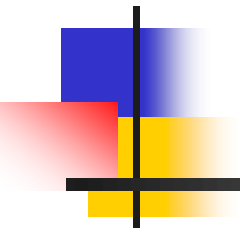


第5章 网络层概述

刘志敏

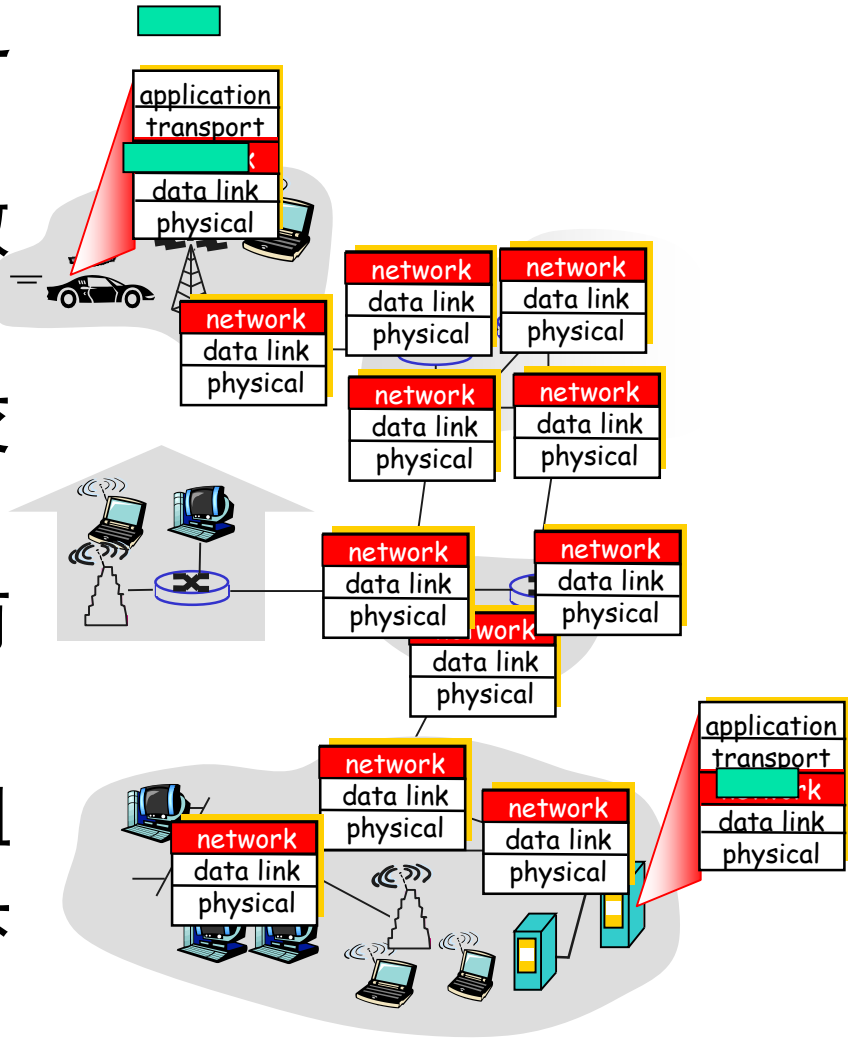
liuzm@pku.edu.cn

提纲

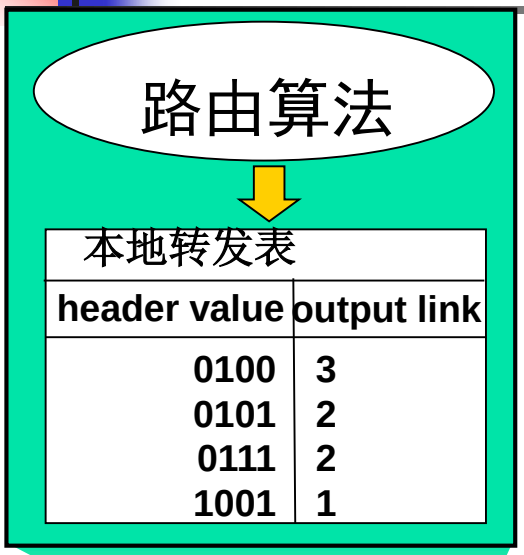
- 网络层的功能
- 路由器结构及工作原理
  - 转发及路由
  - 路由器的结构

网络层

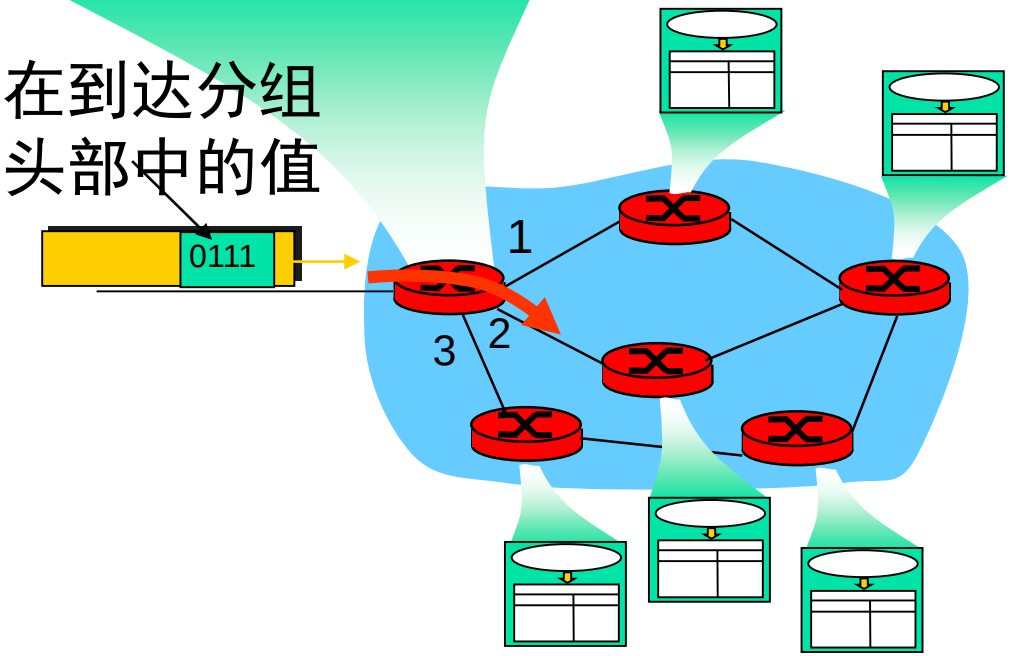
- 在发送主机与接收主机之间传输报文
- 发送端，将报文封装为数据报文
- 接收端，将接收的报文交给传输层
- 每个主机及路由器上都有网络层协议
- 路由器对经过的数据分组，选路并转发（检查其头部信息，查表并转发



网络层两大主要功能：路由与转发



- **转发forwarding:** 路由器将分组由输入端口移送到适当的输出端口



- **路由routing:** 决定分组由源主机到目的主机的路径
  - 依赖于路由算法

网络层为传输层提供的功能

- 网络层设计目标：
  - 为传输层提供服务，与路由器的技术无关
  - 对于传输层，可以屏蔽路由器的数量、类型和拓扑结构
  - 网络地址采用统一的编址方案
  - 网络拥塞控制、保证服务器质量、网络互联
- 网络层提供面向连接的服务还是非连接的服务？
- OSI的网络层提供两种服务
  - 面向连接——虚电路（virtual circuit):
    - 首先发出连接请求，与目的结点建立连接；然后，数据传输；最后拆除连接
  - 无连接——数据报(datagram)
    - 每个分组头都包含目的地址；每个分组在途经节点被单独处理；来自同一数据流的分组可以选择不同路径

虚电路（VC：Virtual Circuits）

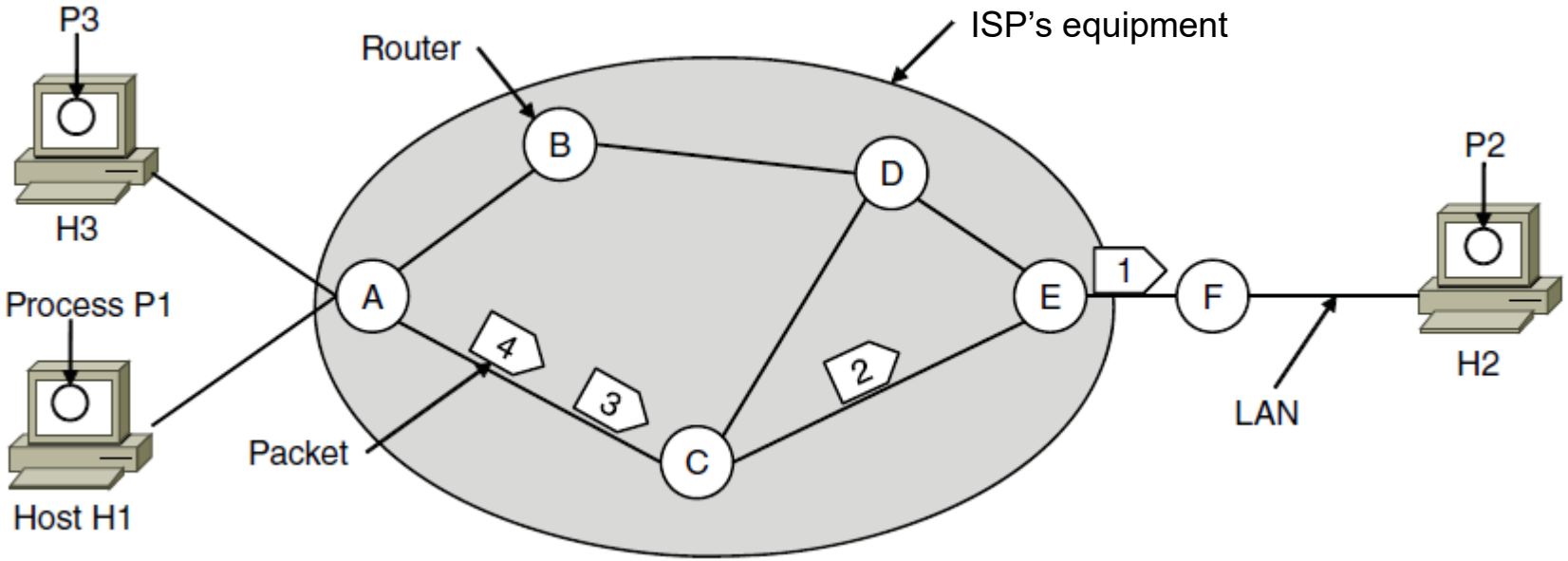
“由源到目的的通路类似于电话网的电路”

- 性能优异
- 网络的作用类似于提供由源到目的的路径

- 在数据传输之前建立呼叫，之后拆除电路
- 每个分组携带VC标识(并非是主机地址)
- 每个路由器都维护源主机到目的主机的连接状态
- 将链路及路由器的资源(带宽及缓存)分配给VC(资源是专用的，因此可保证服务质量)
- 在VC上，网络可实施流量控制和差错控制

面向连接的服务

- H1呼叫H2建立VC，路由器建立表中的第1条记录
- H3呼叫H2建立VC，路由器建立表中的第2条记录，为保证VC标识的唯一性，分组经过路由器交换后VC标识改变



A's table

H1	1
H3	1

In

C	1
C	2

Out

C's Table

A	1
A	2

E	1
E	2

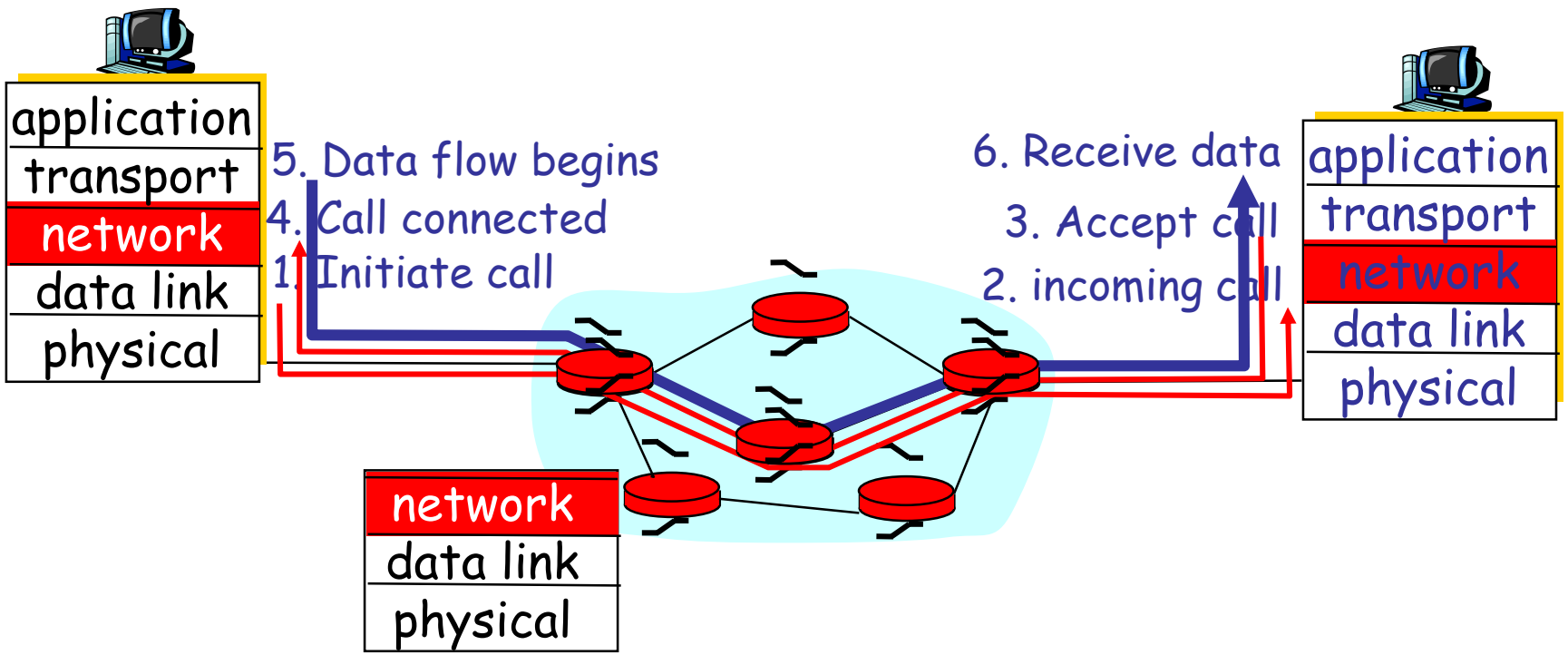
E's Table

C	1
C	2

F	1
F	2

虚电路VC：呼叫需要信令协议

- 用于建立、维持、拆除VC
- ATM, 帧中继(frame-relay), X.25等采用VC
- 互联网分组交换，一般不采用VC



VC 实现

组成每一VC的信息

- 由源到目的的路径
- VC标识，在每条路径的每条链路上是唯一的
- 沿着路径上的每台路由器，保存转发表

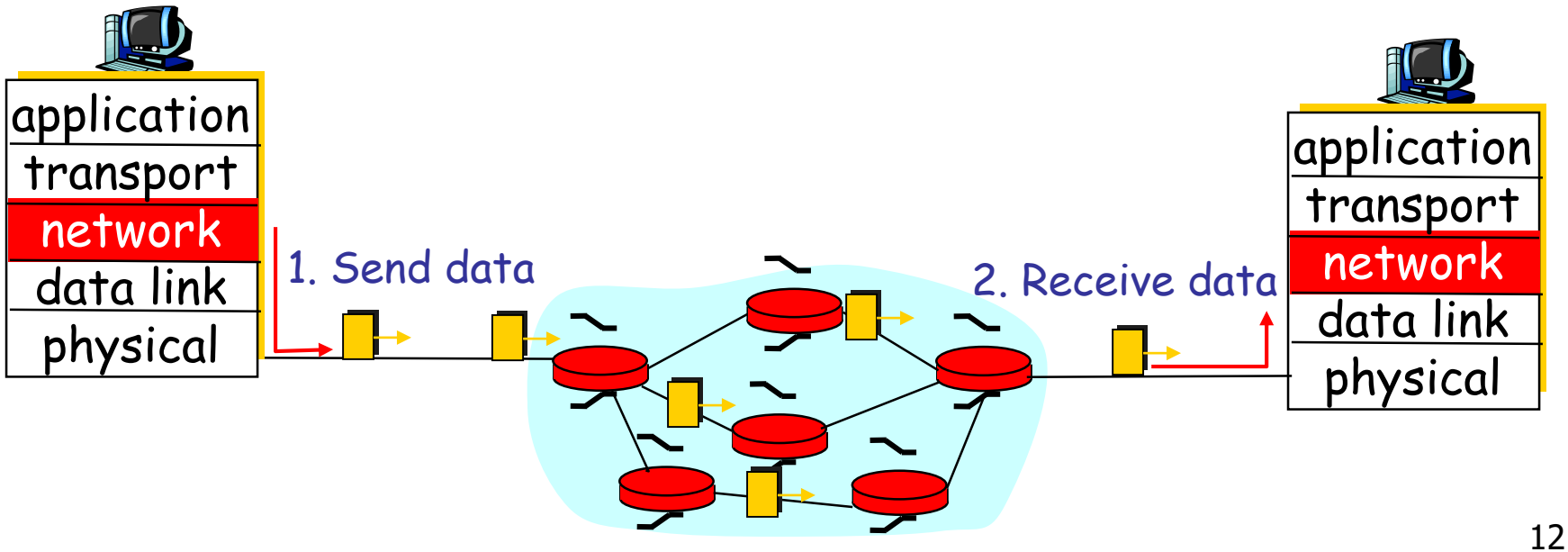
- 属于该VC的每个分组，携带一个VC标识（而非目的地址）
- 在每段链路上的VC标识可以改变
  - 新的VC标识来自于转发表

虚电路的特点

- 在一条链路上可以建立多条逻辑信道
- 一条VC由多段链路上的逻辑信道级联而成
- 分组依靠逻辑信道号（LCN）选择路由，因LCN只有局部意义，减少了分组头的开销和处理复杂度
- 能有效防止拥塞
  - 建立VC阶段
  - 在VC上的流量控制

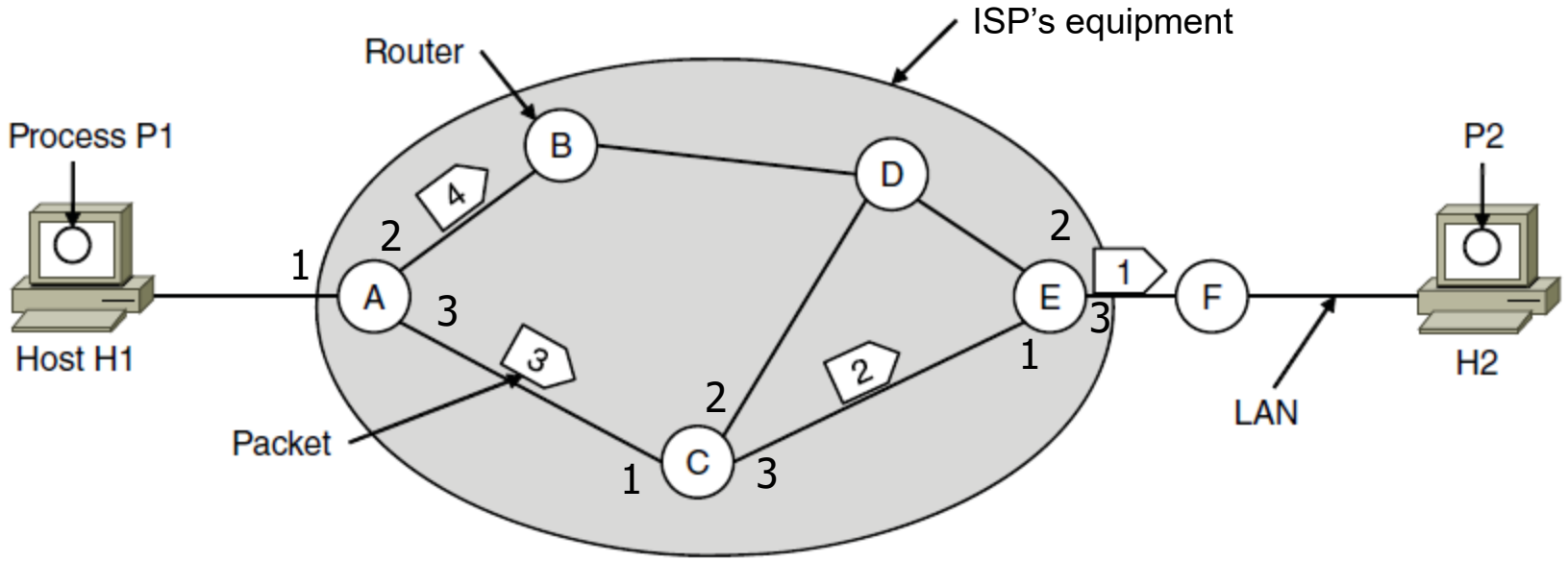
数据报：互联网模式

- 网络层没有呼叫建立过程
- 路由器：不维护端到端的连接状态
  - 没有网络层“连接” 这一概念
- 基于目的地址对分组进行路由选择
  - 一对源地址——目的主机的分组，可以选择不同路径



无连接的服务

- 只要更新转发表，分组的路由就改变了



A's table (initially)

目的	接口
A	--
B	2
C	3
D	2
E	3
F	3

A's table (later)

目的	接口
A	--
B	2
C	3
D	2
E	2
F	3

C's Table

目的	接口
A	1
B	1
C	-
D	3
E	3
F	3

E's Table

目的	接口
A	1
B	2
C	1
D	2
E	--
F	3

数据报的特点

- 每个分组的选路是独立的，利于网络资源的利用
- 分组在转发过程中，遇到一个节点或一条链路发生故障，可以重选路由，只需改变某一路由表项
- 分组头需要包含地址字段，开销增大了
- 各分组经过的路径可能不同，有可能出现先发后到现象
- 分组必须有生存时间限制，当生存期满时则被抛弃，避免在网络内死转而耗费资源

虚电路服务与数据报服务的对比

对比的方面	虚电路服务	数据报服务
思路	可靠通信应当由网络来保证	可靠通信应当由用户主机来保证
连接的建立	必须有	不需要
地址	仅在连接建立阶段使用，每个分组使用短的虚电路号	每个分组都有完整地址
路由方式	属于同一条虚电路的分组均按照同一路由转发	每个分组独立选择路由进行转发
当路由器故障时	所有通过故障的结点的虚电路均不能工作	故障的结点可能丢失分组，一些路由可能会发生变化
分组的顺序	总是按发送顺序到达终点	到达终点时不一定按发送顺序
差错控制和流量控制	可以由网络负责，也可以由用户主机负责	由用户主机负责

问题：关于VC与数据报的优缺点

- 假设路由器临时故障不能正常运行。 在VC与数据报的两类网络体系结构下，处理这种故障各需要采取哪些措施，哪种更有利？
- 发端声明其峰值信息速率，要求网络保证由源到目的节点的性能指标，如果网络无法满足需求的速率，则不允许节点访问网络。这种业务在VC与数据报的两类网络体系结构下，哪种更易实现？
- 试各举两例计算机业务，哪类适于数据报，哪类适于虚电路
- 数据报方式，分组独立路由，路由过程独立；虚电路方式，分组沿着预先指定的路径路由。这是否意味着虚电路无需具备单个分组独立路由的能力？

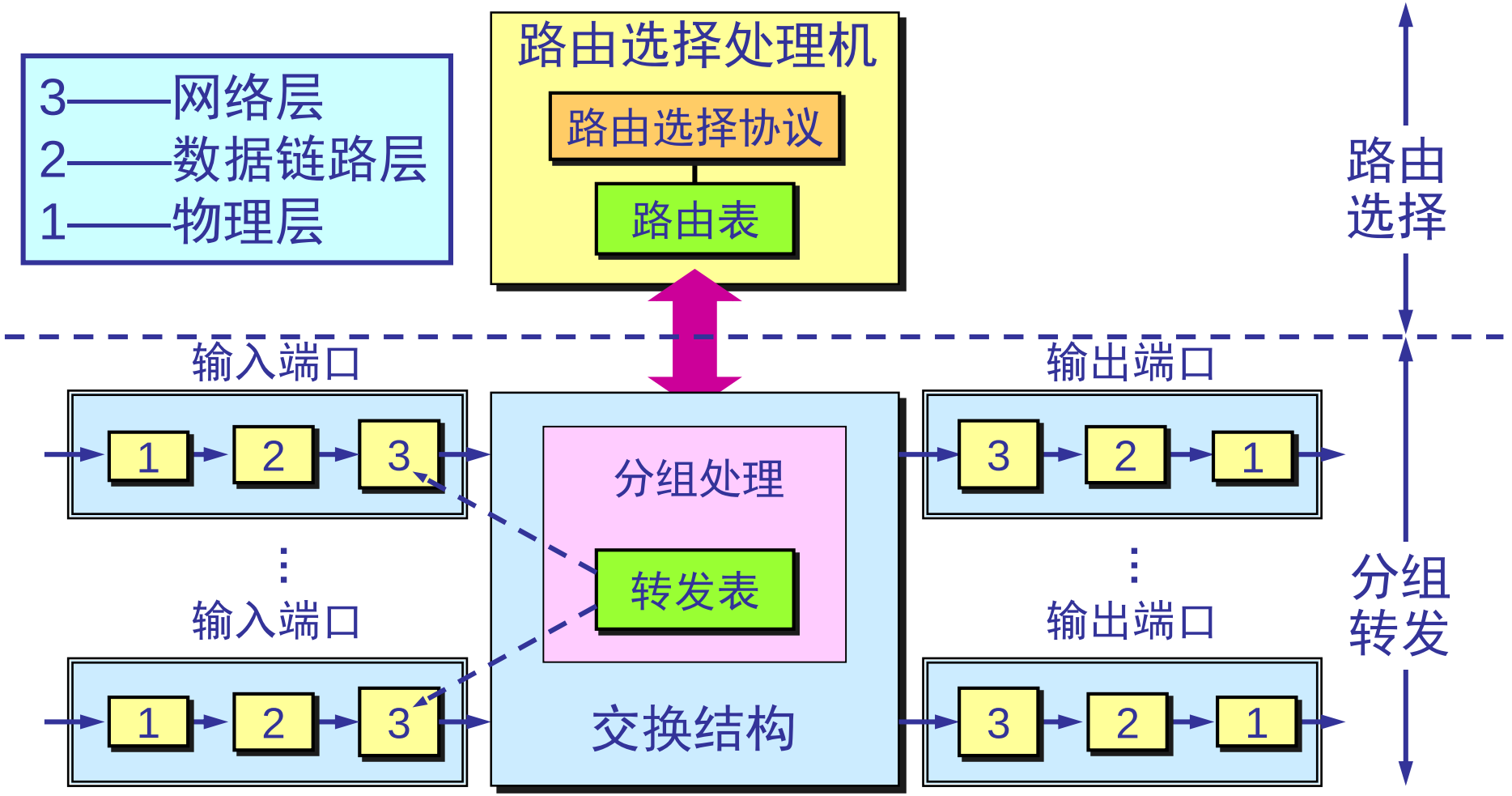
提纲

- 网络层的功能
- 路由器结构及工作原理
  - 转发及路由
  - 路由器的结构

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。

路由器结构及作用

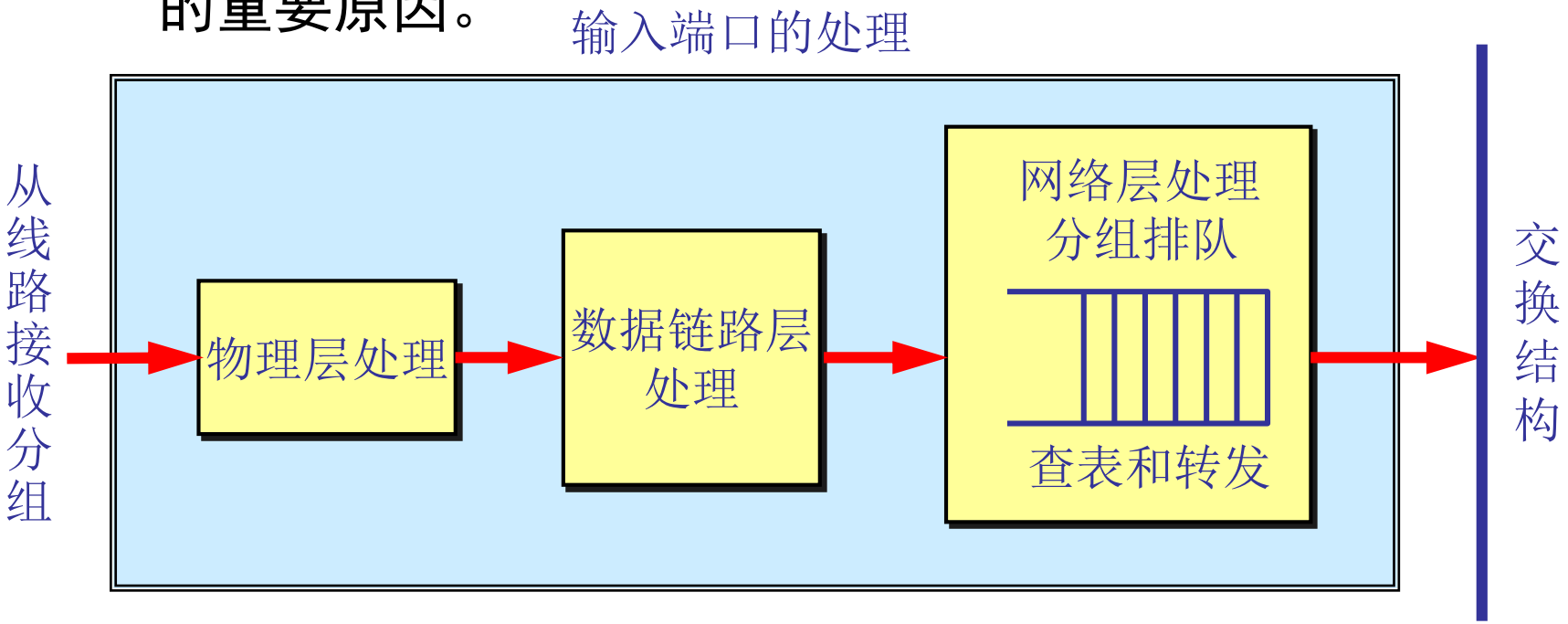

3——网络层
2——数据链路层
1——物理层
路由选择处理机
路由选择协议
路由表
路由选择
输入端口
输出端口
1
2
3
分组处理
转发表
3
2
1
输入端口
输出端口
交换结构
分组转发

“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分组从适当的端口转发出去
- “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

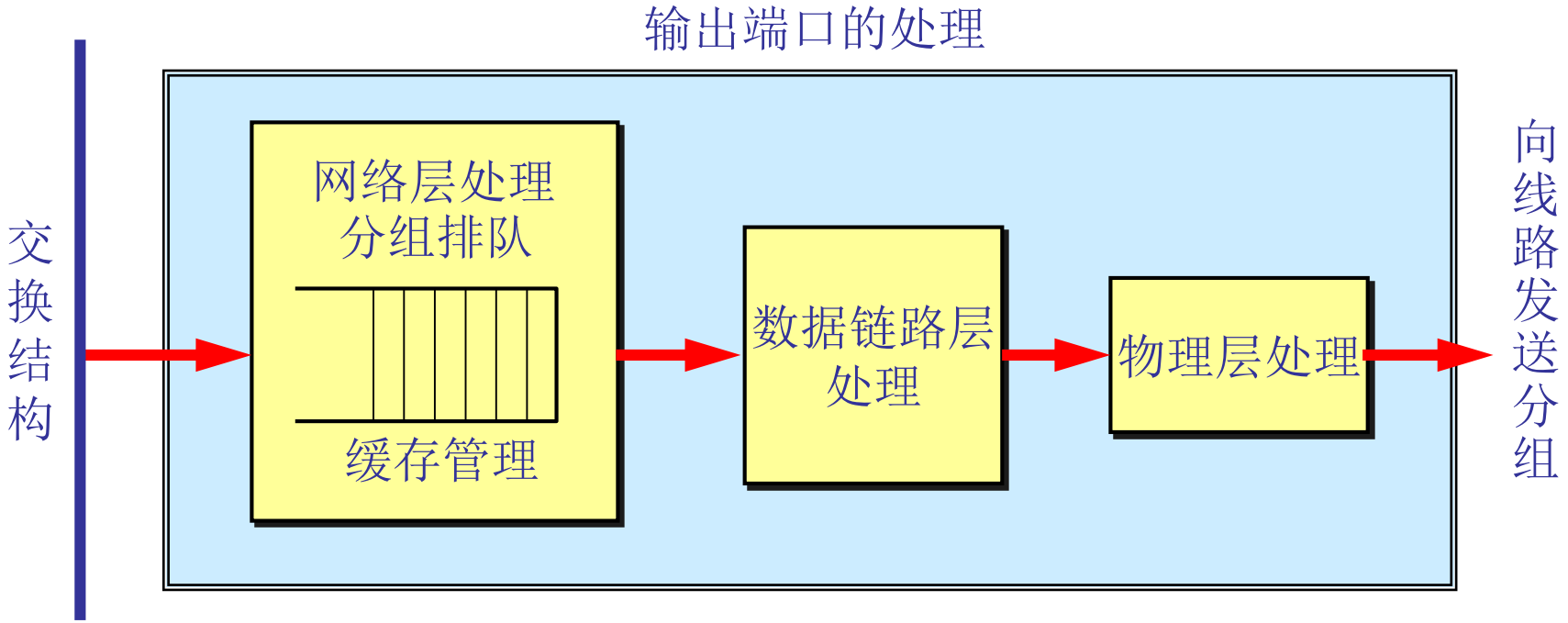
输入端口对收到分组的处理

- 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。这会产生一定的时延。
- 若路由器处理分组的速率低于分组进入队列的速率，则队列的存储空间最终减少到零，导致后续分组因没有存储空间而被丢弃。
- 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。

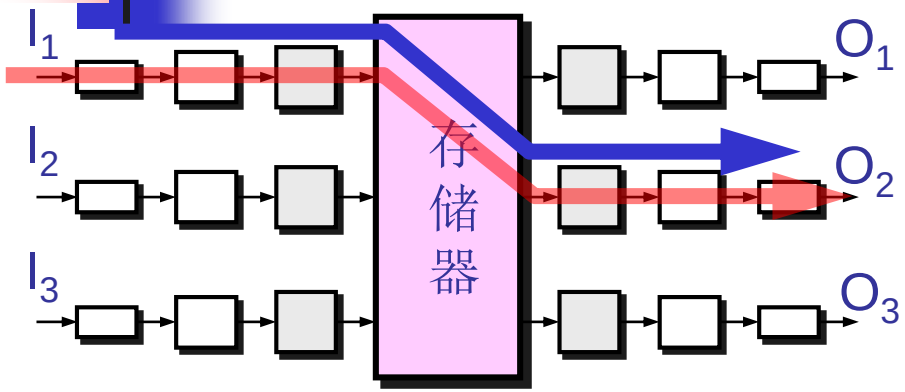


输出端口将分组发送到线路

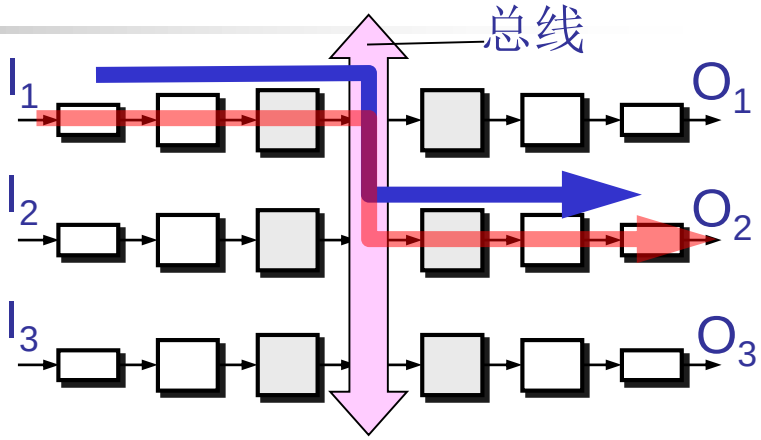
- 先缓存由交换结构接收的分组。数据链路层处理模块将分组加上链路层的首部和尾部，交给物理层后发送到外部线路。



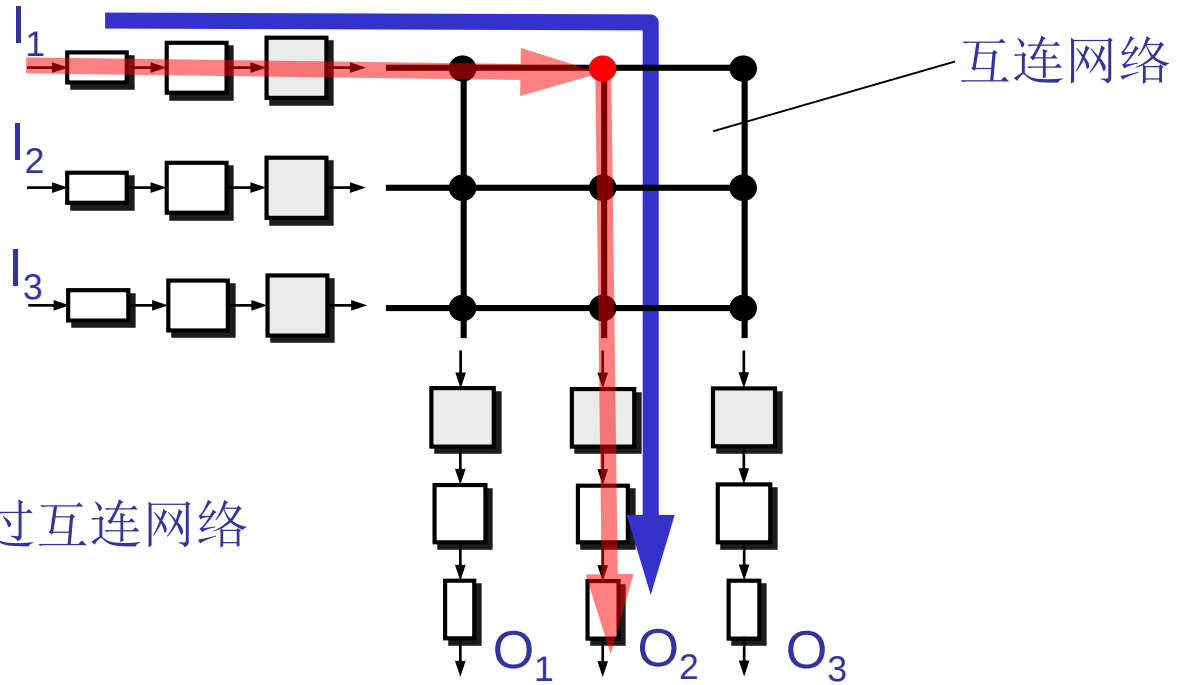
交换结构



(a) 通过存储器



(b) 通过总线

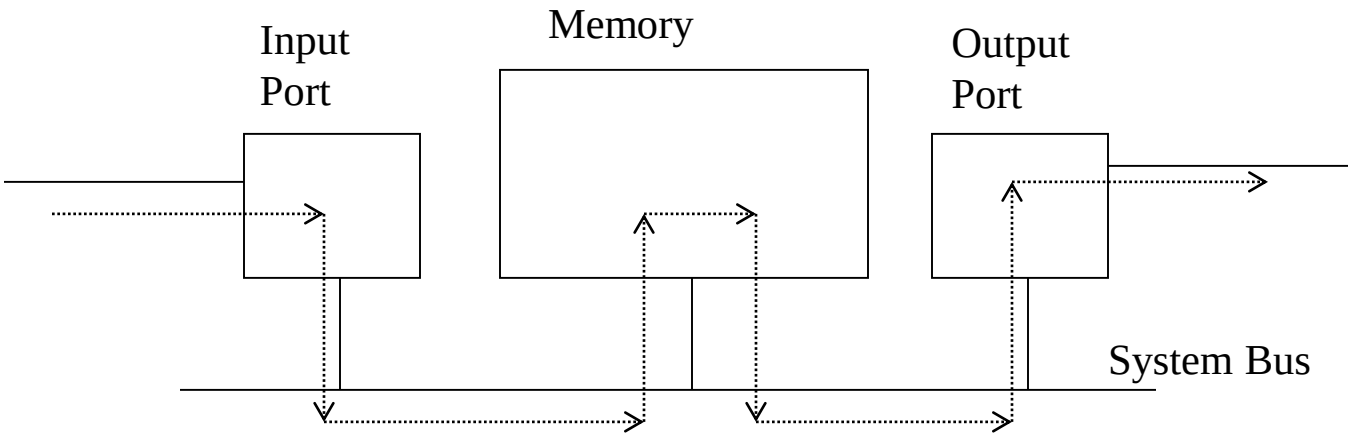


(c) 通过互连网络

交换结构1：经内存交换

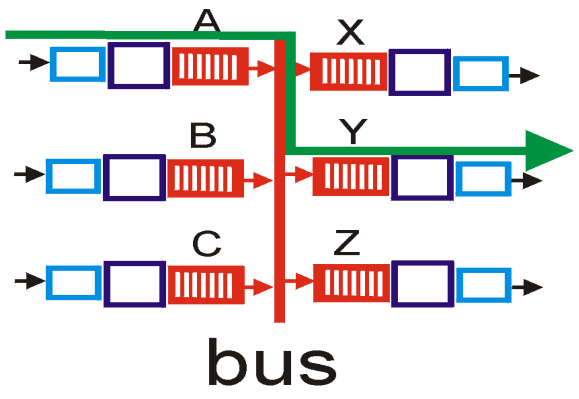
第1代路由器:

- 传统的计算机，带有交换功能，在CPU的直接控制下实施交换
- 将分组复制到系统内存中
- 存储器的带宽限制了速度，若B为总线速度，则转发的吞吐量为B/2
- 现代的路由器，由输入卡上的CPU执行分组的查表及存储



交换结构2：经总线交换

- 分组由输入口到输出口经一个共享总线
- 总线竞争：交换速度受限于总线速度
- 一般的速度高达1Gbps
- 32Gbps bus，Cisco 5600：满足接入及企业网络网的需求



交换结构3：由网络实现交换

- 克服总线带宽的限制
- Banyan 网络及其它互联网络，开发初期是为了互联多个处理器
- 设计优势：将数据报划分为固定长度的信元，加上标签后通过互联网络交换.
- Cisco 12000：经过交换网络，交换速率达到60 Gbps

Internet 路由结构

- 路由器的功能结构
  - 数据通道：支持前向判决，背板和输出端口调度
  - 控制：交换路由表、系统配置管理

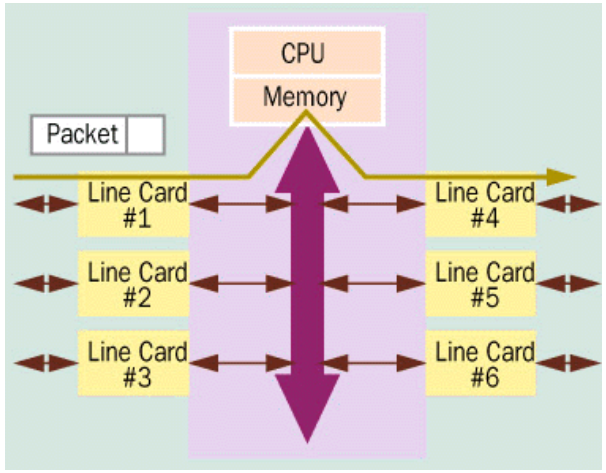


图 A：共享总线、CPU及存储器，线卡；

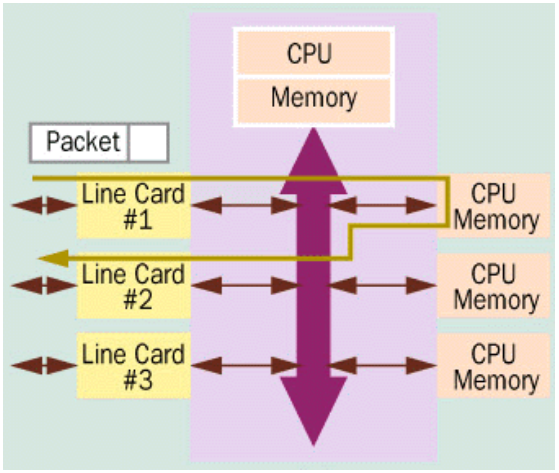


图 C：设置分立的CPU；

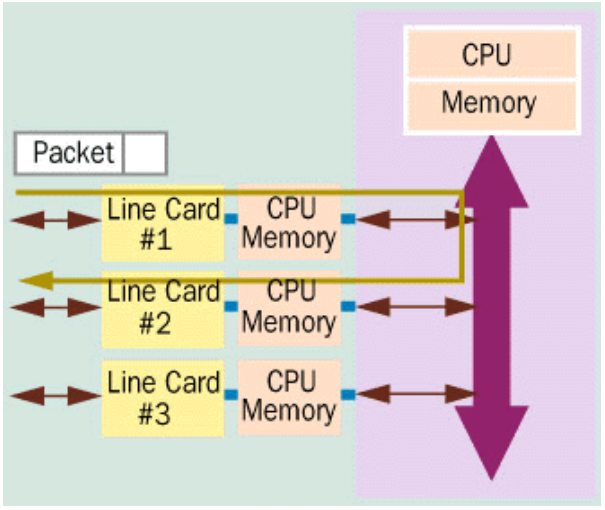


图 B：多个CPU同时并行处理；

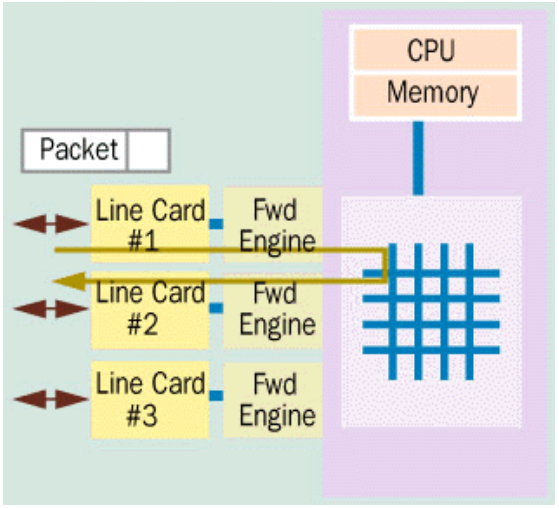
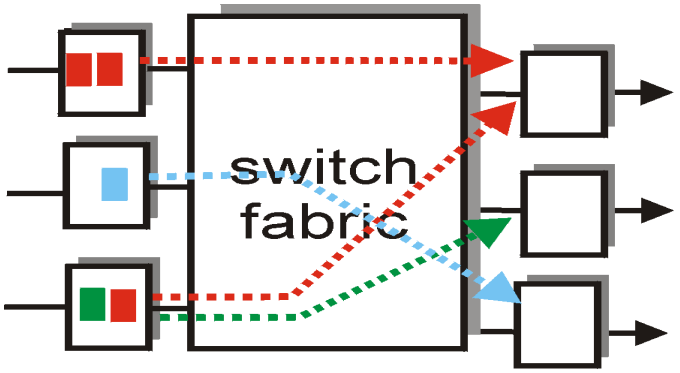


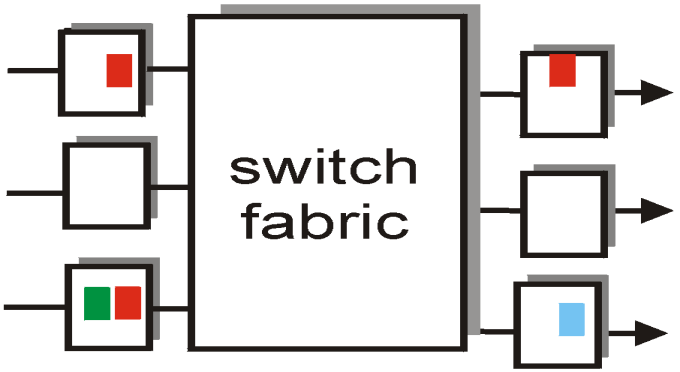
图 D：阵列板结构交叉开关

输入端口的排队

- 当交换速率低于输入速率的总和时，分组将在输入端口排队
- **线路前部 (HOL Head-of-the-Line blocking)阻塞：** 排队的分组必须等待通过交换结构
  - 竞争同一个输出端口
- **排队导致分组延迟、输入缓存的溢出以及分组丢失**
- 解决HOL阻塞的方法？ N. Mckeown 1997," A fast switched backplane for a Gigabit switched router"



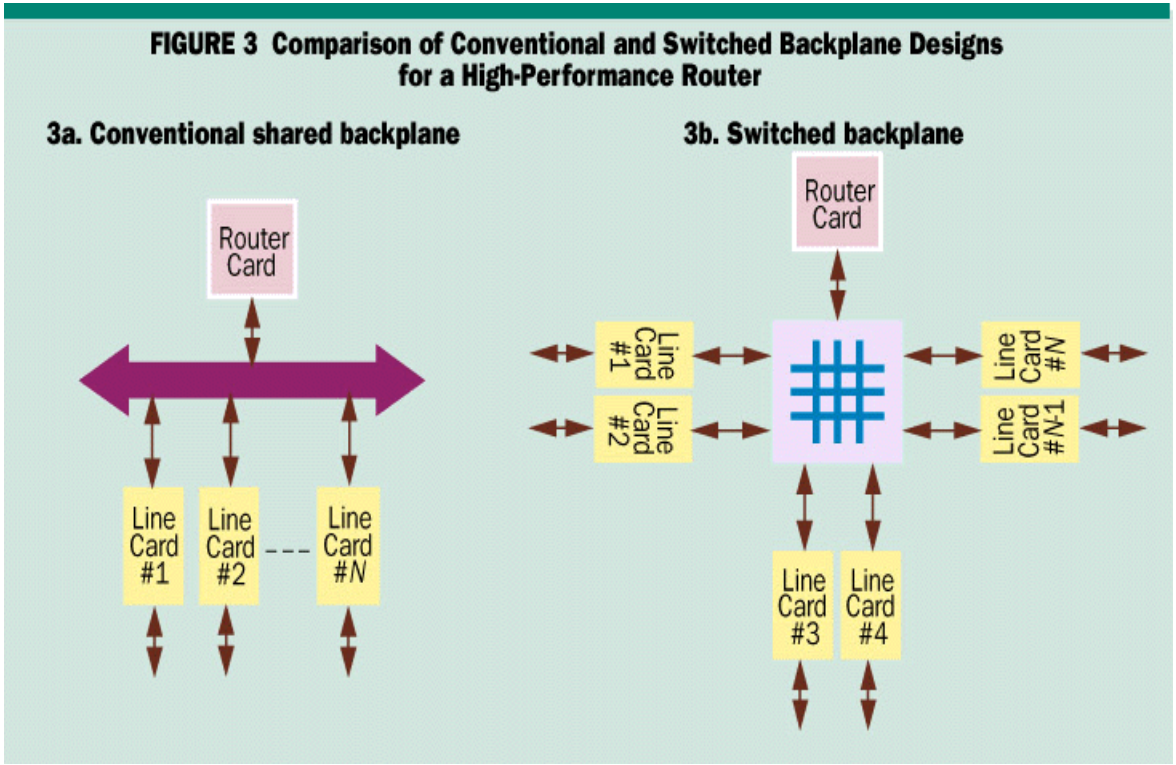
output port contention
at time t - only one red
packet can be transferred



green packet
experiences HOL blocking

阵列结构

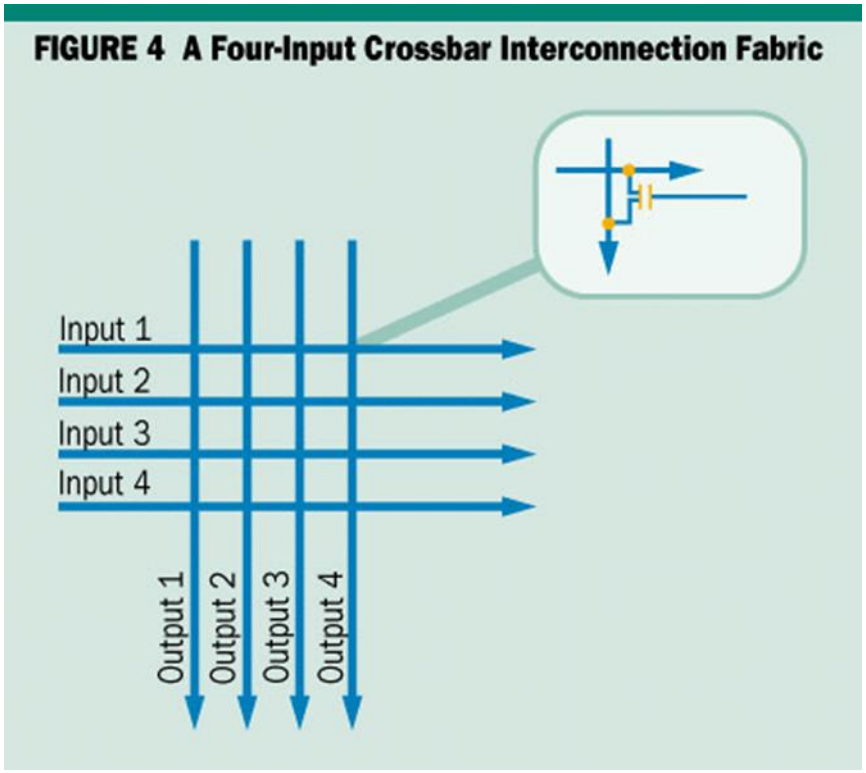
- 高速阵列交换技术
  - 线卡到中心交换机为端到端连接，每根线路有一个发射端，通过控制信号设定是否连接
  - 阵列交换机支持多数据线同时传输，大幅提高系统的聚合带宽



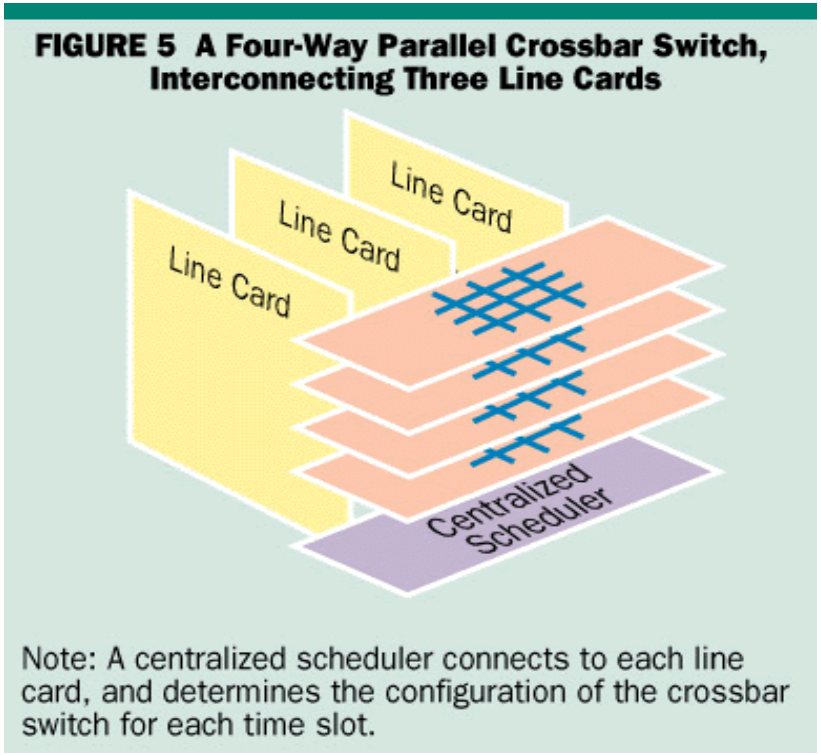
共享背板与阵列交换比较

阵列结构

传统阵列



平行阵列

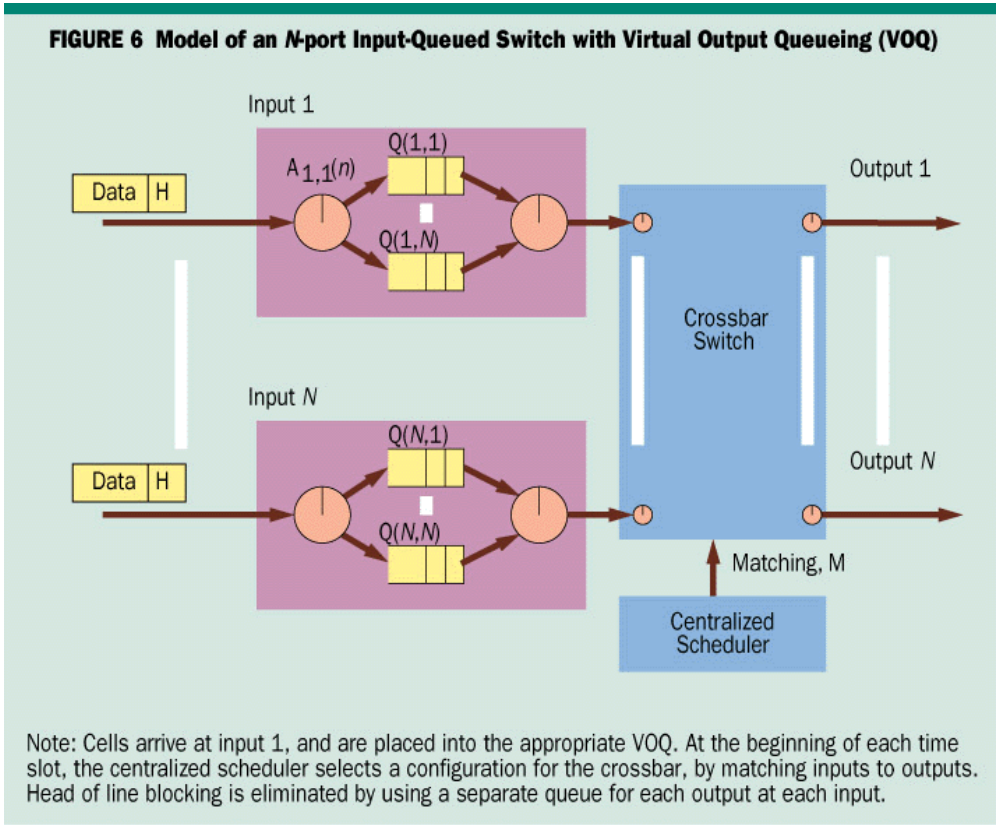


高速交换：定长分组

- 高性能的路由器在传送数据分组到达背板之前，将可变长度的分组分割为固定长度单元，称为cell。在发送到输出线路之前，再将cell重组为可变长度的分组。
- Cell机制下，每一个时隙末尾，调度算法检查正在等待传输至阵列板的分组。之后选择一个配置，决定下个时隙输入端口和输出端口的连接方式
- 通过分配算法保证各端口间的公平性

HOL：解决阻塞

- 解决方法：虚拟输出队列VOQ
  - 在输入端，对不同的输出建立FIFO；到达的cell在对应输出端排队。在每个时隙开始时，中心调度算法检测所有输入端口队列，找到不冲突的输入输出端口匹配。
- 合理应用VOC能够将吞吐量从60%提高至100%



解决输入输出阻塞问题

- 优先级机制
  - 根据紧急性设定优先级，优先级高的优先接入阵列交换板。阻止低优先级的用户影响高优先级的用户。
  - 使用协议如RSVP来限制高优先级的业务流量进入交换机。最高优先级的业务量较小，保持固定的时延。
- 加速原理
  - 令阵列板运行速度高于外部连线速率，如两倍于外部线速率，称之为加速2。
  - 对于N端口的设备，若要保证没有数据分组在输入端排队，至少需要加速N。
- 实际上，VOQ队列加速2即可合理控制交换机延时

缓存区长度如何设计?

- 计算缓存长度的方法：RFC 3439
  平均缓存量B
  = 平均往返时延RTT×链路容量C
  - 例C = 10 Gps，RTT=250ms，需缓存2.5 Gb

- 近期的建议：若有*N*个流经过一条链路，则缓存量为

$$RTT \times C / \sqrt{N}$$

小结

- 网络层提供的服务：
  - 与路由器的技术无关
  - 屏蔽路由器的数量、类型和拓扑结构
  - 采用统一的编址方案
- 面向连接的服务：虚电路
- 面向非连接的服务：数据报
- 路由与转发的区别
- 未讨论的其他功能
  - 路由算法与协议
  - 网络互联
  - 统一编址
  - 网络拥塞控制
  - 保证服务质量