

分组交换

华中科技大学电子信息与通信学院
通信工程系
陈京文

Email: jwchen@hust.edu.cn

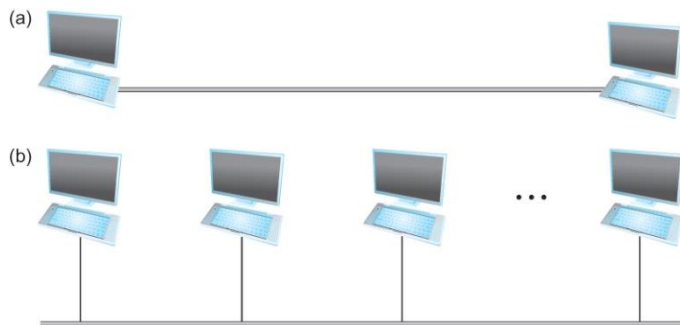
2020.10.9



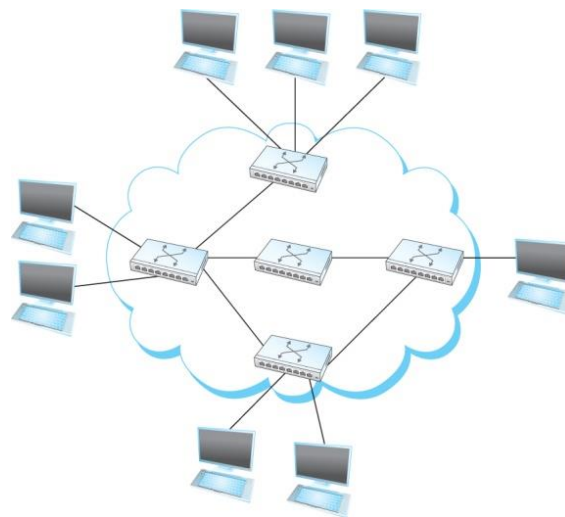
内容提要

- 分组交换机制
 - 数据报交换，虚电路交换
- 数据链路层交换机
 - 网桥，以太网交换机

交换网络



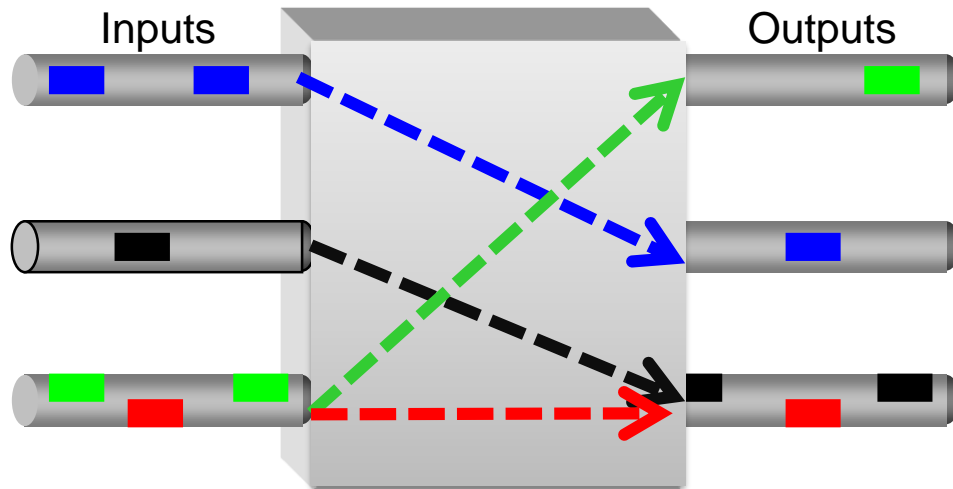
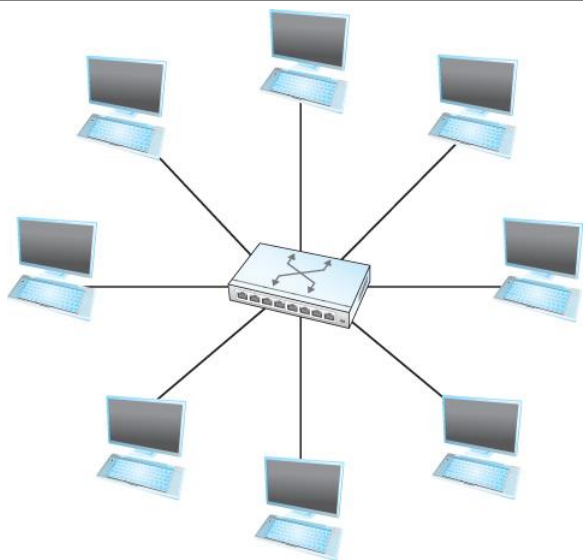
直接链接网络



交换网络

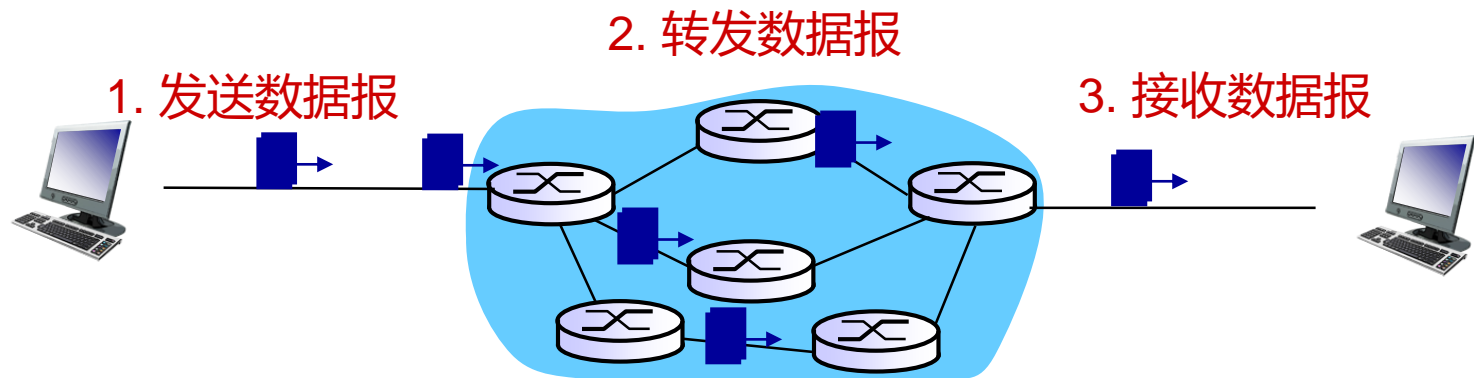
- 交换网络
 - 主机之间无直接链路
 - 交换节点为主机之间提供间接连接，即提供数据传输中继
- 相对于直接链接网络的优点
 - 添加新的主机时，对于网络中现有主机的性能没有影响
 - 可以搭建支持大量主机的网络
 - 可以搭建覆盖较大地理区域的网络

分组交换



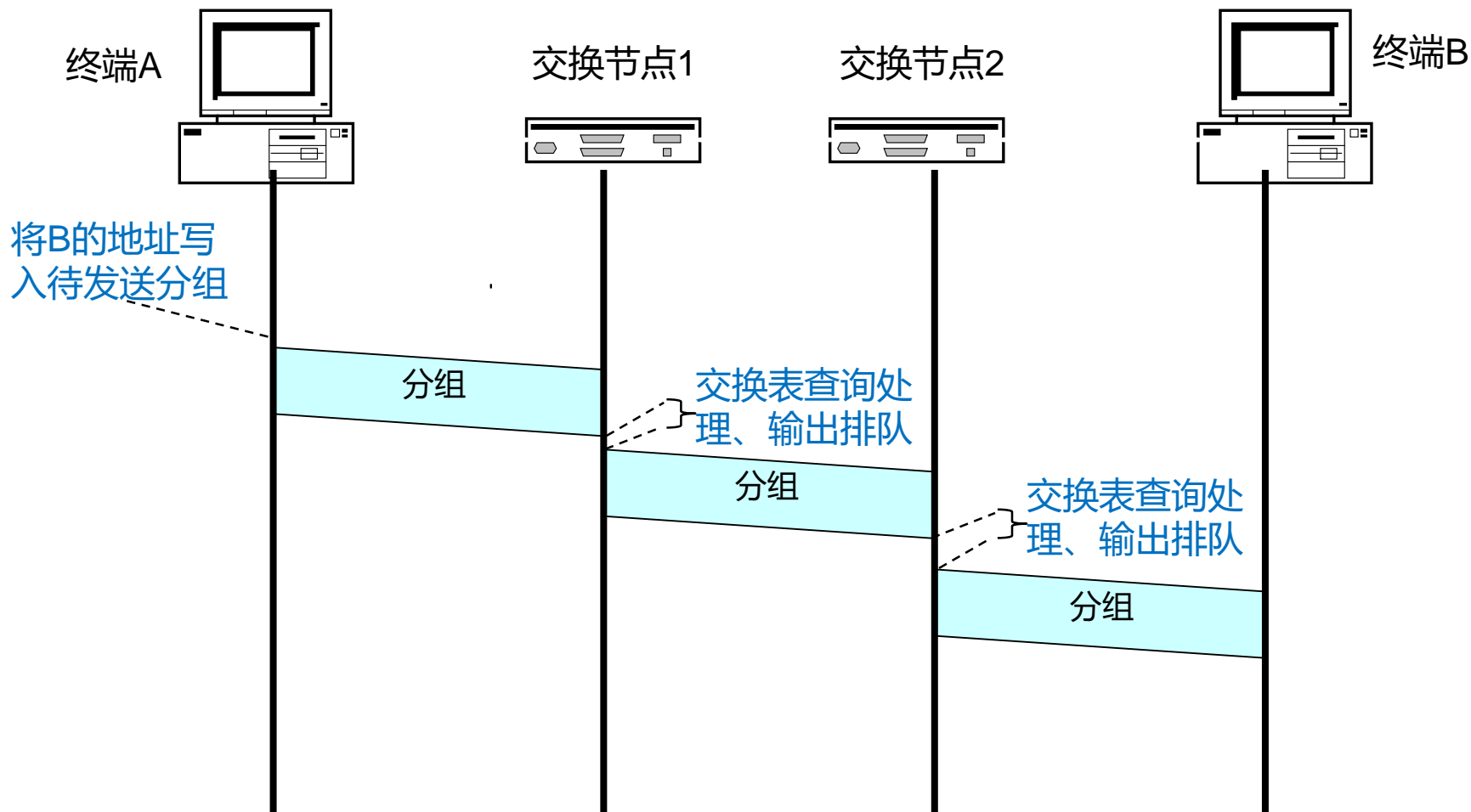
- 分组交换：交换节点从一个端口接收分组，然后从一个或多个端口传输出去
 - 这一过程需要确定合适的输出端口
- 根据**确定输出端口的方式**进行分类
 - 采用交换表(switching table)：数据报交换(datagram switching), 虚电路交换(virtual circuit switching)
 - 源节点确定并给出：源路由(source routing)

数据报交换

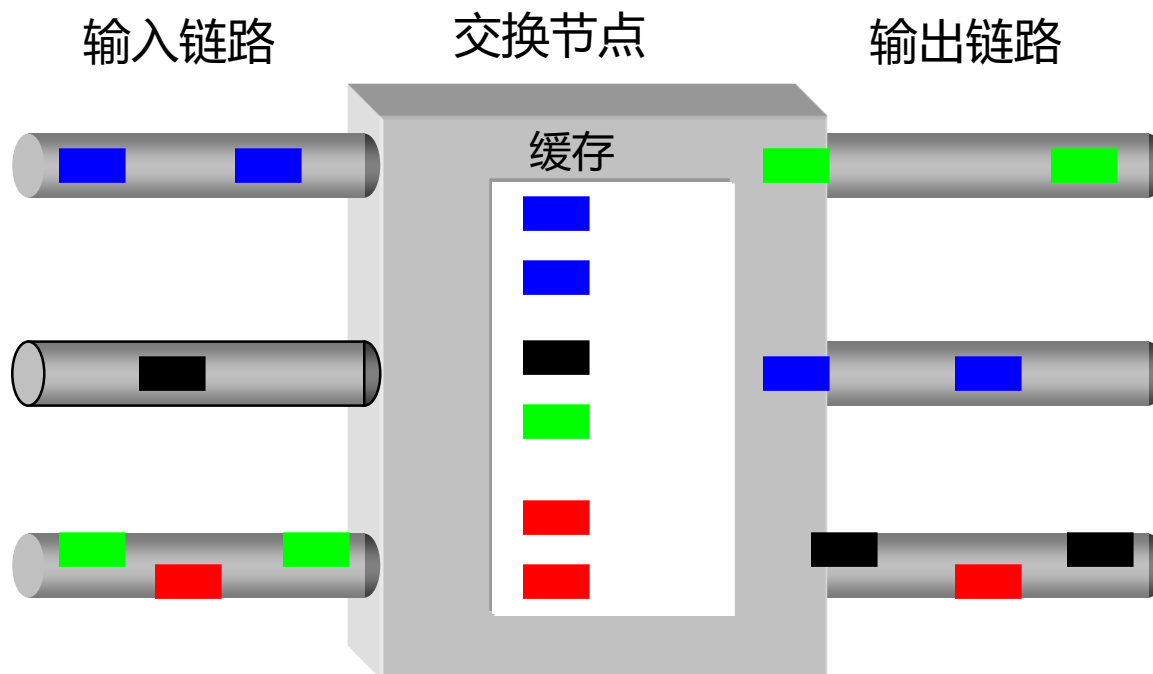


- 发送数据报时，源主机在首部中填入目的地址
- 收到数据报时，交换节点
 - 将目的地址作为索引，查询其交换表，以确定输出端口
 - 然后存储转发这个数据报

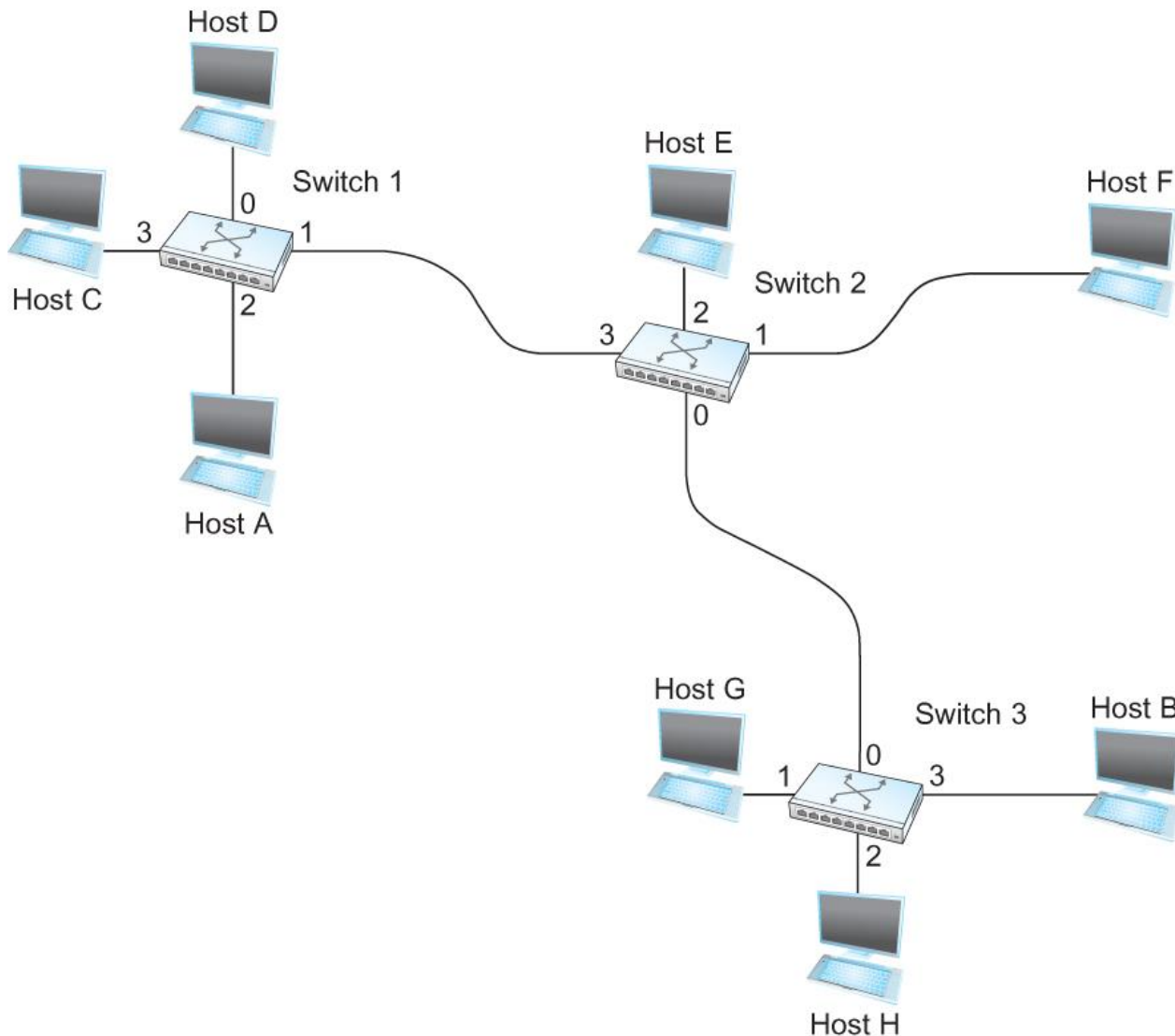
数据报交换时序过程



数据报交换节点：存储转发



数据报交换表示例



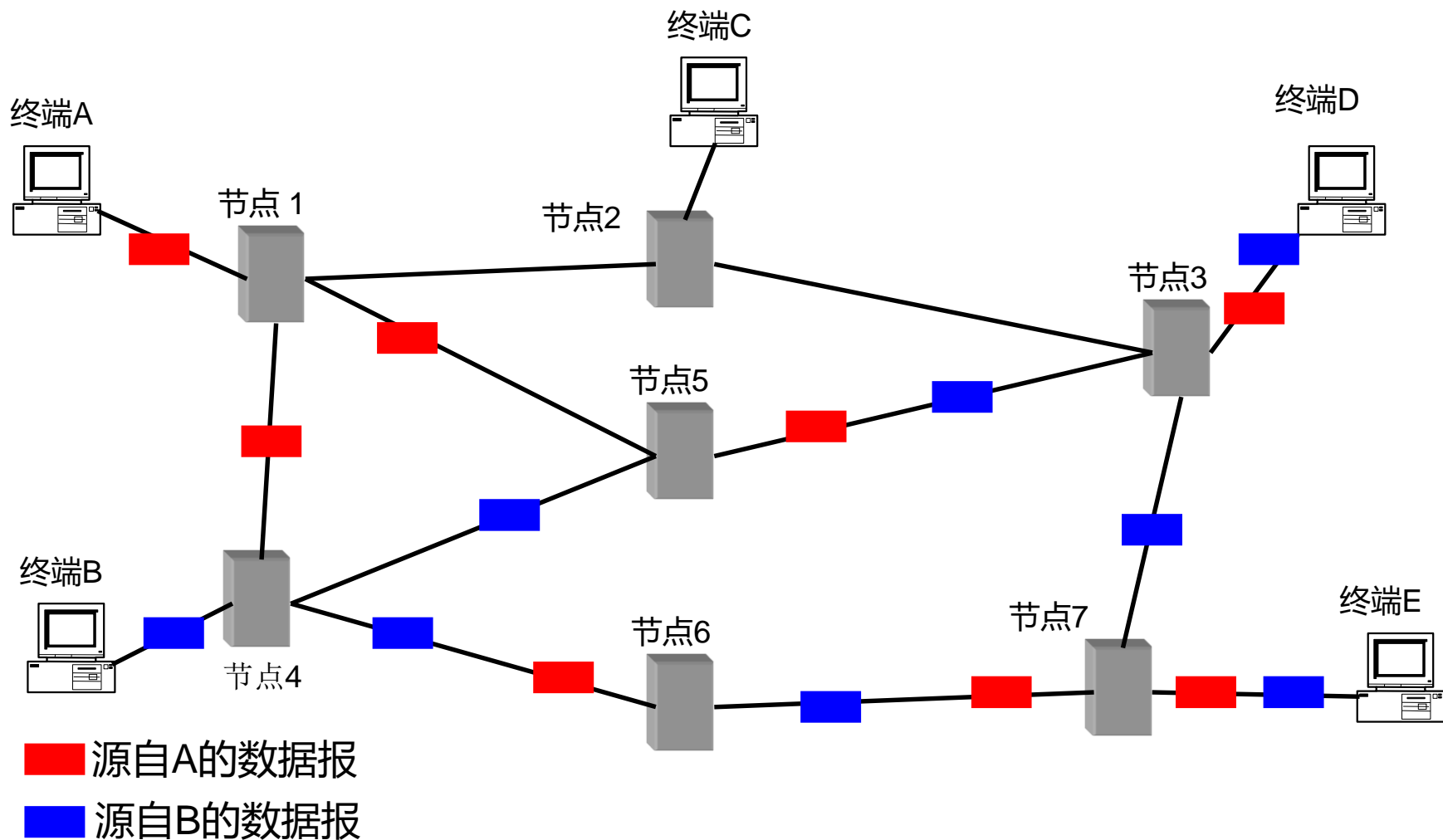
目的地	输出端口
A	3
B	0
C	3
D	3
E	2
F	1
G	0
H	0

交换节点2的交换表

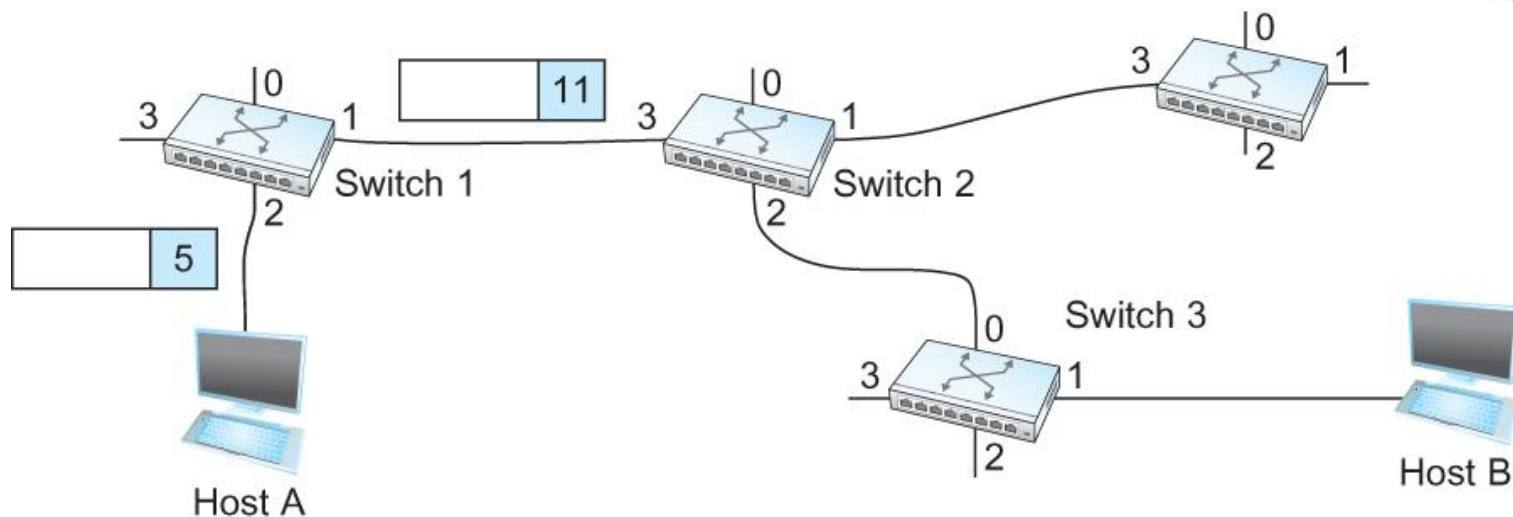
数据报转发

- 所有数据报，即使是来自于同一源节点并去往同一目的地，都是被交换节点**根据其交换表独立地转发**
- **数据报转发路径**
 - 并无可见形式，而是**逻辑上表现为沿途节点交换表中对应于同一个目的地的条目的集合**
 - 也无数据报转发路径的显式设置
- 数据报交换节点**运行特定的协议/算法，以建立和维护其交换表**
 - 相关协议算法的目标：如何求解交换表
 - 求解时考虑的因素：**拓扑， ...?**

数据报交换网络

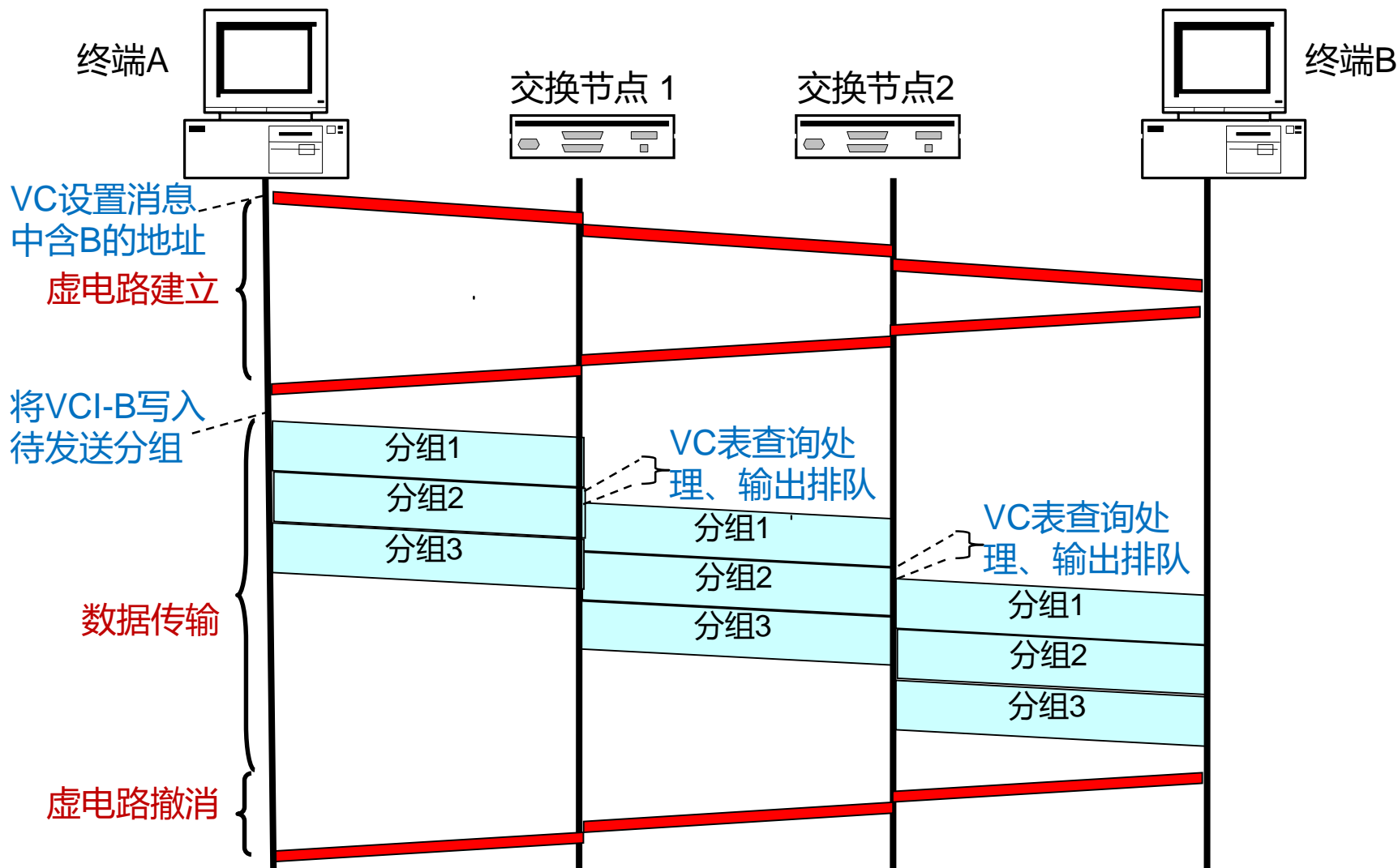


虚电路交换

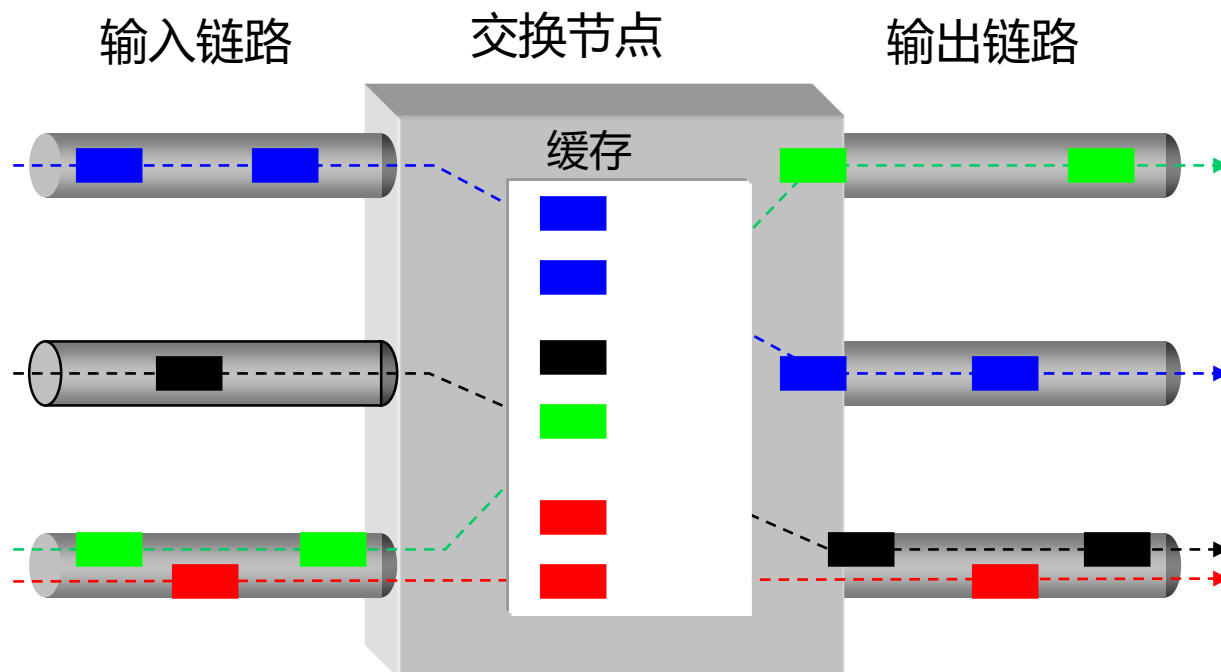


- 传输分组之前，主机需建立一条至远端主机的**虚电路(virtual circuit, VC)**，即一条逻辑连接
- 发送分组时，源主机只需在分组首部填入**虚电路号(virtual circuit identifier, VCI)**，而非完整的目的地址
- 收到分组时，交换节点
 - 采用分组首部中**VCI** (加上输入端口)作为索引，**查询其虚电路表(virtual circuit table)**以确定输出端口
 - 然后**存储转发**这一分组

虚电路交换时序过程



虚电路交换节点：存储转发



虚电路表

- 每条记录(条目)包括至少四个字段
 - 输入端口
 - 输入虚电路号(VCI)
 - 输出端口
 - 输出虚电路号(VCI)
- 虚电路号只有本地意义 —— 有效性局限于给定交换节点的给定输入端口
 - 并无全局意义 —— 即对于相应端到端虚电路经过的交换节点及其它端口均无意义
- 对于一个虚电路交换节点而言, 二元组<输入端口, 输入VCI>唯一地标定了通过该节点的一条虚电路

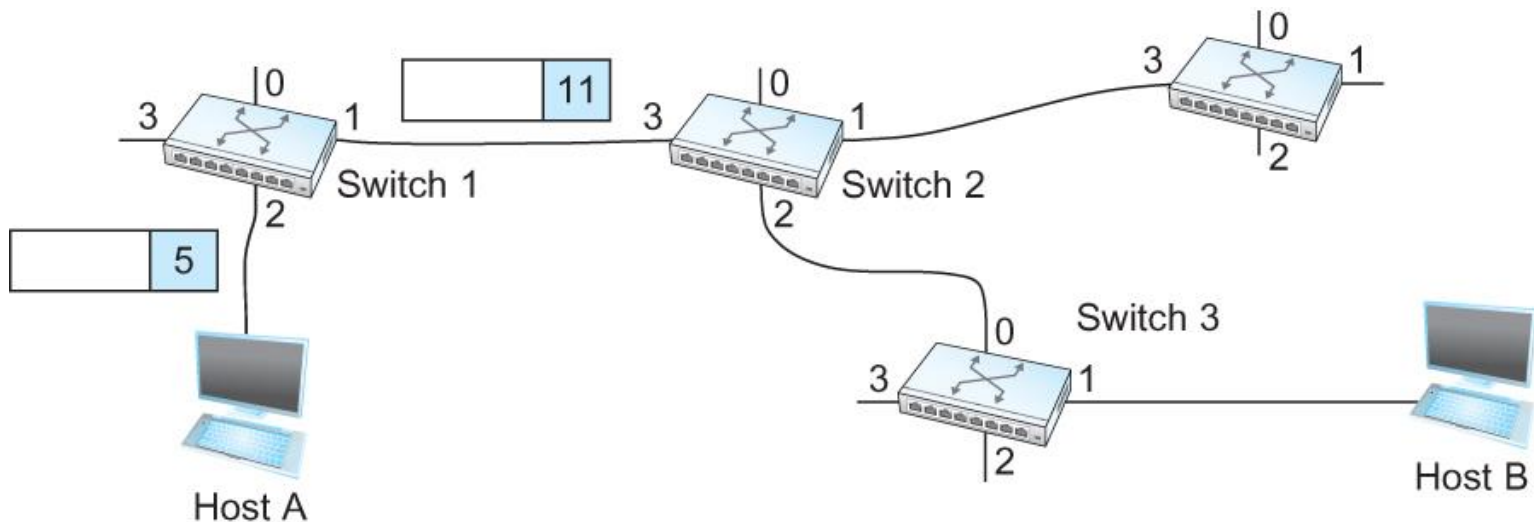
虚电路表示例

交换节点1中虚电路表

IfIn	VCIn	IfOut	VCOut
2	5	1	11
...

交换节点2中虚电路表

IfIn	VCIn	IfOut	VCOut
3	11	2	7
...



IfIn	VCIn	IfOut	VCOut
0	7	1	4
...

交换节点3中虚电路表

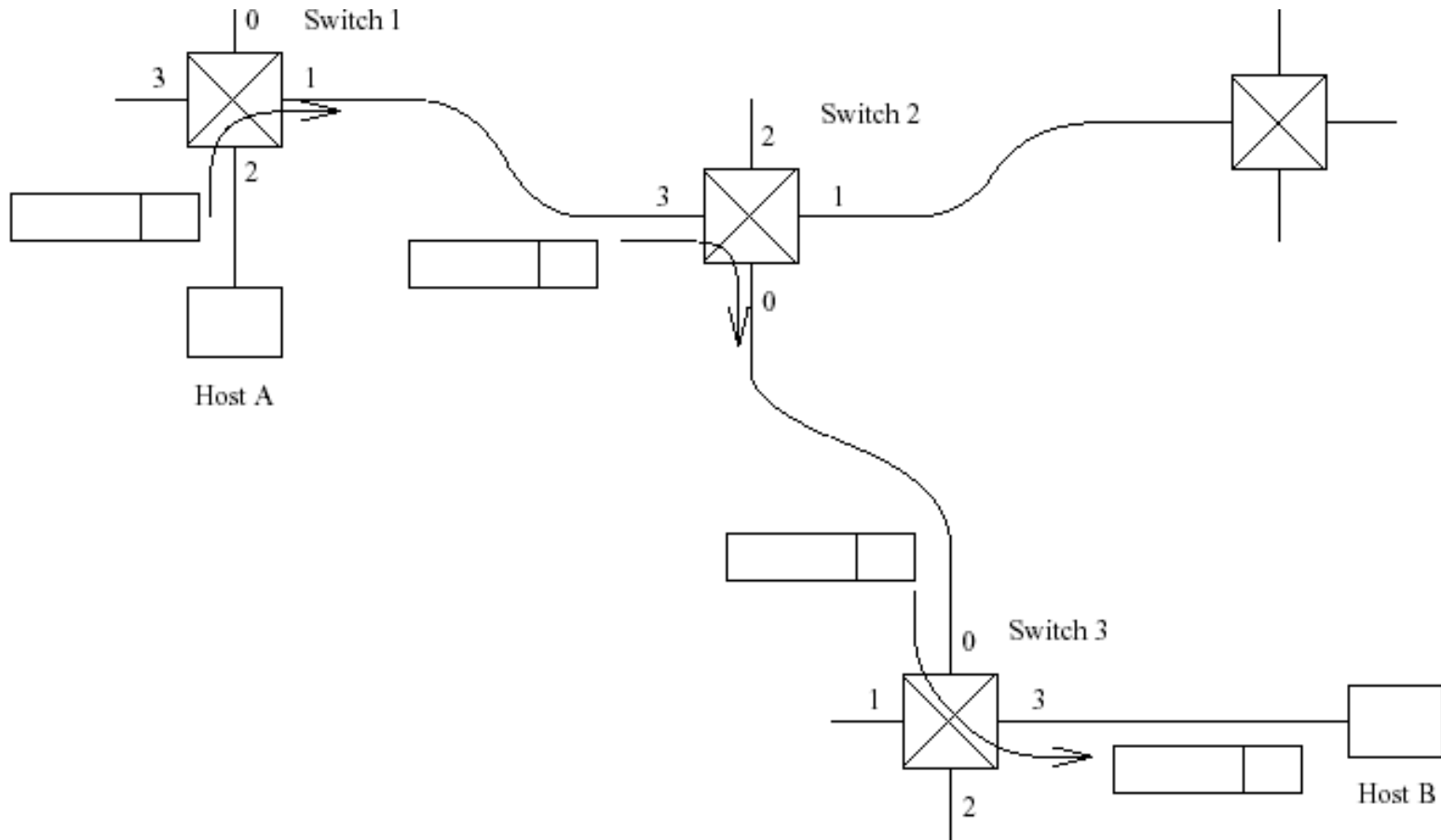
虚电路

- 两种类型
 - 永久虚电路(permanent virtual circuits, PVC)
 - 交换虚电路(switched virtual circuits, SVC)
- PVC
 - 由网络管理员手工或启动信令过程设立
 - 长期存在
- SVC
 - 由终端主机启动信令过程动态建立
 - 短期存在/有效

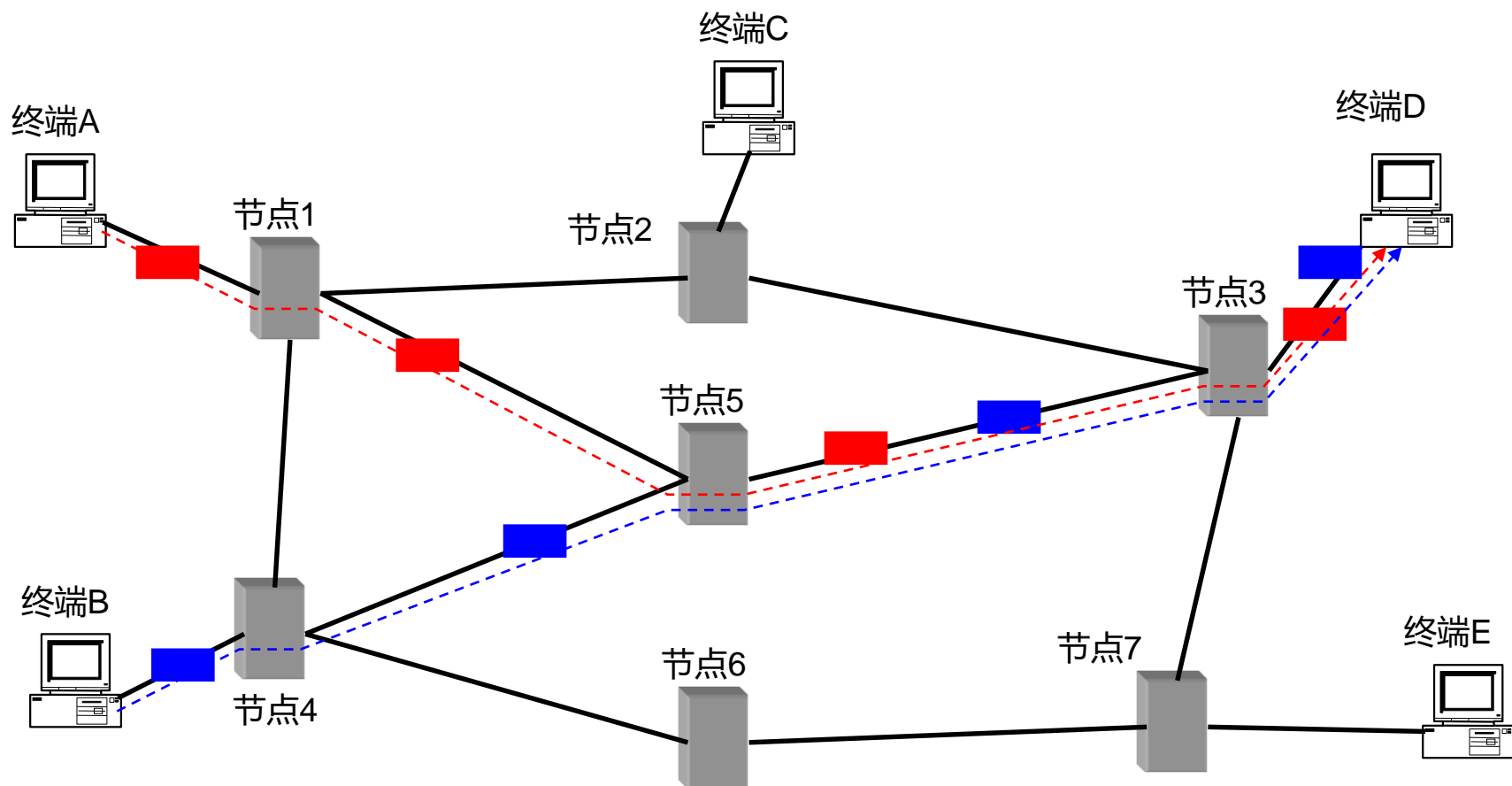
SVC建立过程

- 源主机(caller): 向目的地主机发送一条VC设置消息, 其中包含目的主机的完整地址
- 中间交换节点 (收到VC设置消息后)
 - 根据目的地地址, 确定输出端口
 - 在虚电路表中创建一条VC条目, 其中输入VCI选择在该设置消息到达端口未使用的VCI值, 输出VCI字段空缺
 - 通过输出端口向下游(朝向目的地主机)交换节点转发VC设置消息
- 目的地主机(callee) (收到VC设置消息后)
 - 选择一个在设置消息到达端口未使用的VCI值, 作为其输入VCI
 - 回复一条VC设置确认消息, 包含上述选定输入VCI
- 中间交换节点 (收到VC设置确认消息后)
 - 将下游节点/主机VC设置确认消息中的输入VCI, 填入先前建立的VC条目中空缺的输出VCI字段
 - 向上游节点(朝向源主机)转发VC设置确认消息, 其中输入VCI替换为先前选择的输入VCI值
- 源主机 (收到VC设置确认消息后): 获知用于随后数据传输的输出VCI

SVC建立示例



虚电路交换网络



分组转发路径与VC

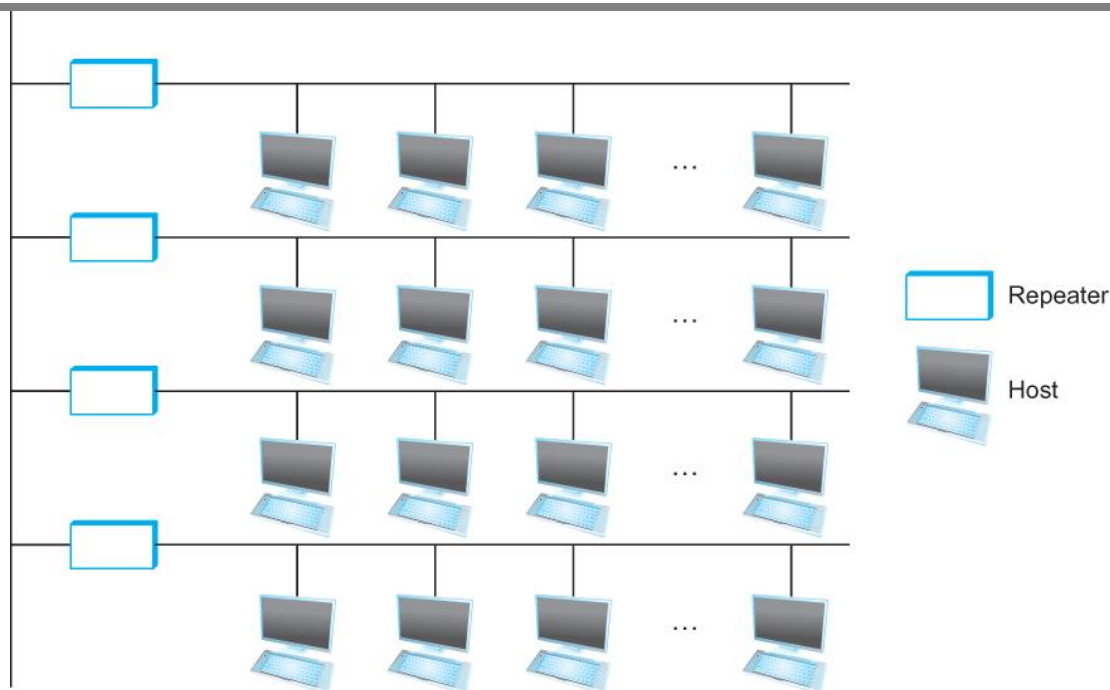
- 虚电路交换节点根据其虚电路表转发分组
 - 每个分组与一条虚电路关联，由分组首部的VCI及输入端口标定
- 虚电路为一条显式建立的逻辑连接
 - 实际表现形式为虚电路所穿过的全部交换节点的虚电路表对应条目中二元组<输入端口, 输入VCI>的串接
 - 对应于该虚电路对应分组的网络传输路径
- 建立虚电路时，交换节点需要确定如何安放该虚电路，即选择下一跳节点
 - 这一问题类似于数据报交换网络中交换表条目的建立



内容提要

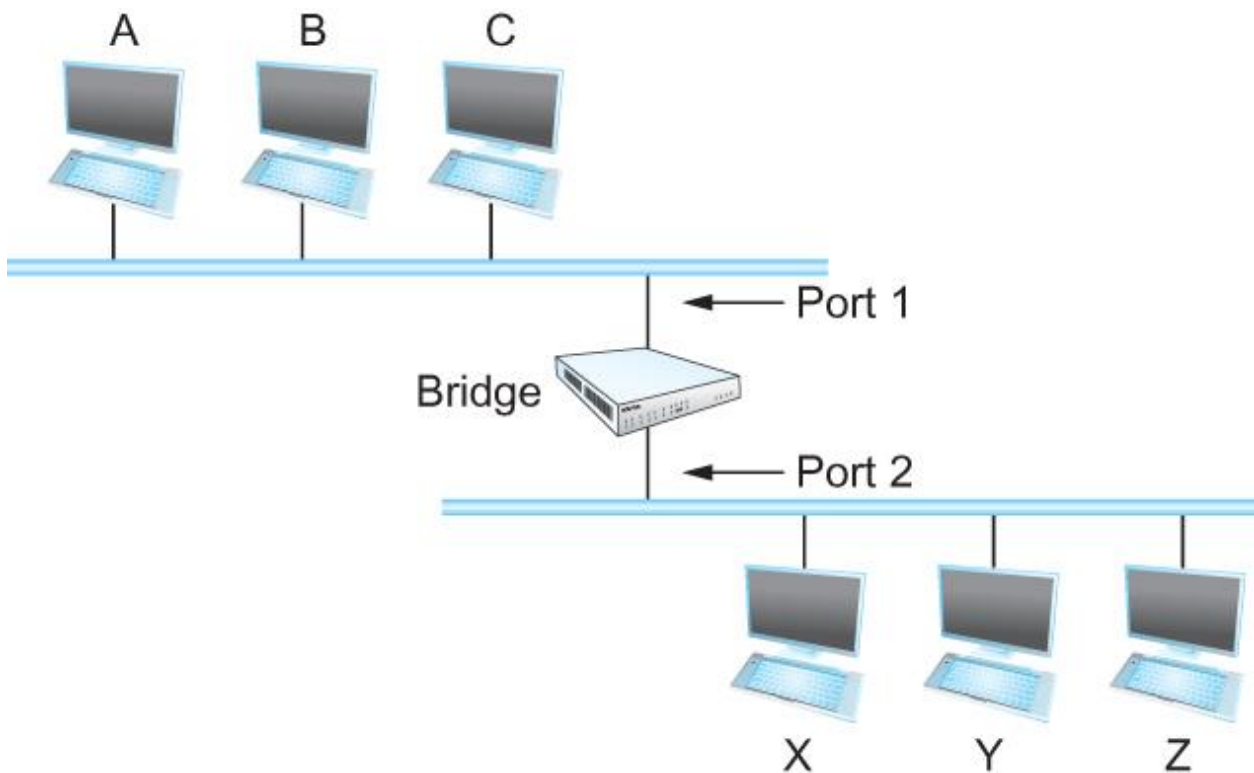
- 分组交换机制
 - 数据报交换，虚电路交换
- 数据链路层交换机
 - 网桥，以太网交换机

常规共享式局域网的局限



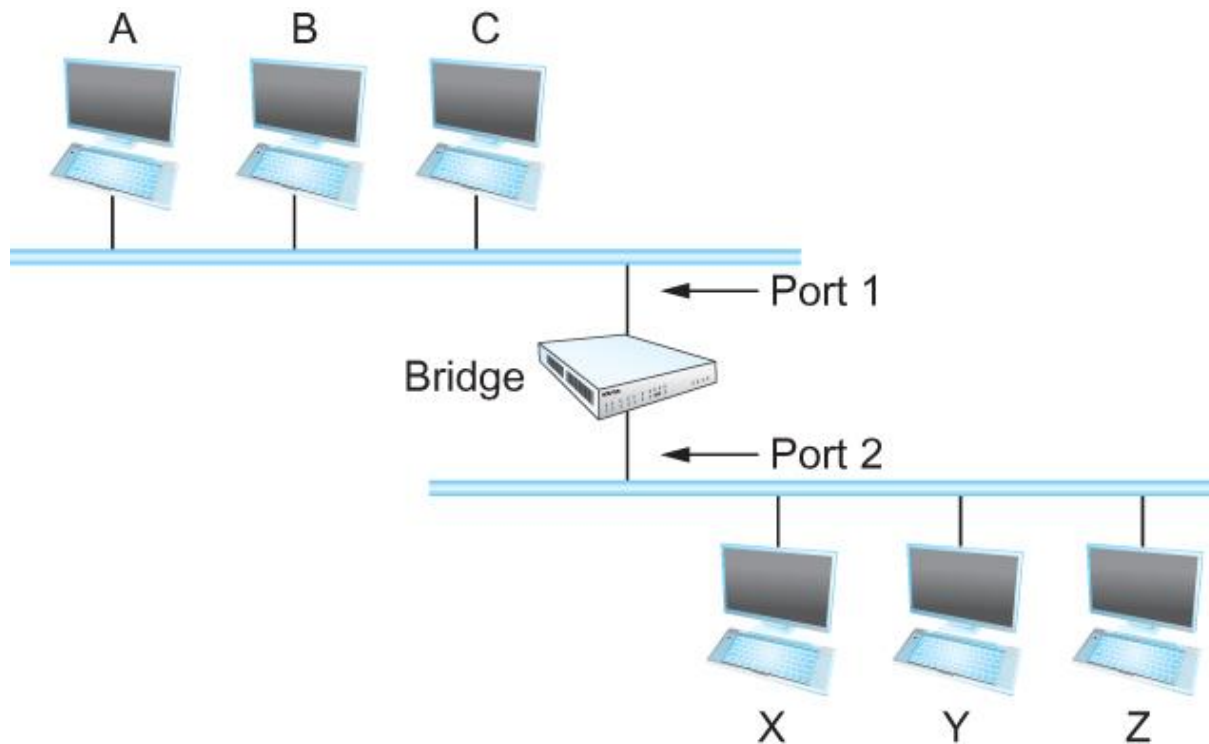
- 常规共享式以太网
 - 最长2500m，最多4个中继器(repeater) —— 限制了局域网规模
- 扩大局域网规模的解决方案
 - 共享式局域网的年代，主要采用网桥(Bridge)
 - 目前采用基于交换机(Switch)的交换式局域网，但沿用了网桥相关生成树协议去除拓扑中的环路

采用网桥扩展局域网



- 网桥连接多个局域网，以建立较大规模的局域网
- 基本网桥(Basic bridge)
 - 将收到的帧向所有其它端口转发 —— 效率低

学习网桥



Host	Port
A	1
B	1
C	1
X	2
Y	2
Z	2

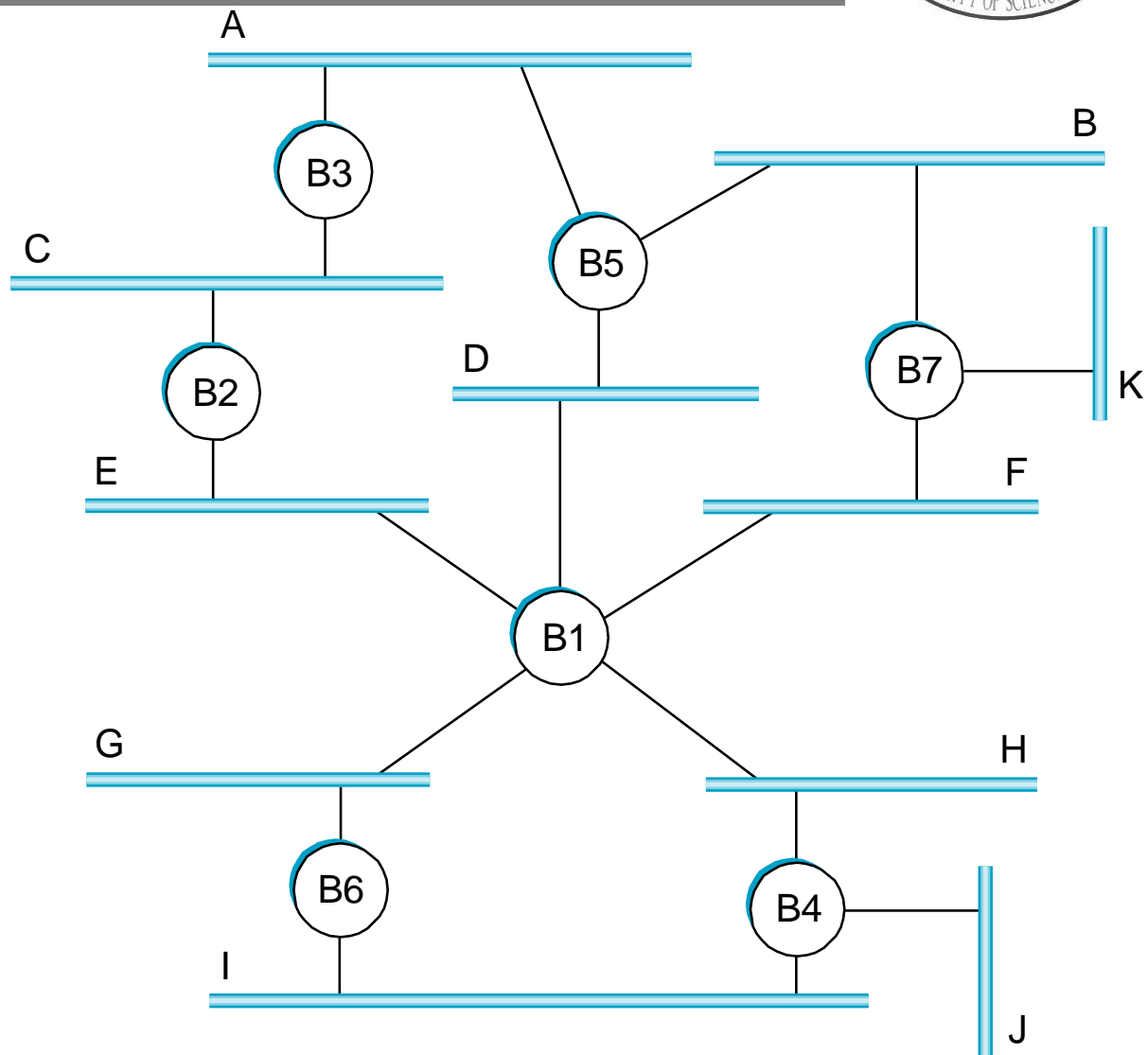
- 学习网桥(Learning bridge)采用选择性转发
 - 维护一个转发表
 - 根据转发表确定转发端口，不向无关端口转发

学习网桥(续)

- 学习连接到一个端口的主机地址
 - 通过查看经该端口收到的帧的源地址
- 通过学习所收到的帧的源地址，**建立和维护转发表**
 - 启动时，转发表为空
 - 运行过程中，**通过学习添加条目**
- **超时机制**
 - 每一条目都存在有效时间，超时则删除，即采用**软状态(soft state)**
- **转发帧**
 - 以目的地址为索引，**查询转发表**
 - 如果无对应条目，**洪泛** —— 向所有其它端口发送该帧
 - 如果存在对应条目，**选择性转发**

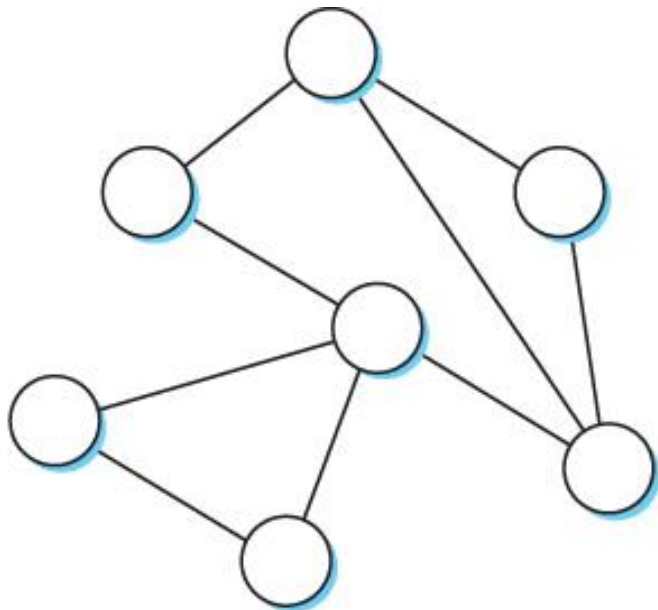
桥接局域网中的环路

- 出于容错考虑额外部署的网桥，会产生环路
- 环路问题：环路致使帧单播和广播往复循环
- 解决方案：生成树 (spanning tree) 算法



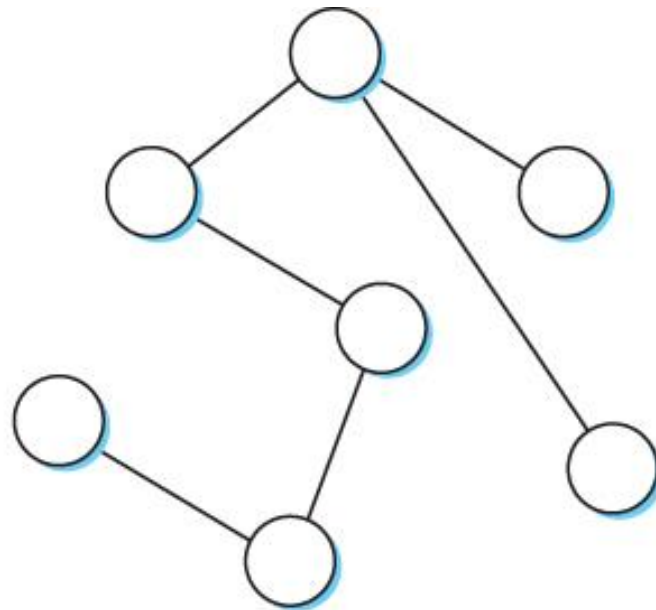
生成树

(a)



(a) 有环图

(b)



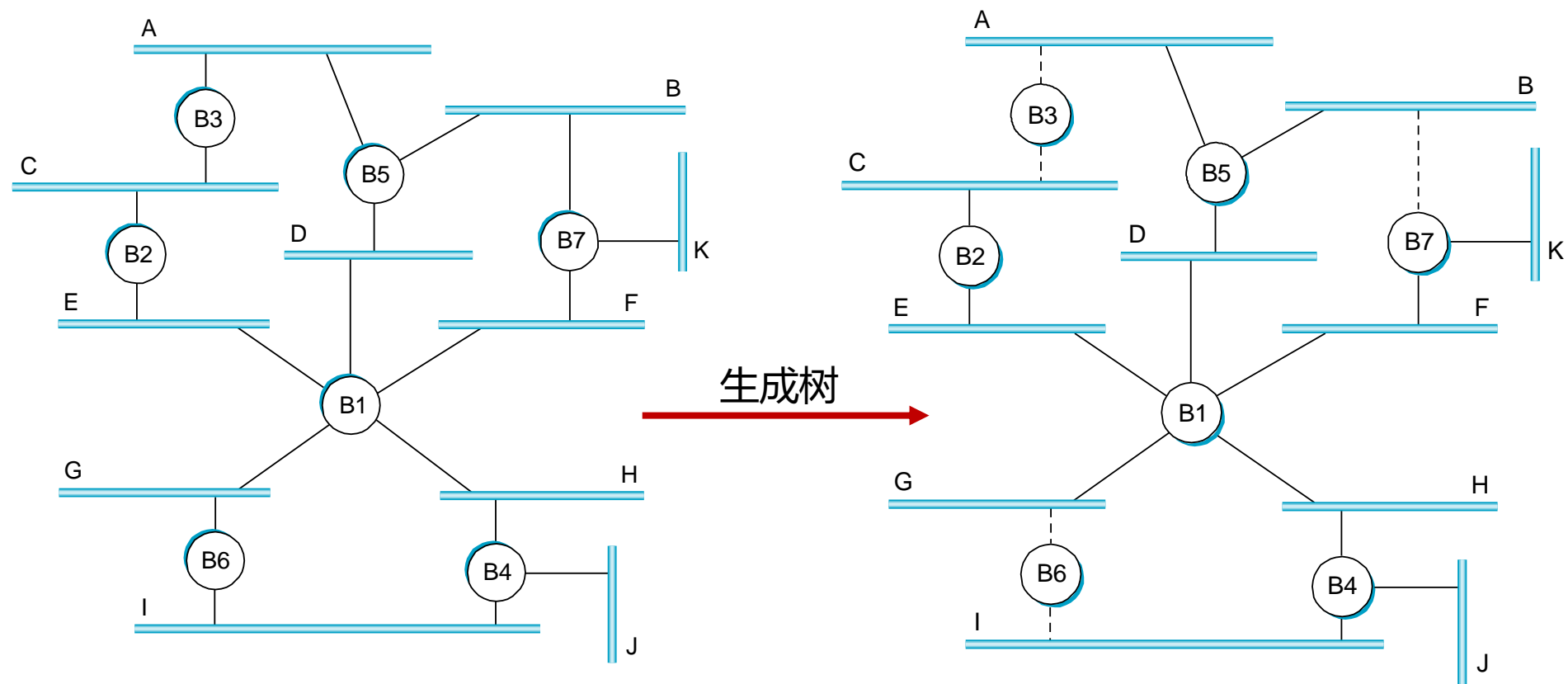
(b) 一棵生成树

- 生成(Spanning) —— 涵盖所有节点
- 树 —— 没有环路，任何两个节点之间只有一条路径

生成树算法

- IEEE 802.1标准的一部分
- 只有**选定网桥(并非全部)**转发帧
 - 这些**选定网桥**形成一棵生成树
 - 生成树上的网桥，只在选定端口转发帧，不一定所有端口都转发
- **构建生成树**
 - 每个网桥有一个**唯一的编号(ID)**，最小编号网桥作为根
 - 对于每个连接的局域网段，**选择距离根网桥最近的网桥作为该网段指定网桥(designated bridge)**
 - 由于指定网桥是与特定局域网段关联，因此一个网桥作为指定网桥的角色，是与端口(链接至不同局域网段)关联
- **帧转发**
 - 根网桥总是向其所有端口转发帧
 - 其它网桥只向其作为指定网桥的端口转发帧

桥接局域网生成树示例

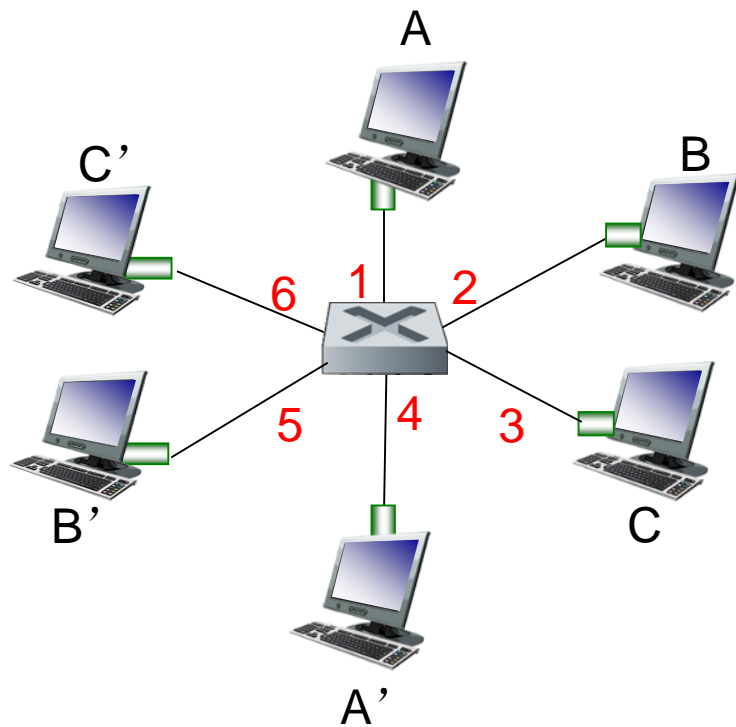


生成树算法细节

- 每个网桥最初都认为自己是根
- 网桥之间交换配置消息
 - 配置消息的源网桥ID
 - 发送该配置消息的网桥认为的根网桥
 - 发送该配置消息的网桥至根网桥的距离(跳数)
- 每个网桥为其每个端口记录当前最佳配置消息
- 每个网桥一旦了解到(通过配置消息交换)自己既不是根、也不是指定网桥, 将停止生成任何配置消息
 - 系统处于稳定状态后, 只有根网桥产生配置消息, 只有指定网桥转发配置消息
- 根网桥持续周期性发送配置消息
- 一个网桥如果在一定时间内没收到任何配置消息, 它将生成和散发声称自己为根的配置消息

以太网交换机

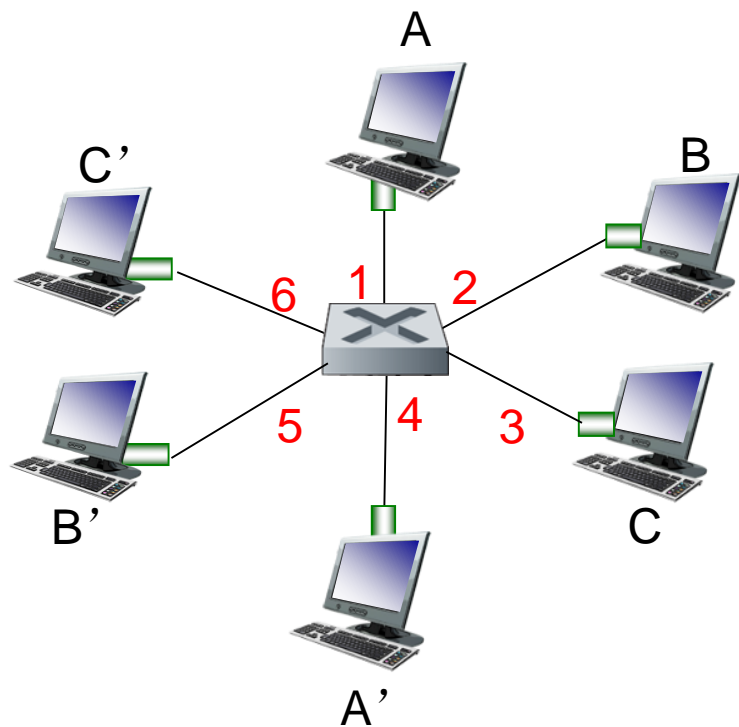
- 每台主机通过一条**专用链路**链接至交换机
 - 所有节点采用以太网协议
- **主机发送帧时没有冲突**
 - 每条链路为一个**独立的冲突域** (collision domain)
 - 链路专用：无其它节点竞争共享
- 交换机存储转发所收到的以太网帧
- **交换**：右图示例中，A—A'和B—B'帧发送同时进行
 - 不同于集线器和共享电缆链接的以太网



6端口交换机

交换表

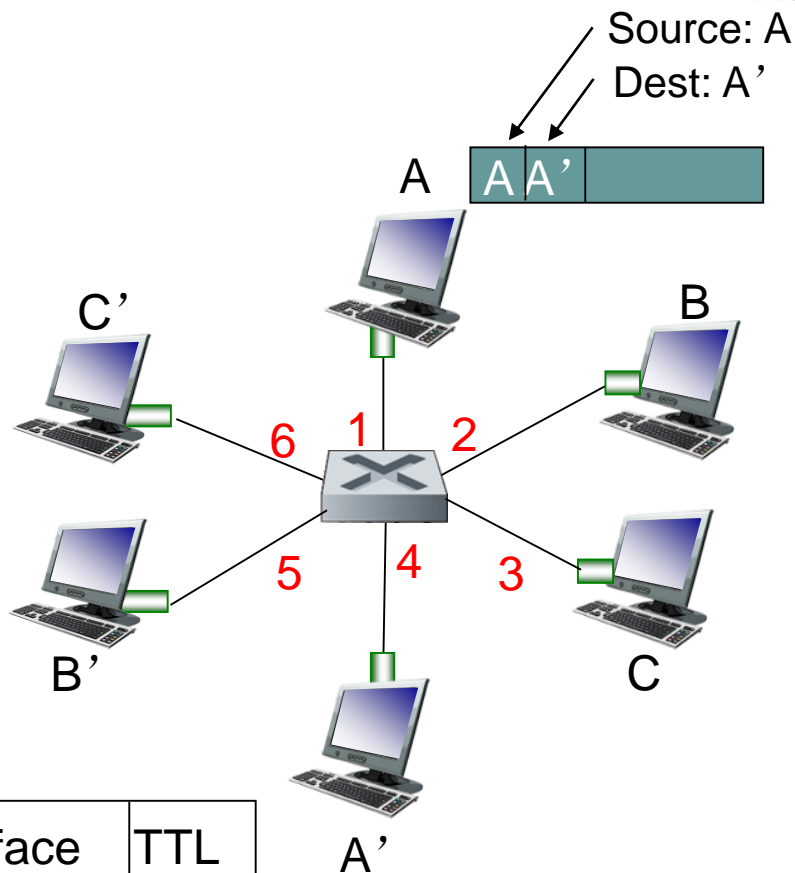
- 右图所示网络中，交换机如何知道A'可通过端口4到达，B'可通过端口5到达？
 - 交换机维护一个交换(转发)表，其中每个条目的字段包括主机MAC地址、可达端口、时间戳(time stamp)
- 交换表中条目的如何建立维护？
 - 与网桥相同 —— 自学习



6端口交换机

自学习

- 交换机学习哪台主机可通过哪个端口到达
 - 收到帧时，交换机学习该帧源主机位置 —— 输入端口
 - 将源主机地址、输入端口记录到交换表条目中



交换表
(初始为空)

MAC addr	interface	TTL
A	1	60

帧转发

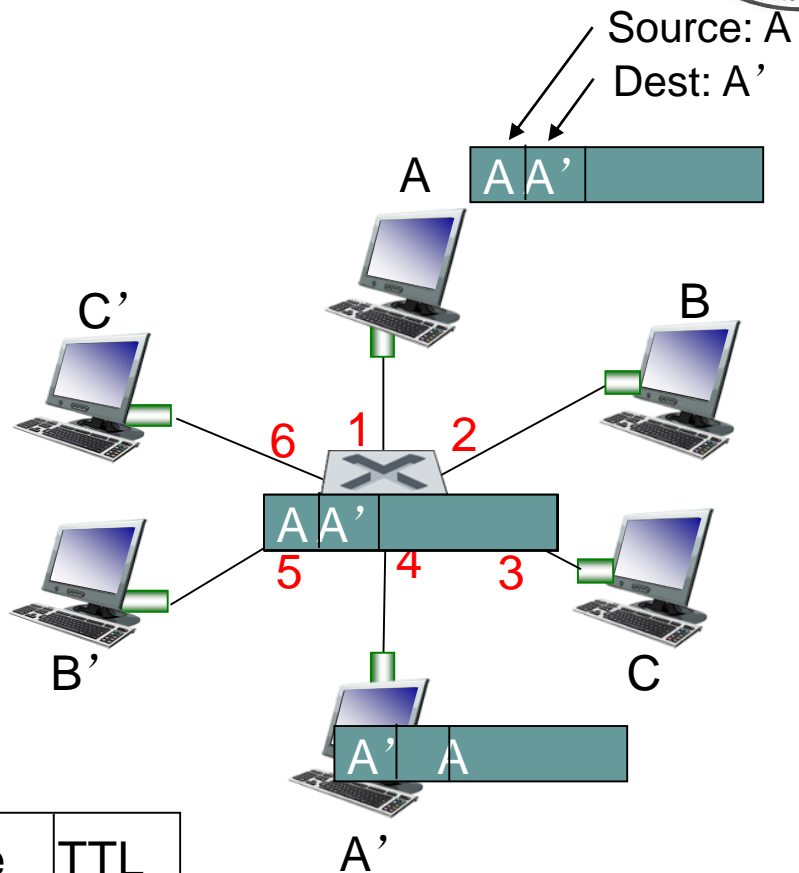
收到帧时:

1. 记录该帧输入端口与其源主机的关联
2. 以该帧目的地址为索引，查询交换表
3. **if** 交换表中存在该帧目的地址对应条目
 then {
 if 该条目输出端口与该帧输入端口相同
 then 丢弃该帧
 else 采用该条目给出的输出端口转发该帧
 }
 else 洪泛该帧

向除了输入端口之外的所有其它端口
转发该帧

自学习和帧转发示例

- 目的地址为A', 位置未知:
洪泛
- 目的地址为A, 位置已知:
通过一个端口选择性发送

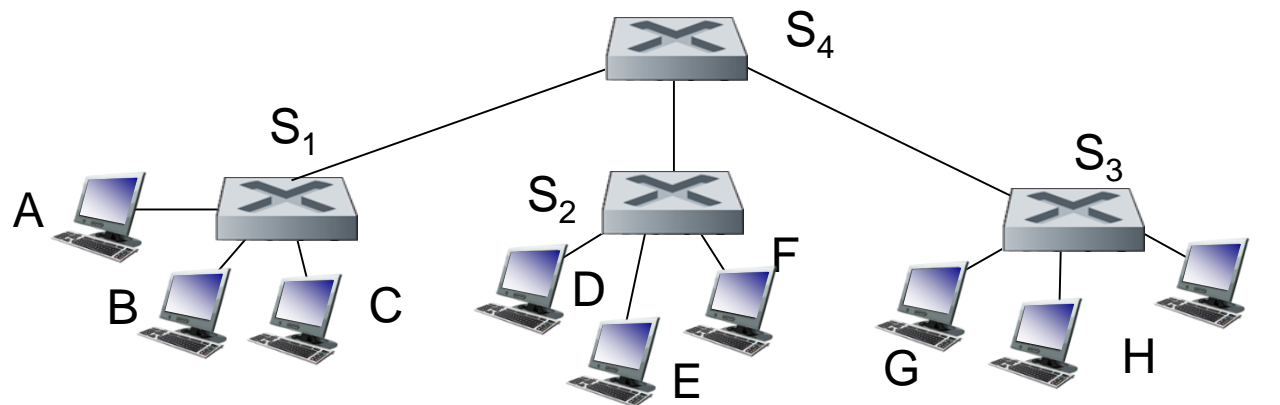


交换表
(初始为空)

MAC addr	interface	TTL
A	1	60
A'	4	60

交换机互连

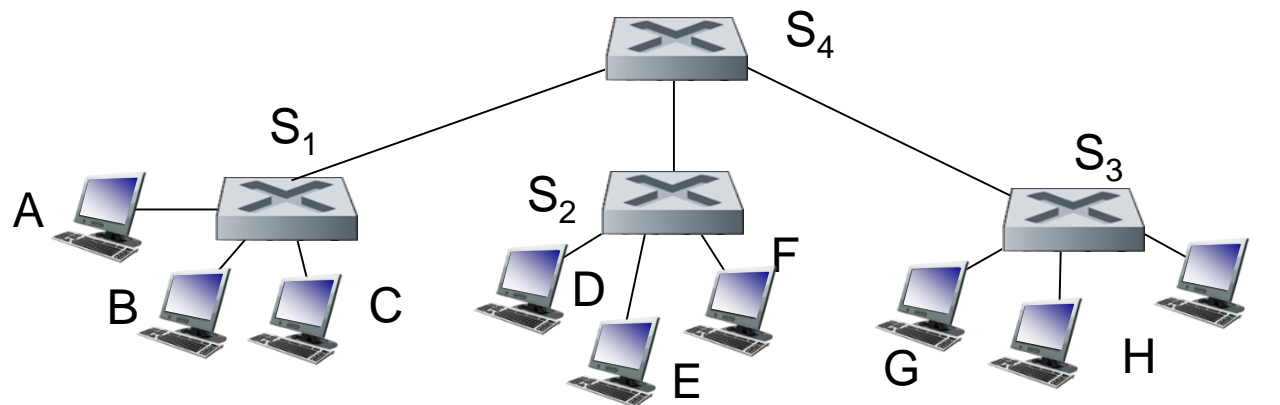
- 交换机可以互相连接，组建更大规模的交换式以太网



Q: 主机A向主机G发送帧，交换机S₁ 如何知道通过S₄、进而通过S₃转发目的地为G的帧？

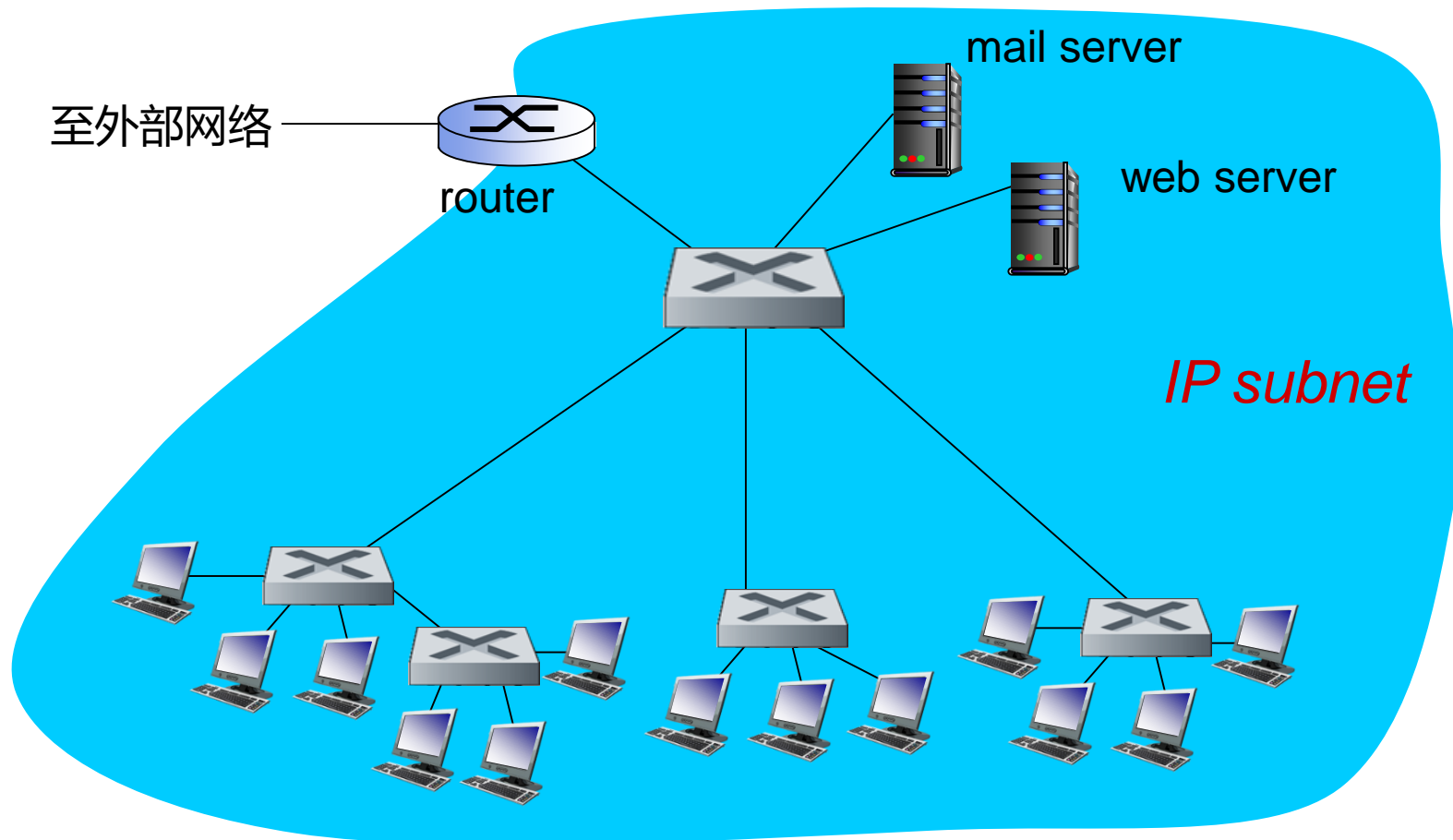
多交换机自学习示例

设C首先向I发送帧，I收到后向C回复帧



Q: 给出交换机 S_1 , S_2 , S_3 , S_4 中帧转发过程和交换表的变化?

机构网络示例



以太网交换机技术特点

- **数据链路层交换节点**：存储转发以太网帧
 - 采用以太网协议算法
 - 根据目的MAC地址，选择性转发
- **透明**
 - 终端主机感觉不到以太网交换机的存在
- **自学习建立交换表**
 - 以太网交换机基本功能无需配置 —— 即插即用(Plug-and-play)
- **相当于多端口以太网网桥**

小结

- 分组交换技术：数据报交换，虚电路交换
 - 工作原理，相似性，不同点
- 数据链路层交换设备：网桥，以太网交换机
 - 工作原理，两类设备之间的联系
- 参考文献
 - 教材3.1节
 - [KR12] 5.4节