CH4 : Network Layer: The Data Plane

Outline

- data plane, control plane
- two control-plane approaches
- 分布式交换
- HOL
- Per-router control plane
- CAs（填空）
- Switching fabrics
- Output ports
- Scheduling mechanisms
- Scheduling policies
- IP datagram format
- IP fragmentation, reassembly
- Subnets
- CIDR
- DHCP
- DHCP client-server scenario
- NAT

- IPv6（32-bit address space了解这个即可）

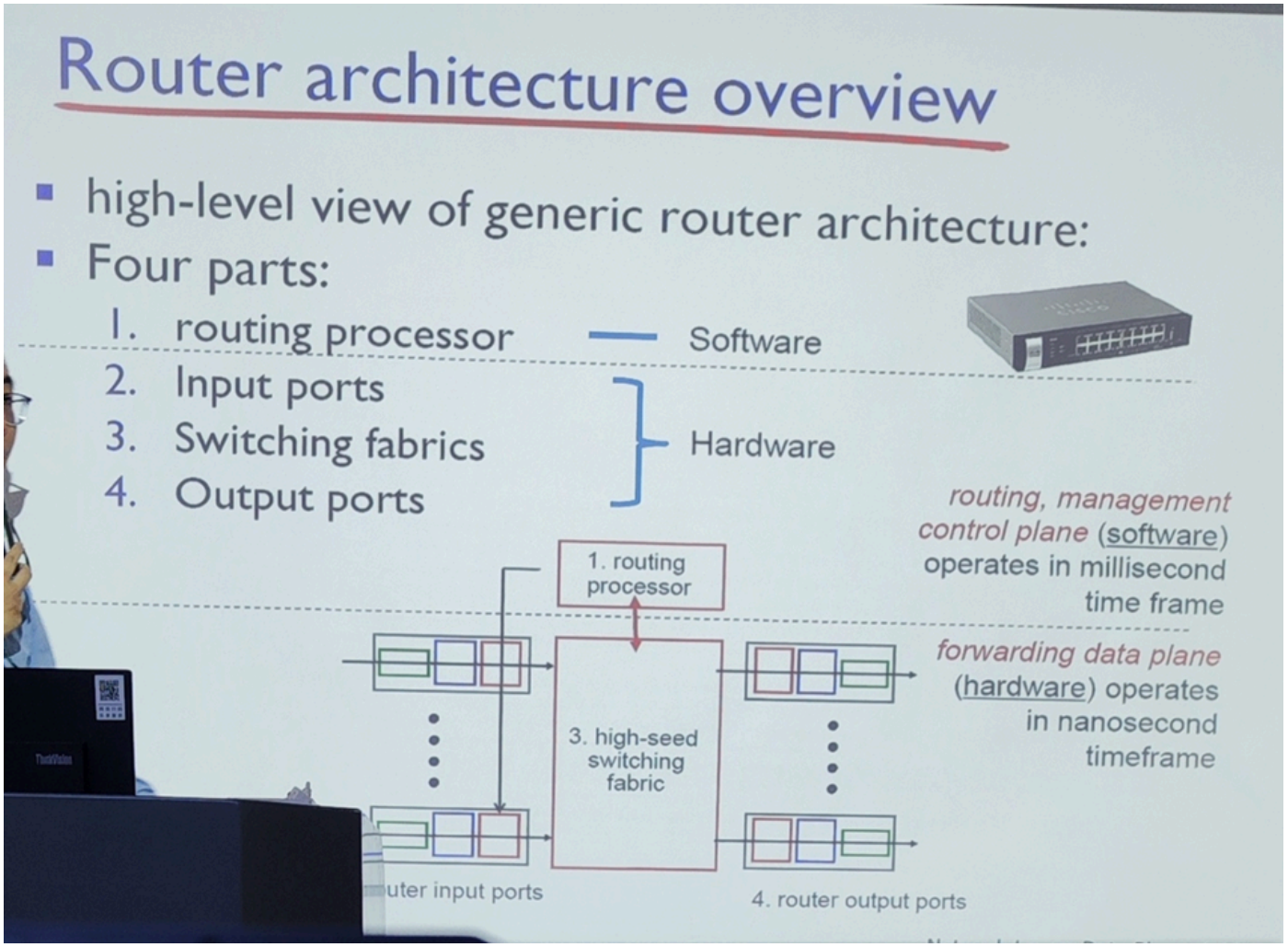
two control-plane approaches:

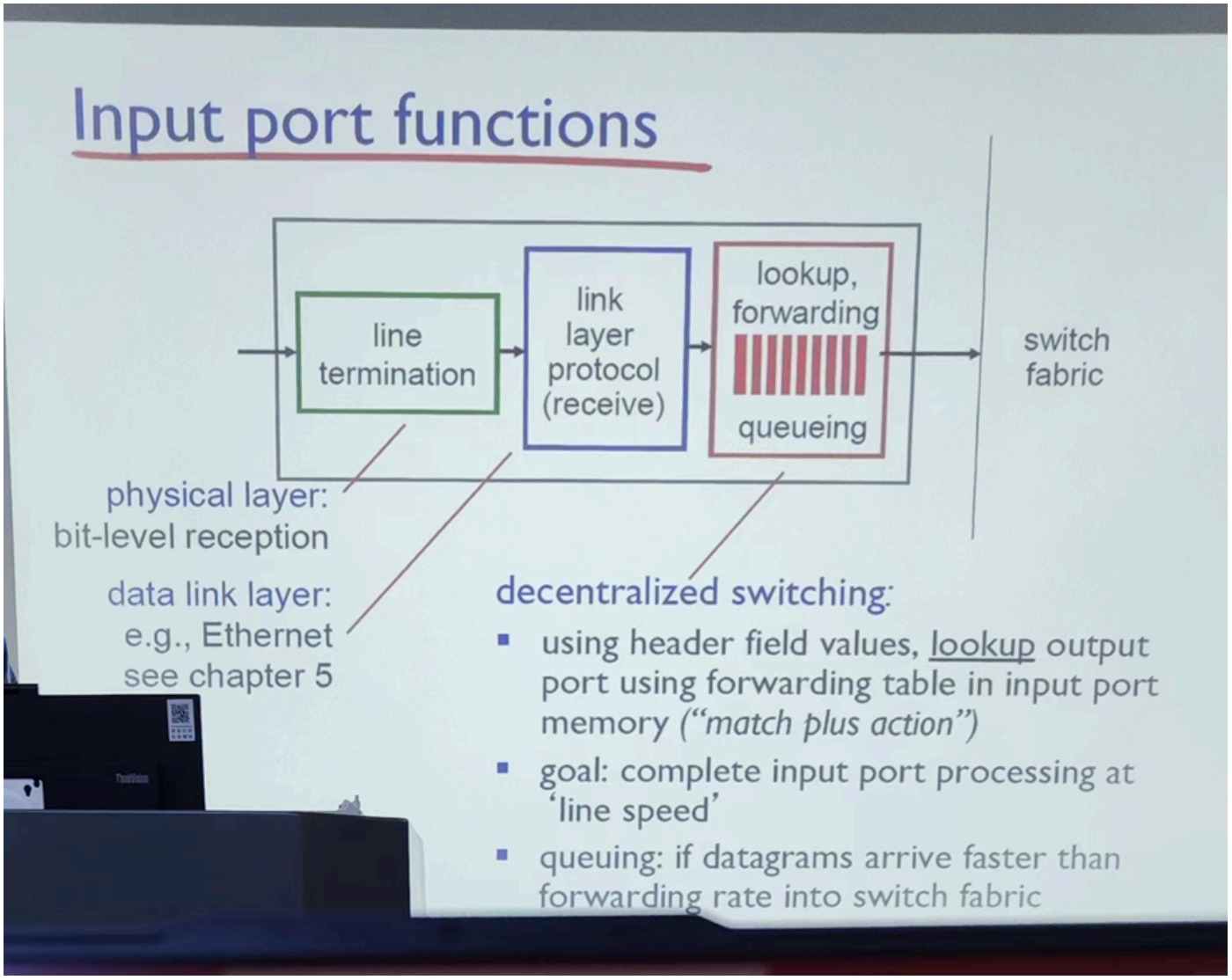
- traditional routing algorithms: implemented in routers
- software-defined networking (SDN): implemented in (remote) servers

Network layer: data plane

在计算机网络中，数据平面（Data Plane）是指处理网络数据包的部分，它负责在网络设备（如路由器或交换机）上执行数据包的转发操作。数据平面决定了当数据报文到达路由器的输入端口时，如何将其转发到路由器的输出端口。

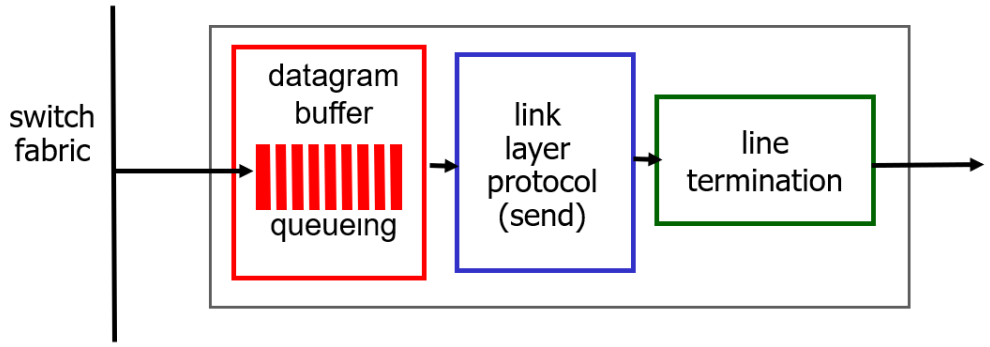
Router architecture overview





- 分散式交换（Decentralized Switching）是一种网络交换技术，它使用输入端口内存中的转发表来根据数据包的头部字段值查找输出端口（"match plus action"）。这种技术的目标是以线速完成输入端口的处理。在分散式交换中，每个交换机的输入端口都维护一个转发表，该表存储了与输入端口关联的输出端口的映射。当数据包到达输入端口时，交换机会根据数据包的头部字段值（如目的地址、虚拟局域网标识符等）在转发表中进行查找。一旦找到匹配项，交换机会根据转发表中的指示，将数据包转发到相应的输出端口。
- Head-of-the-Line (HOL) blocking是一种现象，指的是队列中位于队列前端的数据报文阻塞了后续数据报文的转发，导致后续数据无法向前移动。当数据报文到达交换机或路由器的输出端口时，它们会进入相应的输出队列等待转发。如果队列中的第一个数据报文（位于队列的前端）无法立即转发，例如由于输出链路忙于传输其他数据报文或正在进行某种处理操作，那么后续的数据报文将被阻塞在队列中，无法继续向前移动。
- A distinct (typically remote) controller interacts with local control agents (CAs) 用Cas实现交互[填空判断]（很赚钱）

Output port 🚀

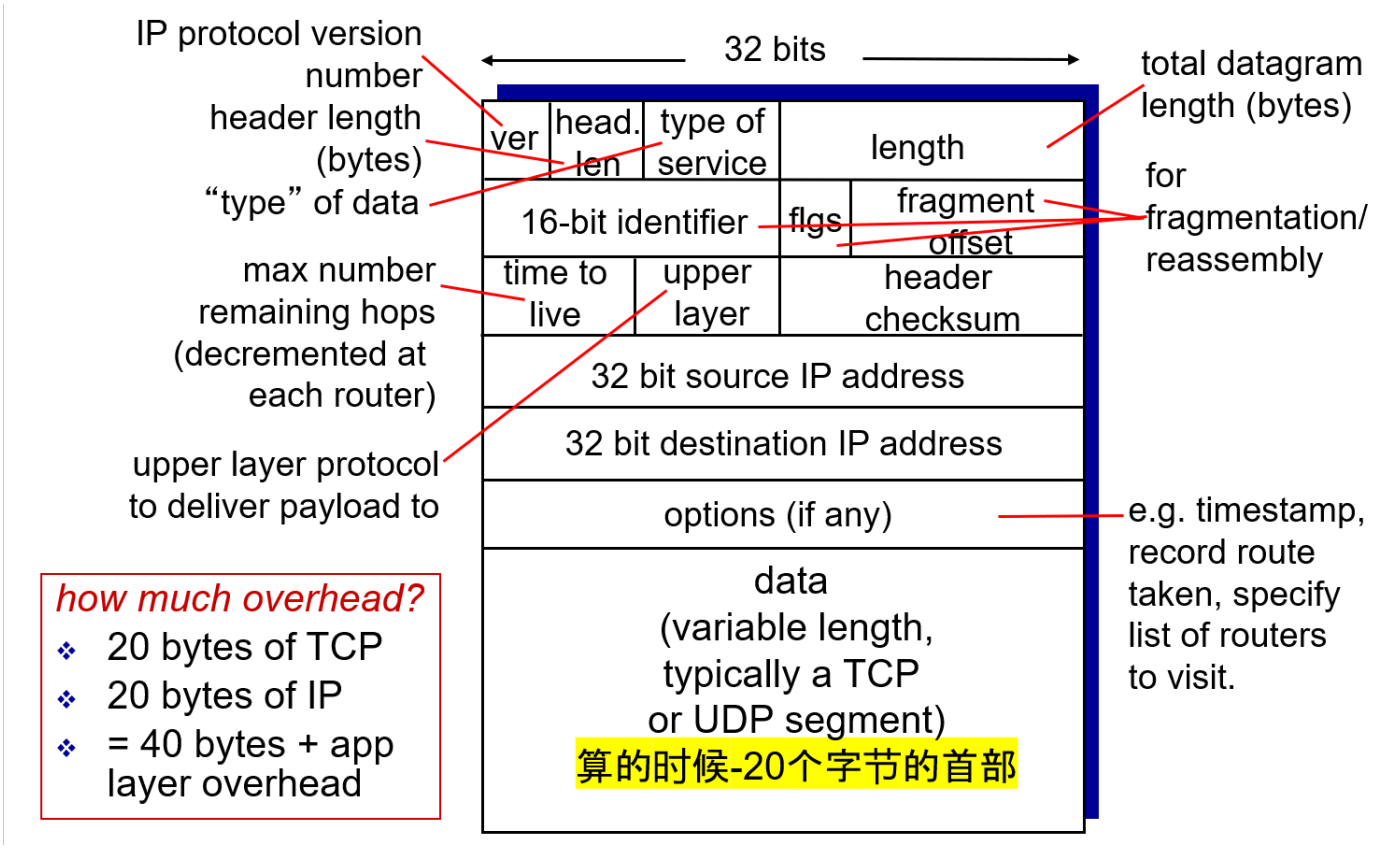


- buffering required when datagrams arrive from fabric faster than the transmission rate
    - Datagram (packets) can be lost due to congestion, lack of buffers
- scheduling discipline chooses among queued datagrams for transmission

Scheduling mechanisms

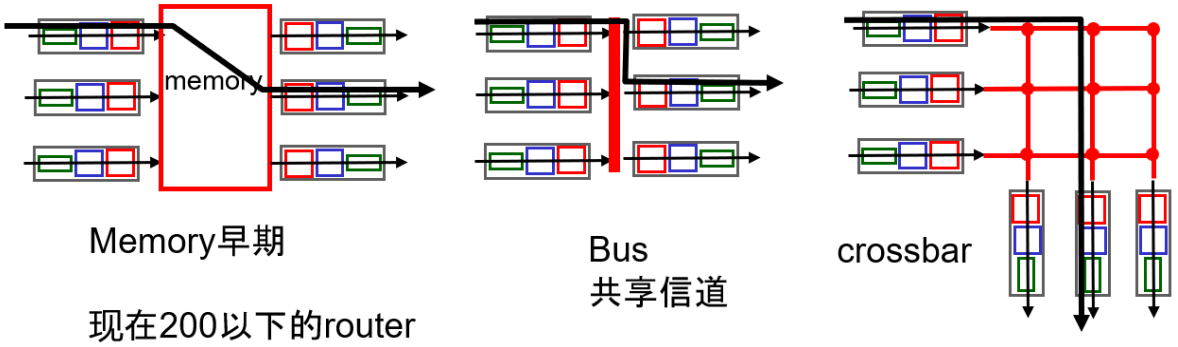
- FIFO/ FCFS (first in first out) scheduling: send in order of arrival to queue
- priority scheduling基于优先级的调度
- Round Robin (RR) scheduling:个排个的，但是一次只发一个，每次按顺序发
- Weighted Fair Queuing (WFQ):each class gets weighted amount of service in each cycle

IP datagram format



Switching fabrics

transfer packet from input buffer to appropriate output buffer



IP fragmentation, reassembly 🚀

在网络中，网络链路具有最大传输单元（MTU），即链路级帧的最大可能大小。不同类型的链路具有不同的MTU。

当一个大的IP数据报在网络中传输时，它可能会被分割（"fragmented"）成多个较小的数据报。这意味着一个IP数据报会变成多个数据报。

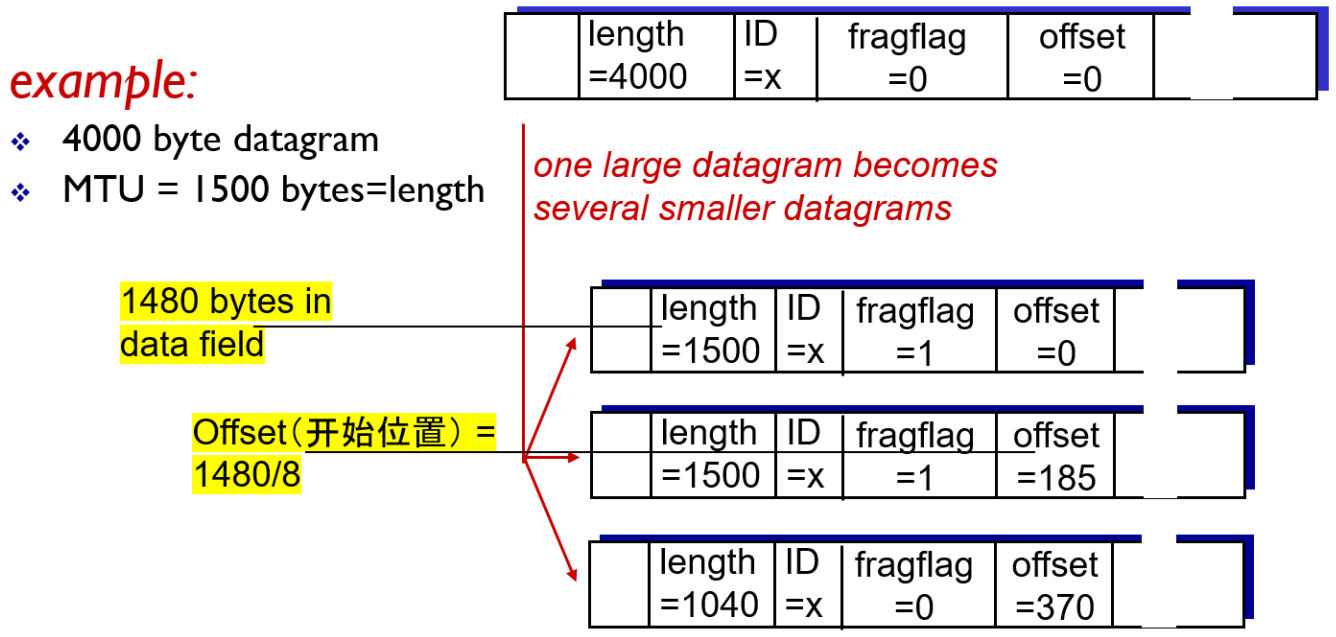
这种分割是由于不同链路的MTU限制。如果IP数据报的大小超过了链路的MTU，那么它需要被分割成更小的片段，以适应链路的传输要求。

在网络中，这些分割后的数据报会独立传输，并在最终目的地进行"reassembled"（重新组装）。这意味着在最终目的地，接收方将接收到的分割数据报重新组装成原始的大IP数据报。

为了正确地重新组装这些数据报，IP头部中的一些标识位被使用来识别和排序相关的片段。这些标识位包括标识（Identification）、片偏移（Fragment Offset）和更多片段（More Fragments）等字段。

通过使用这些标识位，目的地可以正确地识别和重新组装原始的IP数据报，确保数据的完整性和正确性。

需要注意的是，IP分片和重新组装是在网络层（IP层）进行的操作，而不是在传输层（如TCP或UDP）进行的。这意味着传输层协议对于IP分片和重新组装是透明的，它们只在网络层进行处理。



- 图片中的 "flagflag" 是指 IP 数据包的 "分片标志" (Fragment Flag)。该标志位于 IP 数据包的头部，用于指示该数据包是否已被分片。如果分片标志位为 0，则表示该数据包未被分片；如果分片标志位为 1，则表示该数据包已被分片。
- 偏移量(Offset)以 8 字节为单位, 所以 /8

Subnets子网

三类ip地址：

- A 类地址：1.0.0.0 到 126.255.255.255

- B 类地址：128.0.0.0 到 191.255.255.255
- C 类地址：192.0.0.0 到 223.255.255.255

广播链路。这里出现了几个子网呢？3 个子网 223. 1. 1. 0/24、223. 1. 2. 0/24 和 223. 1. 3. 0/24 类似于我们在图 4-18 中遇到的子网。但注意到在本例中还有其他 3 个子网：一个子网是 223. 1. 9. 0/24，用于连接路由器 R1 与 R2 的接口；另外一个子网是 223. 1. 8. 0/24，用于连接路由器 R2 与 R3 的接口；第三个子网是 223. 1. 7. 0/24，用于连接路由器 R3 与 R1 的接口。对于一个路由器和主机的通用互联系统，我们能够使用下列有效方法定义系统中的子网：

为了确定子网，分开主机和路由器的每个接口，产生几个隔离的网络岛，使用接口端接这些隔离的网络的端点。这些隔离的网络中的每一个都叫作一个**子网**（subnet）。

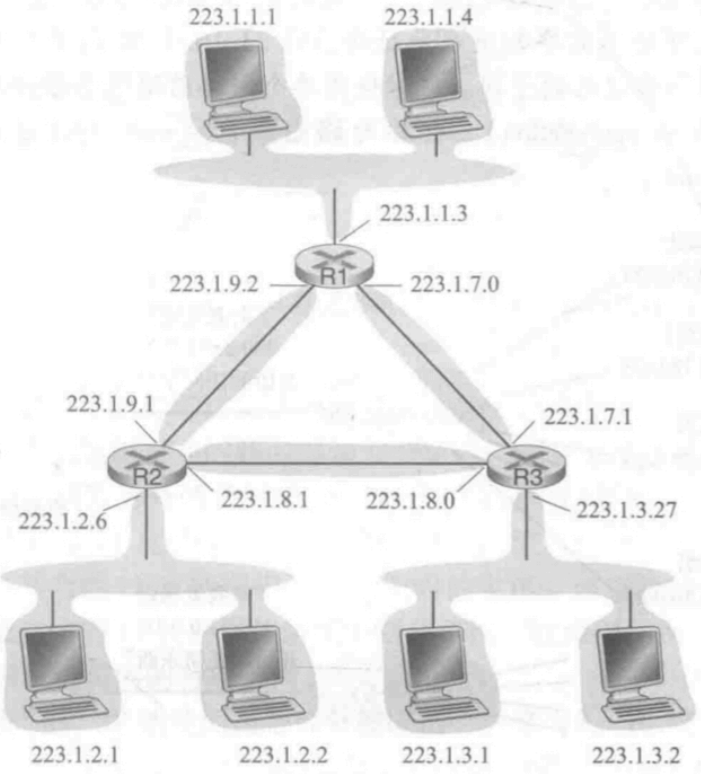
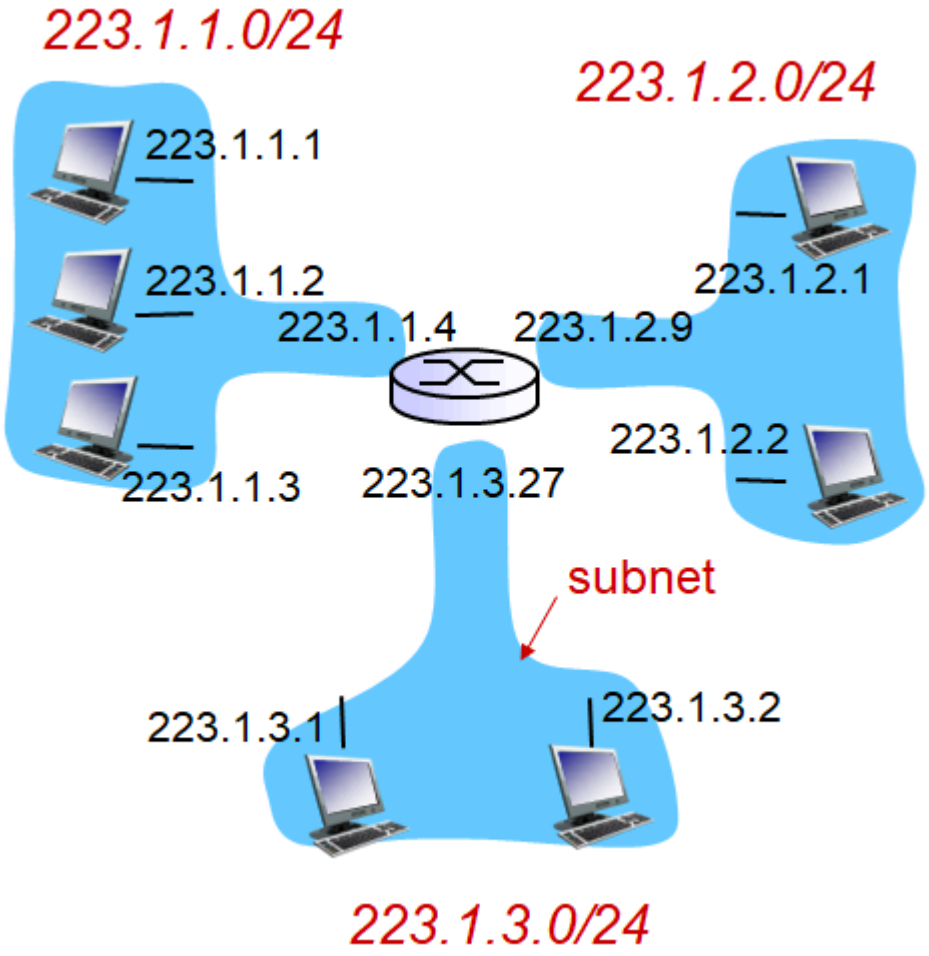


图 4-20　3 台路由器互联 6 个子网

如果我们将该过程用于图 4-20 中的互联系统上，会得到 6 个岛或子网。



CIDR 无类别域间路由

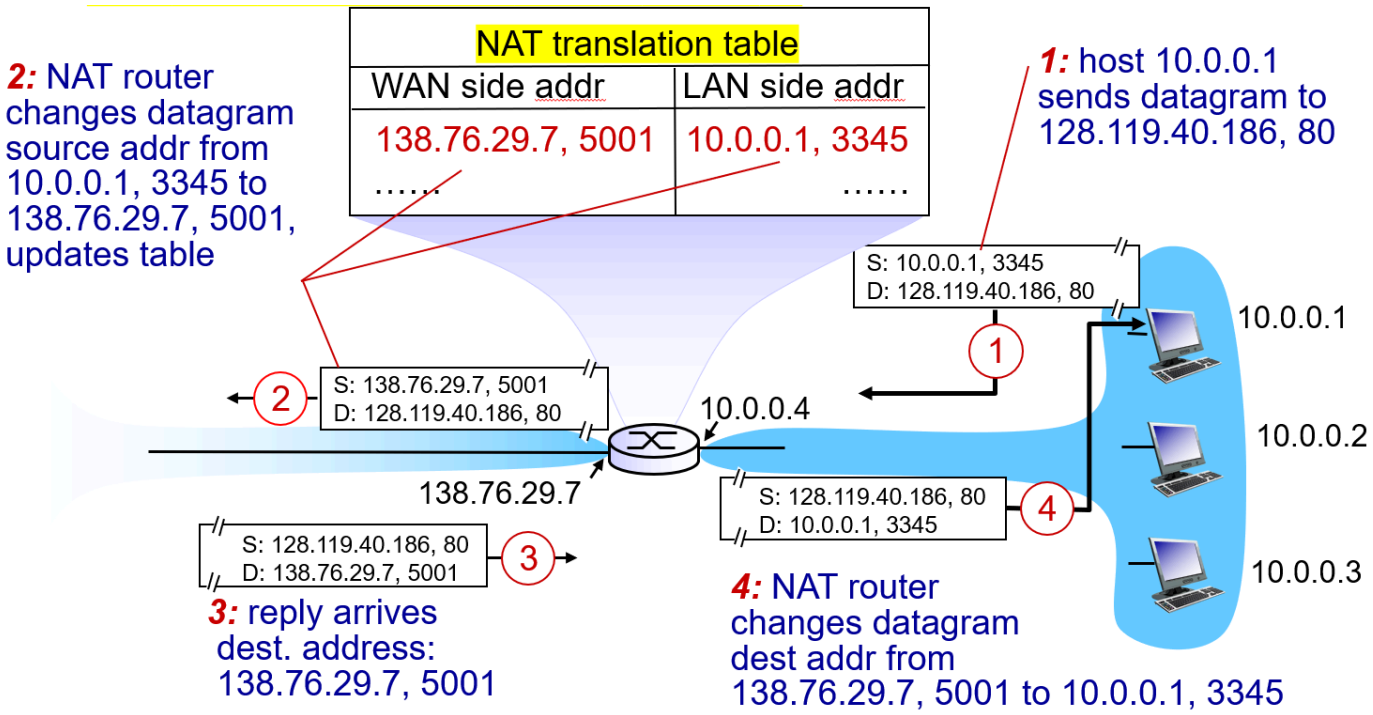
- 使用可变长度的子网掩码划分ip地址
- CIDR中的IP地址格式为a.b.c.d/x，其中a、b、cd是IP地址的四个部分，x表示子网部分的位数 (也称为子网前缀长度)。子网前缀长度指示了IP地址中用于网络部分的位数。
- 通过CIDR表示法，可以对IP地址进行更灵活的划分。例如，对于IP地址192.168.0.0，可以使用CIDR表示为192.168.0.0/24，其中/24表示子网前缀长度为24位。这意味着前24位用于网络部分，剩余8位用于主机部分。

DHCP:Dynamic Host Configuration Protocol

- DHCP（Dynamic Host Configuration Protocol）是一种网络协议，用于自动分配IP地址和其他网络配置参数给计算机和设备。
- 过程：
  - 主机广播 "DHCP Discover" 消息（可选）：当主机启动或连接到网络时，它可以选择发送一个广播消息，称为 "DHCP Discover"。这个消息用于通知网络中的DHCP服务器主机的存在，并请求IP地址和其他网络配置。

- DHCP服务器响应 "DHCP Offer" 消息（可选）：当DHCP服务器收到 "DHCP Discover" 消息后，它可以选择回复一个 "DHCP Offer" 消息。这个消息包含可用的IP地址和其他网络配置参数，供主机选择。
- 主机发送 "DHCP Request" 消息请求IP地址：主机从收到的 "DHCP Offer" 消息中选择一个提供的IP地址，并发送一个 "DHCP Request" 消息，请求该DHCP服务器提供所选的IP地址和其他配置。
- DHCP服务器发送 "DHCP Ack" 消息分配地址：如果DHCP服务器接收到主机的请求，并且所请求的IP地址和配置参数仍然可用，它会回复一个 "DHCP Ack" 消息，确认分配的IP地址和其他配置。这表示主机已成功获得所需的网络配置，并且可以使用该IP地址和其他参数进行通信。

NAT: network address translation 网络地址翻译



IPv6

- 32 bit