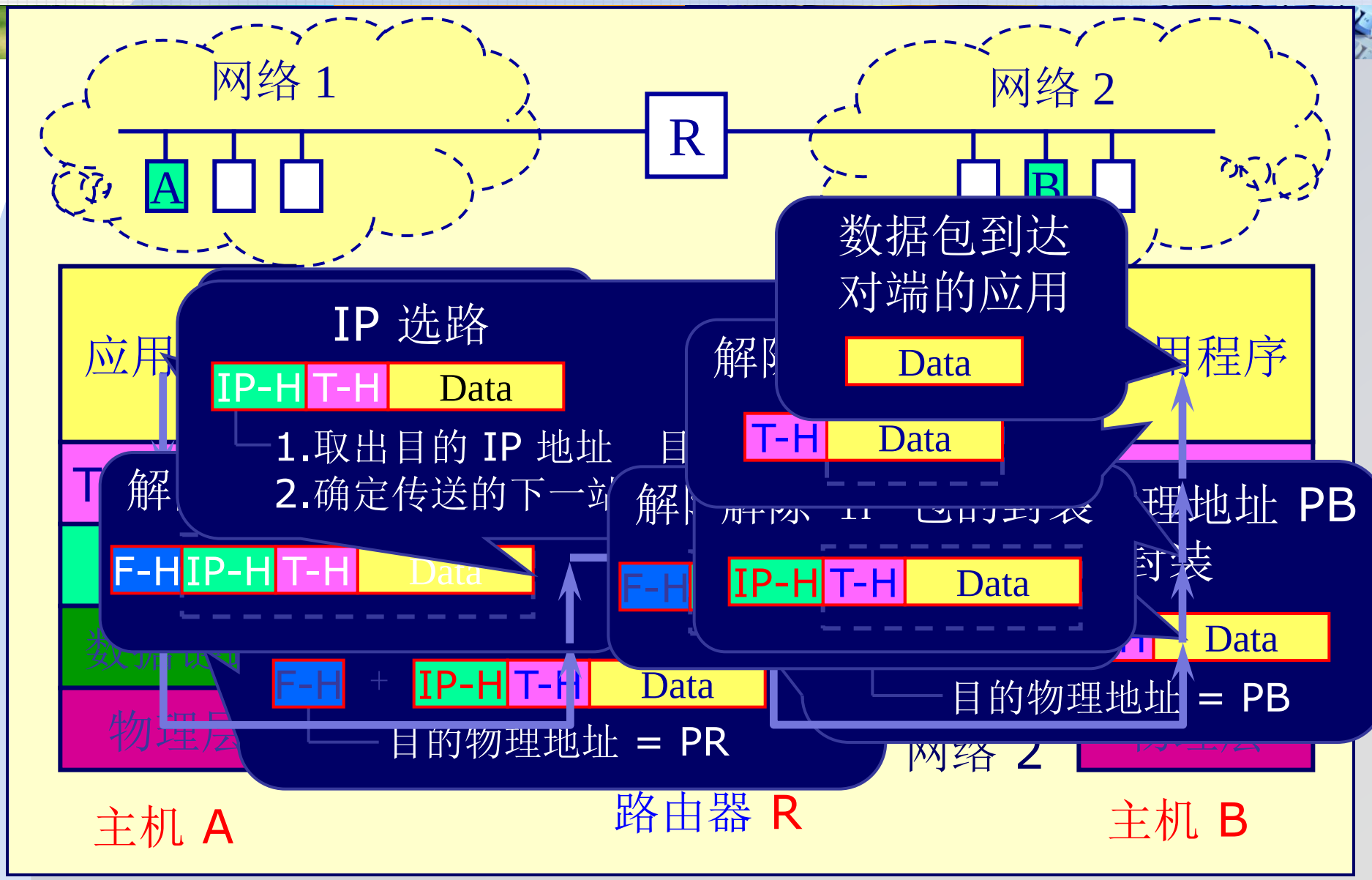
3——网络层
2——数据链路层
1——物理层



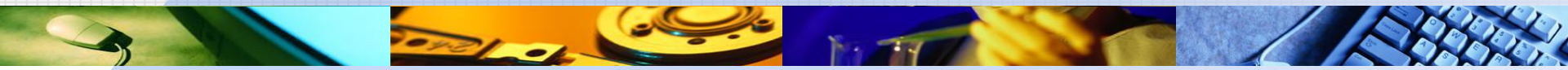
相关概念

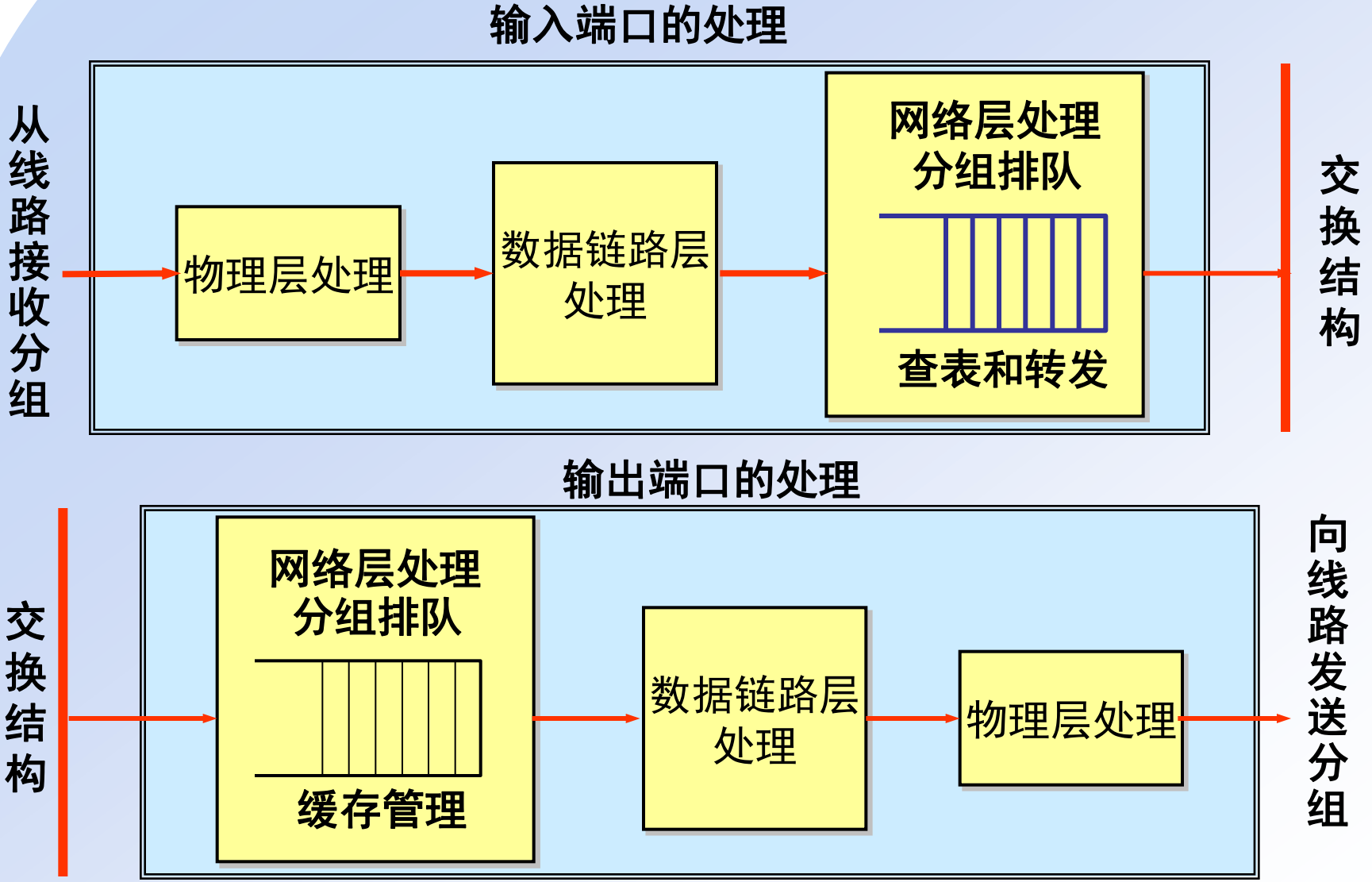
- “转发”(forwarding)：就是路由器根据转发表将用户的 IP 数据报从合适的端口转发出去。
- 路由选择(routing)：则是按照分布式算法，根据从各相邻路由器得到的关于网络拓扑的变化情况，动态地改变所选择的路由。
- 路由表是根据路由选择算法得出的。而转发表是从路由表得出的。
- 在讨论路由选择的原理时，往往不去区分转发表和路由表的区别。

IP分组转发过程

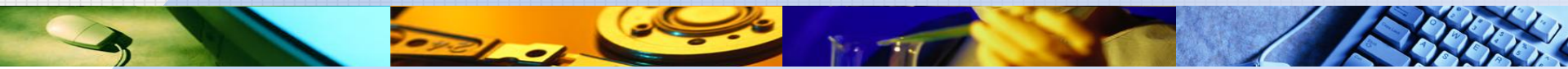

网络 1
网络 2
R
A
B
IP 选路
IP-H T-H Data
1.取出目的 IP 地址
2.确定传送的下一站
F-H IP-H T-H Data
F-H IP-H T-H Data
目的物理地址 = PR
数据包到达
对端的应用
Data
T-H Data
IP-H T-H Data
Data
目的物理地址 = PB
主机 A
路由器 R
主机 B

路由器对IP分组的处理

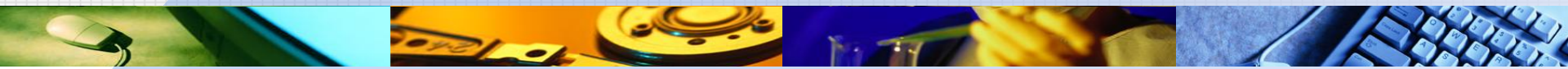




主题 2

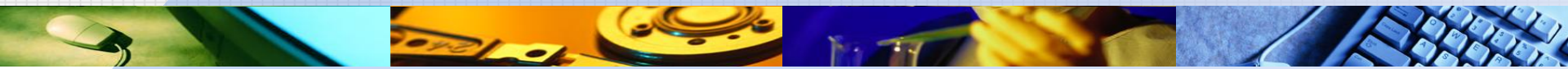


1. 路由器的结构和工作原理
2. 路由选择协议概述
3. RIP协议
4. OSPF协议
5. BGP协议



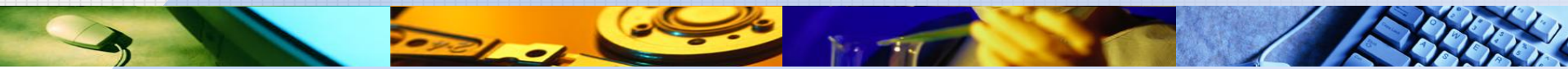
路由器中的路由转发表是怎样得出的？

路由转发表是怎样产生的？



1. 直连路由
   - **直连路由是路由器自动发现并安装的路由信息，即直连路由不需进行配置维护。**
2. 静态路由
   - **静态路由是由网络管理员在路由表中设置的固定的路由条目**
3. 动态路由
   - **路由器执行路由协议**
   - **动态路由是网络中的路由器之间相互通信、传递路由信息、利用收到的路由信息执行路由算法，更新路由表的过程。**

路由协议分类



根据是否在一个自治系统（Autonomous System，AS）内部使用进行分类:

- 内部网关协议（IGP）
- 外部网关协议（EGP）

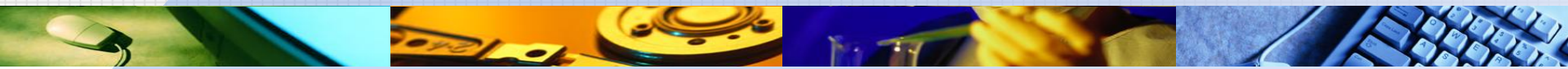
自治系统

- 因特网将整个互联网划分为许多较小的**自治系统 AS**。
- 一个自治系统最重要的特点就是它有权自主地决定在本系统内应采用何种路由选择协议。
- 一个自治系统的所有路由器在本自治系统内都必须是连通的。
- 因特网可视为多个自治系统的集合。

全球的互联网被分成很多个AS 自治域，每个国家的运营商、机构、甚至公司等都可以申请AS号码，AS号码是有限的，最大数目是65536。目前BGP全球路由表已经有24万左右，设备的路由能力一定要强一点

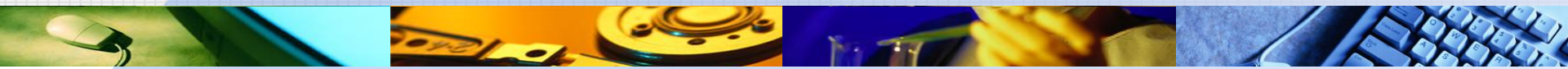
国内申请到的AS号码很少，仅是几个运营商持有，不超过30个。
中国电信：
AS36678 CTUSA – 中国电信美国公司
AS23724 中国电信IDC
AS4816 广东电信
AS4815 上海电信
AS9394 铁通
AS4847 CNIX-AP China Networks 中国互联网交换中心
AS4808 北京网通
AS4837 网通骨干
AS45093 主机屋

内部网关协议



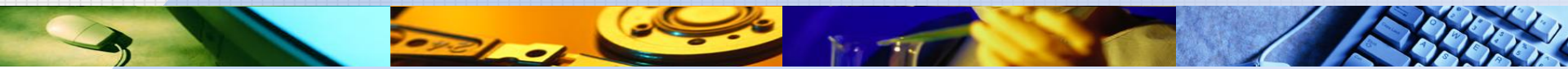
- *内部网关协议* I*GP（Interior GateWay Protocol）*：
  - **在一个自治系统内部运行的路由协议**
  - **与互联网中其他自治系统中采用的路由选择协议无关**
  - **目前最流行的是RIP协议、OSPF协议等。**

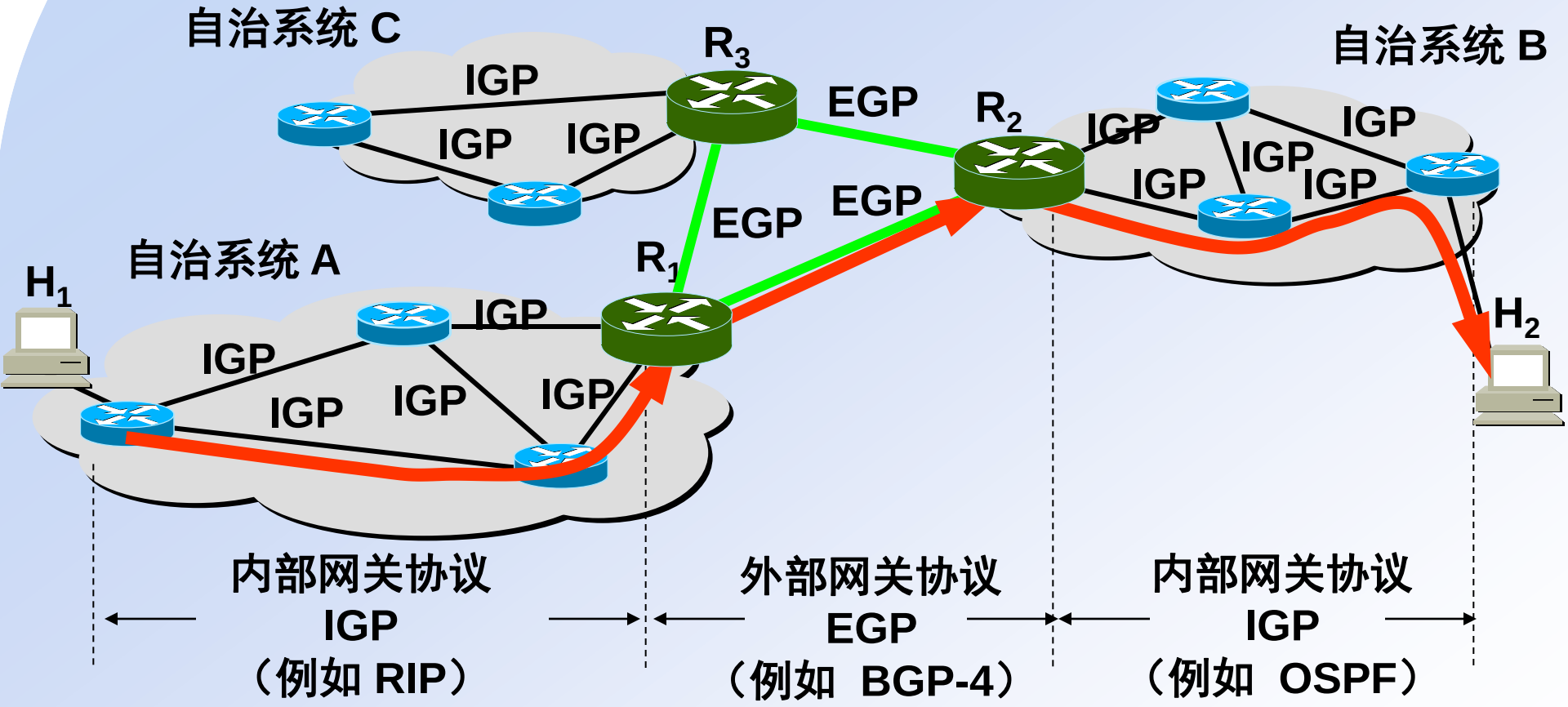
外部网关协议



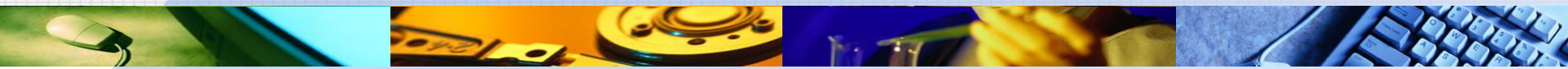
- *外部网关协议EGP（Exterior Gateway Protocol）*：
  - **用于不同自治系统之间的路由协议**
  - **外部网关协议起着连接不同自治区域并在各个自治区域间转发路由数据包的桥梁作用。**
  - **典型的外部网关协议是边界网关路由协议BGP（Border Gateway Protocol）。**

互联网中路由协议的应用





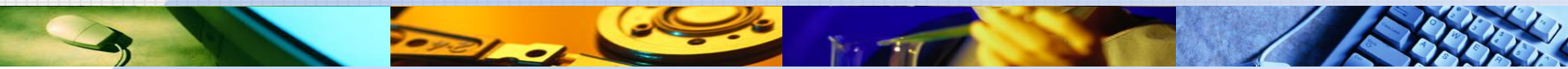
路由协议设计目标



- 正确性
- 稳定性
- 简单性
- 公平性
- 健壮性
- 最优性

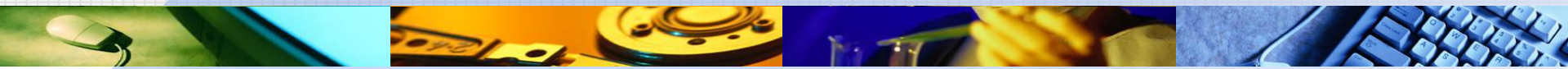
- **公平性与最优性之间矛盾**

链路的度量和代价



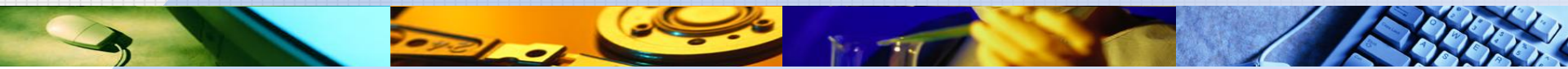
- 在研究路由选择时，需要给每一条链路指明一定的**度量(metric)**。
  - 如链路距离、数据率、链路容量、是否要保密、传播时延等。
- 路由选择往往是由一个或几个因素综合决定的一种**代价（cost）**

路由选择算法



- 不存在一种绝对的最佳路由算法。
- 所谓“最佳”只能是相对于某一种特定要求下得出的较为合理的选择而已。
- 实际的路由选择算法，应尽可能接近于理想的算法。
- 路由选择是个非常复杂的问题
  - 它是网络中的所有结点共同协调工作的结果。
  - 路由选择的环境往往是不断变化的，而这种变化有时无法事先知道。
- 路由算法分两类
  - 基于距离向量的路由选择算法
  - 基于链路状态的路由选择算法

基于距离向量的路由选择算法

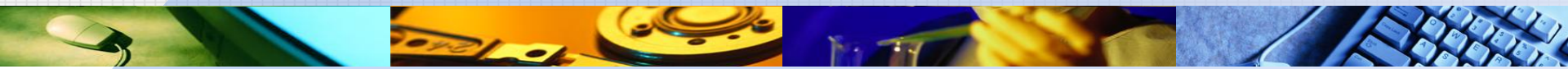


- 距离也称为“跳数” ，每经过一个路由器，跳数就加1。
  - **从一路由器到直接连接的网络的距离定义为 1。从一个路由器到非直接连接的网络的距离定义为所经过的路由器数加 1。**
- 使用“距离”作为度量值，来计算到达目的地要经过的路由器数。
- 基于距离向量路由选择算法的路由协议包括RIP、IGRP等。
- 基于距离向量路由选择算法的扩展—》基于路径的路由选择算法
  - **BGP协议**

基于链路状态的路由选择算法

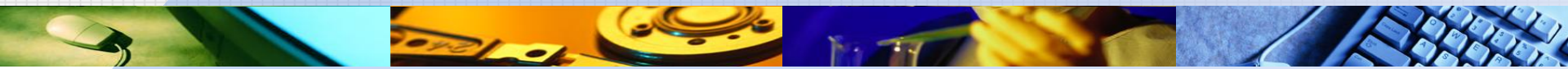
- 链路状态路由选择算法，也称为最短路径优先算法SPF（Shortest-path fast）。
- 它在路由选择过程中使用“代价”作为度量单位，而一般作为代价的网络参数有速度、费用、可靠性等。
- 基于链路状态路由选择算法的路由协议包括OSPF、IS-IS等。

主题 3



1. 路由器的结构和工作原理
2. 路由选择协议概述
3. RIP协议
4. OSPF协议
5. BGP协议

RIP协议原理：知道所有！

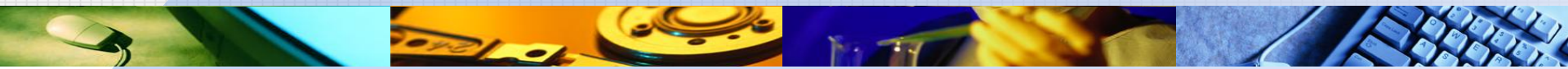


工作原理：

- 路由信息协议（ RIP） 是内部网关协议 IGP中最先得到广泛使用的协议。
- RIP 是一种基于距离向量的路由选择协议。
- RIP 协议要求网络中的每一个路由器都要维护从它自己到其他每一个目的网络的距离记录。

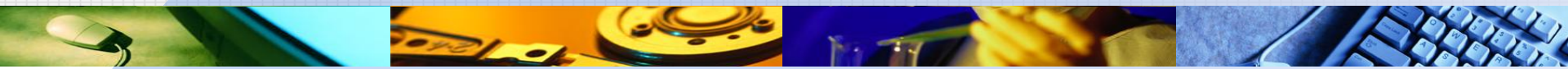
- RIP认为一个好的路由就是它通过的路由器的数目少，即“距离短”。
- RIP 允许一条路径最多只能包含 15 个路由器。
  - “距离”的最大值为16 时即相当于不可达。可见 RIP 只适用于小型互联网。
- RIP 不能在两个网络之间同时使用多条路由。RIP 选择一个具有最少路由器的路由，哪怕还存在另一条高速(低时延)但路由器较多的路由。

RIP协议要点



- 仅和相邻路由器交换信息。
- 交换的信息是当前本路由器所知道的全部信息，即自己的路由表（到本自治系统中所有网络的最短距离，以及到每个网络应经过的下一跳路由器）。
- 按固定的时间间隔交换路由信息，例如，每隔 30 秒。

路由表建立



- 路由器在刚刚开始工作时，只知道到直接连接的网络的距离（此距离定义为1）。
- 以后，每一个路由器也只和数目非常有限的相邻路由器交换并更新路由信息。
- 经过若干次更新后，所有的路由器最终都会知道到达本自治系统中任何一个网络的最短距离和下一跳路由器的地址。

距离向量算法

收到相邻路由器（其地址为 X）的一个 RIP 报文：

(1) 先修改此 RIP 报文中的所有项目：将“下一跳”字段中的地址都改为 X，并将所有的“距离”字段的值加 1。

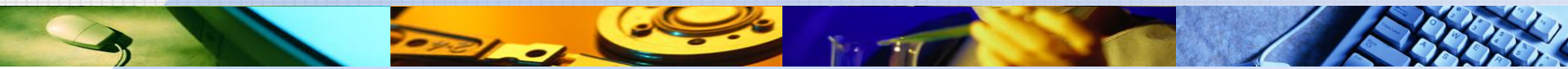
(2) 对修改后的 RIP 报文中的每一个项目，重复以下步骤：

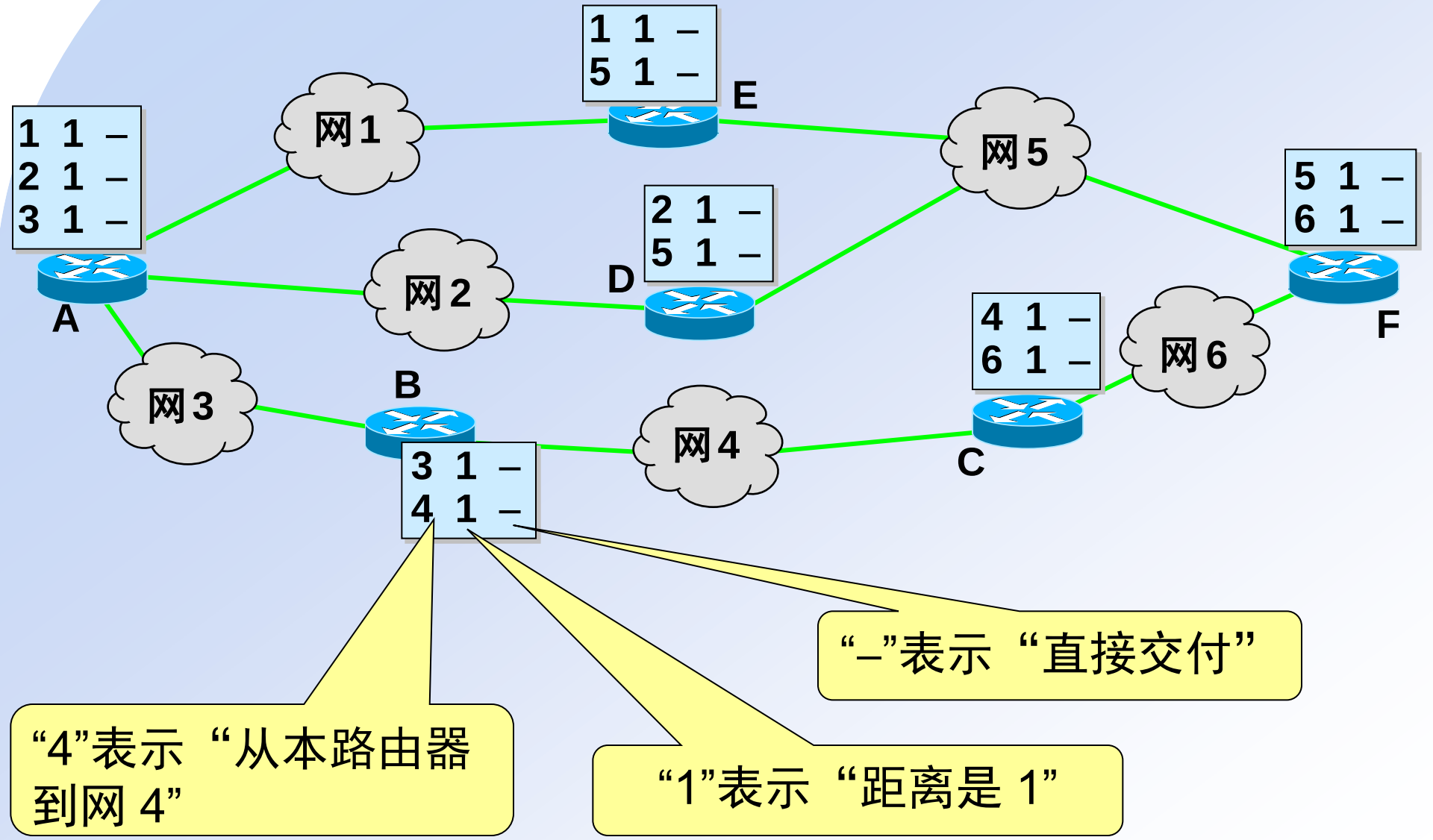
若项目中的目的网络不在路由表中，则将该项目加到路由表中。否则，若下一跳字段给出的路由器地址是同样的，则将收到的项目替换原路由表中的项目。否则，若收到项目中的距离小于路由表中的距离，则进行更新，否则，什么也不做。

(3) 若 3 分钟还没有收到相邻路由器的更新路由表，则将此相邻路由器记为不可达的路由器，即将距离置为16（距离为16表示不可达）。

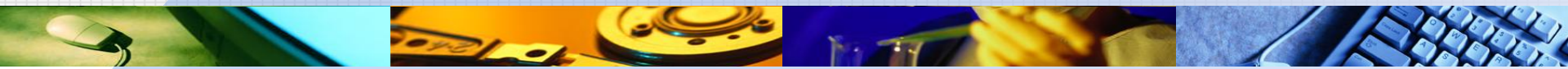
(4) 返回。

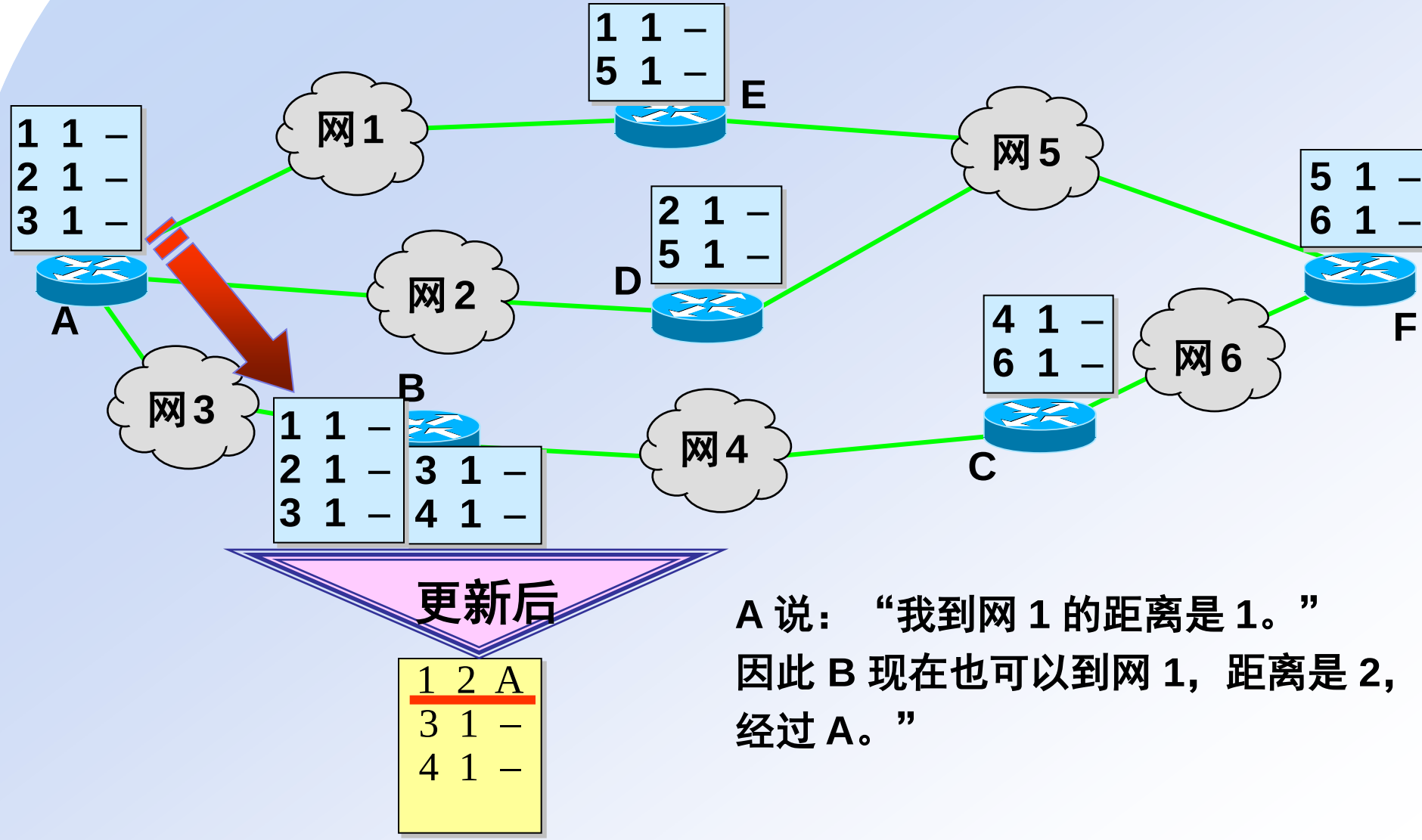
RIP协议示例：初始时有相邻路由器信息





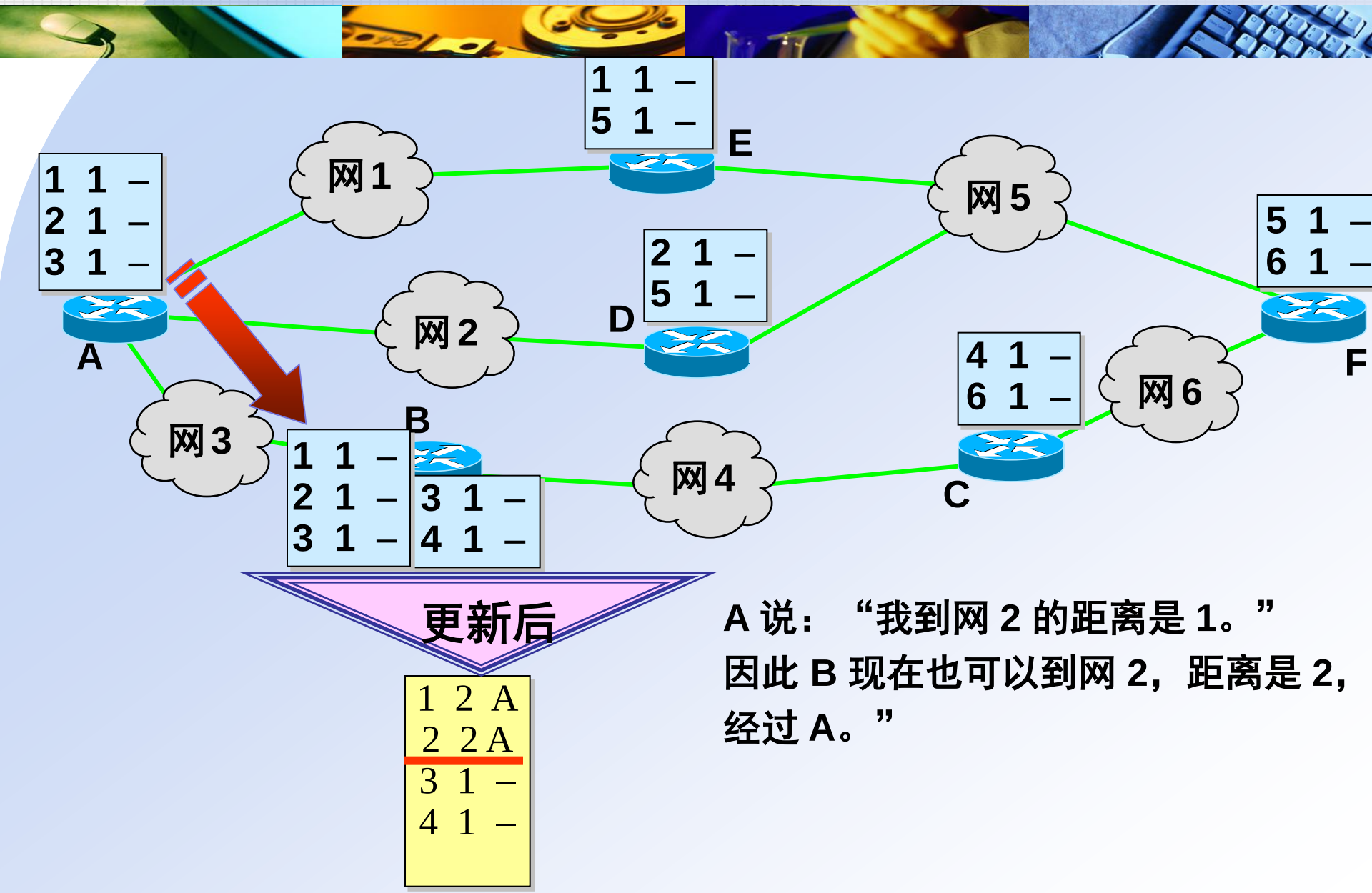
RIP协议示例：B获得A\C信息





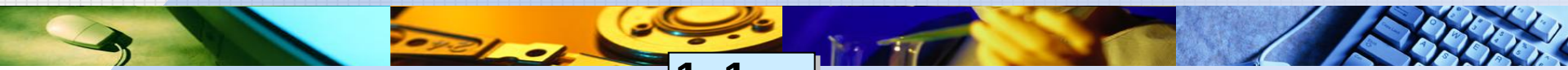
A 说：“我到网 1 的距离是 1。”
因此 B 现在也可以到网 1，距离是 2，经过 A。”

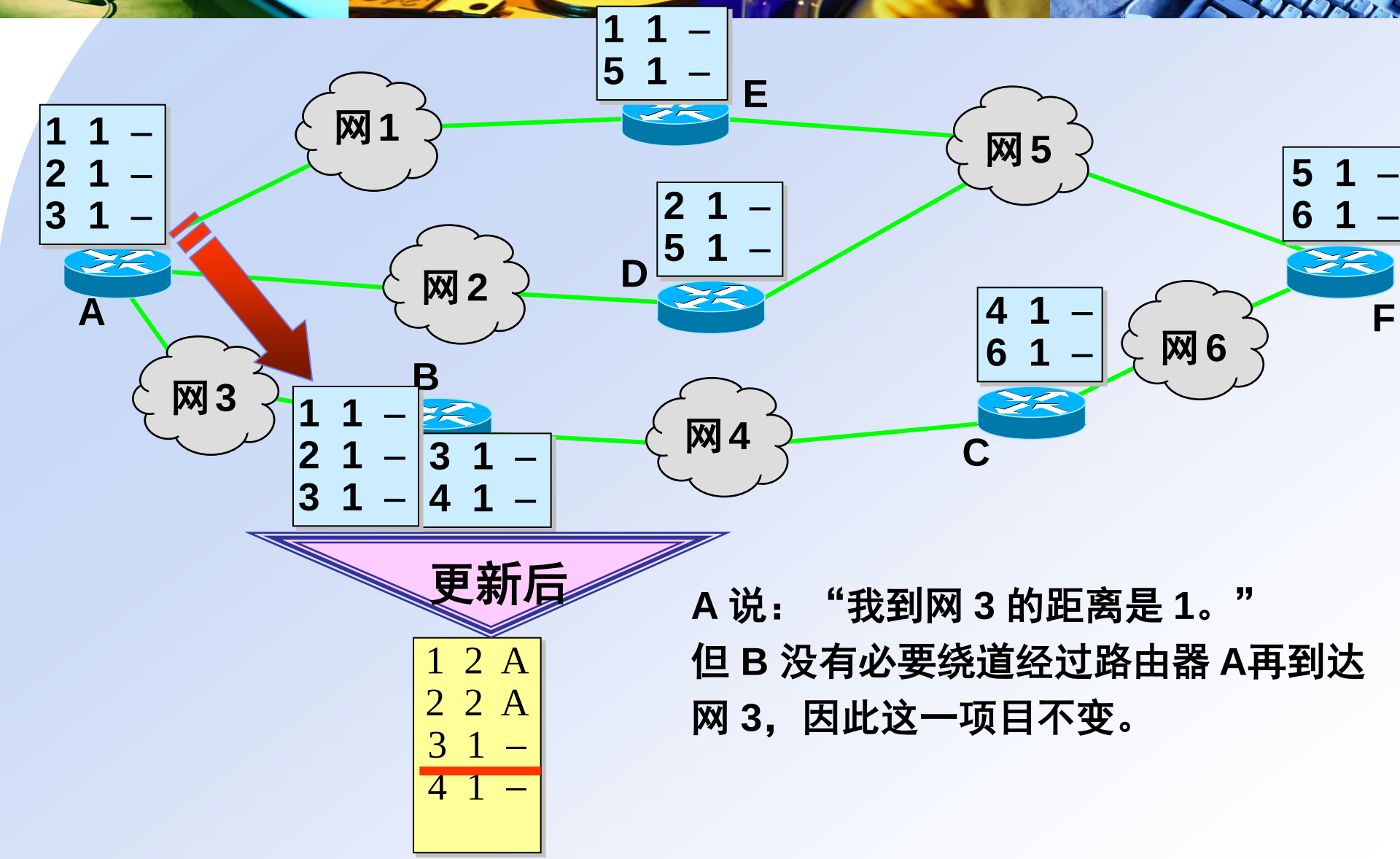
RIP协议示例：B获得A\C信息



A 说：“我到网 2 的距离是 1。”
因此 B 现在也可以到网 2，距离是 2，经过 A。”

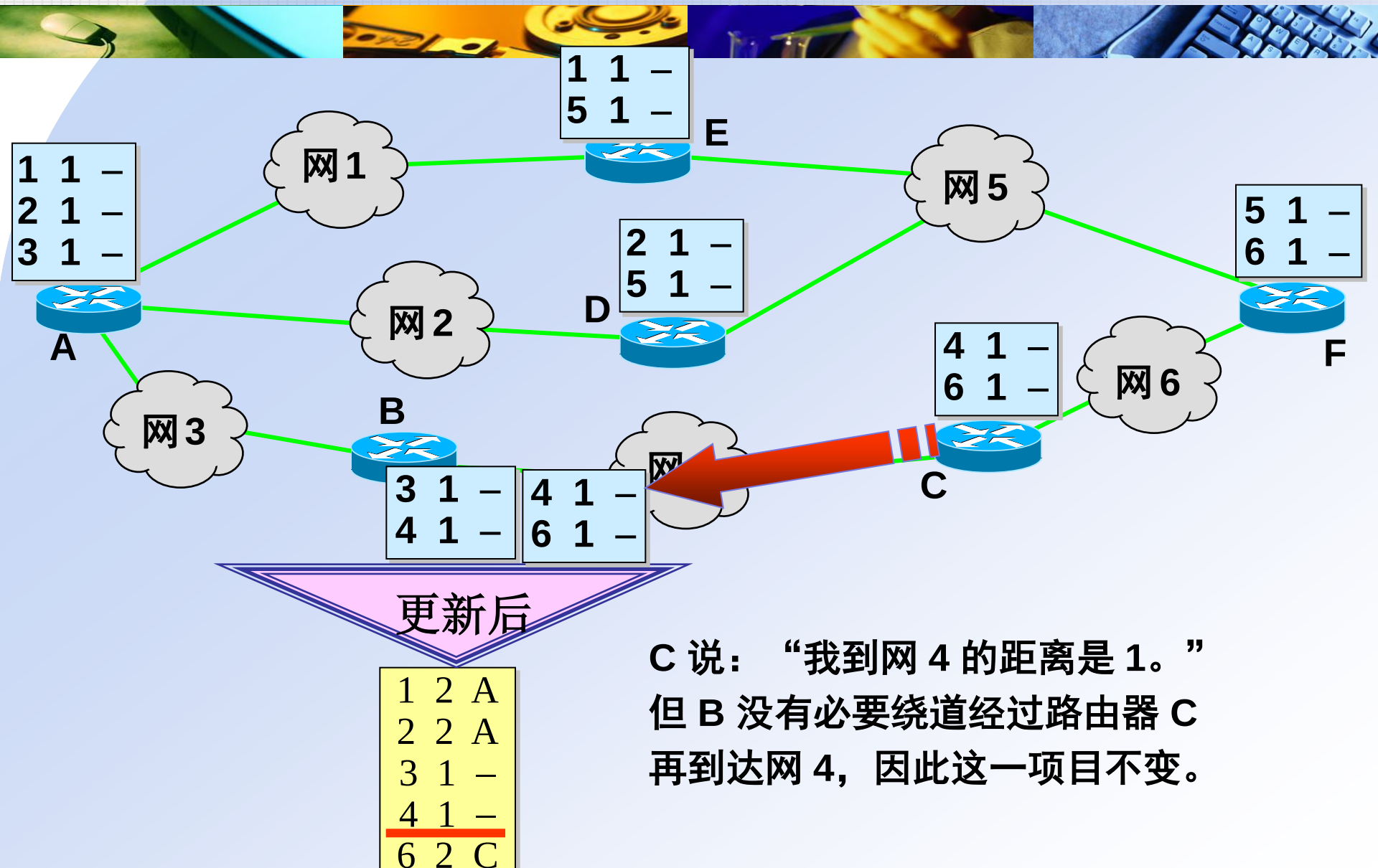
RIP协议示例：B获得A\C信息





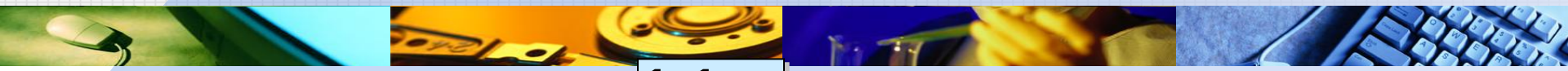
A 说：“我到网 3 的距离是 1。”
但 B 没有必要绕道经过路由器 A再到达网 3，因此这一项目不变。

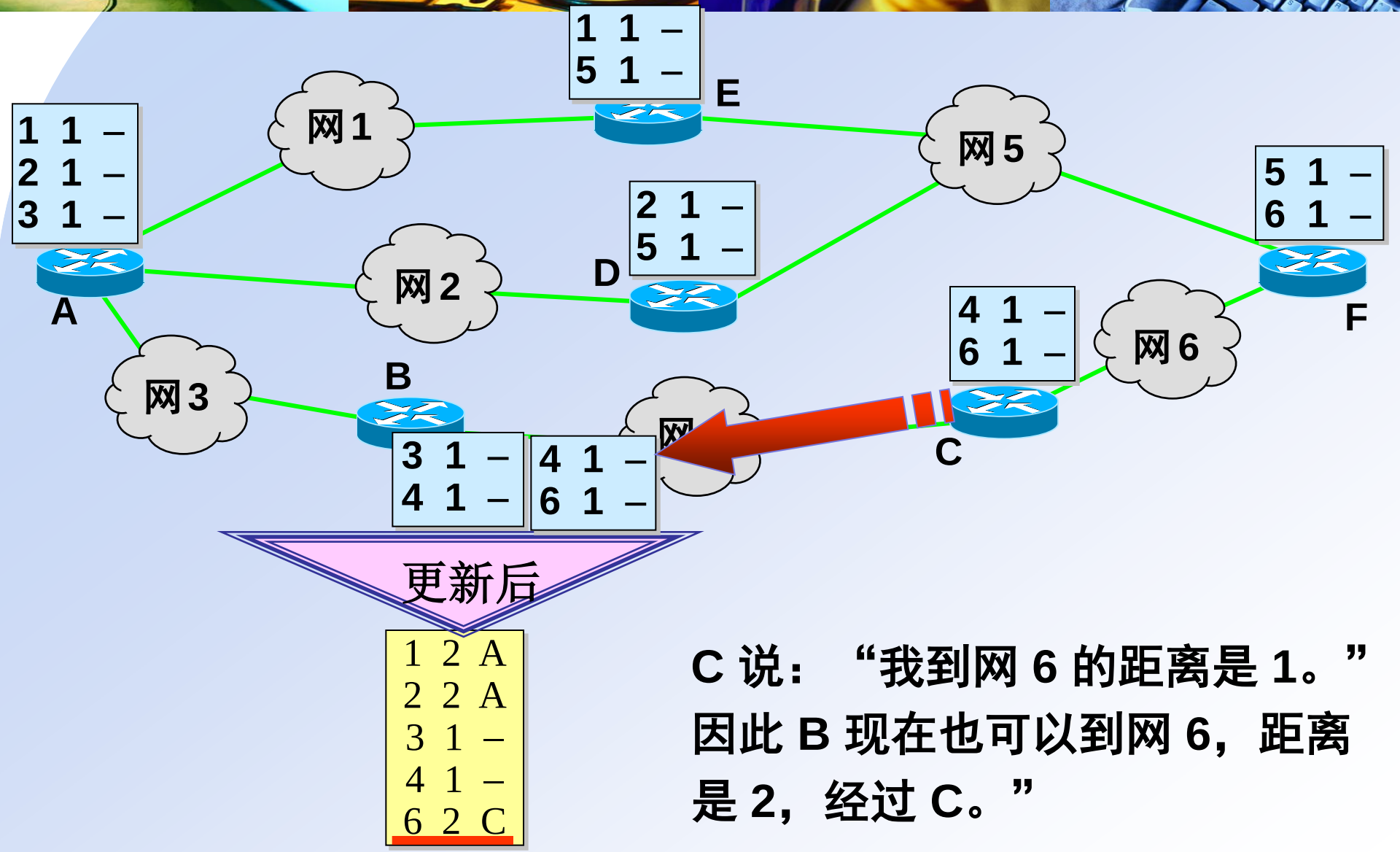
RIP协议示例：B获得A\C信息



C 说：“我到网 4 的距离是 1。”
但 B 没有必要绕道经过路由器 C
再到达网 4，因此这一项目不变。

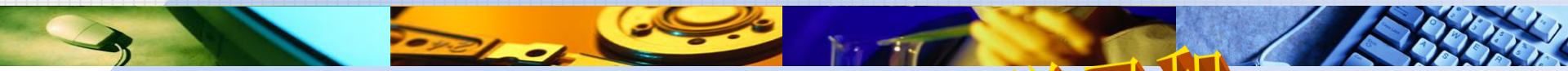
RIP协议示例：B获得A\C信息

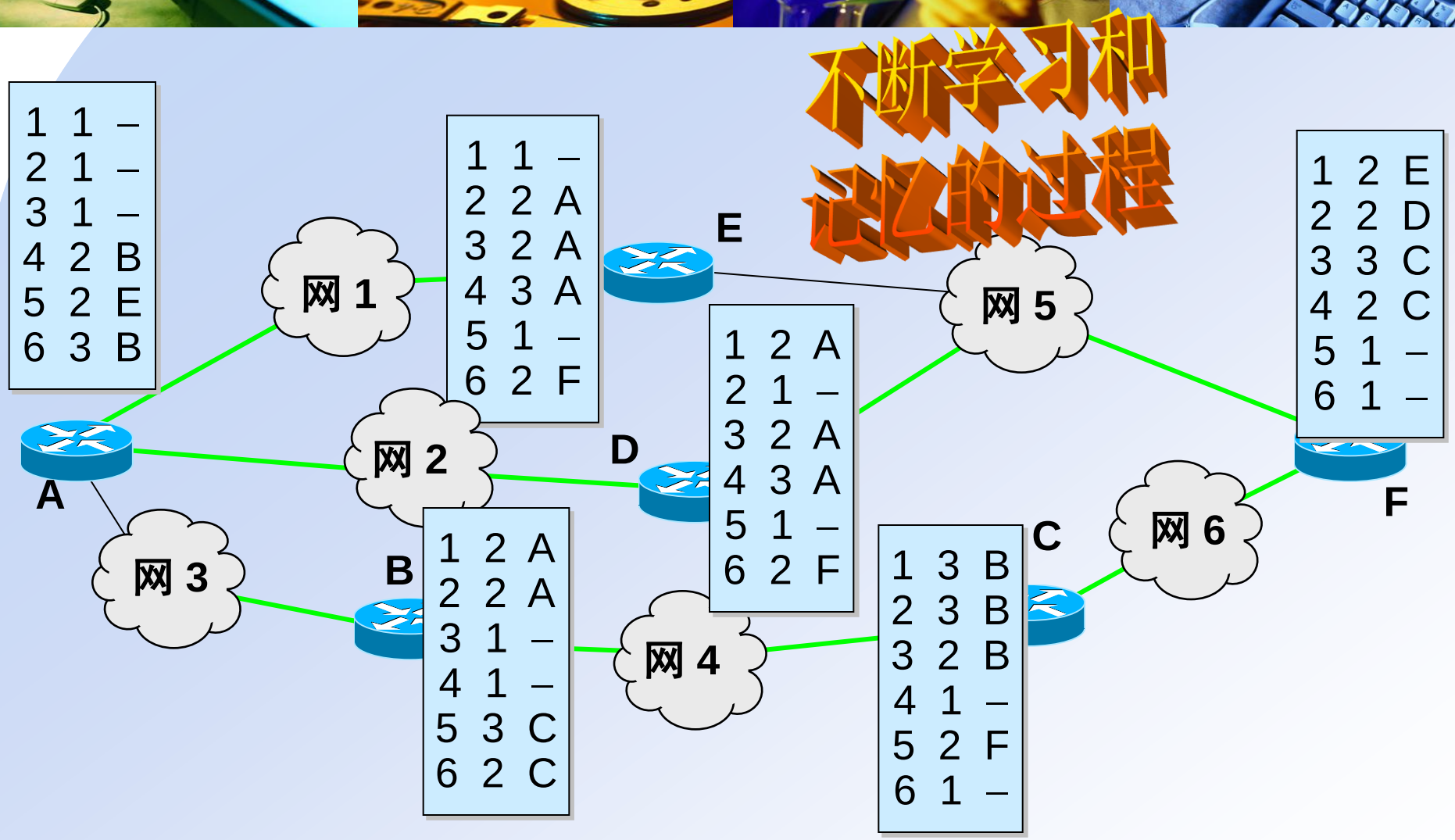




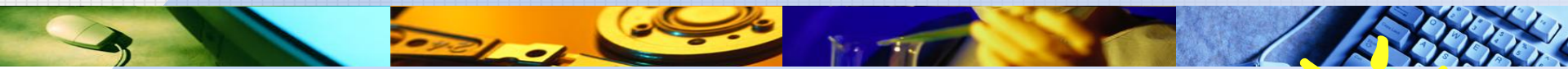
C 说：“我到网 6 的距离是 1。”因此 B 现在也可以到网 6，距离是 2，经过 C。”

RIP协议示例：路由表均更新





距离向量路由算法—好消息反映迅速



好消息：

A接通了！

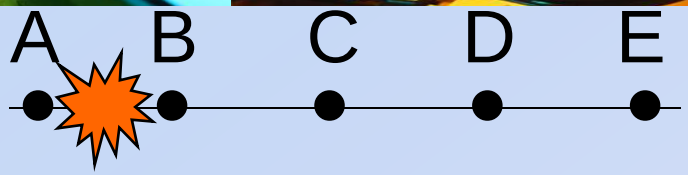
A　　B　　C　　D　　E

●——●——●——●——●

∞	∞	∞	∞	初始时
1	∞	∞	∞	第1次交换后
1	2	∞	∞	第2次交换后
1	2	3	∞	第3次交换后
1	2	3	4	第4次交换后

A开始接通后，经过4次传递，各站都知道到A的路由。

距离向量路由算法—坏消息反映迟钝



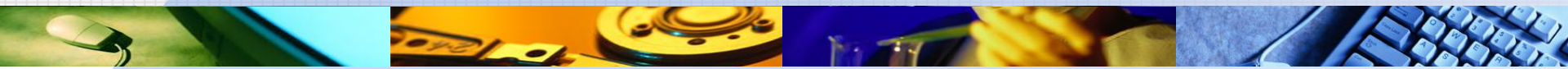
1	2	3	4	初始时
3	2	3	4	第1次交换，认为可从C转
3	4	3	4	第2次交换后
5	4	5	4	第3次交换后，
5	6	5	6	第4次交换后
7	6	7	6	第5次交换后，
7	8	7	8	第6次交换后
	… …			
∞	∞	∞	∞	

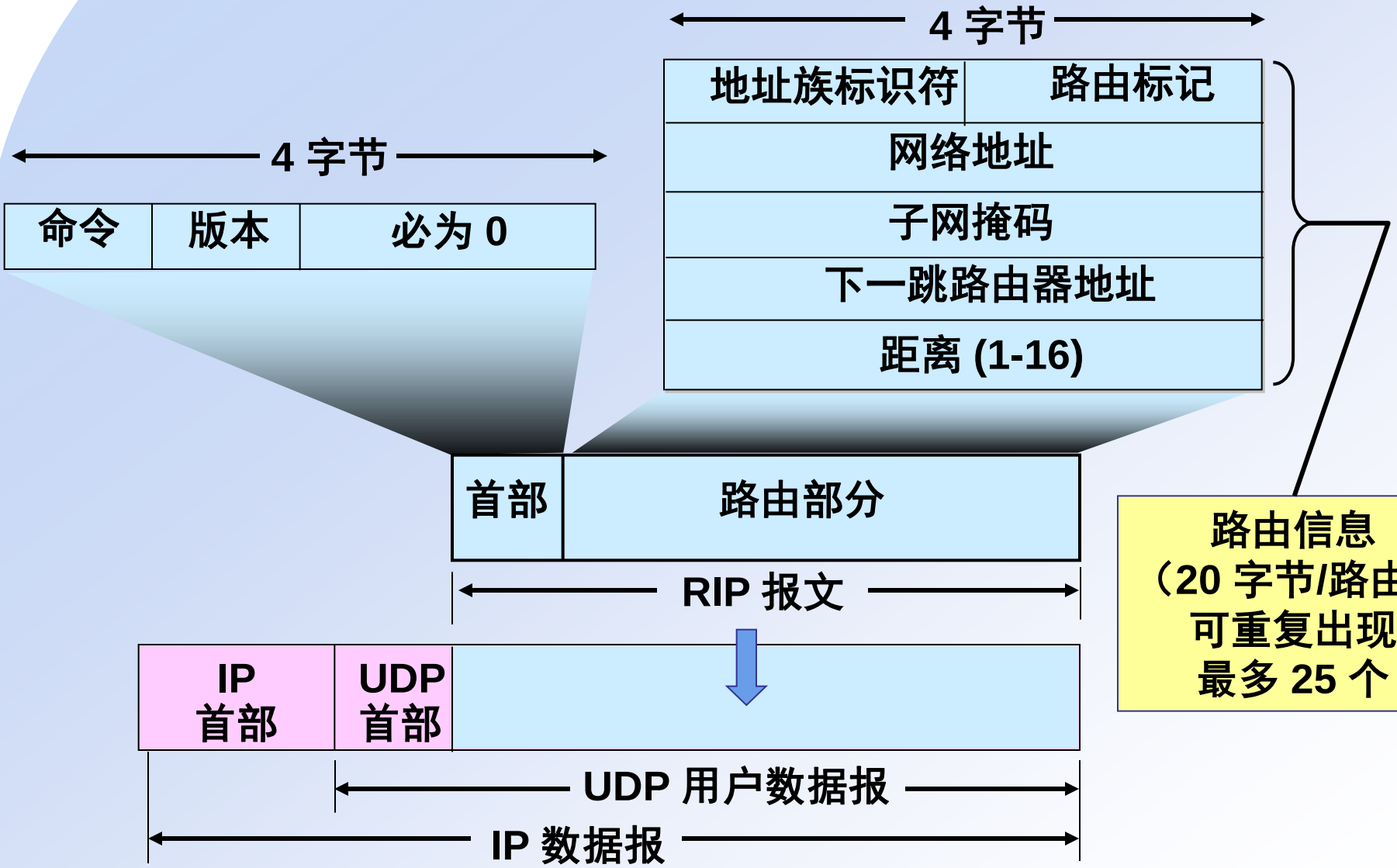
A失效后，消息传递极慢

RIP协议路由收敛慢

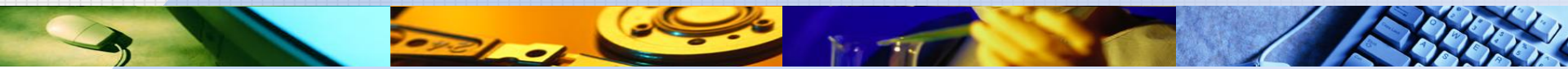
- 存在的问题：路由收敛慢
- 产生问题的根源：距离向量算法产生路由循环问题。
- RIP协议对此改善的机制
  - **有限跳段数**
  - **水平分割**
  - **毒性逆转**
  - **触发更新**
  - **抑制计时**

RIP协议报文结构




4 字节
地址族标识符
路由标记
网络地址
子网掩码
下一跳路由器地址
距离 (1-16)
4 字节
命令
版本
必为 0
首部
路由部分
RIP 报文
IP
首部
UDP
首部
UDP 用户数据报
IP 数据报
路由信息
（20 字节/路由）
可重复出现
最多 25 个

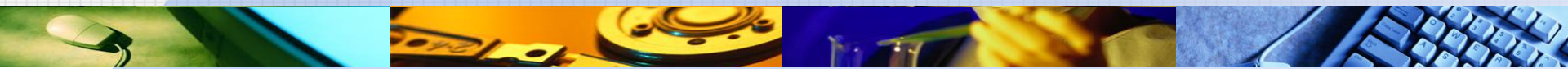
RIP协议报文



- RIP基于UDP，使用UDP端口号520。
- RIP消息可以分为两类：
  - **请求路由信息消息，路由器可以通过发送请求路由信息消息获得某个路由器的全部路由或者部分路由表项**
  - **路由信息消息**

    路由信息消息发送的条件：（1）收到请求路由信息消息（2）每隔30秒定期发送。
- RIP消息都具有一个统一的格式
  - **命令（COMMAND）字段指示RIP消息的类型（request或response）**
  - **路由部分由若干个路由信息组成，每个路由信息需要用 20 个字节。**
    - 地址家族标识（address family identifier）字段，使得RIP协议也可以在别的网络层协议下使用，而不是局限在TCP/IP环境中。
    - 路由标记填入自治系统的号码再后面指出某个网络地址、该网络的子网掩码下一跳路由器地址以及到此网络的距离
  - **没有长度字段，这是因为下层的UDP有封装功能，从而可以知道消息的边界。**

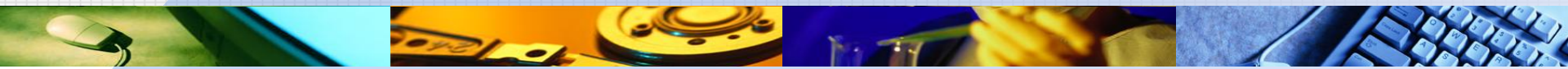
RIP协议特点



- RIP 协议最大的优点就是**实现简单，开销较小。**
- RIP使用**距离作为度量**，很多时候路由选择不够优化，不能采取一种动态的方法（比如根据网络延迟或负载）来选择路由。
- 尽管RIP采用了很多措施，仍存在**路由收敛慢：**当网络出现故障时，要经过比较长的时间才能将此信息传送到所有的路由器。
- RIP 限制了网络的规模，它能使用的**最大距离为 15**（16 表示不可达）。
- 路由器之间**交换的路由信息是**路由器中的**完整路由表**，因而随着网络规模的扩大，开销也就增加。

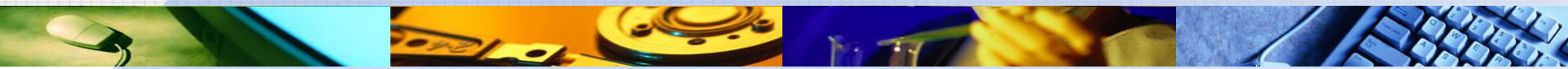
RIP协议目前的版本是RIPv2， 适用于小规模的网络

主题 4



1. 路由器的结构和工作原理
2. 路由选择协议概述
3. RIP协议
4. OSPF协议
5. BGP协议

OSPF简介



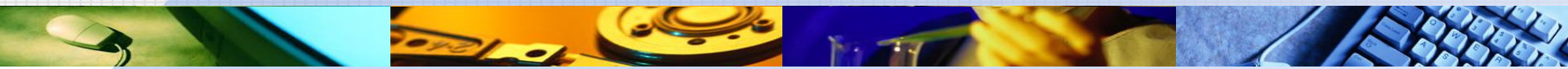
- OSPF是open shortest path first（开放最短路径优先协议）的缩写。
  - **“开放”表明 OSPF 协议不是受某一家厂商控制，而是公开发表的。**
  - **“最短路径优先”是因为使用了 Dijkstra 提出的最短路径优先（Shortest Path First ，SPF) 算法计算路由。**
- 基于分布式链路状态的内部网关协议。
- 用于在单一自治系统(autonomous system,AS)内决策路由，适用于较大规模的AS内部路由。
- OSPF v2已成为互联网标注协议（RFC 2328）。

采用链路状态协议的OSPF要点

与RIP协议相比，采用链路状态协议的OSPF要点：

- 每个路由器使用**洪泛法（flooding)**向本自治系统中所有路由器发送信息。
- 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**（路由器所知道的部分信息），“链路状态”信息包括：
  - 本路由器都和哪些路由器相邻
  - 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。

OSPF基本思想



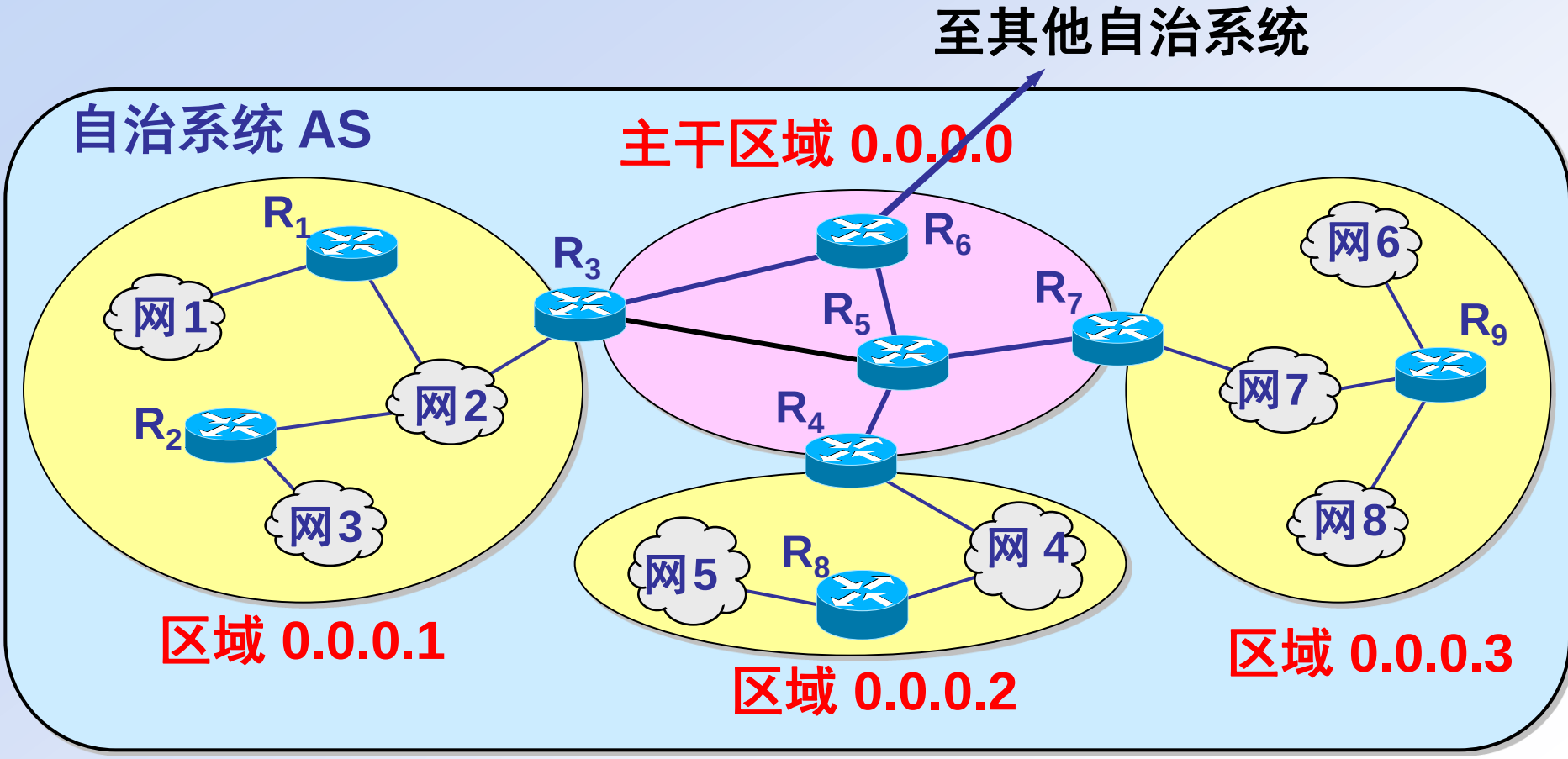
- 由于各路由器之间频繁地交换链路状态信息，因此所有的路由器最终都能建立一个**链路状态数据库**。

- 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**）。

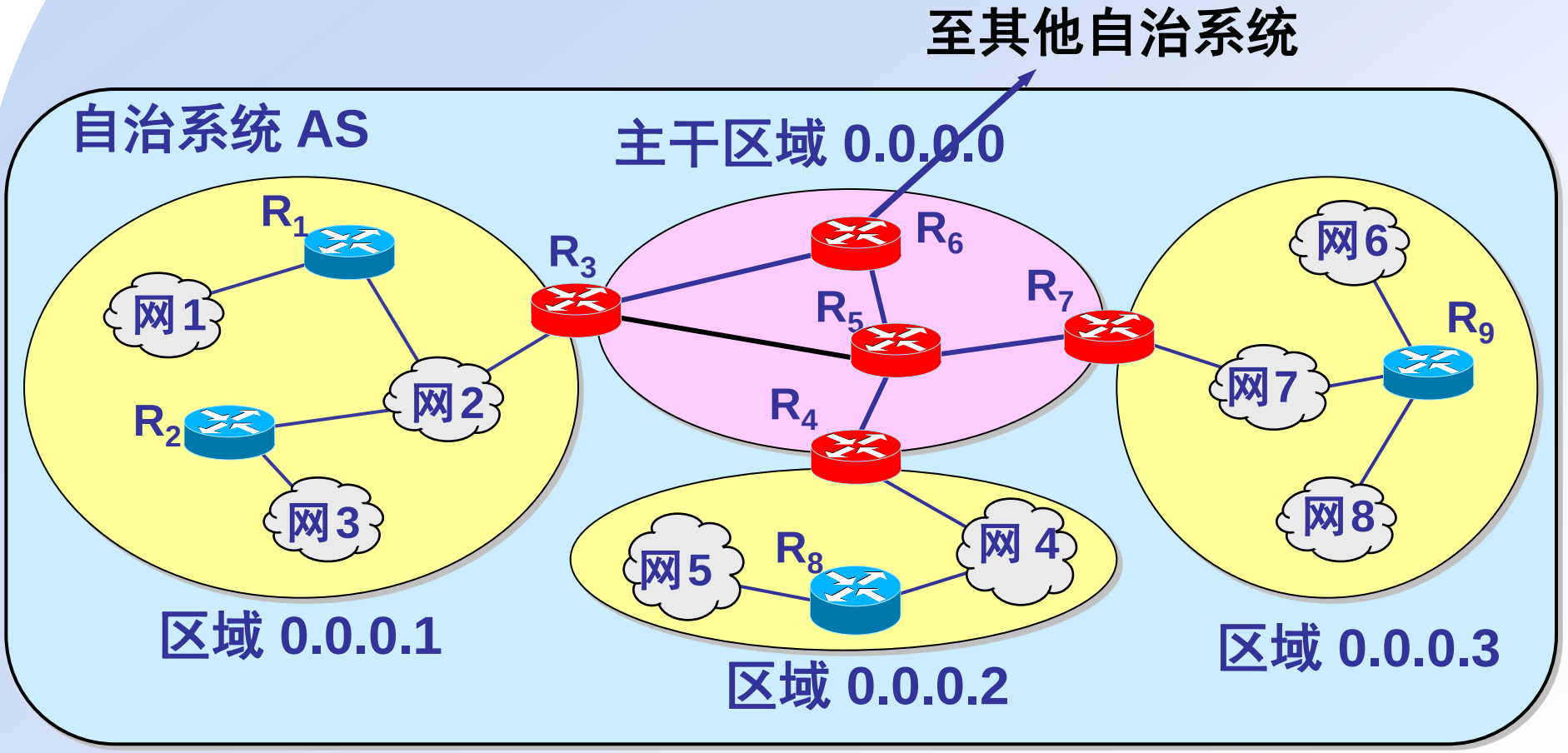
- 路由器执行**Dijkatra**最短路径算法，建立起完整的路由表。

OSPF划分区域

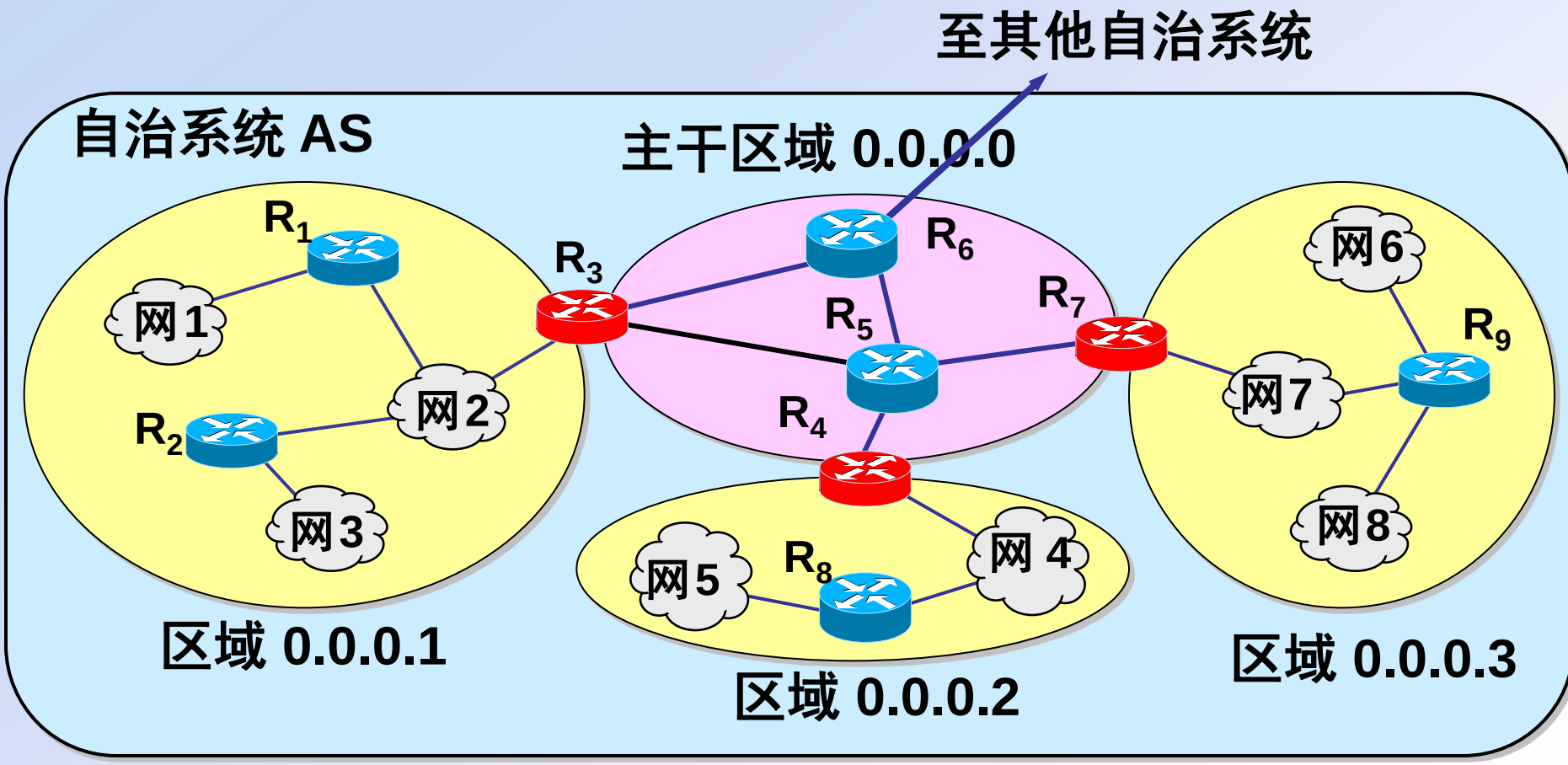
OSPF支持层次化路由，可将一个大的自治域系统划分如干更小的**区域（area），每个区域有一个32为的标识符。**



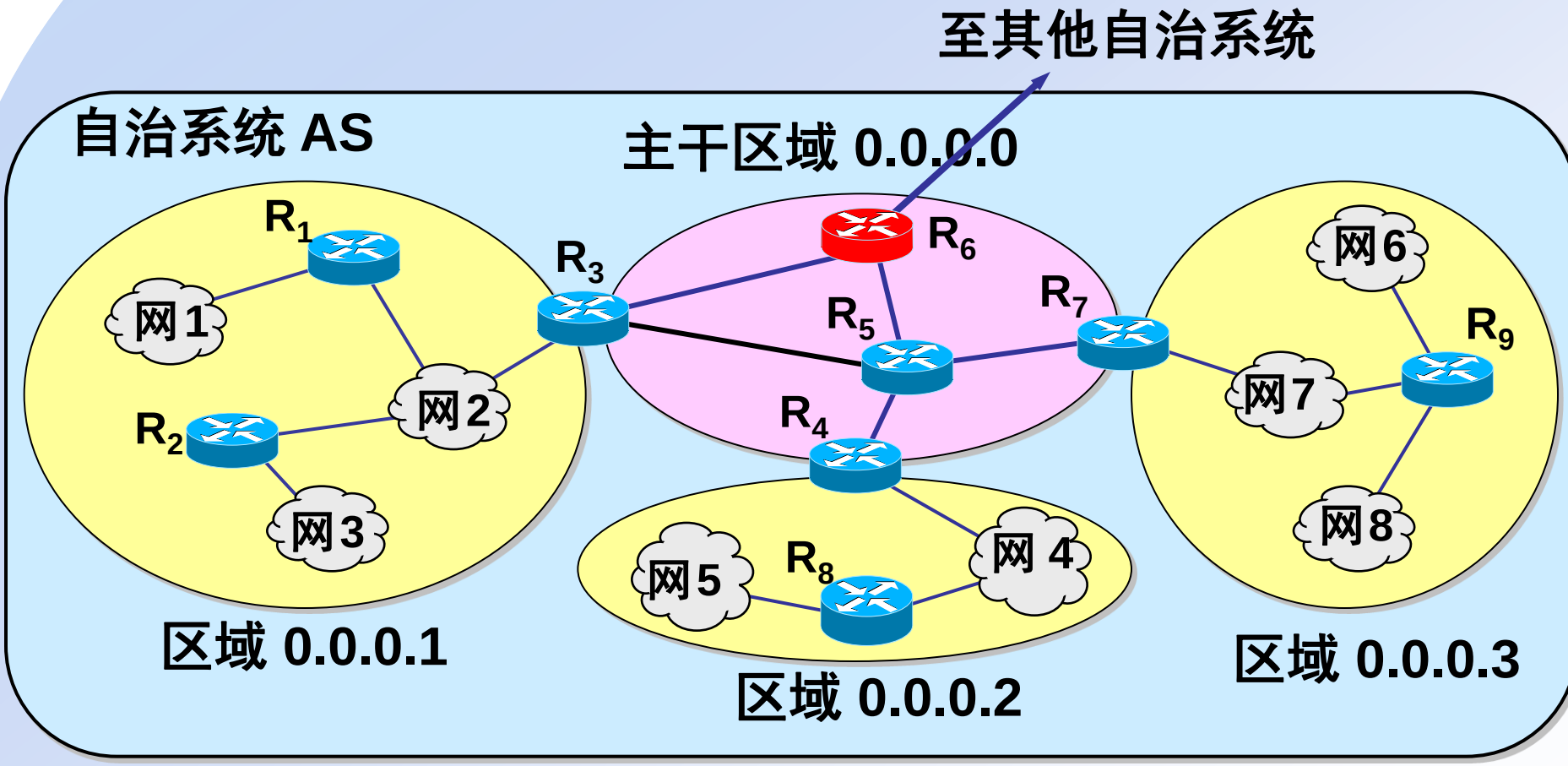
OSPF主干路由器


至其他自治系统
自治系统 AS
主干区域 0.0.0.0
R1
网1
网2
R2
网3
区域 0.0.0.1
R3
R6
R5
R7
R4
网5
R8
网 4
区域 0.0.0.2
网6
R9
网7
网8
区域 0.0.0.3

OSPF区域边界路由器

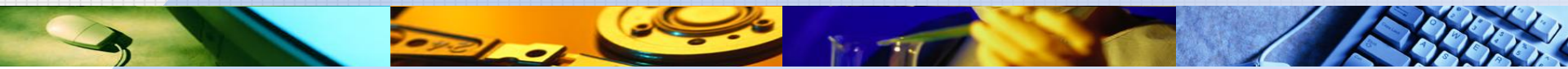


OSPF自治系统边界路由器



自治系统边界路由器是专门和本自治系统外的其他自治系统交换路由信息。

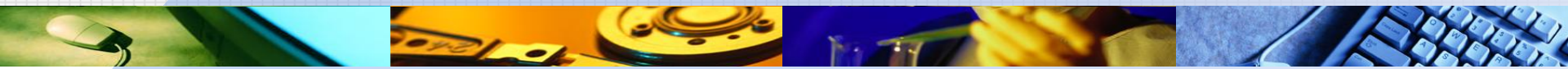
OSPF区域划分的好处



- 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适用于大规模自治系统的内部路由协议

OSPF数据传送

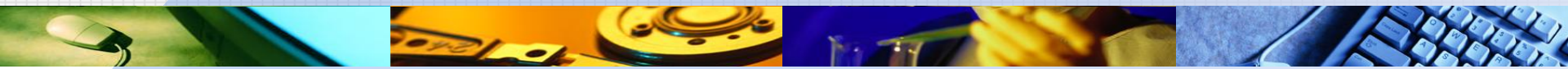


- OSPF 不用 UDP 而是直接用 IP 数据报传送，可见 OSPF 的位置在网络层。

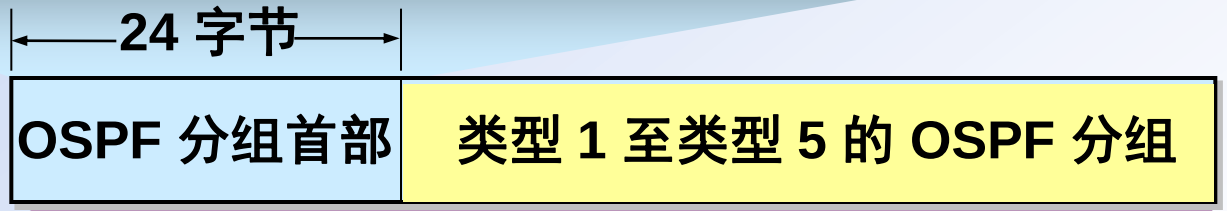
- OSPF 构成的数据报很短。这样做可减少路由信息的通信量。

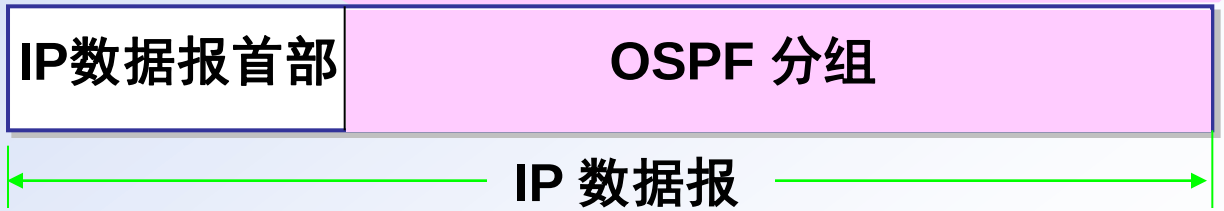
- 数据报很短的另一好处是可以不必将长的数据报分片传送。分片传送的数据报只要丢失一个，就无法组装成原来的数据报，而整个数据报就必须重传。

OSPF报文结构




比特 0
8
16
31
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

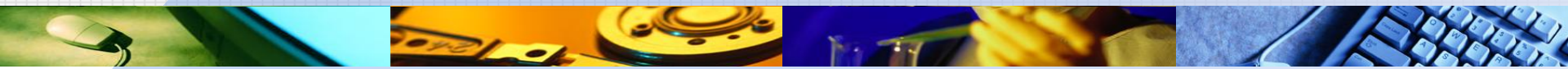

IP数据报首部
OSPF 分组
IP 数据报

OSPF报文结构：分组类型

- 类型1，问候分组：用来发现和维持邻站的可达性。
- 类型2，数据库描述分组：向邻站给出自己的链路状态数据库中所有链路状态项目的摘要信息。
- 类型3，链路状态请求分组：向对方请求发送某些链路状态项目的详细信息。
- 类型4，链路状态更新分组：用洪泛法对全网更新链路状态。
- 类型5，链路状态确认分组：对链路更新分组的确认。

OSPF规定：每两个相邻路由器之间每隔10秒交换一次问候分组，以维持路由器之间的可达性。若40秒没收到某相邻路由器发来的问候分组，则认为该路由器不可达，应立即修改链路状态数据库，并重新计算路由表。其他的四种分组是用来进行链路状态数据库的同步。

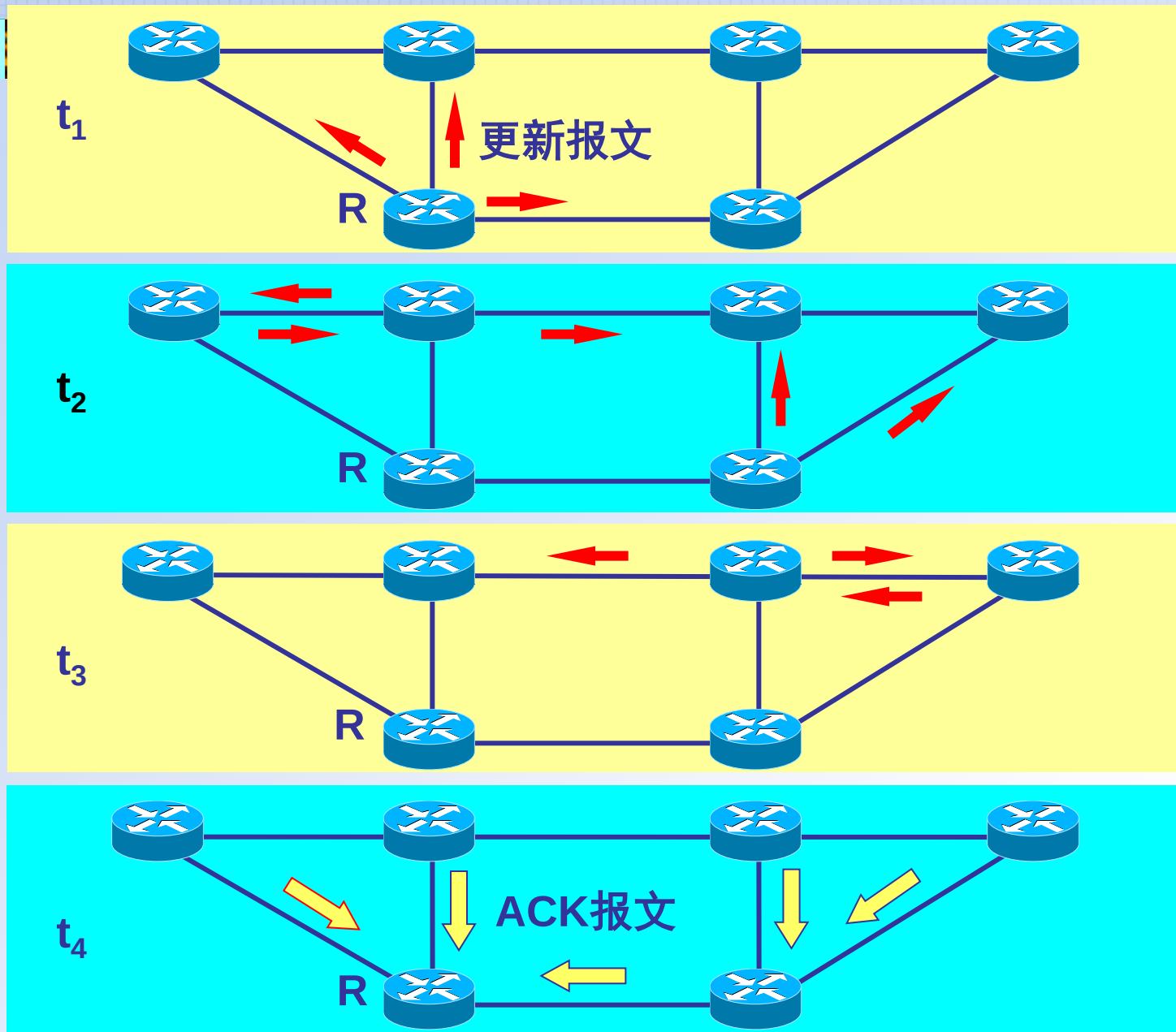
路由器开始工作时链路状态数据库同步过程



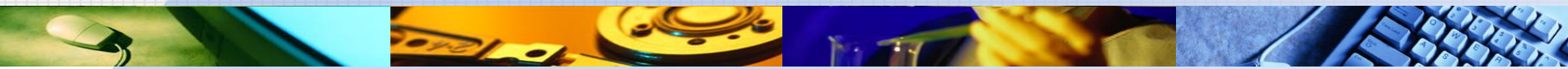
- 路由器使用链路状态请求分组，向对方请求发送自己所缺少的某些链路状态的详细信息。

在网络运行过程中 使用可靠的洪泛法发送更新分组

在网络运行过程中，只要有一个路由器的链路状态发生变化，该路由器就要使用链路状态更新分组，用洪泛法对全网更新链路状态。

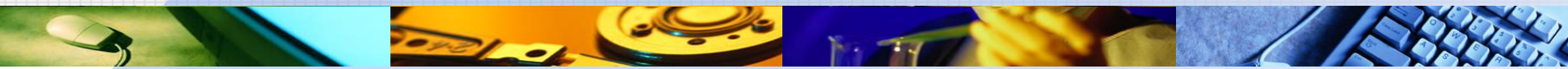


路由器泛洪的时机



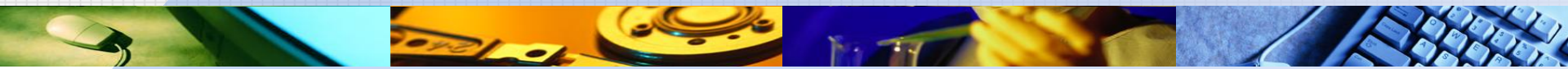
- 新加入路由器或拓扑结构改变时， OSPF通过洪泛过程通告网络上其他路由器新加入的链路状态报文。
- 当网络中的链路状态改变时，通过泛洪方法把更新的本地链路状态信息广播到区域或自治系统中的每个路由器。
  - OSPF让每一个链路状态都带上一个32bit的序号，序号越大状态就越新。
- 定期泛洪：即使链路状态没有发生改变，OSPF路由信息也会自动更新，默认时间为30分钟。

OSPF特点



- **适应范围**：支持各种规模的网络，最多可支持几百台路由器；同时OSPF也支持可变长子网掩码VLSM；
- **快速收敛**：在网络的拓扑结构发生变化后立即发送更新报文，使这一变化在自治系统中同步，当网络拓外改变后迅速收敛，协议带来的网络开销很小，响应网络变化的时间小于 100 ms ；
- **区域划分**：允许自治系统的网络被划分成区域来管理，从而减少了占用的网络带宽；
- **等价路由**：支持到同一目的地址的多条等价路由，实现路径间的负载平衡；
- **支持验证**：支持路由器之间交换的分组的鉴别功能以保证路由计算的安全性；
- **支持不同业务的路由**：OSPF允许管理员给每条路由指派不同的代价，因此OSPF对于不同业务可计算不同路由，可提高网络服务质量。

主题 5



1. 路由器的结构和工作原理
2. 路由选择协议概述
3. RIP协议
4. OSPF协议
5. BGP协议

外部网关协议 BGP

- 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。

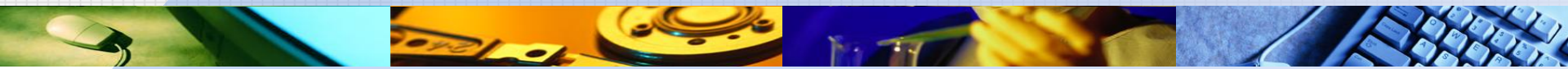
外部网关协议BGP背景

- （1）因特网的规模太大，使得域间路由选择非常困难。
- （2）要寻找最佳路由是很不现实的。
  - **由于各自治系统是运行自己选定的内部路由选择协议，使用自己指明的路径度量，因此，当一条路径通过几个不同的自治系统时，要想对这样的路径计算出有意义的费用是不可能的。**
- （3）域间路由选择必须考虑有关策略。
  - **BGP发言人和自治系统AS的关系。**
  - **一个BGP发言人构造出的自治系统连通图，它是树形结构，不存在回路。**

BGP 发言人(BGP speaker)

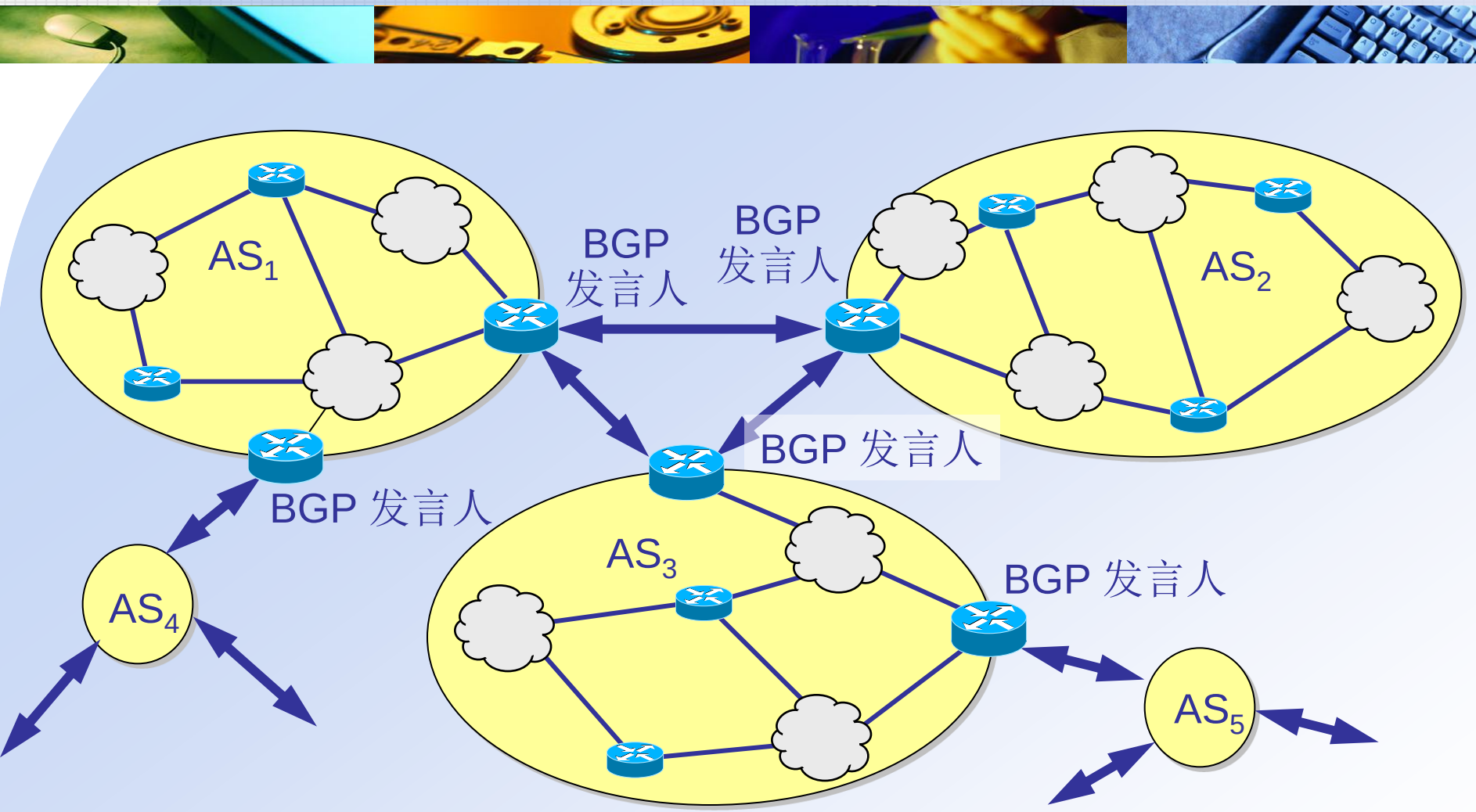
- 每一个自治系统的管理员要选择至少一个路由器作为该自治系统的“ BGP 发言人” 。
- 一般说来，两个 BGP 发言人都是通过一个共享网络连接在一起的，而 BGP 发言人往往就是 BGP 边界路由器，但也可以不是 BGP 边界路由器。

BGP 交换路由信息



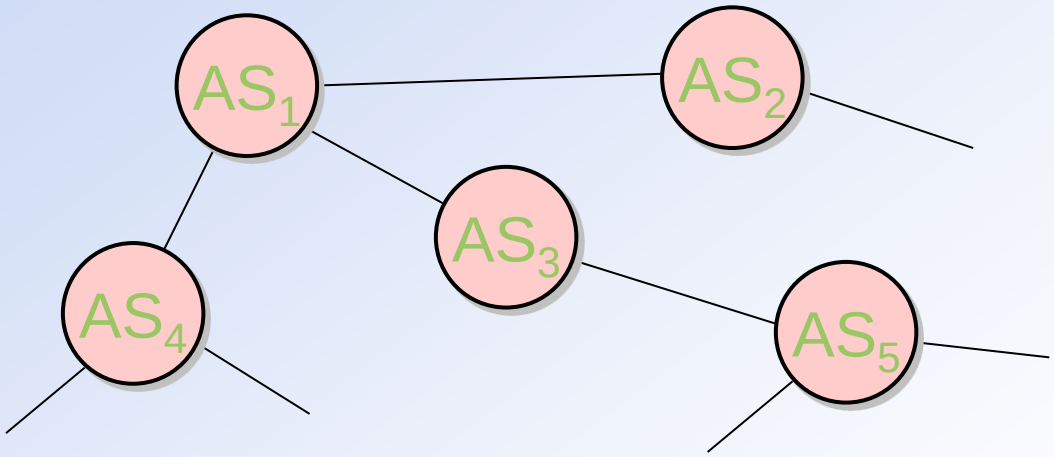
- 一个 BGP 发言人与其他自治系统中的 BGP 发言人要交换路由信息，就要先建立 TCP 连接，然后在此连接上交换 BGP 报文以建立 BGP 会话(session)，利用 BGP 会话交换路由信息。
- 使用 TCP 连接能提供可靠的服务，也简化了路由选择协议。
- 使用 TCP 连接交换路由信息的两个 BGP 发言人，彼此成为对方的邻站或对等站。

BGP 发言人和自治系统 AS 的关系



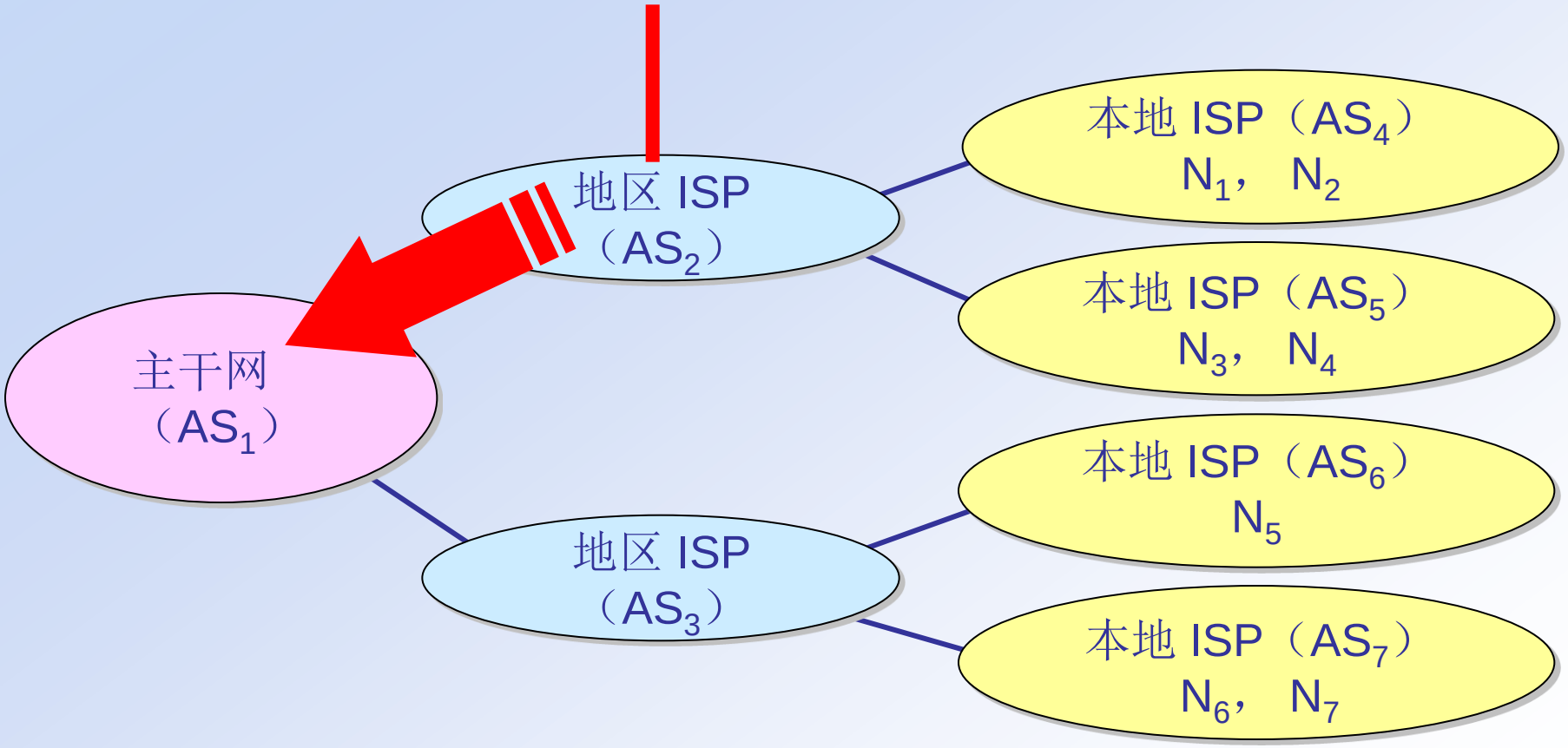
AS 的连通图举例

- 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。



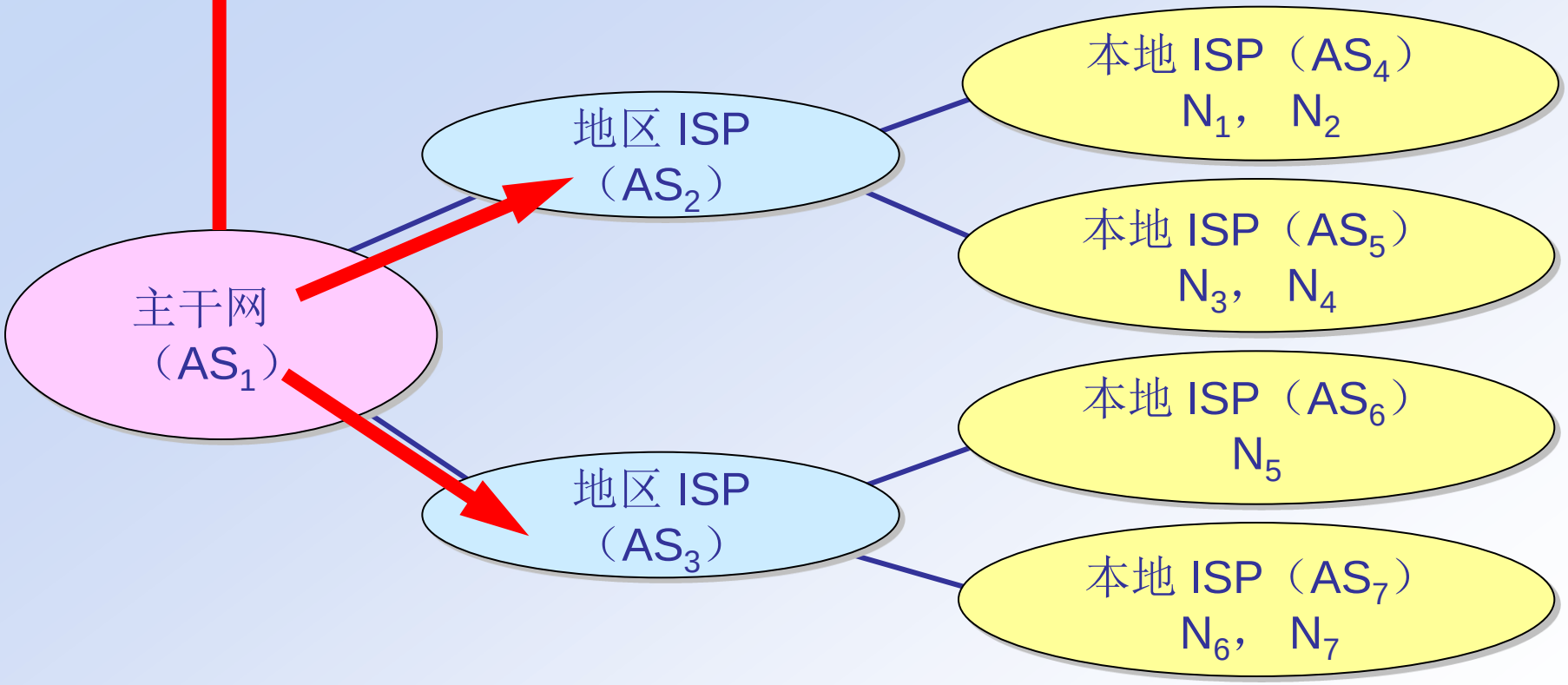
BGP 发言人交换路径向量

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

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-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。

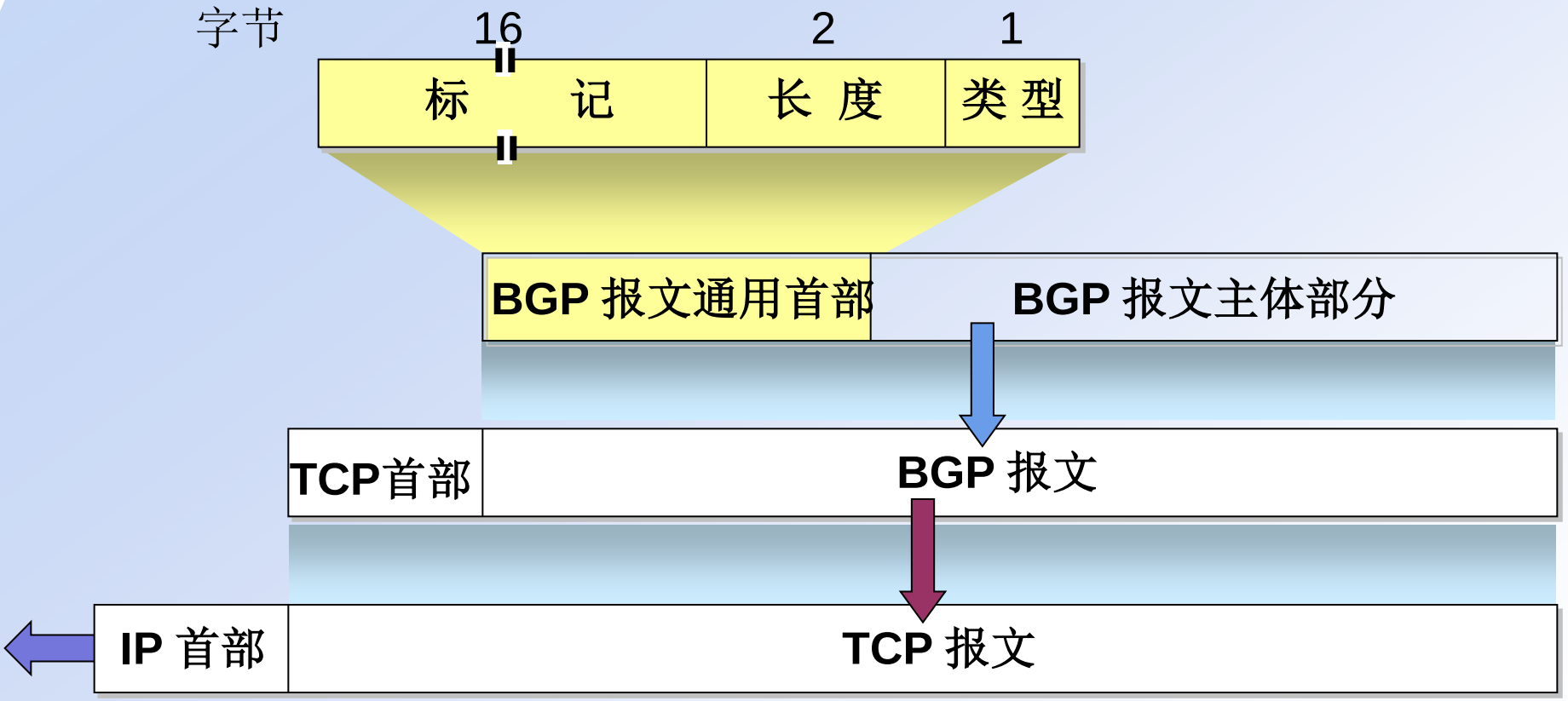
(2) 更新(UPDATE)报文，用来发送某一路由的信息，以及列出要撤消的多条路由。

(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



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。
- **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允许使用**基于策略的选路**，使得路由选择和路由管理都很复杂。