1.3 计算机网络的基本理论与技术

1.3.1 基本网络
1.3.2 命名与定位（Naming & Locating）
1.3.3 IP互连与分组交换
1.3.4 路由与寻址（Routing & Addressing）
1.3.5 数据运输
1.3.6 组播

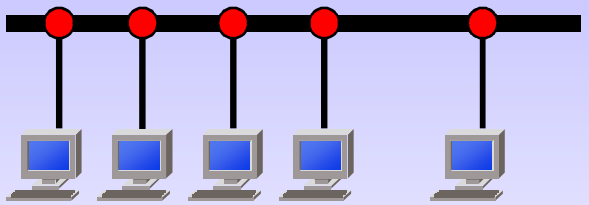
1.3.1 基本网络：之一以太网

- 1975年纯ALOHA原始ethernet：单工竞争系统，基本思想：
  - 无连接，先说后听，想发就发，错了重发；
  - 对数据帧不编号，不要求对方发回确认；不可靠交付，尽力而为
  - 建立在近距离、信道出错概率小-->局域网，出错由高层重发
- 随机接入协议
  - Time sloted ALHOA；Aloha;Taking turns
  - CSMA（载波侦听多路访问）：先听后说+指数退避
    - 1持续CSMA、非持续CSMA、P持续CSMA
  - CSMA/CD：多点接入、载波监听、碰撞检测
    - 信道效率 传送距离越短，发送帧时T_0越长，效率越高
- 以太网优势
  - 可扩展（10M—10G），灵活（多种媒介、全/半双工、共享/交换），便宜、易于安装使用、稳健性好。

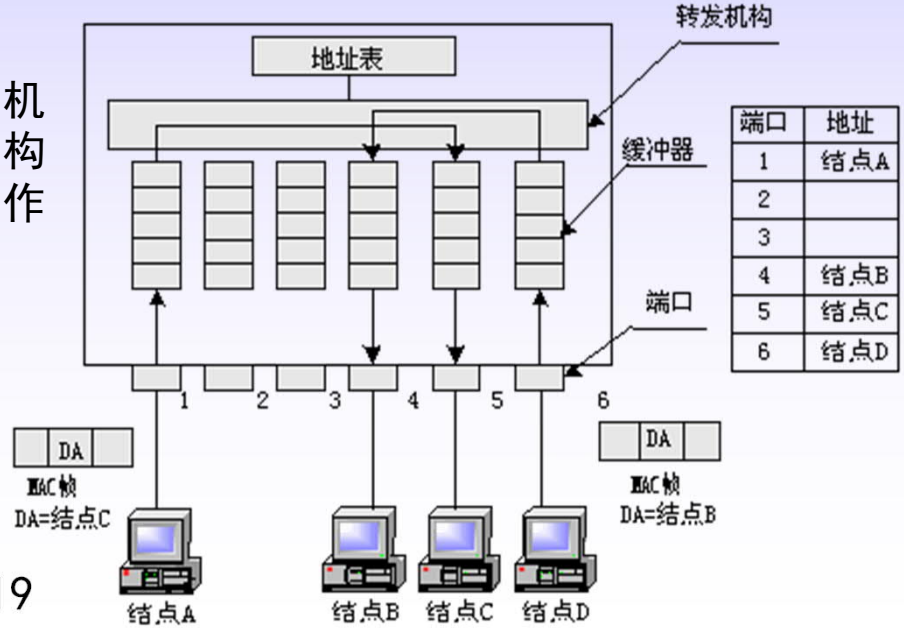
以太网的基本设备

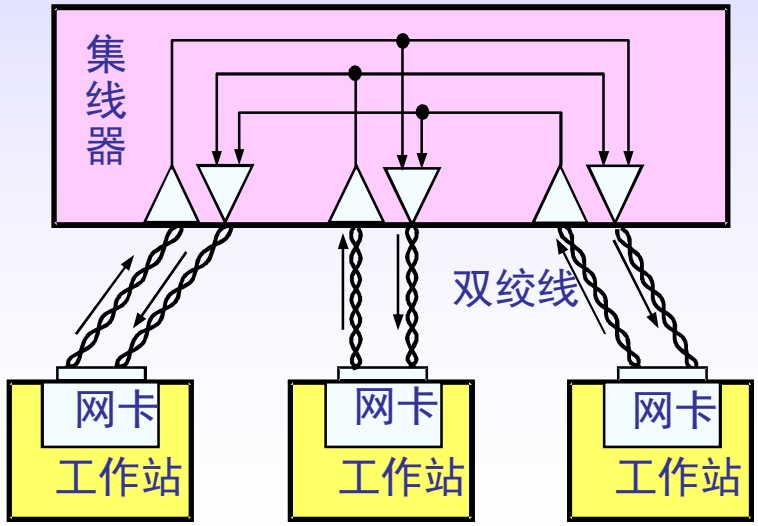
- 集线器（HUB）：物理层互连设备
  - 1进多出，相同速率，无帧缓冲/线障隔离，使用方便
  - 带宽受限 ，广播风暴 ，单工传输，通信效率低
- 交换机（Switch）：链路层互连设备
  - 依帧头存储转发以太帧；隔离广播;构成VLAN；独立带宽，对用户透明，即插即用，自学习
  - 实现方法
    - 直接交换方式
    - 存储转发方式
    - 改进直接交换方式。

共享总线的集线器

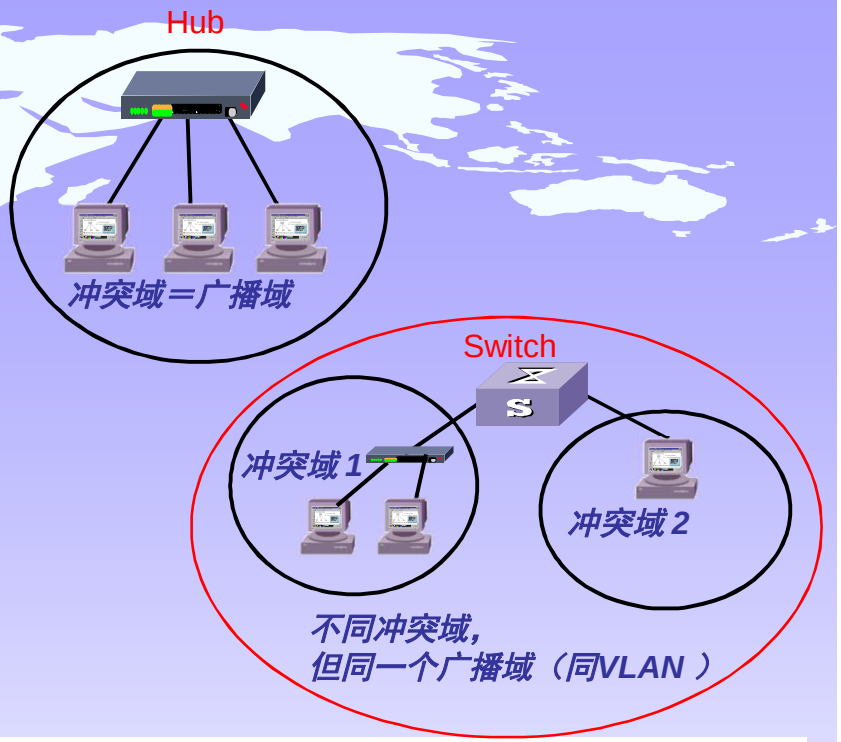


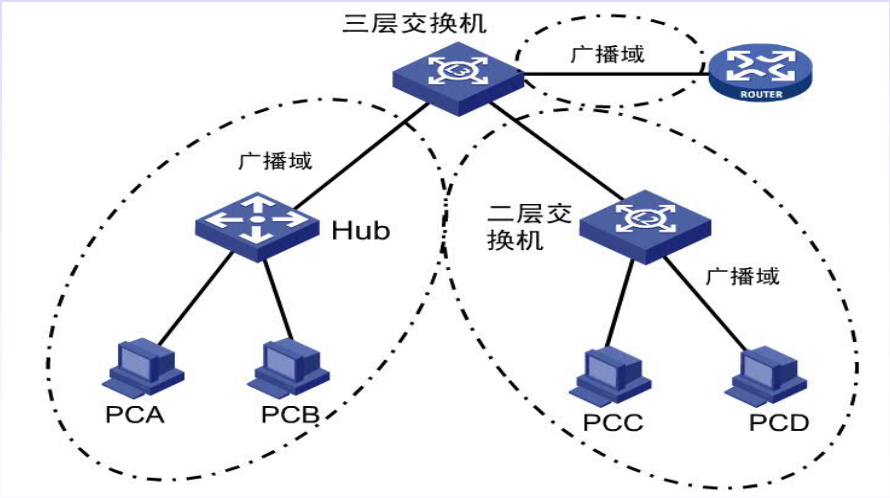
交换机的结构与工作过程



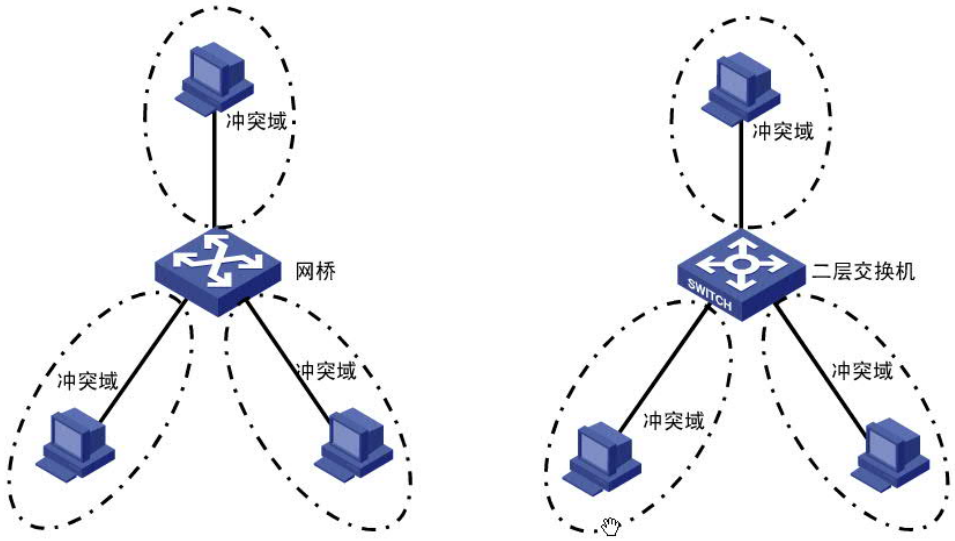


- **直通转发（cut-through）：**
  - 交换机检测到**目标地址后即转发帧**
  - 优点－转发延迟小；缺点－错误率高
- **存储转发(store and forword)**
  - **完整地收到帧**并检查无错后才转发
  - 优点－错误率低；缺点－转发延迟大
- **广播域与冲突域**
  - 同时**共享同一广播帧**的计算机子网
  - 同时**共享同一传播媒体**的计算机子网





广播域：路由器或三层交换机的三层接口处于独立的广播域中，终端主机发出的广播帧终止在这些三层接口上



交换以太网每个端口处于独立冲突域

网卡与MAC地址模式

- 网卡功能
  - 数据的封装与解封
  - 链路管理：CSMA/CD
  - Bit的编码与解码
- MAC地址
  - Unicast:单播帧地址,仅对某个网卡
  - Broadcast:广播帧地址,仅对某个子网
  - Multicast:多播帧地址,组地址
  - 杂收模式：Promiscuous mode:接收总线上所有的可能接收的帧

高速局域网:快速以太网100 Mbps

- 对10 Mbps 802.3 LAN的改进
  - 10Base-T，使用HUB
  - 局域网发展史上重要里程碑
- Fast Ethernet标准
  - 1995年，IEEE通过802.3u标准，实际上是802.3的一个补充。原有的帧格式、接口、规程不变，只是将每比特时间从100ns缩短为10ns。

Name	Cable	Max. segment	Advantages
100Base-T4	Twisted pair	100 m	Uses category 3 UTP
100Base-TX	Twisted pair	100 m	Full duplex at 100 Mbps
100Base-FX	Fiber optics	2000 m	Full duplex at 100 Mbps; long runs

高速局域网:100Base-TX/F

- 100Base-TX
  - 使用2对5类平衡双绞线或150Ω屏蔽平衡电缆，1对 to the hub，1对 from the hub，全双工；
  - 5类双绞线使用125 MHz的信号；
  - 4B/5B编码，5个时钟周期发送4个比特，物理层与FDDI兼容，比特率为 125 * 4/5 = 100 Mbps；
- 100Base-FX
  - 使用2根多模光纤，全双工
- 100Base-T4 和 100Base-TX 统称 100Base-T
- 两种类型的HUB
  - 共享式 HUB，一个冲突域，工作方式与802.3相同，CSMA/CD，二进制指数后退算法，半双工 …
  - 交换式HUB，输入帧被缓存，一个端口构成一个冲突域。

高速局域网：1000Mbps以太网

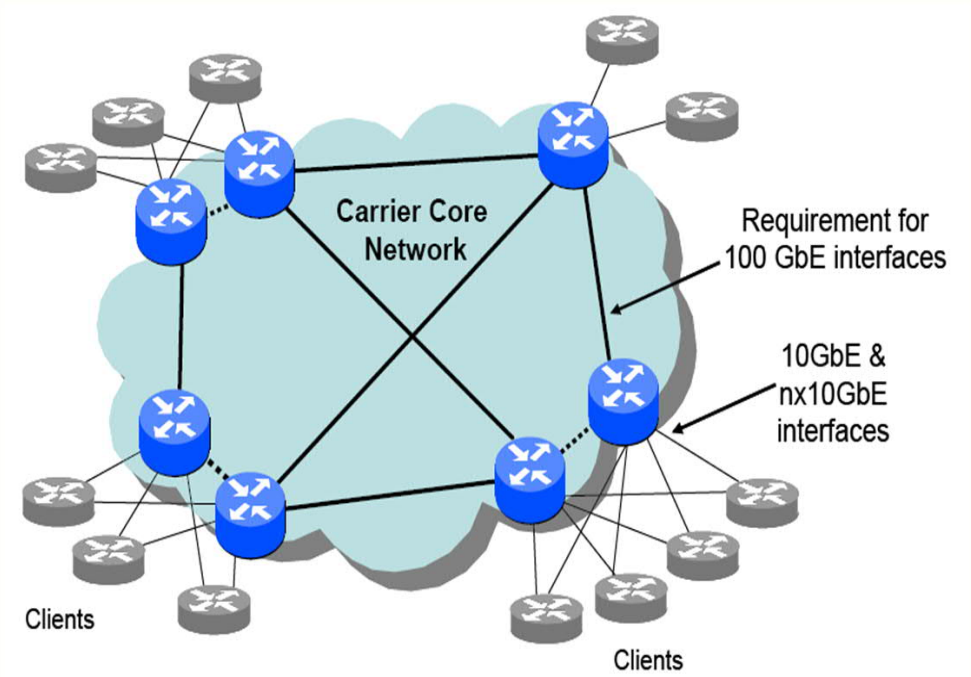
- 工作方式
  - IEEE 802.3定义的10M/100M以太网一致的CSMA/CD帧格式和MAC层协议
  - 以太网交换机（全双工模式）中的千兆端口不能采用共享信道方式访问介质，不使用 CSMA/CD 协议,而只能采用专用信道方式.
  - 在专用信道方式下，数据的收/发能够不受干扰地同步进行.
  - 在半双工方式下仍使用 CSMA/CD 协议
  - 物理层采用已有光纤通道技术；
- PAUSE协议
  - 规范发展完善了PAUSE协议，不采用CSMA/CD协议完成全双工操作。
  - 该协议采用不均匀流量控制方法最先应用于100M以太网中。
- 流控
  - 利用802.3定义的Pause控制帧进行流量控制，要求发送数据节点暂停数据发送，避免缓冲区溢出造成的丢包。
  - 只有在全双工时，才支持Pause流控，半双工时不支持流控。

万兆（10Gbps）以太网

- 2002.6月正式发布802.3ae 10GE标准
  - 只全双工，不支持单工和半双工，也不采用CSMA/CD
  - 不持自协商；提供广域网物理层接口。
- 长距离(40-50KM)网络
  - 扩展了网络的覆盖区域，且标准简化。
  - 支持现存的大量SONET网络兼容
- 两种物理层技术：
  - 局域网物理层LAN PHY；10.000Gbps精确10G；
  - 广域网物理层WAN PHY；入OC-192，异步SONET/SDH
  - 与10M/100M/1000Mbps帧格式完全相同；

100Gbps以太网

Ethernet in carrier networks



- 以太网封装比SONET/SDH更简单且成本更低
- 40 Gbps已成为过渡产品
- 2010年6月22日，IEEE802.3ba和100Gb/s以太网技术标准已经正式获审通过。
- 国内华为、中兴；国外Juniper Network、CISCO；上海贝尔已经开发出了自有标准100Gbps以太网接口路由器

100 GbE standard is needed

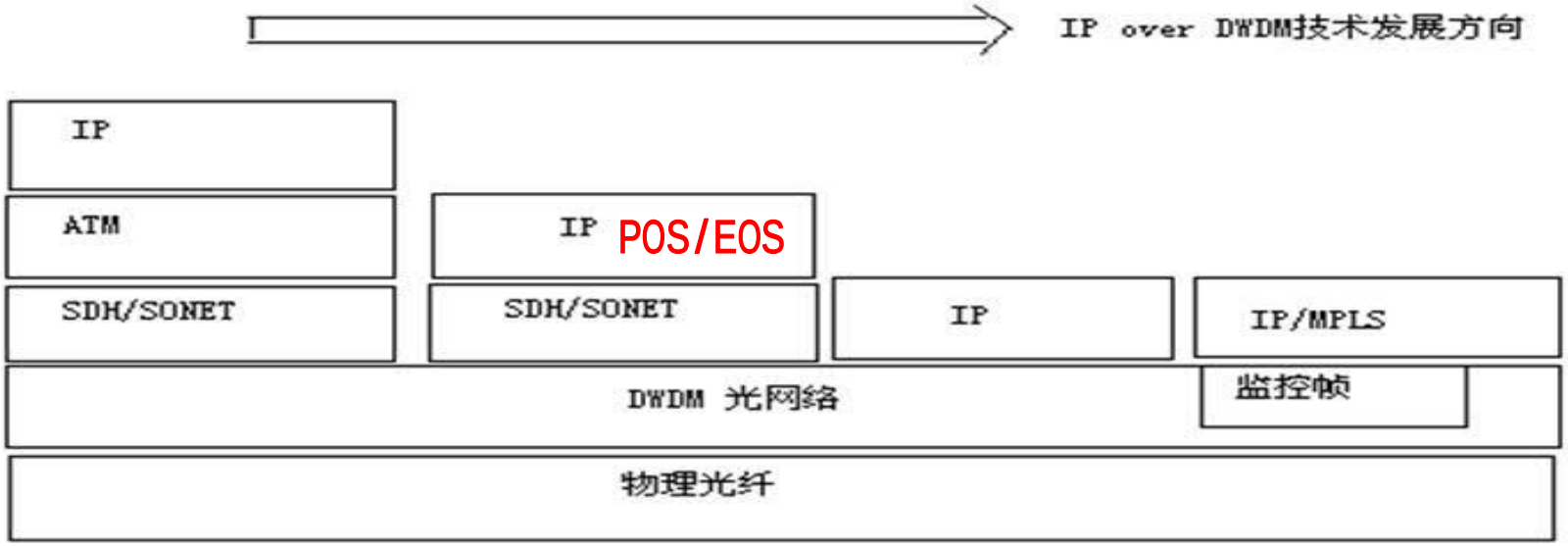
Jumping to 100 GbE Ethernet at 100 Gbps may take place by using several or just one lambda(s):

100 GbE over 10x10Gbps, *or* 4x25Gbps *or* 1x100Gpbs

Different from 10x10GbE !!

之二：POS+SDH/WDM网络主干

- POS技术-Packet Over Sonet/SDH
  - 采用高速光纤传输，以点对点方式提供从STM1(155.520)到STM64(OC-192:9953.280=10Gbps)甚至更高的传输速率
  - 将IP包直接封装到SDH帧中，提高了传输的效率。

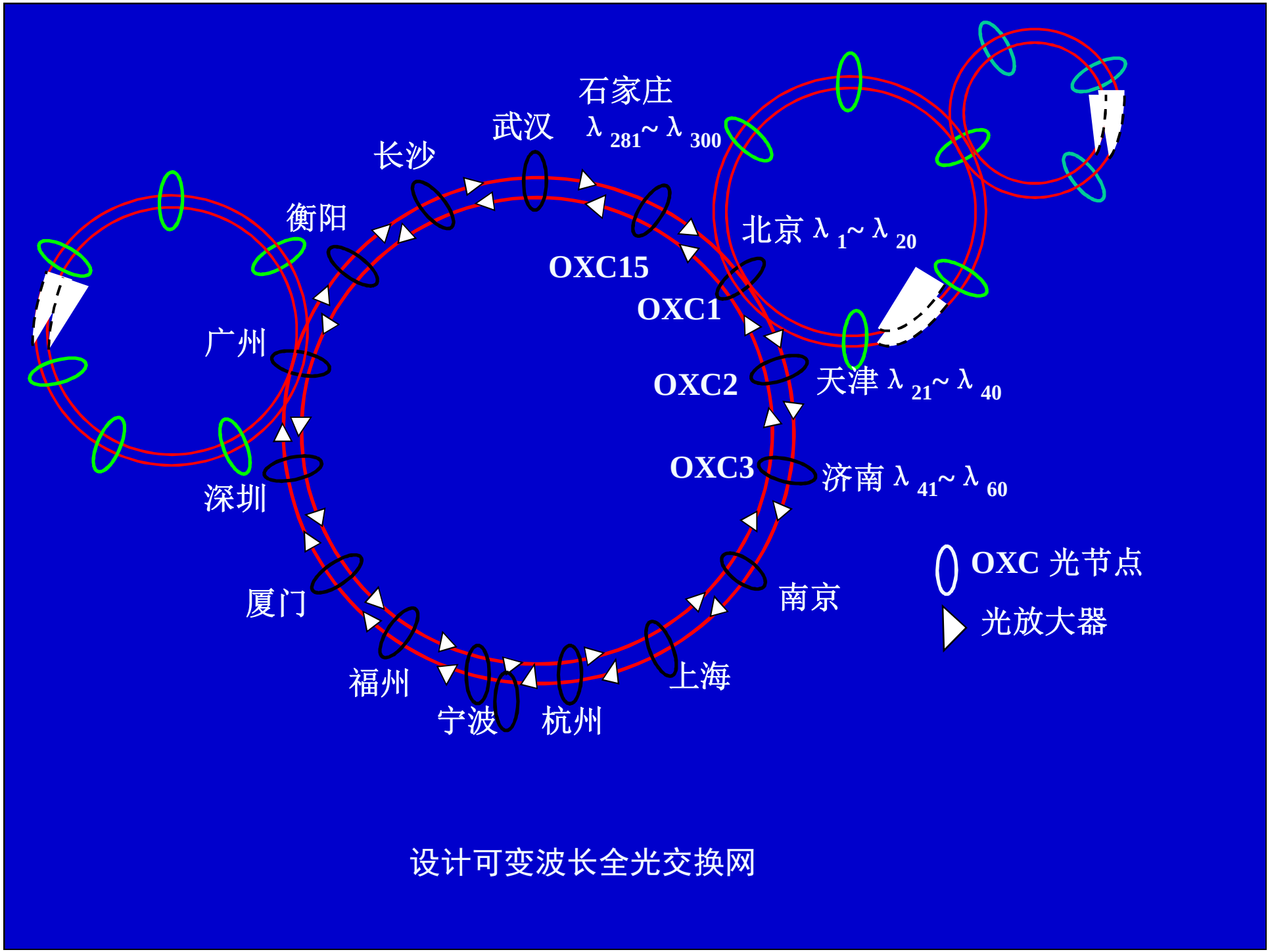


之二：EOS+SDH/WDM网络主干

- Ethernet over SONET/SDH
  - 在SDH/SONET或DWDM上加以太网的二层或三层交换板，这样在传统的SDH或WDM设备中可提供以太网接口100/1G/10G/100Gbps
  - 采用1套EOS设备，放在现有端到端的SDH/SONET(或DWDM)与以太网二层或三层交换机之间，实现端到端的传输连接
  - 在现有的以太网二层或三层交换机上增加EOS(STM-1/STM-16)接口

之四：全光网

- 全光网络：
  - 指光信息流在网络中的传输及交换始终以光的形式实现，而不需要经过光/电、电/光变换。
  - OTN:Optical Transport Network；
  - ASON ：Automatically Switched Optical Network
- 波分复用
  - 使波长本身成为组网（分插、交换、路由）的资源。逐步成熟的光分插复用（OADM）和光交叉联接（OXC）技术，只提供带宽传输的光层开始有组网能力。
  - 近30年间，随着光器件的发展和光系统的演进，光传输系统的容量已从Mbps发展到Tbps，提高了近10万倍。200光纤/光缆；10.92T/光芯



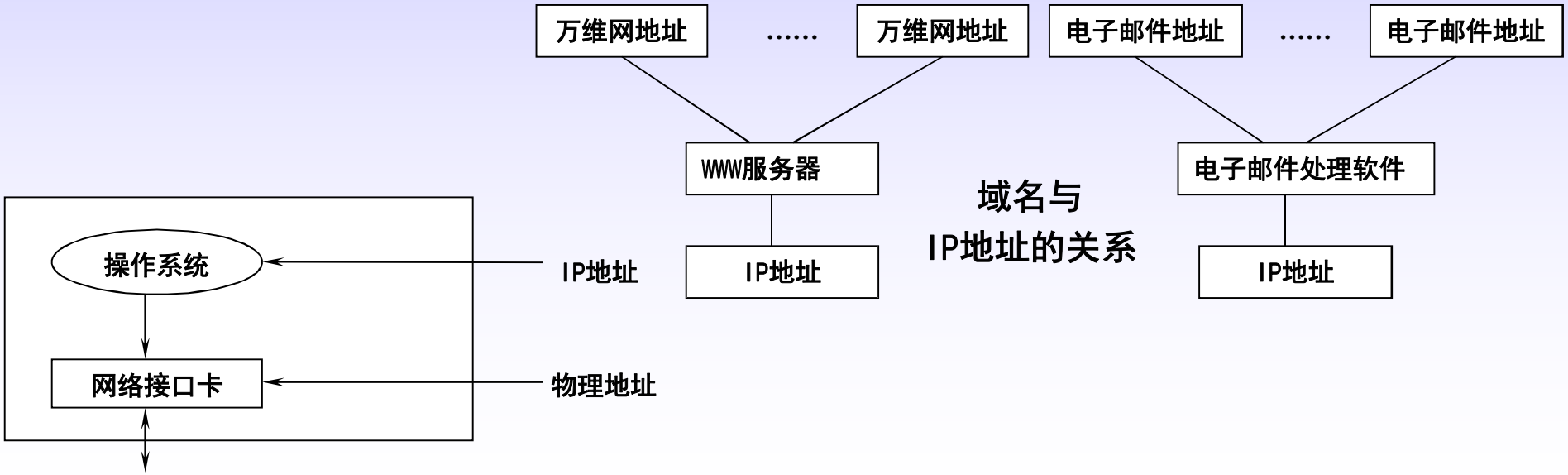
设计可变波长全光交换网

1.3.2 （网络实体）命名与定位

- 直接定位与解析定位
- 因特网是全球单一可寻址的抽象网络
  - 必须连接全球所有计算机或其它设备
  - 必须给每个计算机或设备（路由器等）一个全球唯一ID
  - IP协议希望这样，故有IPv4 = 2^{32}个地址，A/D/C/E类
  - IP 地址结构 = Net号 + Host号
  - 位置与身份合一
- IP地址引发问题（简单,但重载）
  - 不便记忆：→域名产生→解析需要
  - 局域网早先出现：导致 Mac←→IP变换→ARP协议
  - IP地址分配不均不够：NAT，三个段公用私网地址：10/8:1个A+172.16.—172.31/16:15个B+192.168.—192.168/16:1个B = 17621775个IP

因特网的三地址

- 用户/网络及物理地址
  - 域名：公司/机关/团体/个人注册的因特网可访问的世界唯一ID
  - 网络(接口)地址：同一体系结构中的可访问的计算机ID:IP地址
  - 物理地址：同一体系结构中物理媒体可访问的计算机某端口的唯一ID:MAC地址



域名与
IP地址的关系

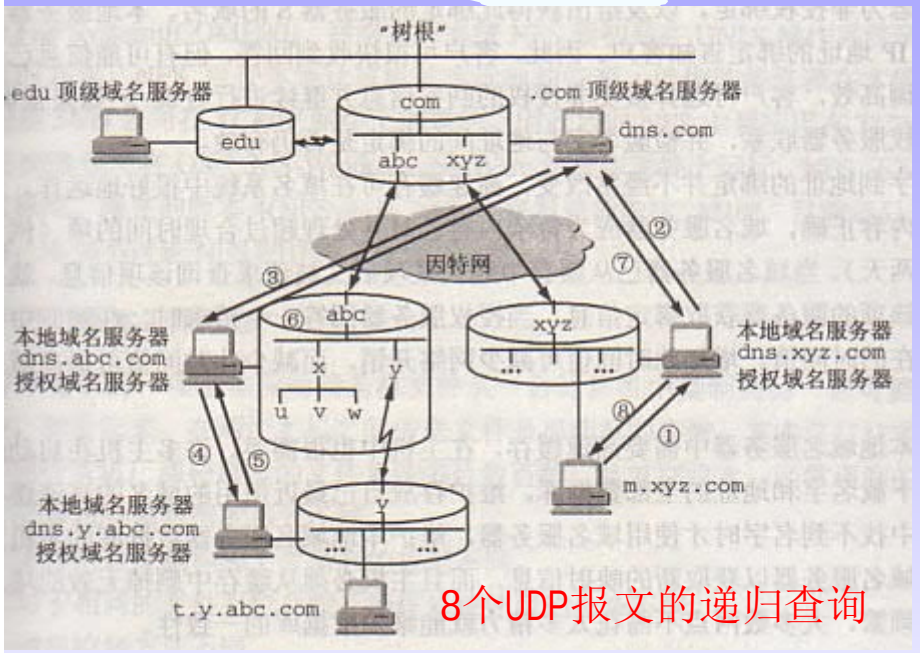
IP地址与物理地址的关系

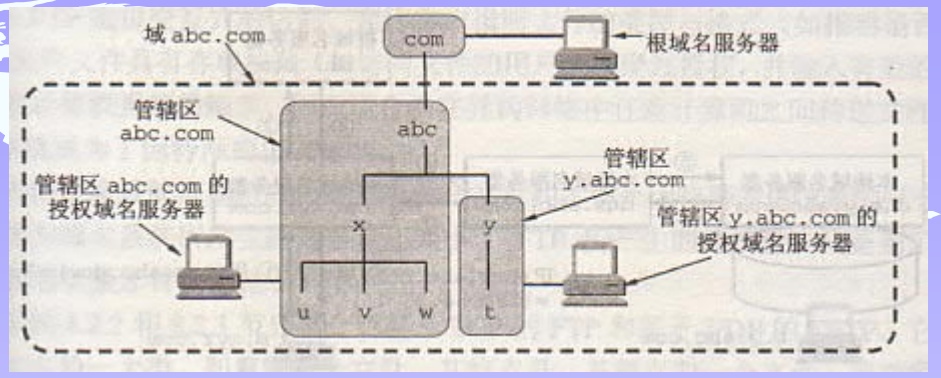
物理地址的配置和作用

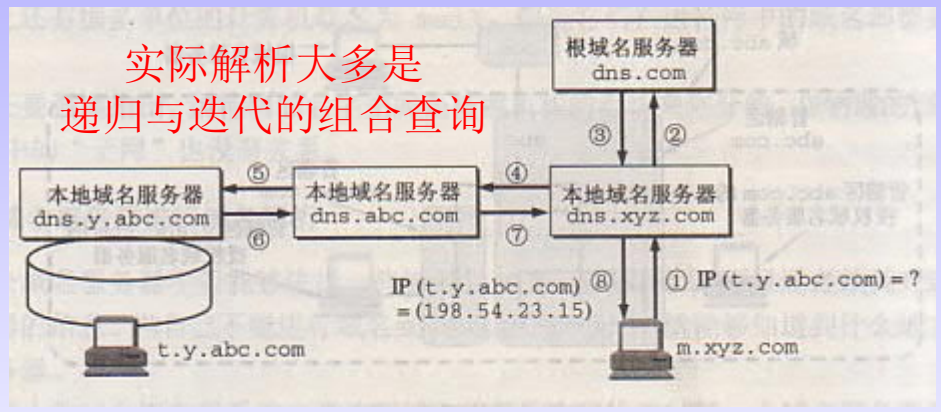
- 作用：
  - 过滤：指明网卡NIC，过滤不属于本主机接收的数据
  - 广播：识别广播地址标志，接收广播报文并送主机
- 地址配置：
  - 固定方式：物理地址由厂商决定。不可改变
  - 可配置方式：通过EPROM内程序进行交互命令配置
  - 动态配置方式：可自动修改和管理物理地址，启动网络时，通过程序配置一个不冲突的物理地址

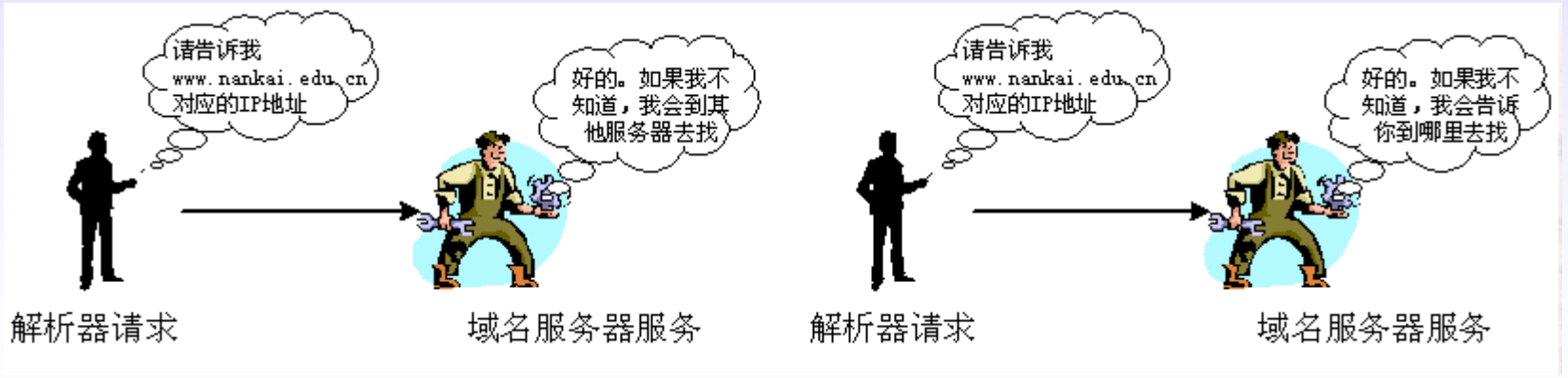
域名解析

域名：标识计算机的名字









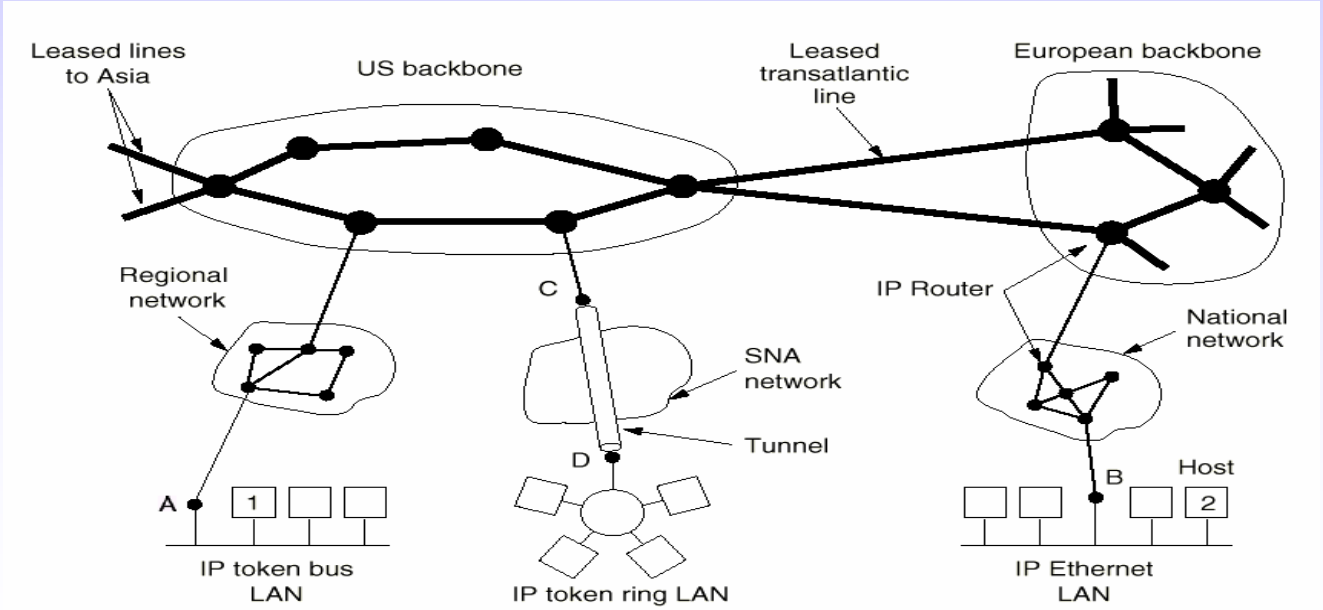
递归查询(给出直接IP地址)　　*迭代查询（给出间接IP地址）*

1.3.3 IP互连与分组交换

- 多个独立网络怎么互连？（历史：包容策略战胜统一策略）
- 多个网络互连的2个主要问题：
  - 异构：
    - 以太网、令牌环、点-点链路及各种交换网络，每个都有自己的地址模式、媒体访问协议及服务模式等
    - 传输介质不同/网络拓扑结构不同
    - 介质访问方式不同/网络编址方式不同
    - 分组长度/有连接/无连接服务的区别
    - 传输控制方式不同
    - 各层协议的功能定义、格式、接口与调用方式不同
  - 可扩展：
    - 网络不断扩大，路由规模指数增（100万—22.7亿个节点）
    - N个路由器（交换机）+1个路由器（交换机）= 网络；车床/汽车/飞机…不行；

Internet网络层协议

- 在网络层，Internet可以看成是自治系统的集合，是由网络组成的网络。
- 网络之间互连的纽带是IP（Internet Protocol）协议。

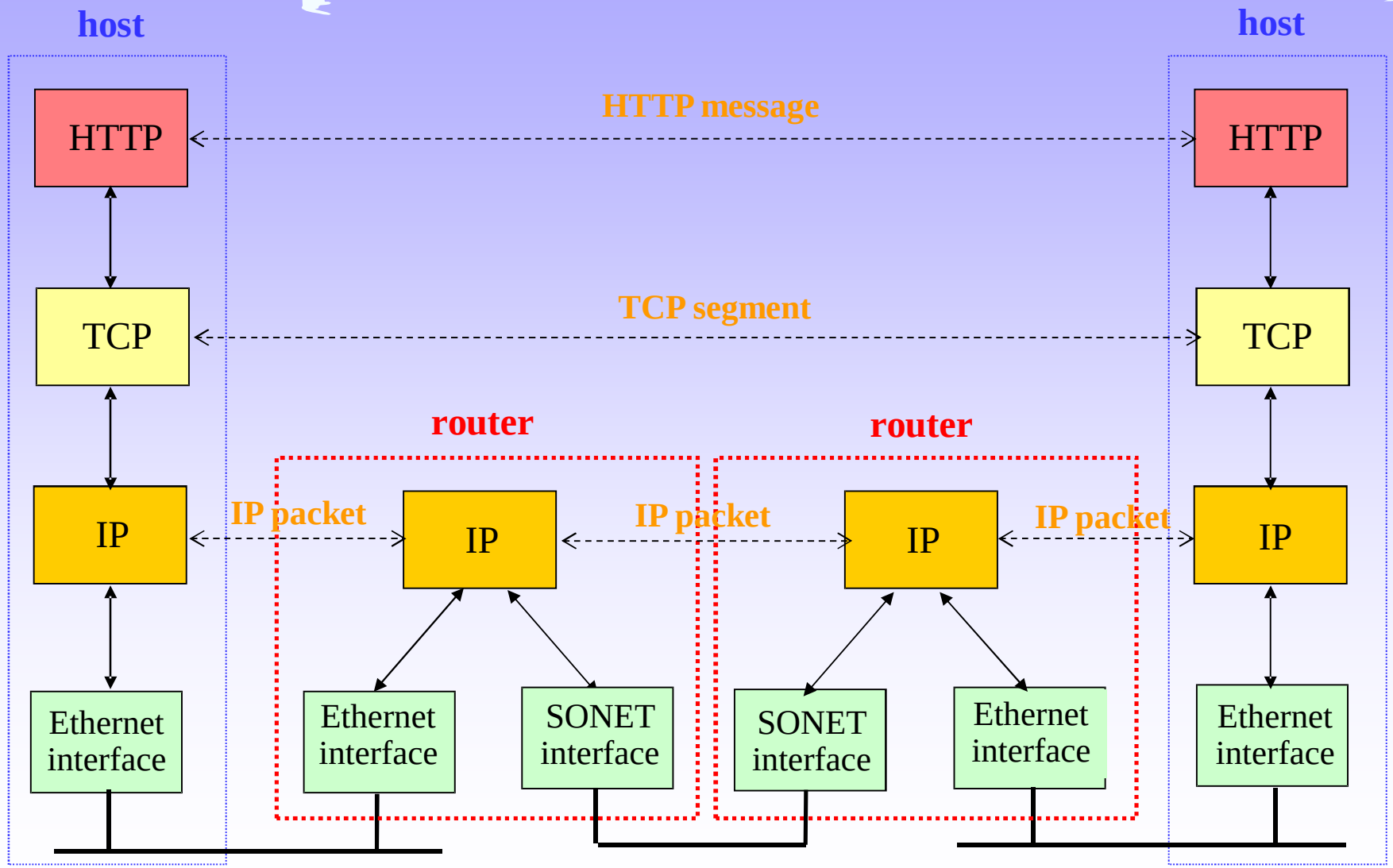


网络互连层的功能

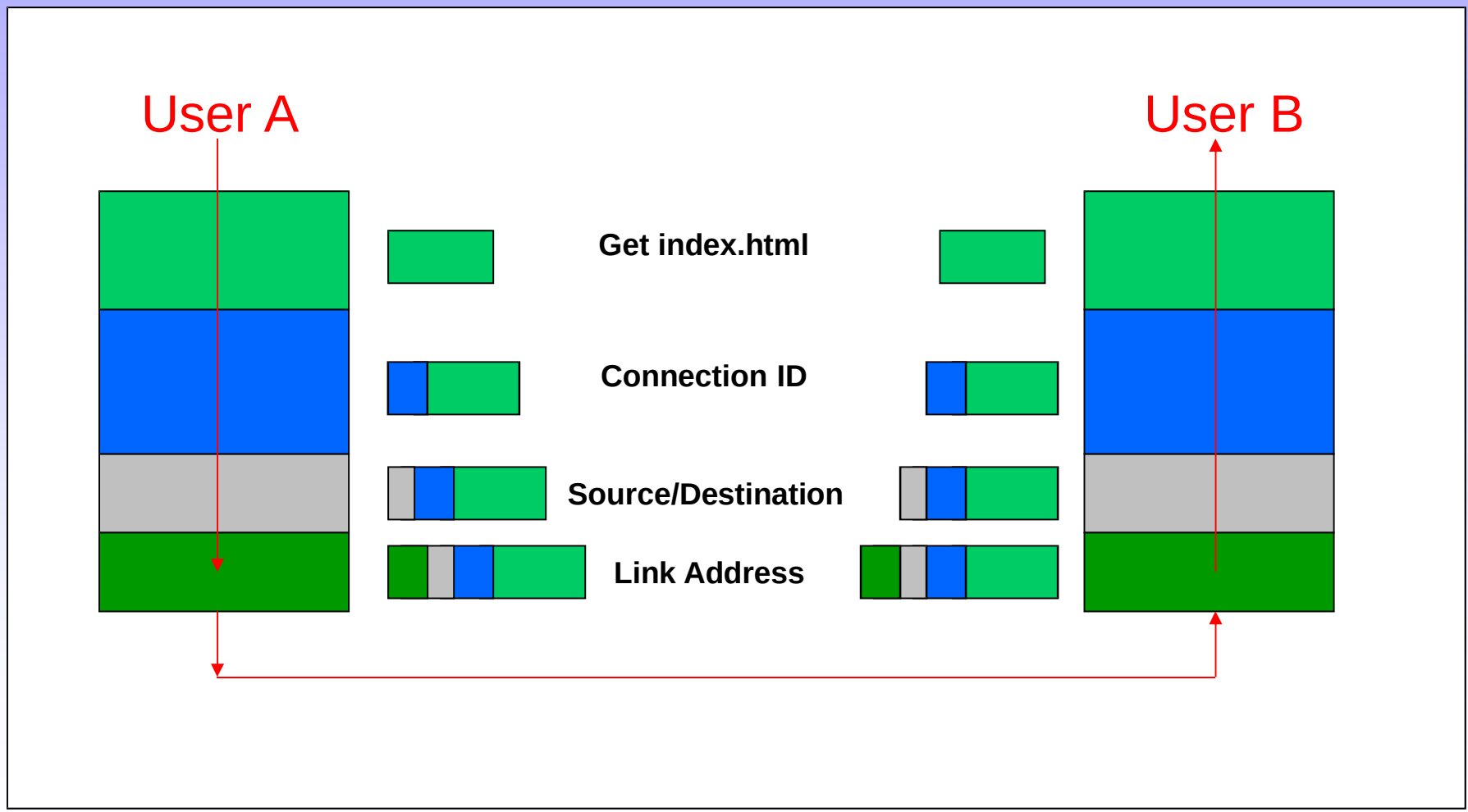
- 功能目标
  - 把多个异构或同构的网络连接成更大网络
  - 直接支持传输层的端到端服务。
- 关键问题
  - 掌握通信子网的拓扑结构，选择路由；
  - Hop by Hop 寻址到目的主机
- 为传输层提供何种服务？
  - 面向连接服务：
    - 将复杂的功能放在网络层
    - 传统电信观点：通信子网应该提供可靠的、面向连接的服务
  - 无连接服务：将复杂的功能放在传输层
    - Internet的观点：通信子网无论怎么设计都是不可靠的，因此网络层只需提供无连接服务。

- 网络层的内部组织
  - 虚电路（virtual circuit）
  - 数据报（datagram）
- 虚电路子网vs数据报子网
  - 路由器内存空间与带宽的权衡
    - 虚电路方式，路由器需要维护虚电路的状态信息；
    - 数据报方式，每个包携完整目/源地址，无状态但浪费带宽；
  - 连接建立时间与地址查找时间的权衡
    - 虚电路需要在建立连接时花费时间（带外信令）
    - 数据报则在每次路由时过程复杂
  - 服务质量QoS（Quality of Service）权衡
    - 电路方式容易保证服务质量适用于实时操作，但比较脆弱。
    - 数据报不保证服务质量，但对于通信线路故障，适应性很强。

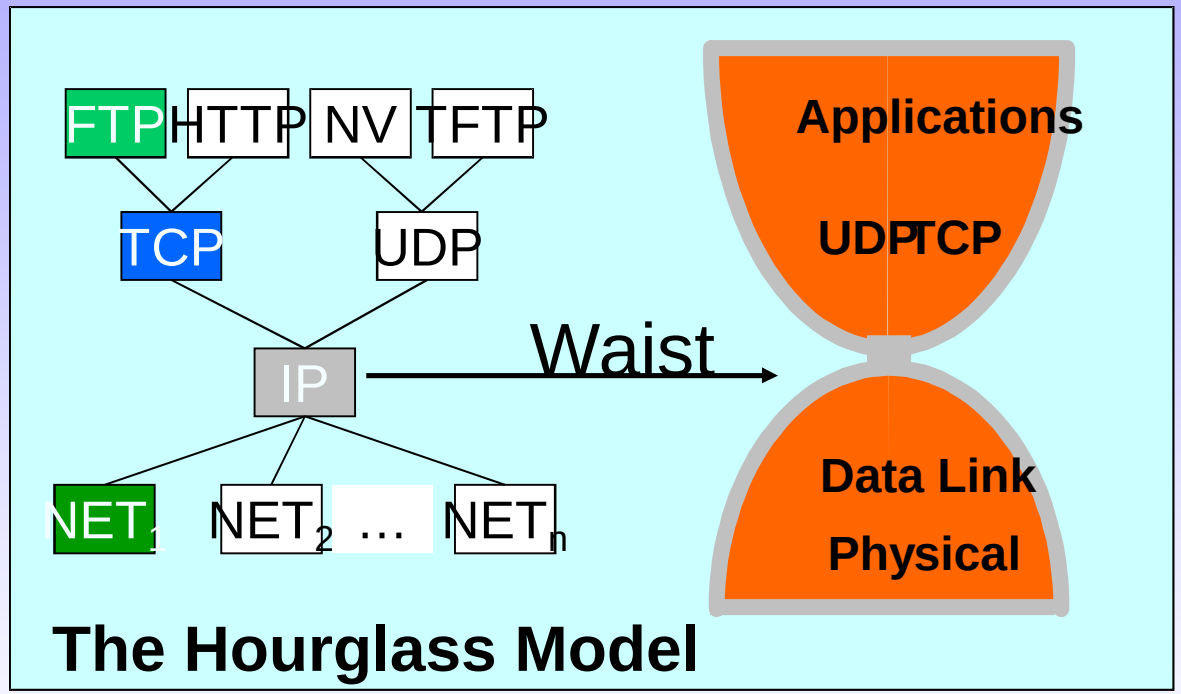
IP Suite: End Hosts vs. Routers

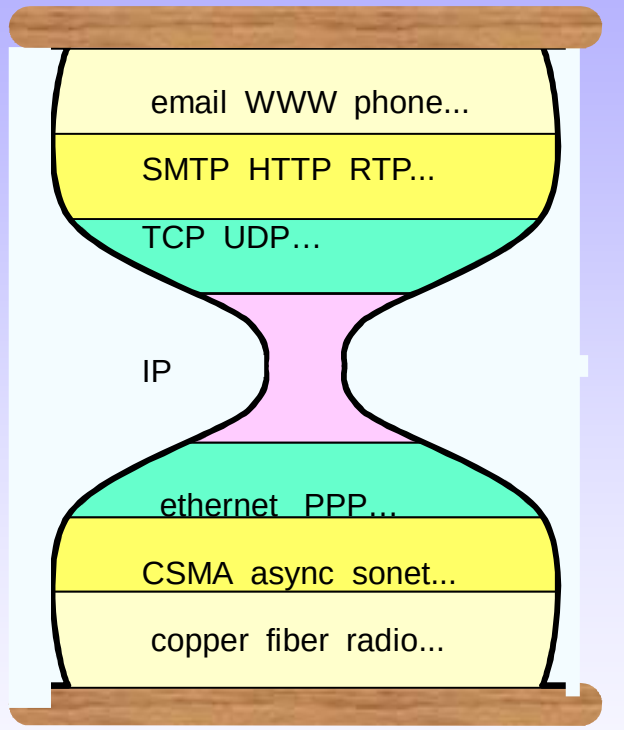


Layer Encapsulation



The “Narrow Waist” of IP

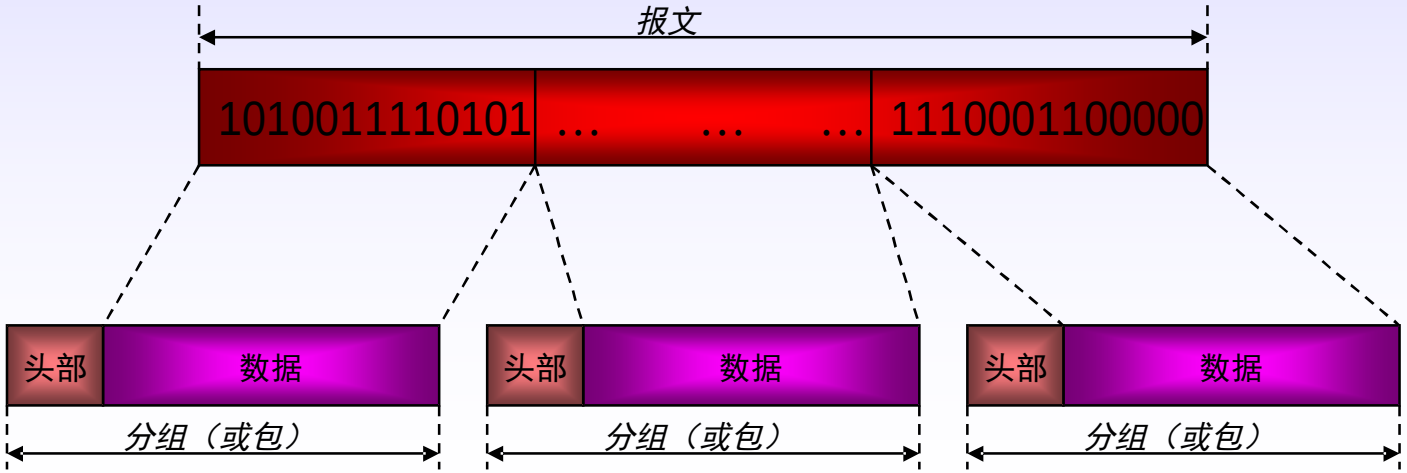




细腰促进了交互

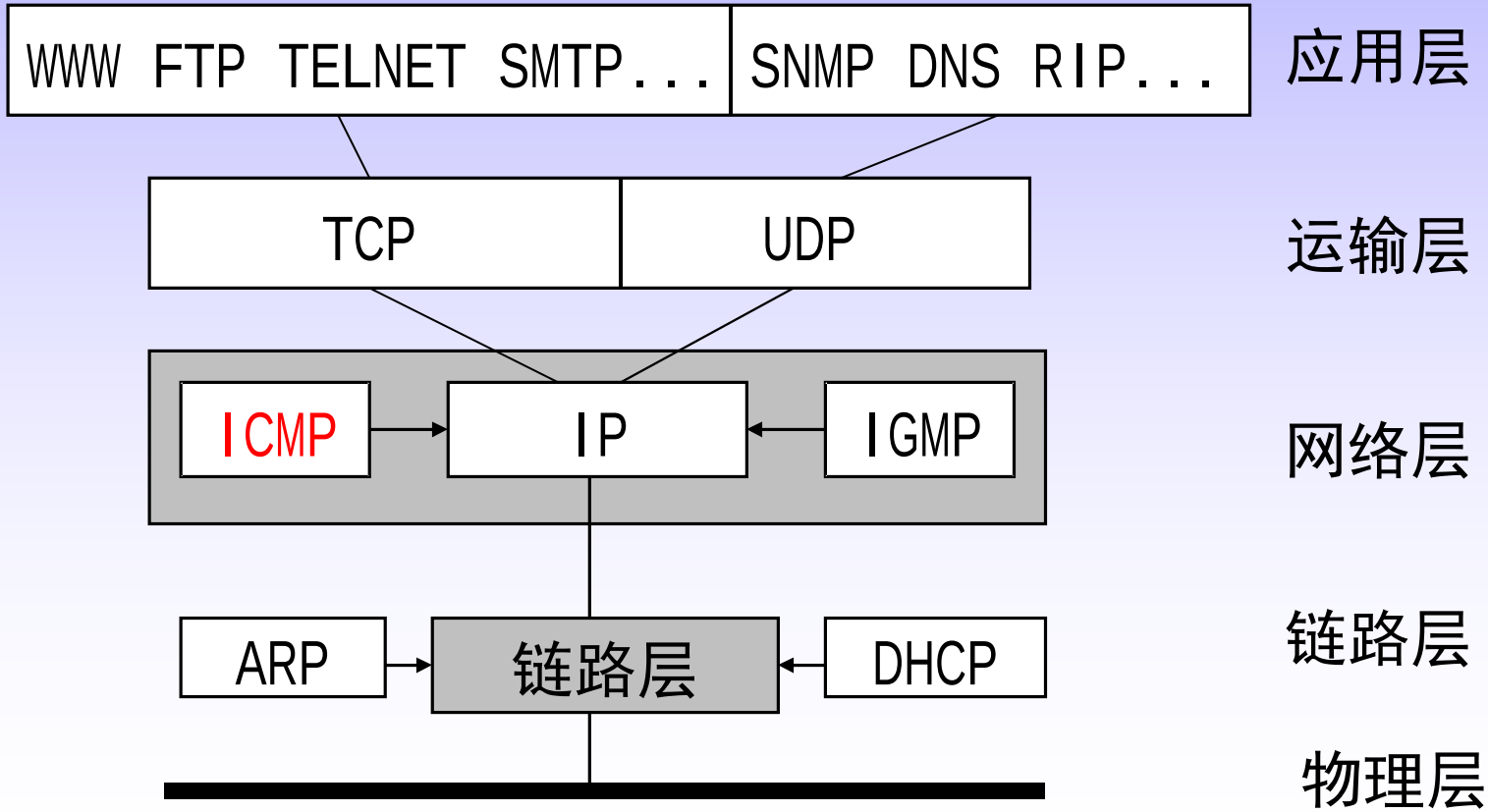
分组交换

- 转发单元包含分组的目的地址
- 没有提前建立状态（以帮助命运共担）
- 作为基本构建单元，必须在其上建类TCP
  - IP是best-effort不可靠，应用上层协议间可靠机制
- 几乎暗指统计多路复用
- 其它可选项
  - 电路交换：信令协议在带外建立整体通路（电话网）
  - 虚拟电路：混合方法，包带“目址”指通路，IP上转发
  - 源路由：全部路由包含在每个数据包中



错误报告协议(ICMP)

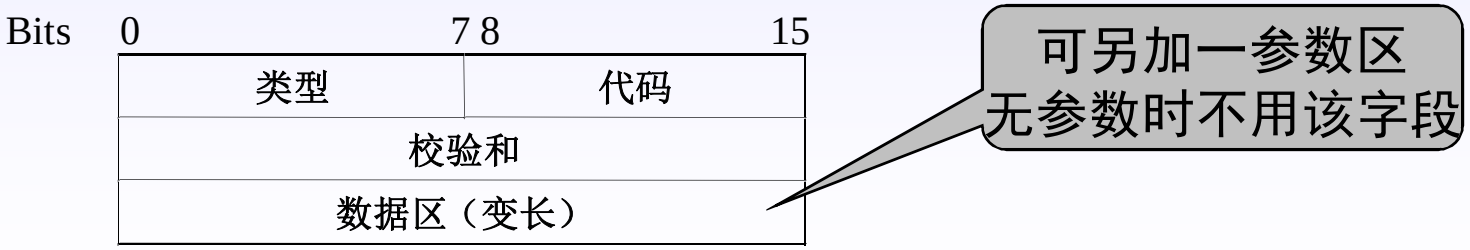
- 协议层次的回顾



ICMP协议

ICMP 报文分类

- 差错和控制报文协议：
  - 报告IP传输中发生的（差错报文+控制报文+测试报文）
  - ICMP报文=头部+数据部分
- ICMP报文封装在IP数据包中进行传输
  - IP头中的包类型=1；
  - ICMP并不是IP的上一层协议，仅用IP的转发功能

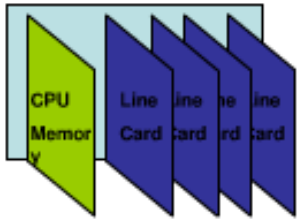


ICMP报文格式

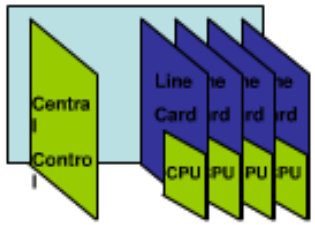
1.3.4 路由与寻址

- IP寻址
  - 每包经每路由器都要决定从哪个端口到下一节点？
  - 路由器需要匹配包之目址与路由表，选择最佳出端口
- 路由器：网络层互连设备，丰富灵活，可扩展
  - 静态路由：手工配置/简单拓扑/直观；不适合大规模/多变网络
  - 动态路由：实时自动动态更新，并算出当前最佳路由
- 问题：
  - 路由器怎样实时知道变化着的网络拓扑？
  - 路由器怎样从全局选择每个包的最优出端口？
- 方法：
  - 通过路由协议获得节点和链路的信息！
  - 通过路由算法选择走向目的地的最优下一出口！

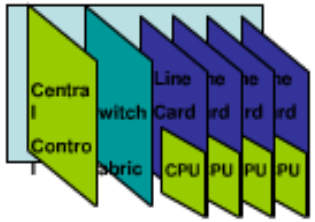
First Generation
Single CPU – multiple line cards
Single electrical backplane



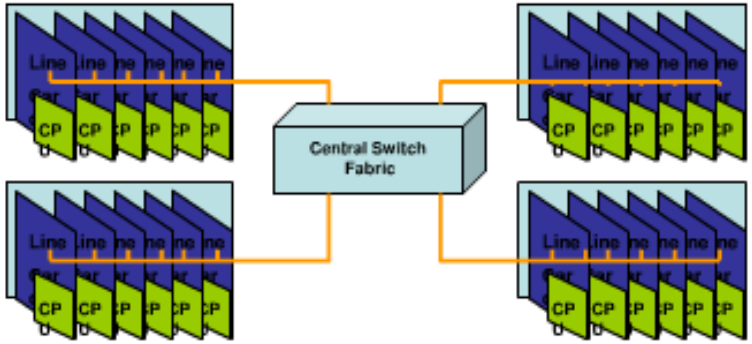
Second Generation
One CPU per Line Card
Central Controller for Routing Protocols

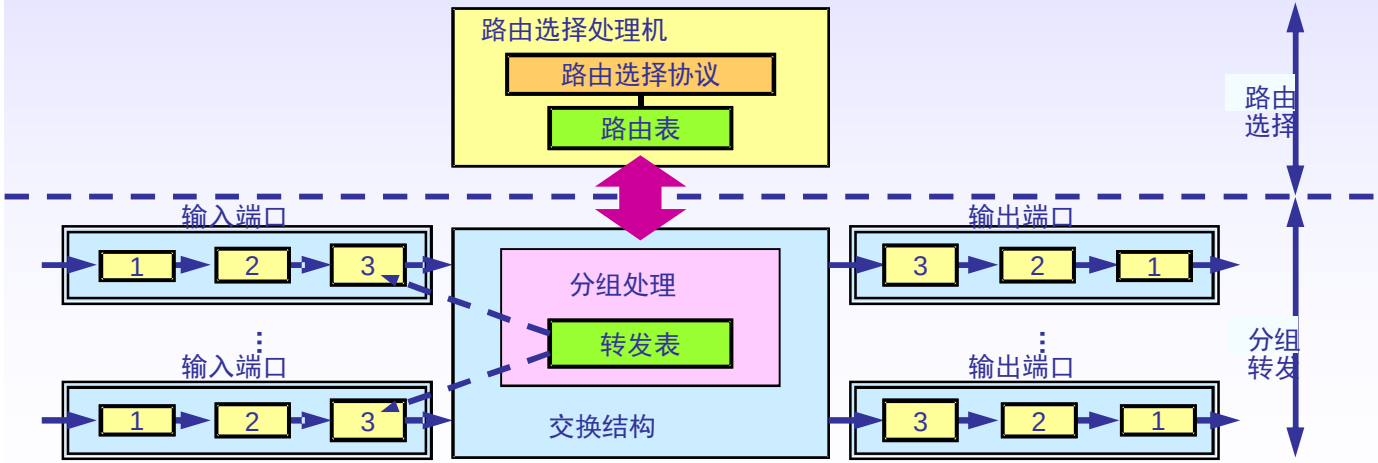


Third Generation
One CPU per Line Card
Central Controller for Routing Protocols
Switch Fabric for inter-connection



Fourth Generation
Multiple shelves of Line Cards
Centralized Switch Fabric
Optical links interconnecting Line Cards and Fabric

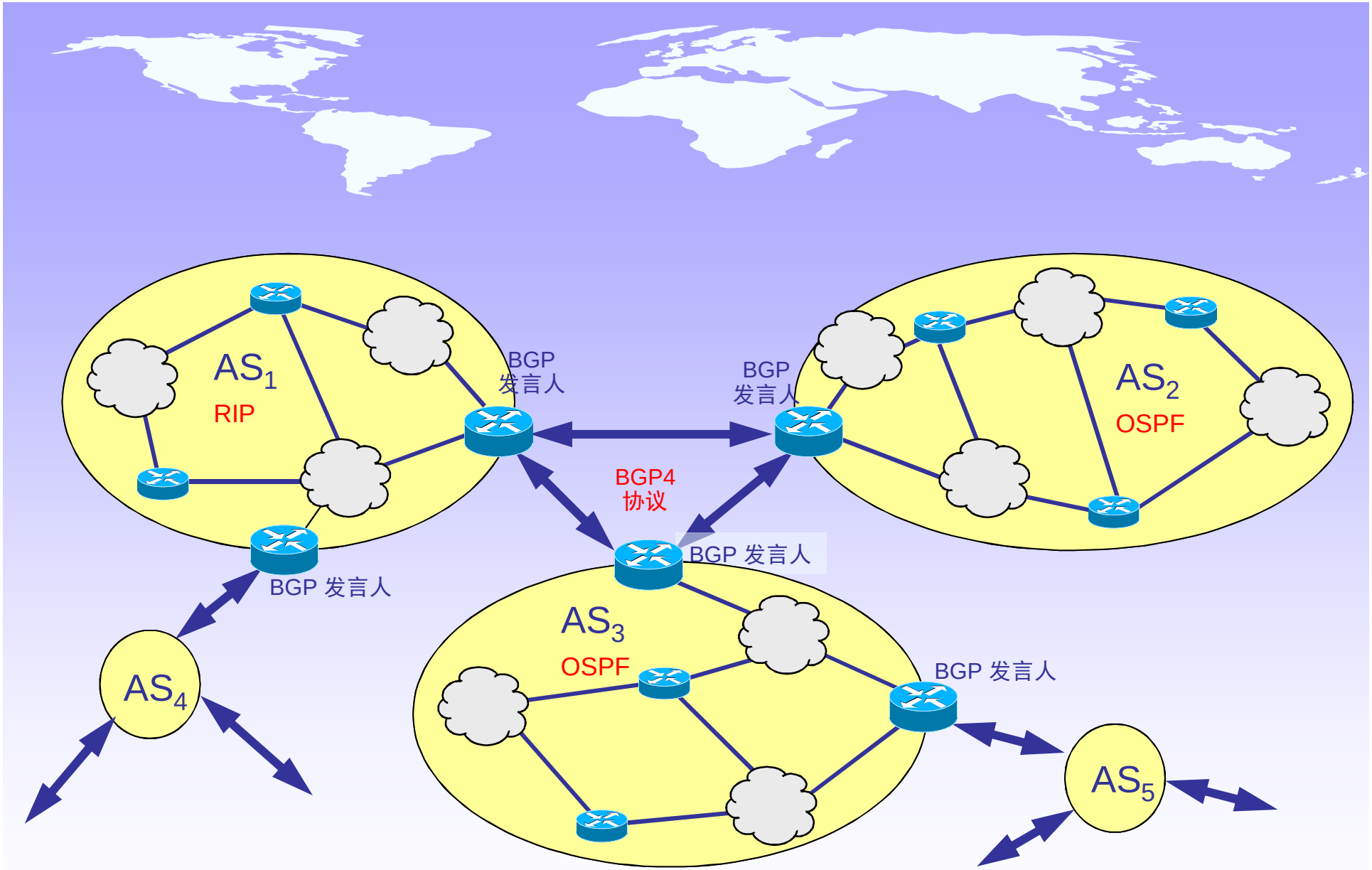




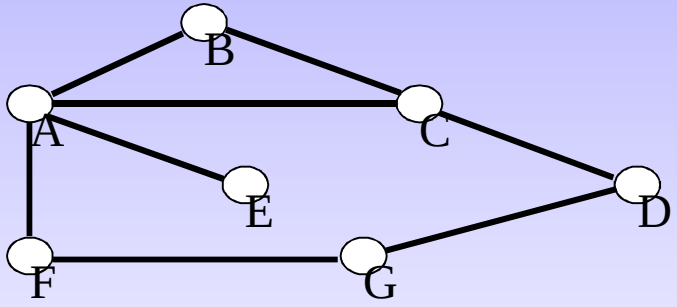
典型路由器的结构

路由协议

- 自治域AS
  - 在一个权威管理机构下运行的网络，分配唯一编号。全球所有AS组成因特网。
- 内部网关协议IGP：在一个AS内运行
  - RIP(Routing Information Protocol)适应小规模网络：定时只向邻居说（如30s），说知道的所有（自路由表）；
  - OSPF（Open Shortest Path First Protocol）适应较大规模网络：（链路）变时向所有人说（区内路由器组播），只说邻居的事（链路状态）
- 外部网关协议EGP：在AS间运行
  - 原因：因特网规模大AS间直接OSPF收敛时间很长；AS间最佳路由不现实（cost意义不统一）；各AS管理策略不同。
  - BGP4：只寻求一条能到达目网的较好路由（不循环），并不最佳；
  - BGP采用路径向量（Path Vector）路由选择协议；
  - 变动触发BGP发言人（对接路由器）间交换彼此的路由信息


AS1
RIP
BGP
发言人
BGP
发言人
AS2
OSPF
BGP4
协议
BGP 发言人
BGP 发言人
AS3
OSPF
BGP 发言人
AS4
AS5

距离向量算法，如RIP采用



Information Stored at Node	Distance to Reach Node						
	A	B	C	D	E	F	G
A	0	1	1	∞	1	1	∞
B	1	0	1	∞	∞	∞	∞
C	1	1	0	1	∞	∞	∞
D	∞	∞	1	0	∞	∞	1
E	1	∞	∞	∞	0	∞	∞
F	1	∞	∞	∞	∞	0	1
G	∞	∞	∞	1	∞	1	0

Information Stored at Node	Distance to Reach Node						
	A	B	C	D	E	F	G
A	0	1	1	2	1	1	2
B	1	0	1	2	2	2	3
C	1	1	0	1	2	2	2
D	2	2	1	0	3	2	1
E	1	2	2	3	0	2	3
F	1	2	2	2	2	0	1
G	2	3	2	1	3	1	0

链路状态算法：如OSPF采用Dijkstra's

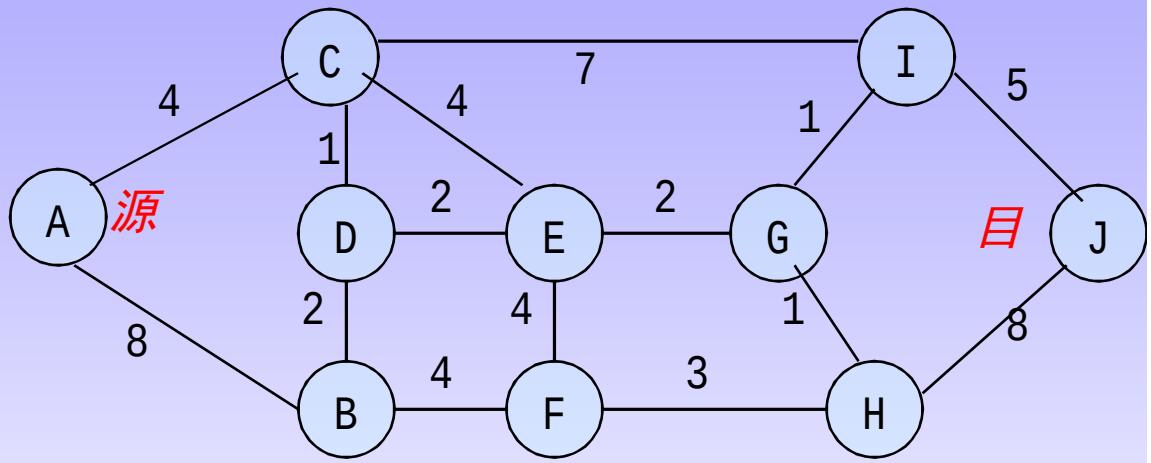
从源A到目J的A算法定义

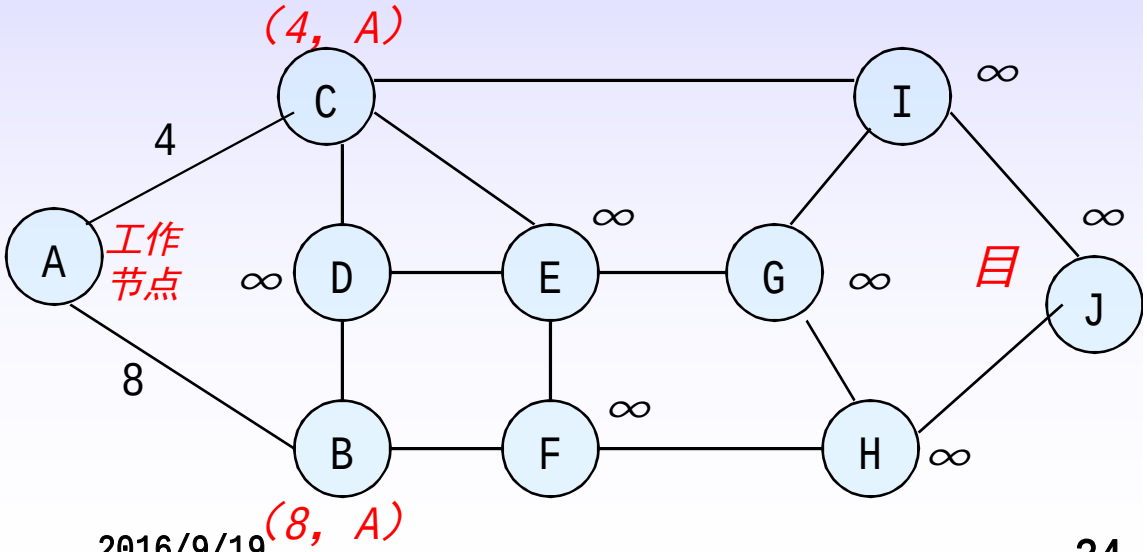
1、***网络每条路径有一权值根据最小代价标准得出***

2、*首选源节点A为工作节点*

3、*所有与A非直连点到A为∞*

4、*为工作点相邻点分配最小代价，若发现有从该点到源点更短路径，则修改代价*





(4, A)

从本节点到工作节点的最小代价

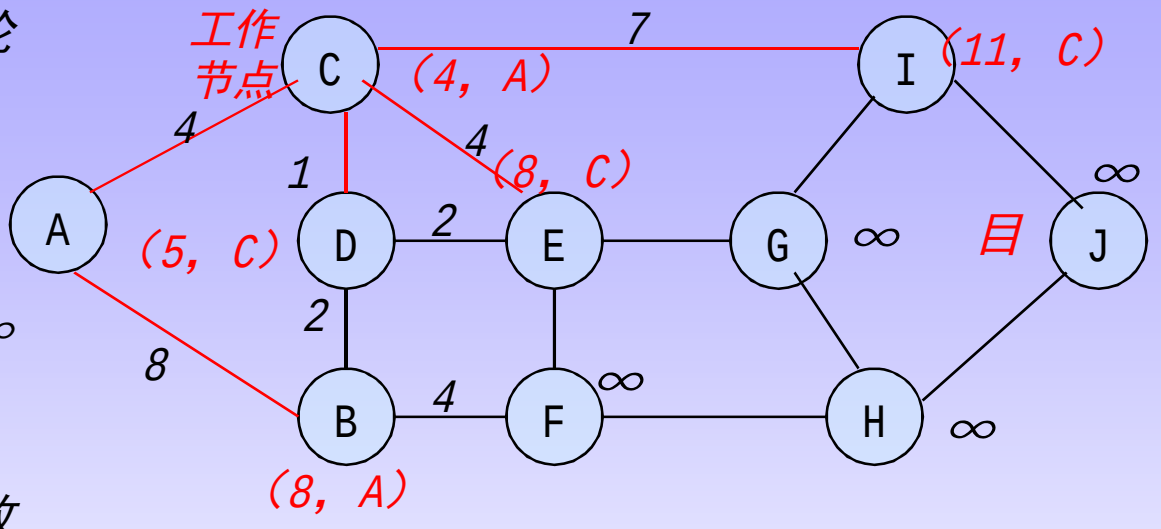
工作节点

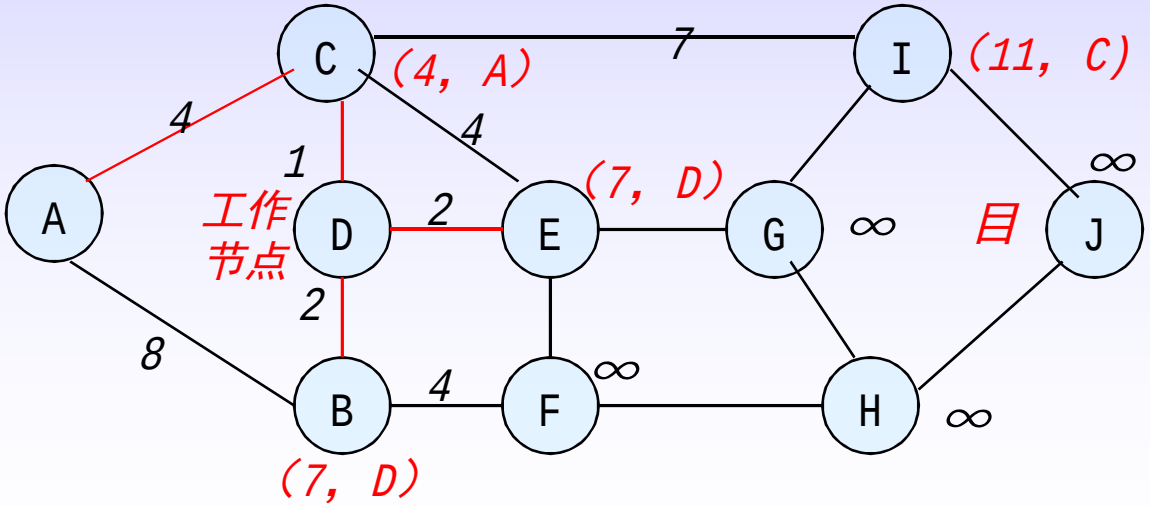
Dijkstra's算法

5、选（D,V）中最小者为下一轮工作节点。(4,A)<(8,A)

6、若某点代价与其相邻点路径上代价之和小于该相邻点的代价，则用小和代替之。如对C之邻点I：4+7<∞,11代∞

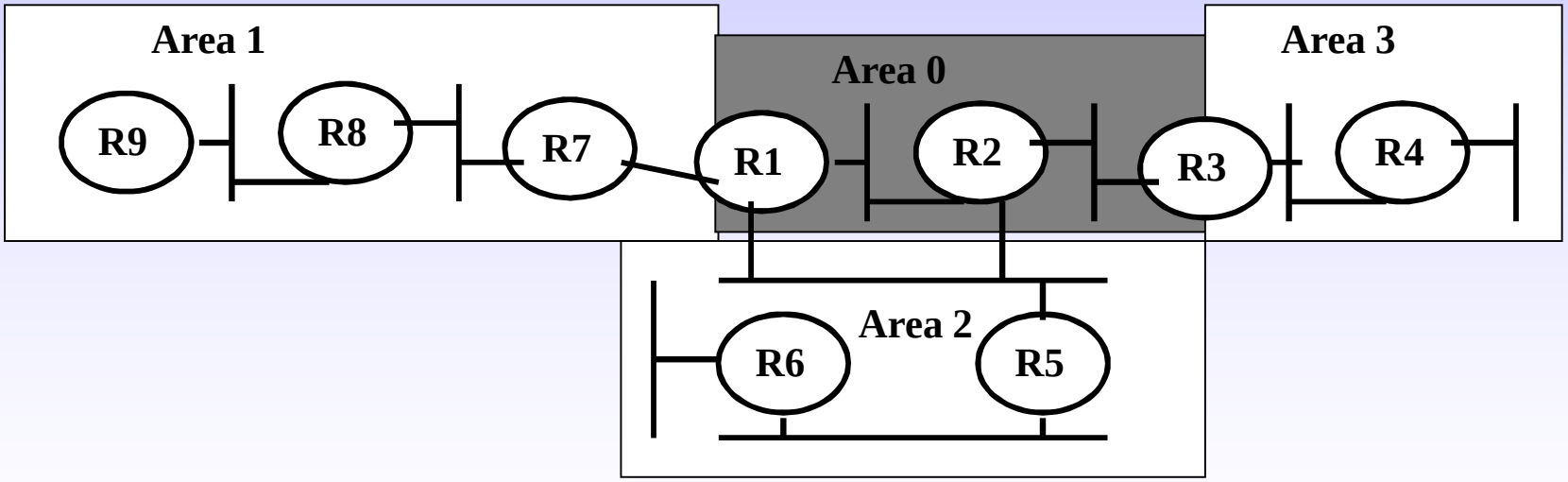
7、选择另一工作点，重复上述过程直到穷尽所有可能性。如：B->D->C->A < B->A,故(8,A)改为(7,D)





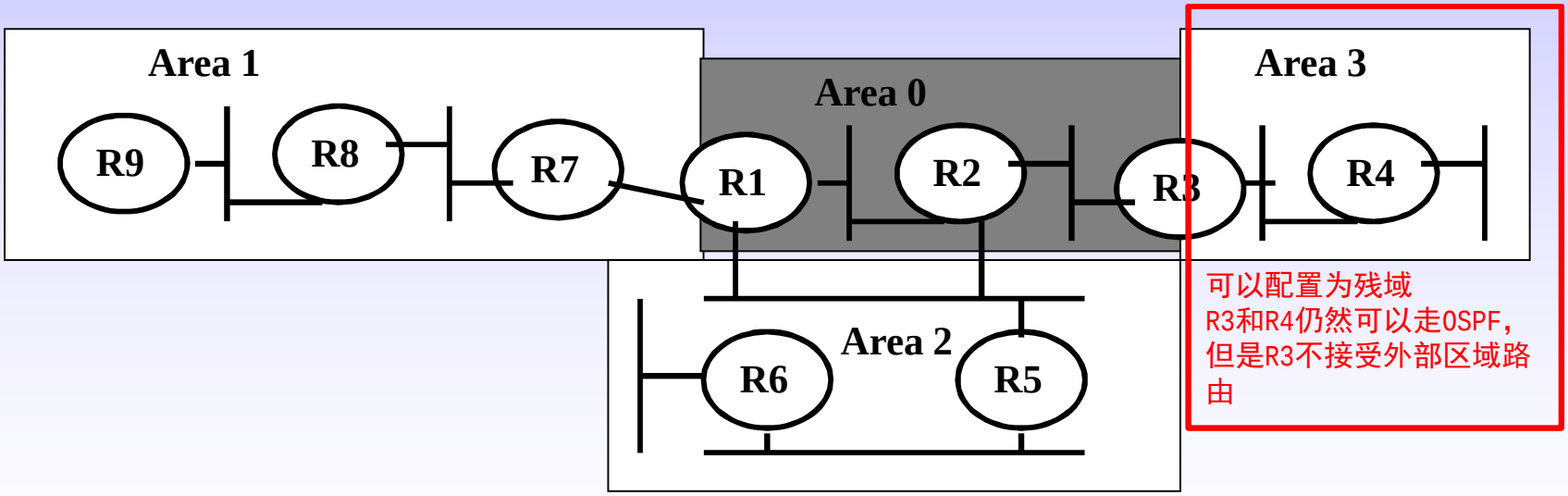
路由区

- 象OSPF一样,把域分成许多区，域内路由协议就能分层,使域增大
- 区:管理上配置成和其它节点交换链路状态信息的路由器集.主干网是一个特别区--Area 0



路由区

- 所有的非主干区域要和主干区域连接、
- *R1,R2,R3连接主干域又属于其他非主干域,是非主干域的区边界路由器 ABR*
- *R1，R2，R3之间交换路由信息（类似单域的OSPF），但是会将非主干域的路由信息汇总*
- *不能直接跟外部网络相连的网络（无ASBR）可以配置为stub area（残域，完全残域），不接收外部域信息，仅仅通过路由器走默认路由*



1.3.5 数据传输

- 问题:怎样实现远程进程间的数据传输
  - 主机-主机的包传输转化成进程-进程通信通道
  - 网络层结构，支持端应用程序--端到端协议
- 什么是连接?
  - 一条连接就是不同系统内的两个实体之间的一个临时性的逻辑关联通路(目IP,源IP,目端口,源端口，TCP/UDP,五元组)
  - 在连接持续期间，每个实体都跟踪从对方到达和发送到对方的PDU，以便调节PDU的流量以及对丢失和损坏的PDU进行恢复。
- 互联网的全部功能，最基本、最小粒度的服务
  - 端到端数据传输

对传输层协议的希望与IP层现实

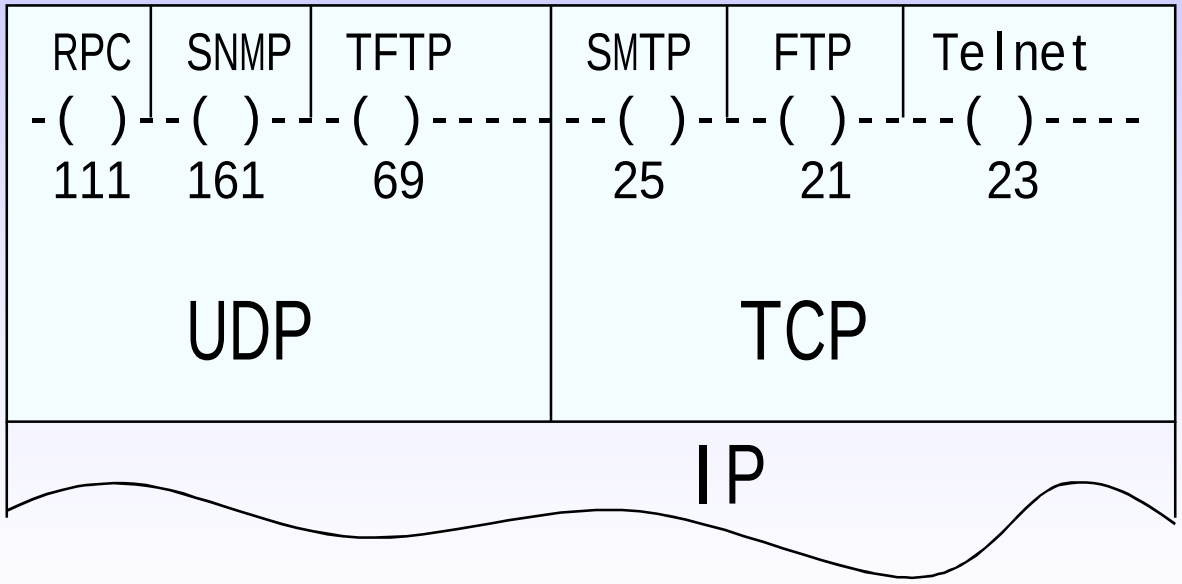
- 希望
  - 保障报文传输
  - 以发送相同的顺序传输报文
  - 每个报文最多传输一个拷贝
  - 支持任意长报文
  - 支持收、发之间的同步
  - 允许收方应用流控发方
  - 支持每个主机上的多个进程
- 现实（IP层提供的服务）
  - 丢包
  - 报文重排序
  - 对给定报文传输重复拷贝
  - 限制报文在某个有限大小
  - 在任意长延迟后传输报文
  - 以上是best-effort 层次上的服务，如IP

I) 简单多路器-Multiplexer
UDP 协议

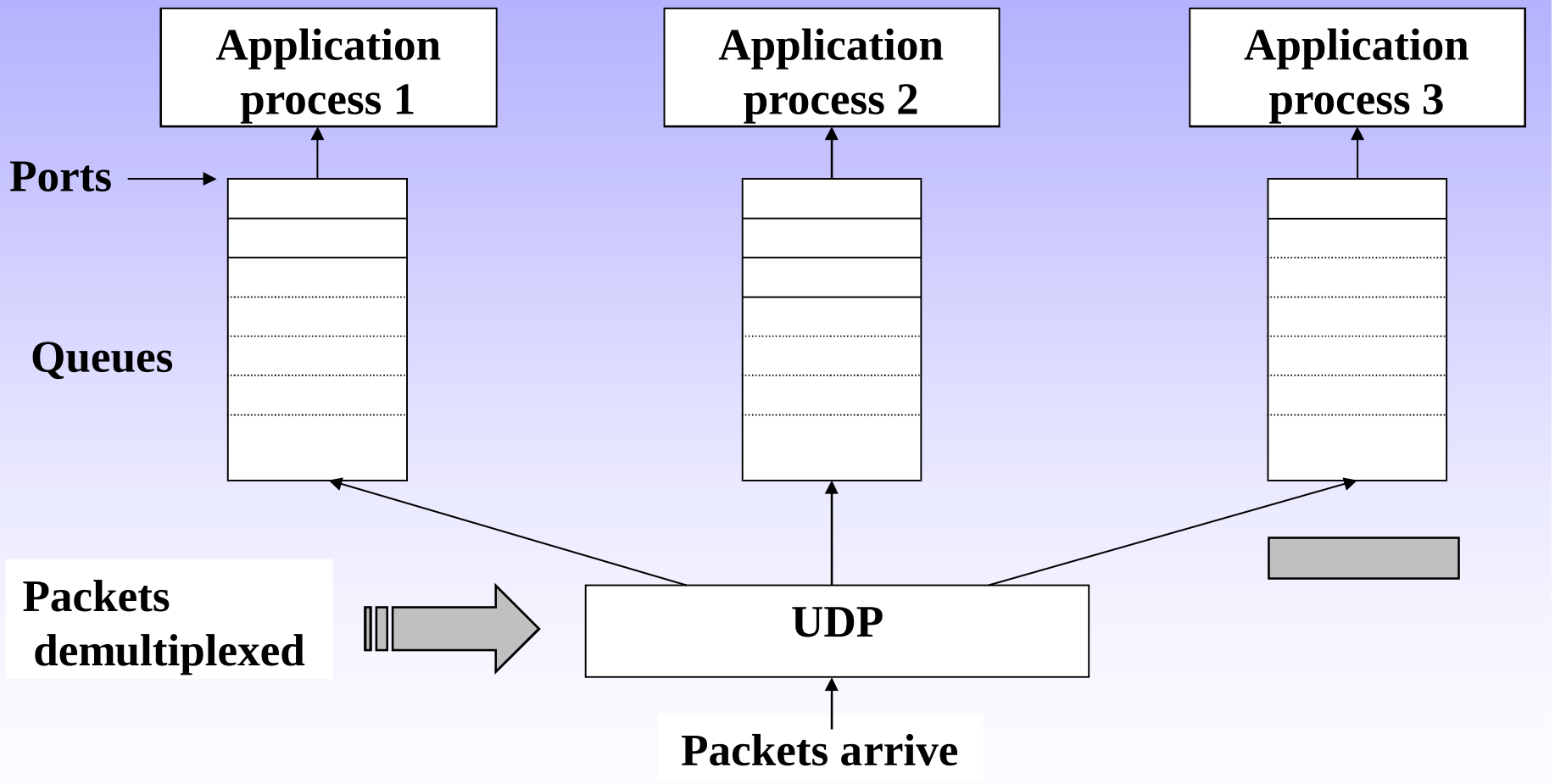
- 最简单的传输层协议
  - 主机-主机的传输服务在IP协议上扩展成进程-进程的直接通信服务
  - 一台主机上可运行多进程，需加一多路开关层,区别它们并共享网络
- 传输层协议
  - best-effort未加任何功能.如User Datagram Protocol-UDP
- UDP:
  - 最小/简单分路协议
  - just port numbers, and an optional checksum
  - no flow control, no congestion control, no reliability or ordering

端口的概念

- 唯一区别全网上的每个进程或目的主机上的进程
  - 端口:间接区别每个进程的抽象定位器-数字
  - 唯一标识=主机IP地址 + 端口号



UDP消息队列


Application process 1
Application process 2
Application process 3
Ports
Queues
Packets demultiplexed
UDP
Packets arrive

UDP协议

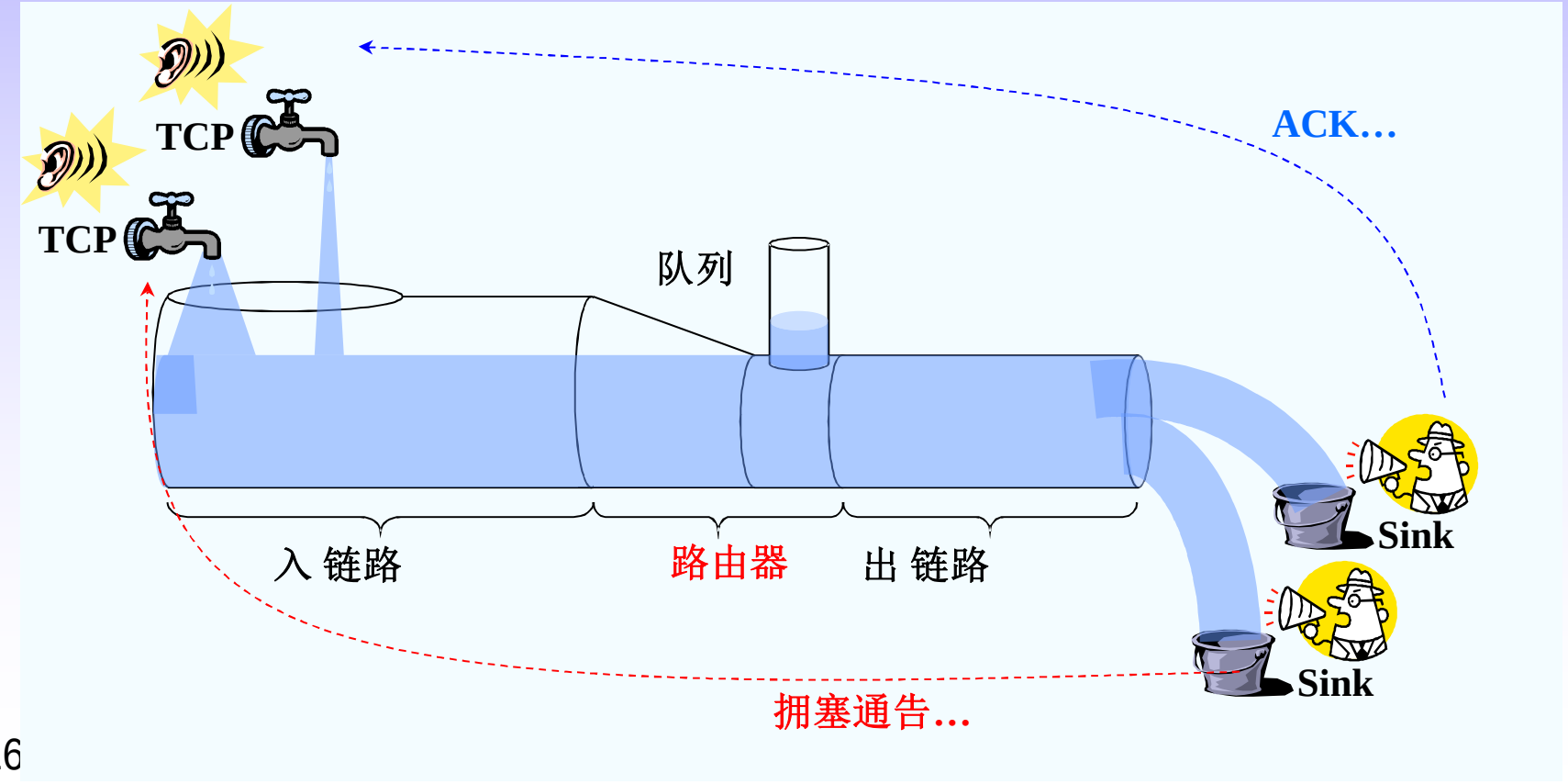
- 提供无连接服务，不保证数据完整到达目的地，减轻了网络的通信负担
- 适应C/S模式的简单请求/响应通信需要
- 应用程序要实施超时重传机制，并对数据包编号，但增加了应用程序的复杂性
- UDP可保留各报文间的边界，不把应用多次发送的数据合并成一个包发出去，且发包后不对该包缓存，这对简单请求/响应很方便
- 组播应用、多数音视频都建立在UDP之上。

II) 可靠字节流协议(TCP)

- TCP:更成熟的传输协议
  - 提供可靠,面向连接,按序字节流
  - 全双工,每个连接支持一对字节流,每个流一个方向
  - 流控机制:允许每个字节流的接收端在给定时间内限制其发送端的数据速率
  - 支持多路输出机制,允许一个主机上同时有多个会话对
  - 还提供拥塞控制机制
- 流控与拥控之差别:
  - 流控:防止发送超过接收者能力（速/量）， 是端到端的发送
  - 拥控:防止过多数据注入到网络中,从而引起交换机或链路超载,拥控是关于主机到网络的发送

TCP friendly

- TCP: a package deal
  - flow control, congestion control, *byte-stream orientation*
  - *total* ordering and *total* reliability



III) 流控制传输协议：SCTP

- RFC 2960,Oct.2000y
  - SCTP：Stream Control Transmission Protocol
  - 最初设计用于在IP上传输电话信令SS7（可靠/边界），把SS7信令网络的一些可靠特性引入IP，以后扩大了一些其它应用
- 信令类需求
  - Multi-homing, Multi-streaming
  - Message boundaries (with reliability*)
  - Improved SYN-flood protection
  - Tunable parameters (Timeout, Retrans, etc.)
  - A range of reliability and order (full to partial to none)along with congestion control
- UDP/TCP很难满足
  - UDP不可靠、无连接、无顺序、有边界；信令需要面向连接/可靠性！
  - TCP有可靠、有连接、有顺序、无边界；信令需要边界性/部分有序！

SCTP 关键特点

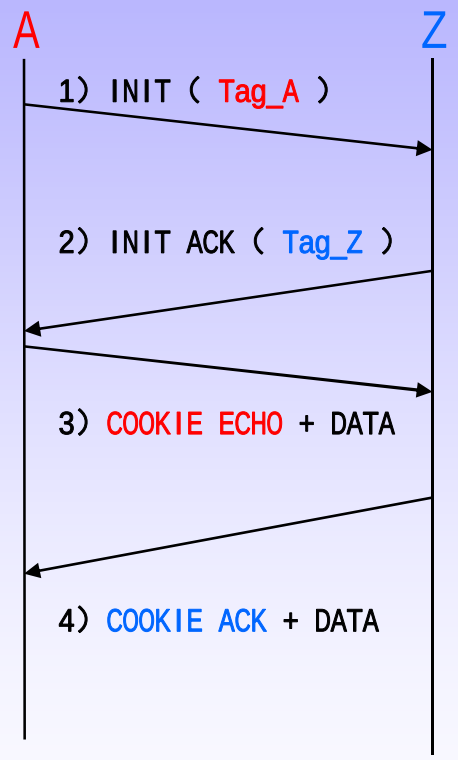
- Multi-homing *improved robustness to failures*
  - In TCP,连接仅在 <IP addr,port> 与 <IP addr, port>之间进行；如果接口down，整个连接down
  - In SCTP,For multi-homed, 每端可列出许多 IP addresses ;如果接口down ，仍可通过任何其它地址保持连接
- *Multi-streaming reduced delay*
  - 部分保序. 消减 Head of Line (HOL) 阻塞
  - In TCP, 所有数据保序; 队列头的丢失导致整个数据段延迟交付
  - In SCTP, 你可发送多达 64K 的独立流, 每个保序流独立,某个流上的丢失并不导致其它流延迟交付
- *Message boundaries preserved easier coding*
  - In TCP 打包并不保留报文的边界
  - In SCTP 保护报文边界,应用层协议容易写入，编码简单！

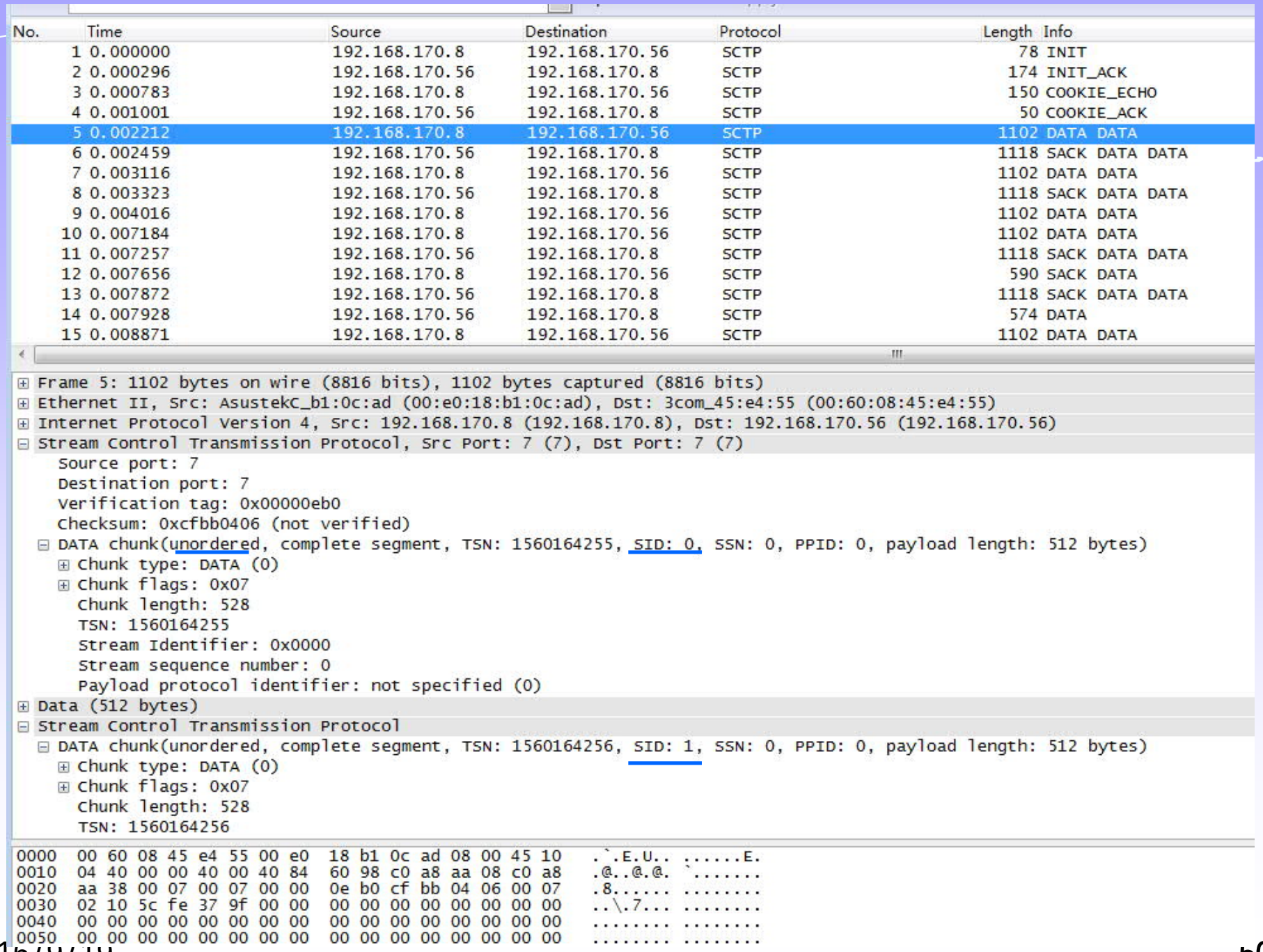
SCTP 关键特点

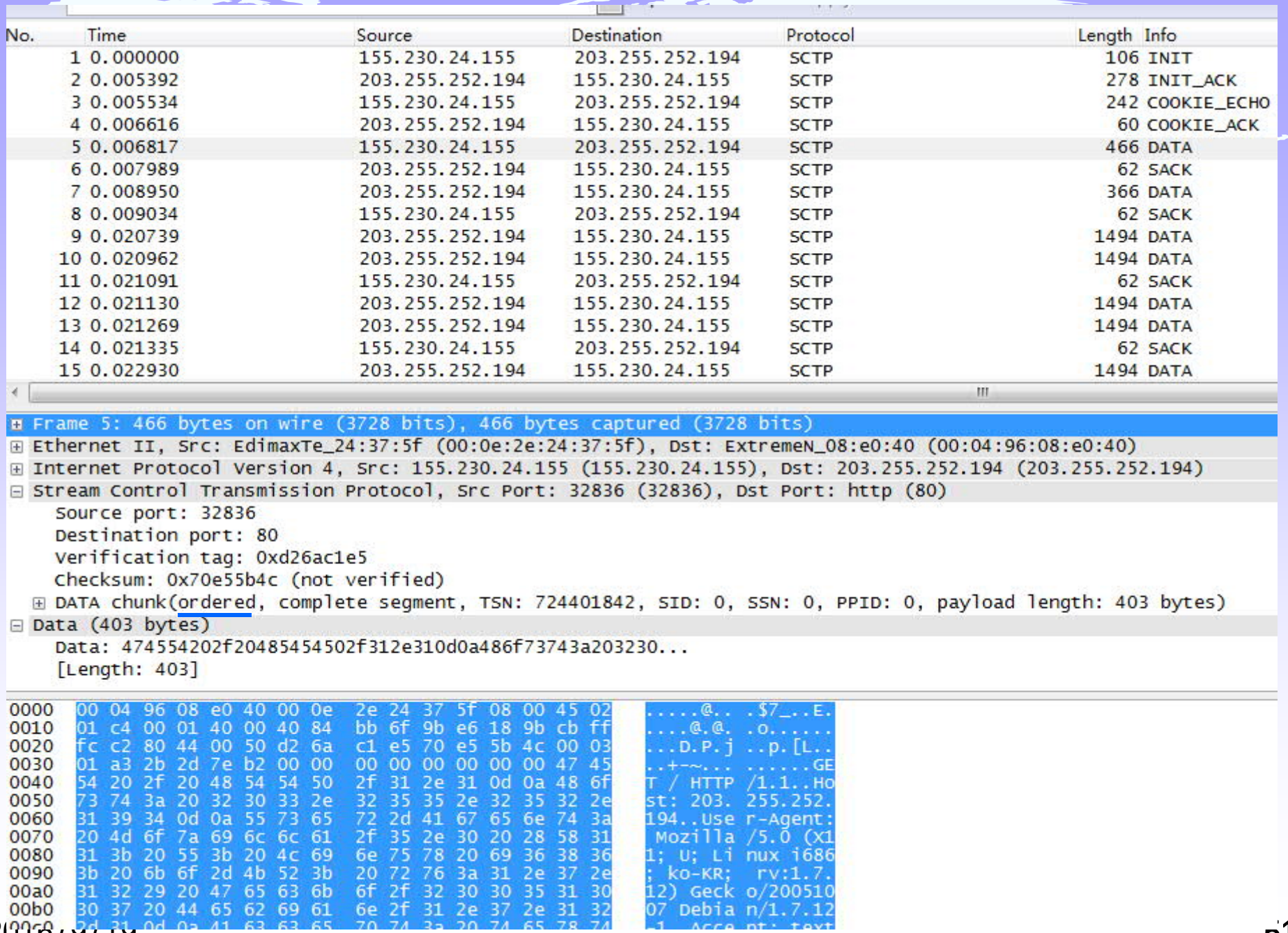
- Improved SYN-flood protection *more secure*
  - TCP 易受 SYN flooding攻击;
  - SCTP 采用四次握手，保护免受SYN flooding攻击
- Tunable parameters (Timeout, Retrans, etc.) *more flexibility*
  - TCP 参数调整只有系统管理员才能进行，实施内核的改变和锁定等
  - SCTP 参数可由socket basis调整
- Congestion controlled unreliable/unordered data *more flexibility*
  - TCP 虽有拥控，但不能做不可靠/失序的交付
  - UDP 虽能做不可靠/失序的交付，但没有拥控
  - SCTP 总有拥控，且能在部分/全范围提供可靠性、保序的服务
  - SCTP，可靠/不可靠数据都能在相同连接上多路

4次握手建立连接

- “A” 发送 INIT 块到 “Z”，块中含验证标志 Tag (Tag_A) （1 to 4294967295随机数）
- “Z” 响应 INIT ACK 块. 其中含自己的验证块 (Tag_Z)
- “A” 发送 COOKIE ECHO 块到 “Z”. 可与 DATA 块绑定
- “Z” 回答 COOKIE ACK 到“A” ，可与 DATA 块绑定




No. Time Source Destination Protocol Length Info
1 0.000000 192.168.170.8 192.168.170.56 SCTP 78 INIT
2 0.000296 192.168.170.56 192.168.170.8 SCTP 174 INIT_ACK
3 0.000783 192.168.170.8 192.168.170.56 SCTP 150 COOKIE_ECHO
4 0.001001 192.168.170.56 192.168.170.8 SCTP 50 COOKIE_ACK
5 0.002212 192.168.170.8 192.168.170.56 SCTP 1102 DATA DATA
6 0.002459 192.168.170.56 192.168.170.8 SCTP 1118 SACK DATA DATA
7 0.003116 192.168.170.8 192.168.170.56 SCTP 1102 DATA DATA
8 0.003323 192.168.170.56 192.168.170.8 SCTP 1118 SACK DATA DATA
9 0.004016 192.168.170.8 192.168.170.56 SCTP 1102 DATA DATA
10 0.007184 192.168.170.8 192.168.170.56 SCTP 1102 DATA DATA
11 0.007257 192.168.170.56 192.168.170.8 SCTP 1118 SACK DATA DATA
12 0.007656 192.168.170.8 192.168.170.56 SCTP 590 SACK DATA
13 0.007872 192.168.170.56 192.168.170.8 SCTP 1118 SACK DATA DATA
14 0.007928 192.168.170.56 192.168.170.8 SCTP 574 DATA
15 0.008871 192.168.170.8 192.168.170.56 SCTP 1102 DATA DATA
Frame 5: 1102 bytes on wire (8816 bits), 1102 bytes captured (8816 bits)
Ethernet II, Src: AsustekC_b1:0c:ad (00:e0:18:b1:0c:ad), Dst: 3com_45:e4:55 (00:60:08:45:e4:55)
Internet Protocol Version 4, Src: 192.168.170.8 (192.168.170.8), Dst: 192.168.170.56 (192.168.170.56)
Stream Control Transmission Protocol, Src Port: 7 (7), Dst Port: 7 (7)
Source port: 7
Destination port: 7
Verification tag: 0x00000eb0
Checksum: 0xcfbb0406 (not verified)
DATA chunk(unordered, complete segment, TSN: 1560164255, SID: 0, SSN: 0, PPID: 0, payload length: 512 bytes)
Chunk type: DATA (0)
Chunk flags: 0x07
Chunk length: 528
TSN: 1560164255
Stream Identifier: 0x0000
Stream sequence number: 0
Payload protocol identifier: not specified (0)
Data (512 bytes)
Stream Control Transmission Protocol
DATA chunk(unordered, complete segment, TSN: 1560164256, SID: 1, SSN: 0, PPID: 0, payload length: 512 bytes)
Chunk type: DATA (0)
Chunk flags: 0x07
Chunk length: 528
TSN: 1560164256
0000 00 60 08 45 e4 55 00 e0 18 b1 0c ad 08 00 45 10 .`.E.U.. ......E.
0010 04 40 00 00 40 00 40 84 60 98 c0 a8 aa 08 c0 a8 .@..@.@. `.......
0020 aa 38 00 07 00 07 00 00 0e b0 cf bb 04 06 00 07 .8...... ....
0030 02 10 5c fe 37 9f 00 00 00 00 00 00 00 00 00 00 ..\.7... ....
0040 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ....
0050 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ....



No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	155.230.24.155	203.255.252.194	SCTP	106	INIT
2	0.005392	203.255.252.194	155.230.24.155	SCTP	278	INIT_ACK
3	0.005534	155.230.24.155	203.255.252.194	SCTP	242	COOKIE_ECHO
4	0.006616	203.255.252.194	155.230.24.155	SCTP	60	COOKIE_ACK
5	0.006817	155.230.24.155	203.255.252.194	SCTP	466	DATA
6	0.007989	203.255.252.194	155.230.24.155	SCTP	62	SACK
7	0.008950	203.255.252.194	155.230.24.155	SCTP	366	DATA
8	0.009034	155.230.24.155	203.255.252.194	SCTP	62	SACK
9	0.020739	203.255.252.194	155.230.24.155	SCTP	1494	DATA
10	0.020962	203.255.252.194	155.230.24.155	SCTP	1494	DATA
11	0.021091	155.230.24.155	203.255.252.194	SCTP	62	SACK
12	0.021130	203.255.252.194	155.230.24.155	SCTP	1494	DATA
13	0.021269	203.255.252.194	155.230.24.155	SCTP	1494	DATA
14	0.021335	155.230.24.155	203.255.252.194	SCTP	62	SACK
15	0.022930	203.255.252.194	155.230.24.155	SCTP	1494	DATA

```
⊞ Frame 5: 466 bytes on wire (3728 bits), 466 bytes captured (3728 bits)
⊞ Ethernet II, Src: EdimaxTe_24:37:5f (00:0e:2e:24:37:5f), Dst: ExtremeN_08:e0:40 (00:04:96:08:e0:40)
⊞ Internet Protocol Version 4, Src: 155.230.24.155 (155.230.24.155), Dst: 203.255.252.194 (203.255.252.194)
⊟ Stream Control Transmission Protocol, Src Port: 32836 (32836), Dst Port: http (80)
    Source port: 32836
    Destination port: 80
    Verification tag: 0xd26ac1e5
    Checksum: 0x70e55b4c (not verified)
  ⊞ DATA chunk(ordered, complete segment, TSN: 724401842, SID: 0, SSN: 0, PPID: 0, payload length: 403 bytes)
⊟ Data (403 bytes)
    Data: 474554202f20485454502f312e310d0a486f73743a203230...
    [Length: 403]
```

```
0000  00 04 96 08 e0 40 00 0e  2e 24 37 5f 08 00 45 02   .....@.. .$7_..E.
0010  01 c4 00 01 40 00 40 84  bb 6f 9b e6 18 9b cb ff   ....@.@. .o......
0020  fc c2 80 44 00 50 d2 6a  c1 e5 70 e5 5b 4c 00 03   ...D.P.j ..p.[L..
0030  01 a3 2b 2d 7e b2 00 00  00 00 00 00 00 00 47 45   ..+-~... ......GE
0040  54 20 2f 20 48 54 54 50  2f 31 2e 31 0d 0a 48 6f   T / HTTP /1.1..Ho
0050  73 74 3a 20 32 30 33 2e  32 35 35 2e 32 35 32 2e   st: 203. 255.252.
0060  31 39 34 0d 0a 55 73 65  72 2d 41 67 65 6e 74 3a   194..Use r-Agent:
0070  20 4d 6f 7a 69 6c 6c 61  2f 35 2e 30 20 28 58 31    Mozilla /5.0 (X1
0080  31 3b 20 55 3b 20 4c 69  6e 75 78 20 69 36 38 36   1; U; Li nux i686
0090  3b 20 6b 6f 2d 4b 52 3b  20 72 76 3a 31 2e 37 2e   ; ko-KR;  rv:1.7.
00a0  31 32 29 20 47 65 63 6b  6f 2f 32 30 30 35 31 30   12) Geck o/200510
00b0  30 37 20 44 65 62 69 61  6e 2f 31 2e 37 2e 31 32   07 Debia n/1.7.12
00c0  2d 31 0d 0a 41 63 63 65  70 74 3a 20 74 65 78 74   -1..Acce pt: text
```

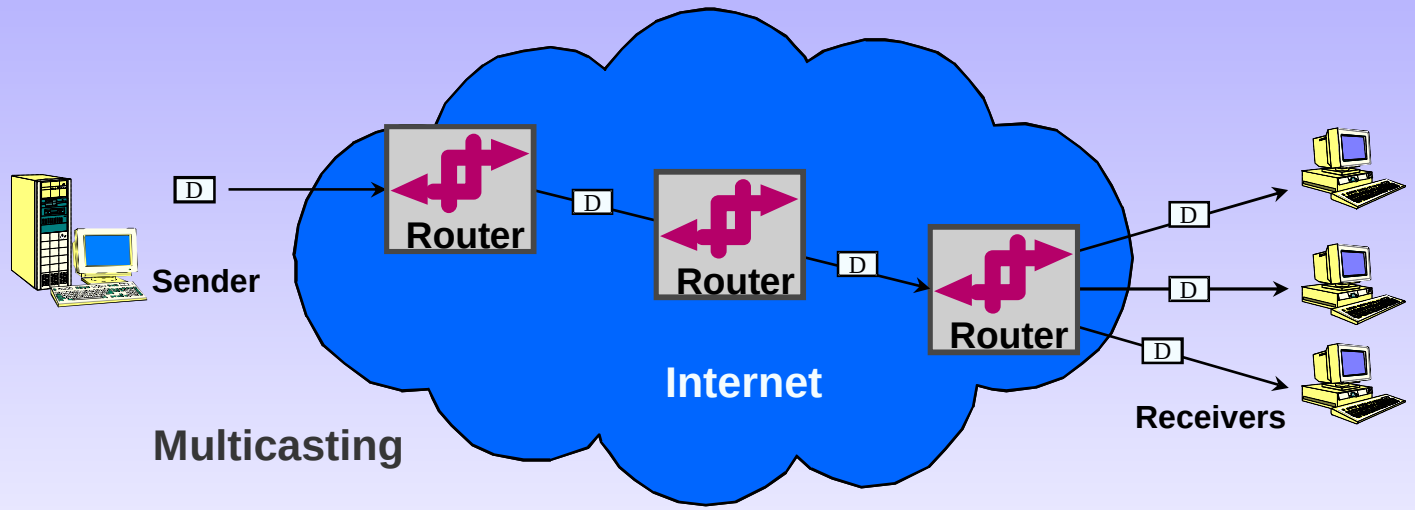
1.3.6 Multicasting-组播

- IP组播的基本定义：
  - 在LAN/WAN上能使IP数据报从一个源同时到多个目的传输过程
  - 接收组成员参加组播会议，应用只需发送一份拷贝到需要接收的组
  - 组播技术让包只寻址到组，而不是单个接收者
  - 开放组播的结点要运行一套能接收组播报文的TCP/IP协议
  - 由IETF推荐的RFC 1112定义的对IP的扩展（ Host Extensions for IP Multicasting ）
- 相关缩略语
  - BSR:Boot Strap Router,自举路由器
  - IGMP:Internet Group Management Protocol,互联网组管协议
  - MBGP:Multi-protocol Border Gateway Protocol, 多协议边界网关协议
  - MSDP:Multicast Source Discovery Protocol,组播源发现协议
  - PIM-SM:Protocol Independent Multicast-Sparse mode,协议独立组播—稀疏模式
  - PIM-DM:Protocol Independent Multicast-Dense mode,协议独立组播—密集模式
  - RP:Rendezvous Point,汇集点
  - RPT: Rendezvous Point Tree , PIM-SM协议共享树
  - SPT: Shortest Path Tree , 最短路径树
  - DR：Designated Router，指定路由器（运行IGMP协议）

IP组播路由的基本问题与方法

- 组播的基本矛盾？
  - 组播IP地址只是一个组成员集合的逻辑名字，不具有单播IP地址可直接全网寻址的功能
  - 组播地址用以区别多个组播会话（session），而不是特别的物理目的地，源不需/也不知道所有相关可直接寻址的地址
  - 组成员可能分散在Internet各个地方，不可能进行CIDR类似的聚类寻址
  - 基本问题：知道目的组的名字，但不知其包发往何处！
- 组播的基本任务：
  - 建构以组播源为根的分发树－找到接收对象
  - 在分发树上传输IP组播流－发送组播流
- 组播的特点
  - 对不同的接收端,许多会话数据流是相同的
  - 组成员可随时进出，导致组播分发树会变化
  - 组播分发树的构造方法随组播协议的不同而变化
- 单播与组播
  - 单播路由关心包发往哪里？
  - 组播路由则关心包来自哪里？故称逆向路径？寻找源在何处？

先进的组播技术

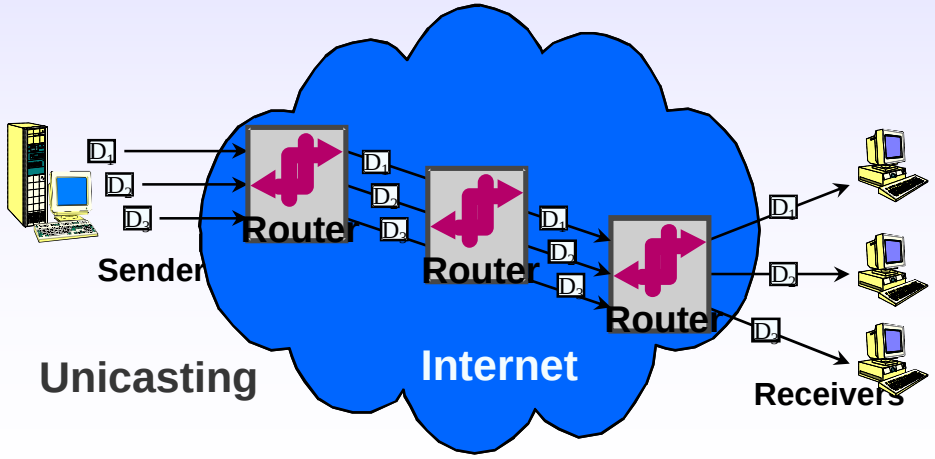


Multicast

-在 **IP** 网上一对多的传输

-支持视频会议,

-**e-learning,** 培训等



I）组播通信模型与协议

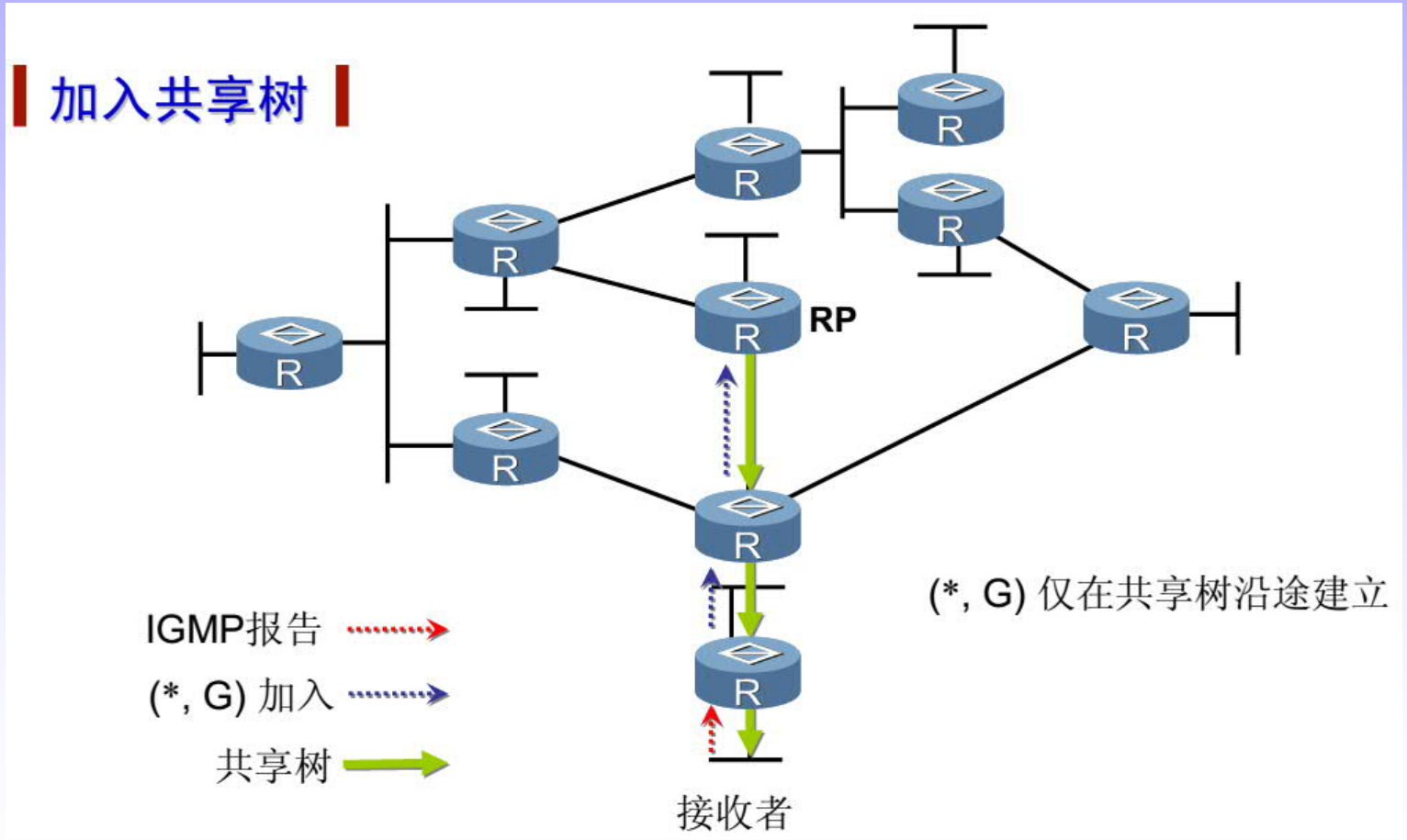
- 构成：
  - 1个核心：分发树为核心
  - 3个发现：源发现、接收者发现、拓扑分离
- 任务：
  - 组播路由协议根据组播源信息、接收者信息、网络拓扑（源和接收者间的连接关系）信息来构造组播分发树
- 协议：
  - 组播协议 ＝ 组播接收者发现协议 ＋ 组播路由协议 ＋ 组播源发现协议 ＋ 组播拓扑分离协议

PIM：组播路由协议

- 常用组播路由协议：PIM-SM/DM
- 任务:根据IGMP掌握的接收者信息，单播路由协议（如OSPF等）掌握的拓扑信息来
  - 完成源发现
  - 构建分发树
- PIM-DM：假设接收者在网上密集分布。首先将数据推到全网，后用协议信令剪枝不需要数据的网段，即为扩散－剪枝方式，其构建的分发树属于源树
- PIM-SM：假设接收者在网上稀疏分布。采用按需发送方式组播数据，即只向那些需要数据的网段转发。该方式首先构建共享树，然后构建源树，共同构建分发树
- RP:是PIM-SM的核心路由器， RP通常为一个或多个组播组服务。
  - 组播用户所直连的路由器采取“ 显式加入 ”机制主动加入以RP为根的共享树
  - 当用户接收到大量组播数据后绕开RP还可切换到源树

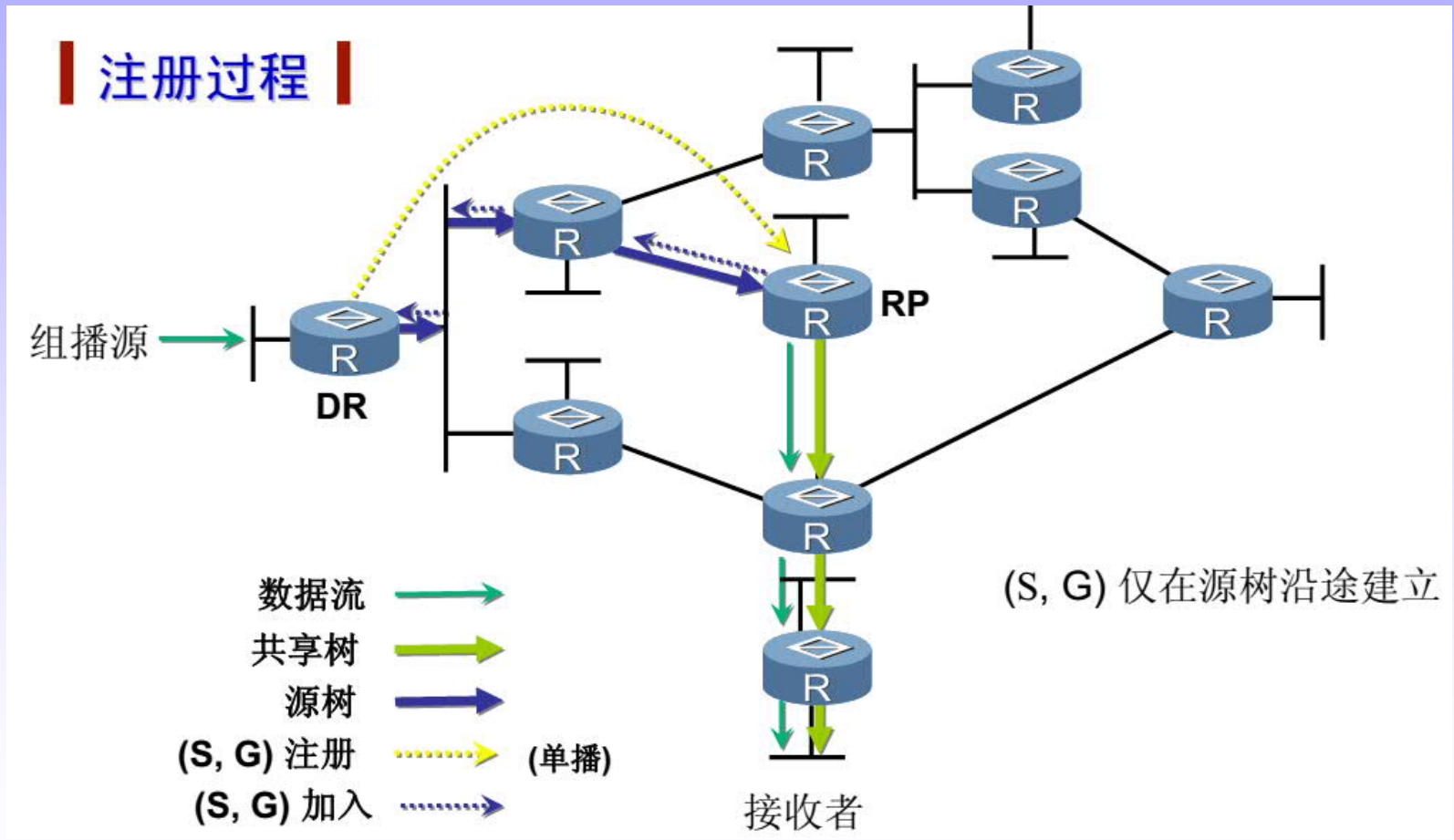
一个核心（分发树）

PIM-SM的分发树



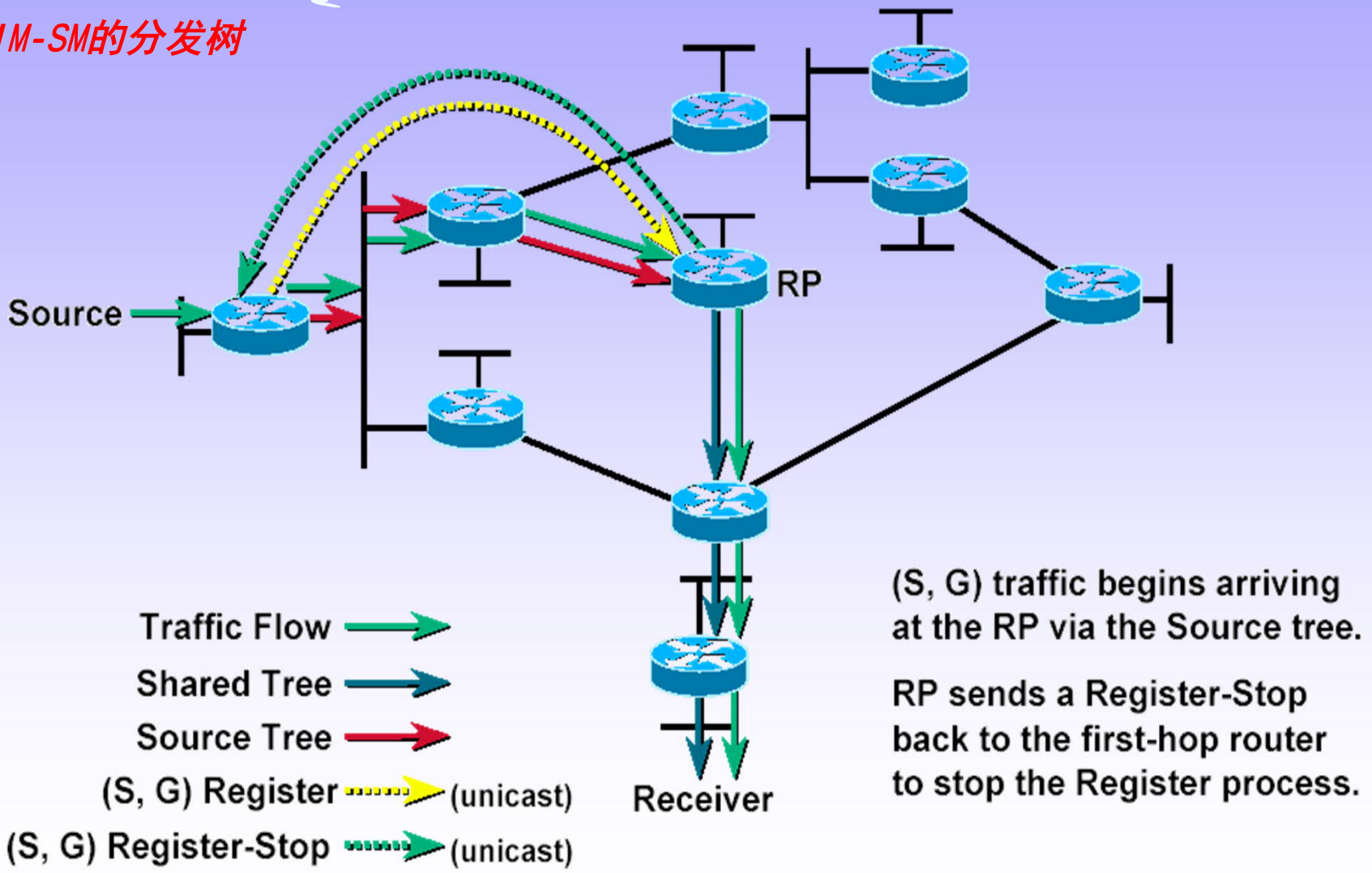
一个核心（分发树）

PIM-SM的分发树



一个核心（分发树）

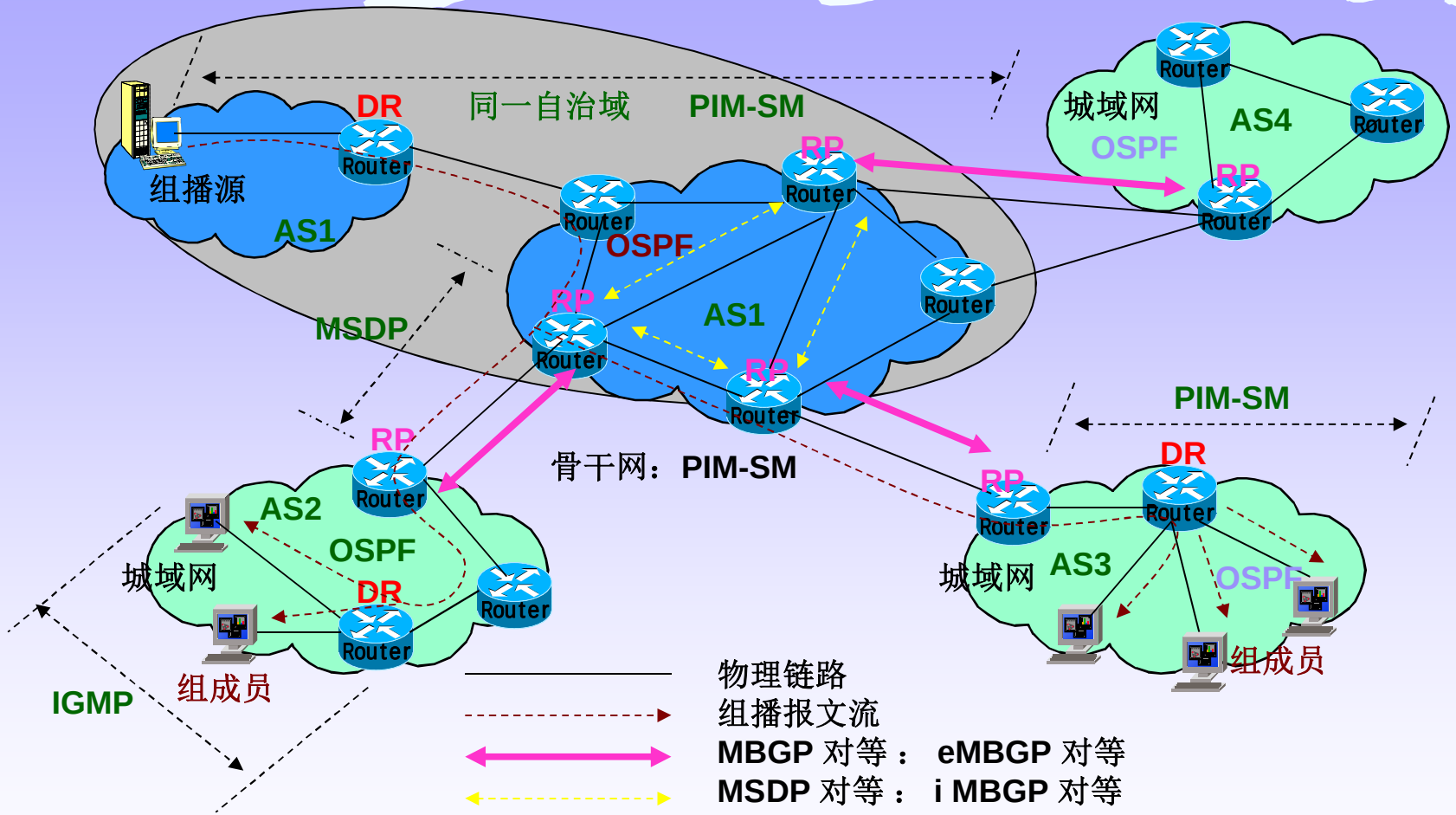
PIM-SM的分发树



三个发现

- 源发现协议
  - PIM-SM：完成AS域内的源发现
  - MSDP：完成AS域外的源信息发现和传播
- 接收者发现协议IGMP（放置在接入网关）
  - 位置：运行在主机和指定路由器DR之间
  - 任务：维护组播组是否有成员？组成员信息？不关心到底是哪些成员，有多少成员
  - 优点：状态信息不因组播成员增加而增加；具有良好可扩展性
- 拓扑分离协议
  - 目的：用于传播AS间的组播拓扑信息，指导跨AS组播分发树建立
  - 原因：
    - PIM-SM在跨AS时，通过BGP协议掌握其它AS的网络拓扑信息
    - 有时需要单播、组播流量沿不同跨域路径转发，或应用不同路由策略
  - MBGP = eMBGP + iMBGP

跨AS组播结构



IGMP： 完成路由器直连网段接收者发现　　**OSPF：** 完成AS内单播拓扑发现

MSDP： 交换AS之间的源信息　　**MBGP：** 交换AS之间的拓扑信息

PIM-SM： 完成AS内的源发现，根据上述信息构建组播分发树，达到域间组播转发

两大类组播路由协议

- **密集模式**组播路由协议
  - 特点：
    - 假设组播组成员**密集分布**于网络上的许多子网，并至少包含一个组成员
    - 都需要大量的带宽
  - **典型**路由协议
    - DVMRP
    - MOSPF
    - PIM-DM
  - 基本方法：
    - 依靠**洪泛技术传播**信息到所有组播路由器

- **稀疏模式**组播路由协议
  - 特点：
    - 假设组播组成员**稀疏分布**于网络上，并至少包含一个组成员
    - **不需要大量的带宽**，组成员可能由ISDN连通
  - 典型路由协议
    - CBT:Core-Based Tree
    - PIM-SM
  - 基本方法：
    - **不用洪泛方式**，因组成员稀疏分布，否则会浪费带宽并引发严重性能下降

II) 主要组播路由算法

- 组播路由协议中使用的主要算法有
  - Flooding
  - Spanning tree
  - Reverse-Path Broadcasting(RPB)
  - Truncated Reverse-Path Broadcasting(TRPB)
  - Reverse-Path Multicasting(MPB)
  - Core-based tree

III) 组播的挑战、问题与发展

- 尽力传送（Best effort）
  - 会产生丢包
  - 不可能有很可靠的数据传输，应有针对性设计，可靠组播有待进一步研究
- 问题
  - 不能避免拥塞
    - 缺乏TCP滑动窗口，且“慢启动”导致拥塞，可尝试检测和避免机制
  - 复制：某些协议会导致偶尔生成重复的包
  - 无序发送：一些协议机制会导致无序发送

组播的挑战

- 没有更多的ISPs和OEM厂商开发有用的组播应用,缺乏组播工具和平台
- 防火墙阻隔
  - UDP能有效防止组成员ACKs的数量爆炸,但Firewall对其失去控制作用.解决办--应用网关?
- QoS:探讨用ATM,RSVP等?
- 基于Internet的组播实际上很少成功案例
- 应用层组播发展迅猛-P2P...IPTV...

组播的安全问题

- 组管理和访问控制
- 真实性（授权与认证）
  机密性和完整性
- 组密钥的分发
- 成员的加入
- 成员的脱离
- 组密钥的更新
- 允许外部审计

- 非法组播源侵入
- 非法组播接收者接收
- 组播核心路由器仿冒
  - BSR仿冒
  - RP仿冒
- 跨网络的非法组播加入
- 审计与计费

组播的发展

- 支持组播的主要高层应用协议:支持可靠数据传输
  - RTP:Real Time Transport Protocol
  - RTCP:Real Time Control Protocol
  - RTP:Real Time Streaming Protocol
  - RSVP:Resource Reservation Protocol
  - RMP: Reliable Multicast Protocol
  - RMF: Reliable Multicast Framework Protocol
  - RAMP: Reliable Adaptive Multicast Framework Protocol
  - Reliable Multicast Transport Protocol
    - Lucent 在其e-cast 用RMTP处理文件传输

习题

- Ch2

25（序列号计算）；40（以太网地址相同)；

- Ch3

14（交换机生成树算法）

- Ch4

4，5（MTU）； 21，22（路由）；40，46（子网）

- Ch5

12，13（TCP序列号）；39，40（TCP协议）