



网络层：IP组播、 移动主机路由及自组织网络

刘志敏

liuzm@pku.edu.cn



再议路由问题

- 路由与转发：

- 路由器执行路由协议与算法、形成路由表及转发表
 - 最小距离路由、距离矢量路由
- 路由器接收分组、查表选择转发端口

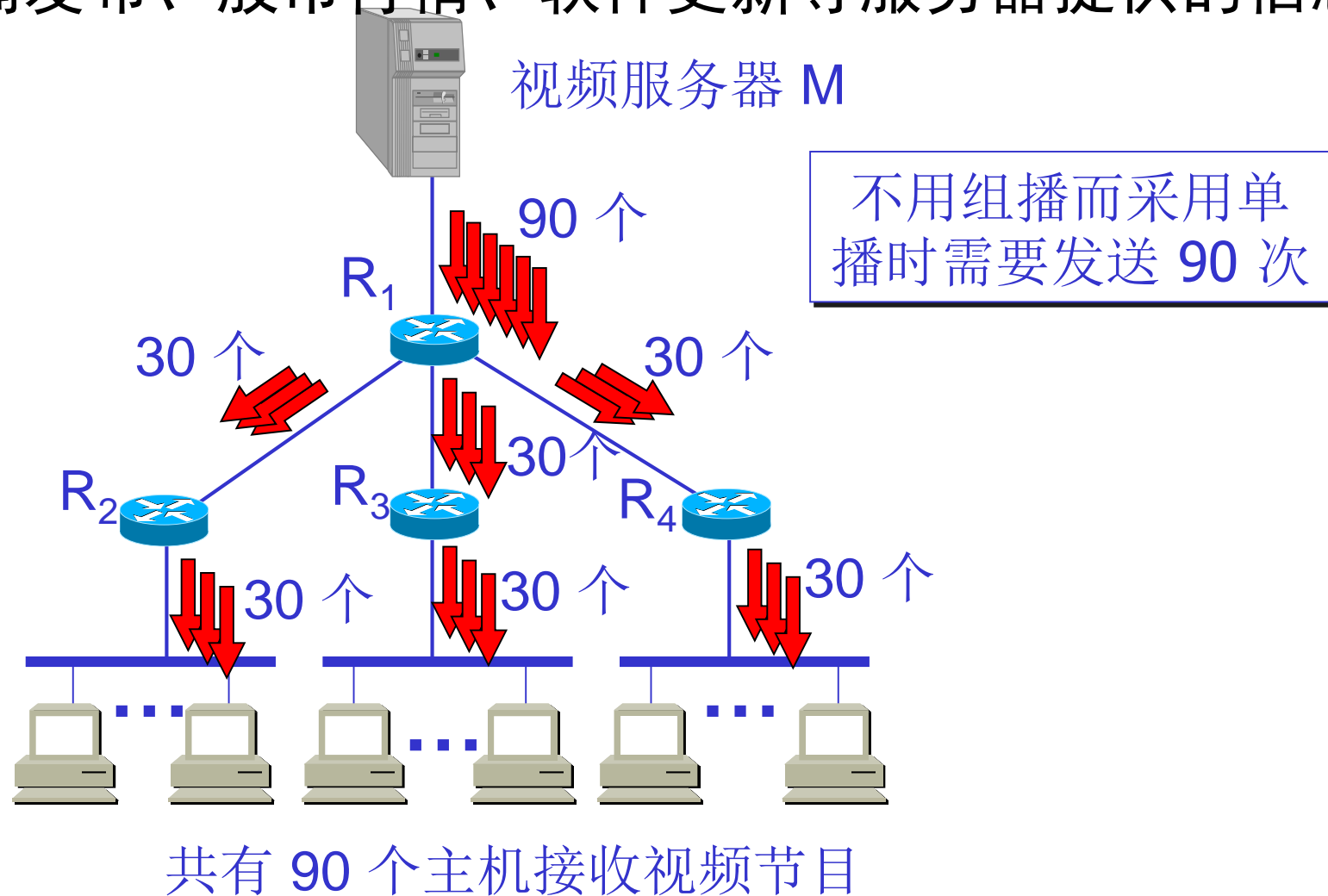
- IP组播及路由

- 移动主机路由

- 自组织网络路由

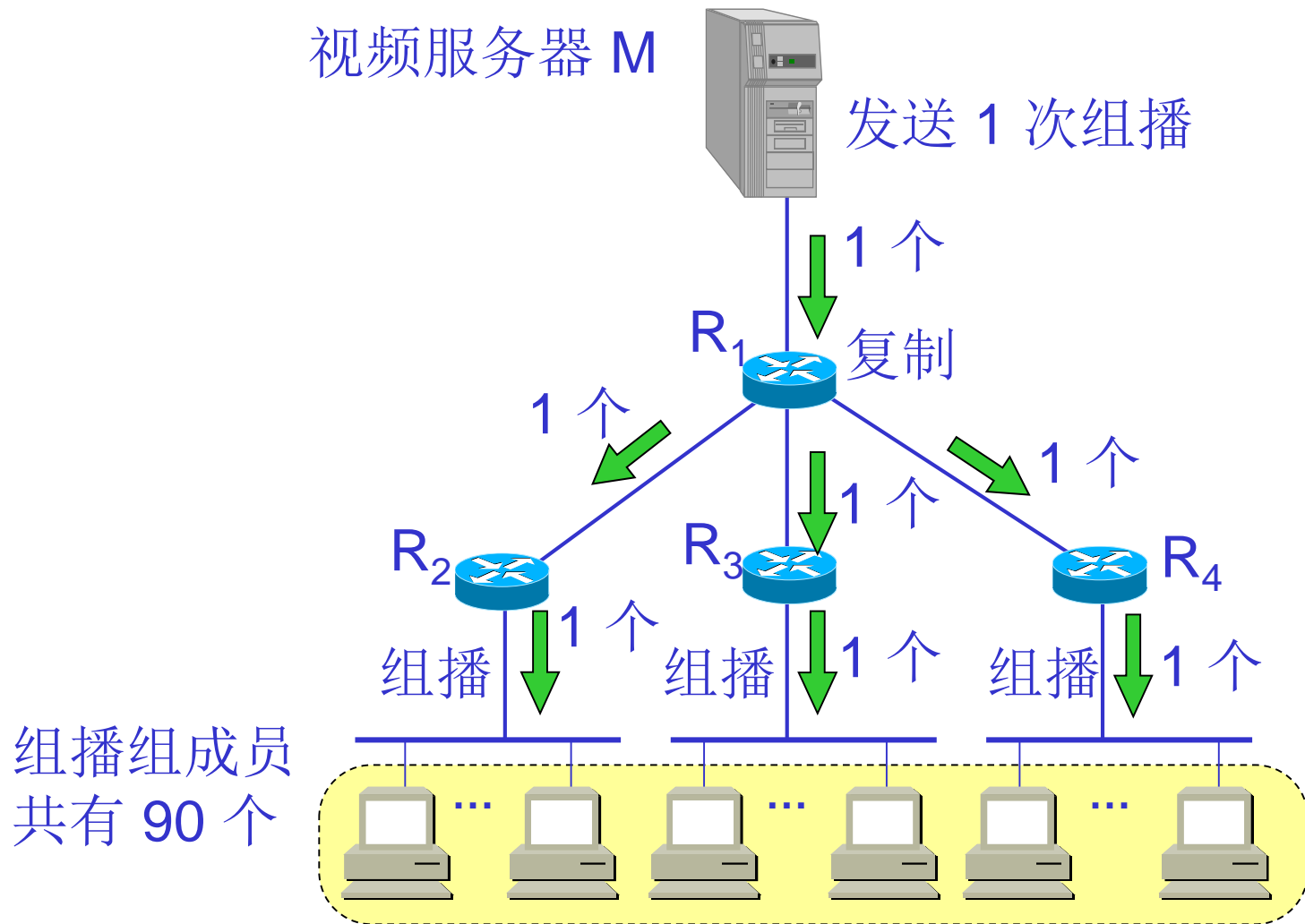
IP 组播 (Multicast)

- 一点到多点的数据交付，如路由器之间的交换路由信息，以及体育赛事、新闻发布、股市行情、软件更新等服务器提供的信息服务



IP 组播的概念

- 组播方式减少了网络中带宽资源的消耗



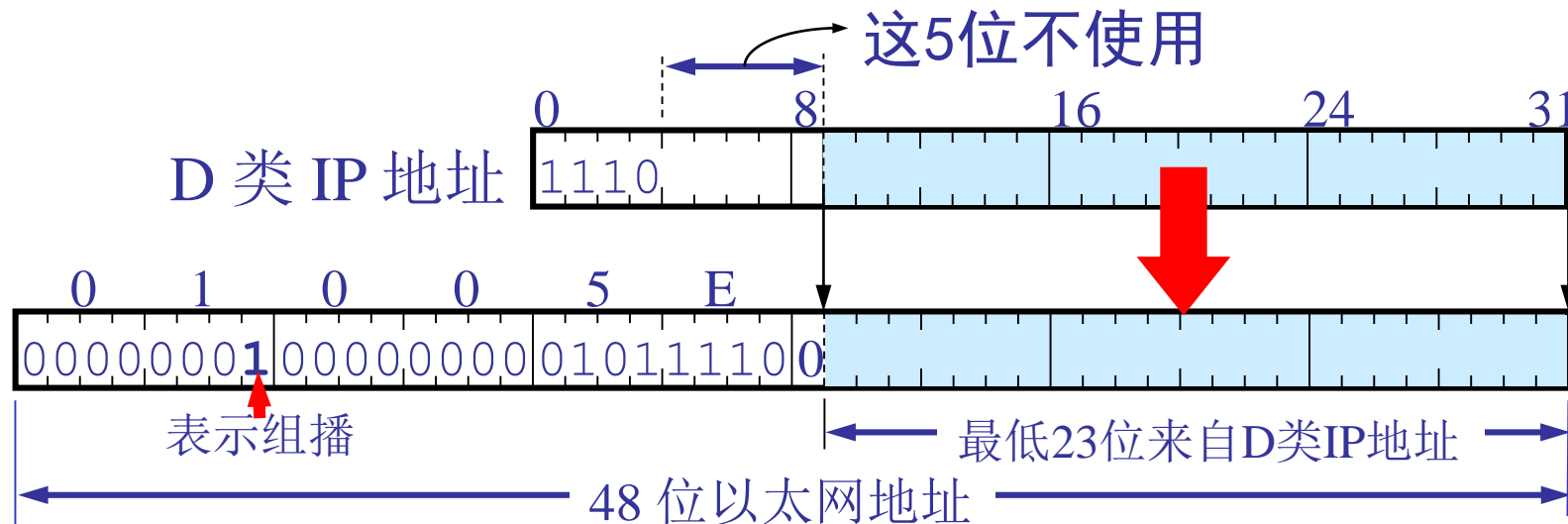


IP 组播的特点

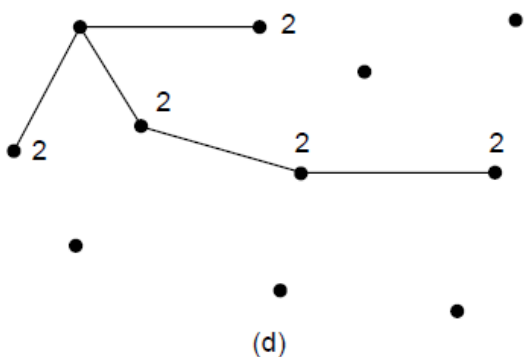
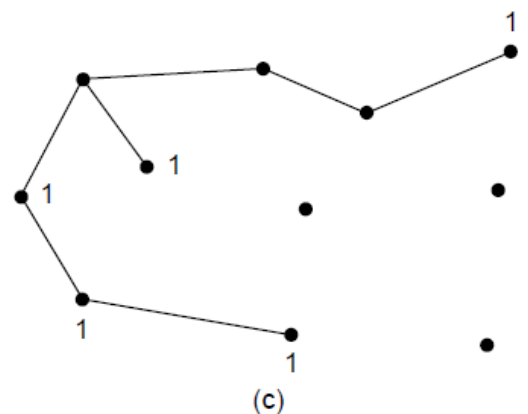
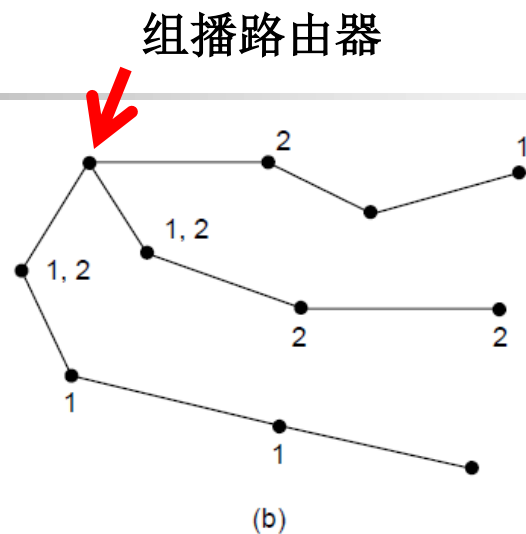
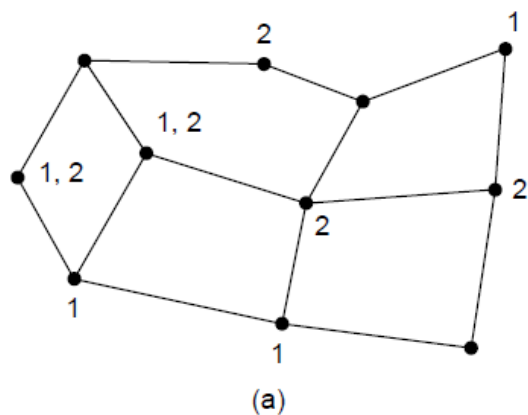
- (1) 组播使用组播地址：用IP的D类地址支持组播。 D类播地址只能用于目的地址表示业务而不能用于源地址。
- (2) 永久组地址：由互联网数字分配机构 IANA (Internet Assigned Numbers Authority) 分配。
- (3) 组播组的成员是动态的
- (4) 使用以太网的硬件进行组播

在局域网上进行硬件组播

- 以太网硬件组播地址：MAC地址的首字节的最低位为1
- 先由硬件接收组播帧，之后再由软件对IP组播地址进行过滤
- 例如：01:80:C2:00:00:00是802.1d网桥多播地址，网桥之间用该地址交换配置信息，之后采用算法生成分布式生成树以消除网络中的环路
- 01-00-5E-00-00-00~01-00-5E-7F-FF-FF为与IP组播地址对应的MAC组播地址，其中最右23位来自D类地址，例如：
 - OSPF IP地址是224.0.0.5，MAC组播地址为01:00:5e:00:00:05
 - RIP 组播地址为224.0.0.9，MAC组播地址为01:00:5e:00:00:09



组播路由及其问题



(a) 网络拓扑图：有2个组播组 (b) 组播路由的一种生成树

(c) 组播组1的生成树 (d) 组播组2的生成树

组播组不同则组播树也不同，每个路由器都需要为每个组播业务维护一个组播树

组播路由器如何获得组播成员信息？



IGMP和组播路由选择协议

- 实现组播，路由器需要获得组播组成员的位置信息，并建立由组播源到接收组播消息的多个信宿的最小代价路由
- IP组播需要两种协议：
 - **网际组管理协议IGMP**(Internet Group Management Protocol):
使路由器获得组播组的成员信息
 - 一个组播组用一个D类地址标识
 - **组播路由选择协议**：使组播路由器之间协同工作，用最小代价将组播数据报传送给所有的组成员



采用IGMP维护组播组成员信息

■ (1) 主机加入或离开组播组：

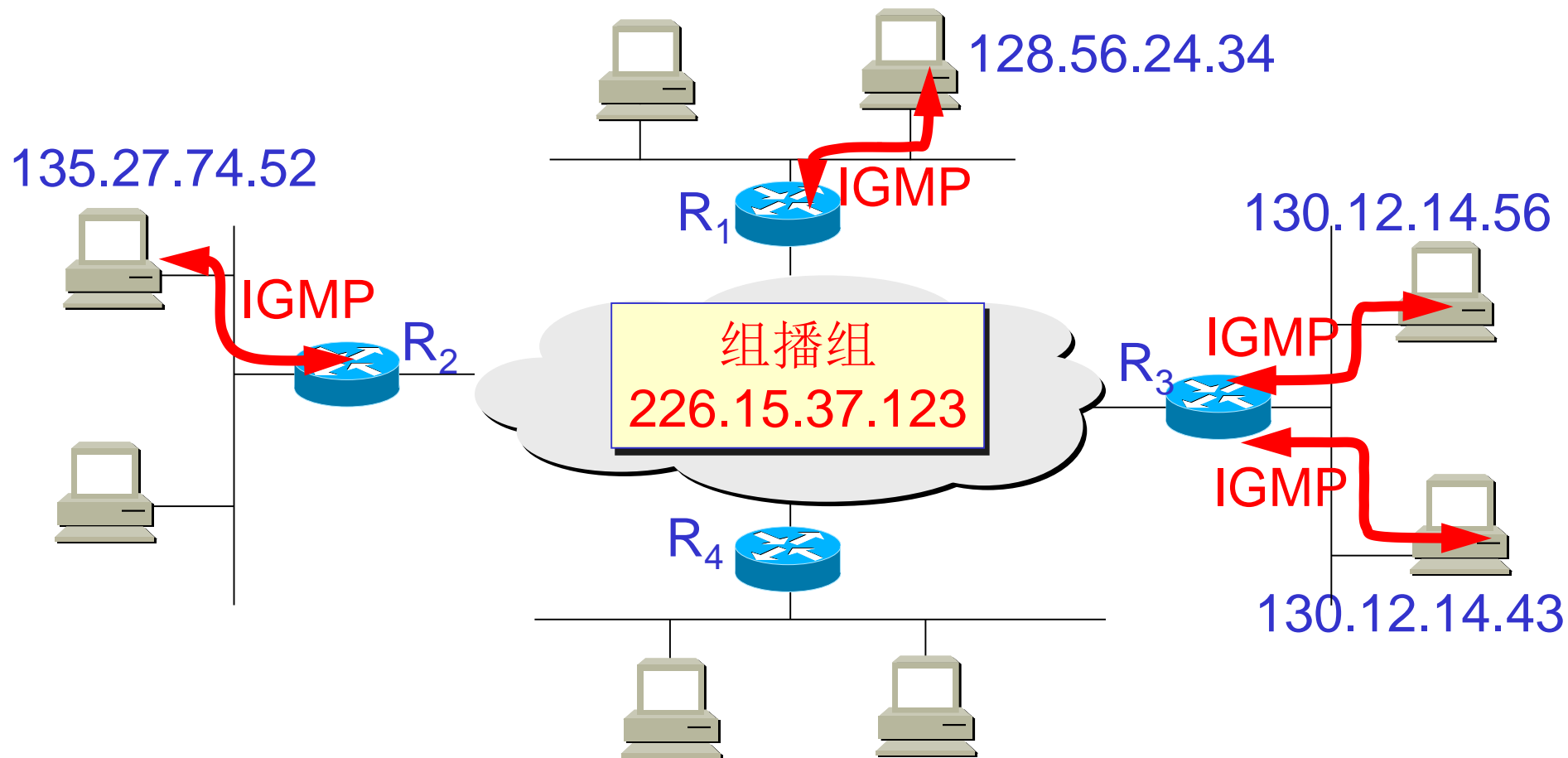
- 当主机加入组播组时，向组播组对应的某一D类地址发送IGMP成员报告报文，本地组播路由器收到IGMP报文后，将组成员消息转发给其他组播路由器。
- 当主机离开组播组时，发送IGMP成员离开报文

■ (2) 组播路由器维护组成员信息：路由器周期性地发送组探测报文，因为组成员是动态的

- 在某个子网上只要有一个主机响应，组播路由器就认为该组是活跃的；
- 若在某个子网上经过几次探测后仍然没有一个主机响应，则路由器不再向其他组播路由器转发成员信息

IGMP使组播路由器获得组播成员信息

- 三类报文：（主机发送的）报告和离开、（路由器发送的）探寻
- 主机如何获得某一组播业务的D类地址？访问该组播服务器Web获得D类地址





IGMP 的具体措施

- 主机和组播路由器间的通信都使用IP组播，IP的目的地址为D类地址
- 组播路由器周期性地发送目的地址为组播地址的询问报文，以探询组成员关系，默认间隔为125s
- 当同一个子网上有几个组播路由器时，可选择其中一个发送询问报文
- 在IGMP询问报文中有一个数值 N ，指明最长响应时间为 $N \times 0.1s$ ；
- 主机随机选择【0~ N 】时间发送成员报告
- 组内每个主机都监听成员报告，只要检测到本组其他主机已经发送了成员报告，则不再发送成员报告

IGMP: 分组格式及目的地址

IGMPv2 packet structure^[8]

+	Bits 0–7	8–15	16–31
0	Type	Max Resp Time	Checksum
32	Group Address		

IGMPv2 destination address^[9]

Message Type	Multicast Address
General Query	All hosts (224.0.0.1)
Group-Specific Query	The group being queried
Membership Report	The group being reported
Leave Group	All routers (224.0.0.2)

来自<https://en.wikipedia.org/wiki/>

类型字段:

成员查询: 0x11,

成员报告: 0x12 (IGMPv1), 0x16 (IGMPv2), 0x22 (IGMPv3)

离开组播组: 0x17

最长响应时间: 响应报告的时间限制, 单位为100毫秒, 该值仅在成员查询报文中有意义

组播地址:

当发送给一个组播组, 或者组播和源指定的查寻时, 指定查询的组播组地址

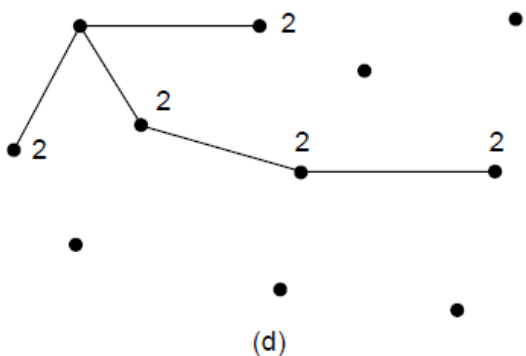
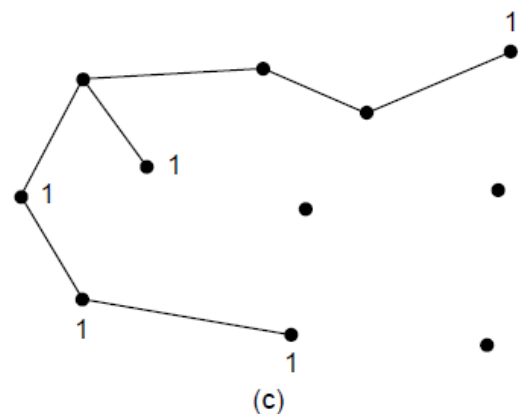
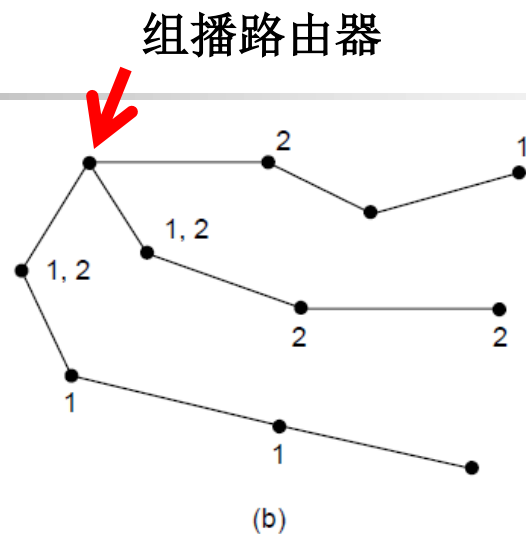
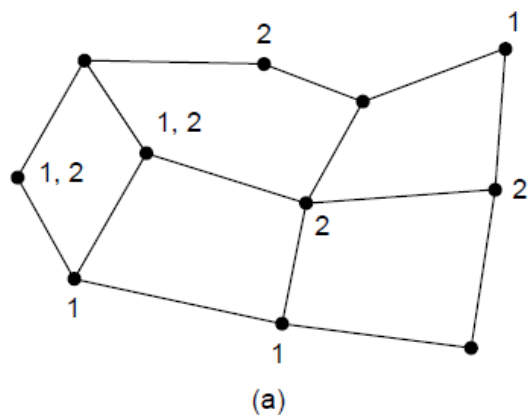
当发送一般探寻时, 地址为0, 该信息紧随IP地址发送



组播路由选择

- 组播组中的成员是动态变化的
- 组播路由选择是要找出以组播源主机为根的组播转发树
- 不同组播组对应于不同的组播树。
- 同一个组播组，对不同的源主机也会有不同的组播树
- 如何转发组播数据报？有以下两种方法
 - 基于生成树：泛洪+剪枝（修剪广播生成树）；每次组播，则建立以组播源为根的生成树
 - 类似移动广播站，总是以组播源主机作为根
 - 基于核心树：以核心为根的生成树，核心相对固定
 - 类似于电台、电视台，根的位置是相对固定的

组播路由及其问题



(a) 网络拓扑图：有2个组播组 (b) 组播路由的一种生成树

(c) 组播组1的生成树 (d) 组播组2的生成树

组播组不同则组播树也不同，每个路由器都需要为每个组播业务维护一个组播树



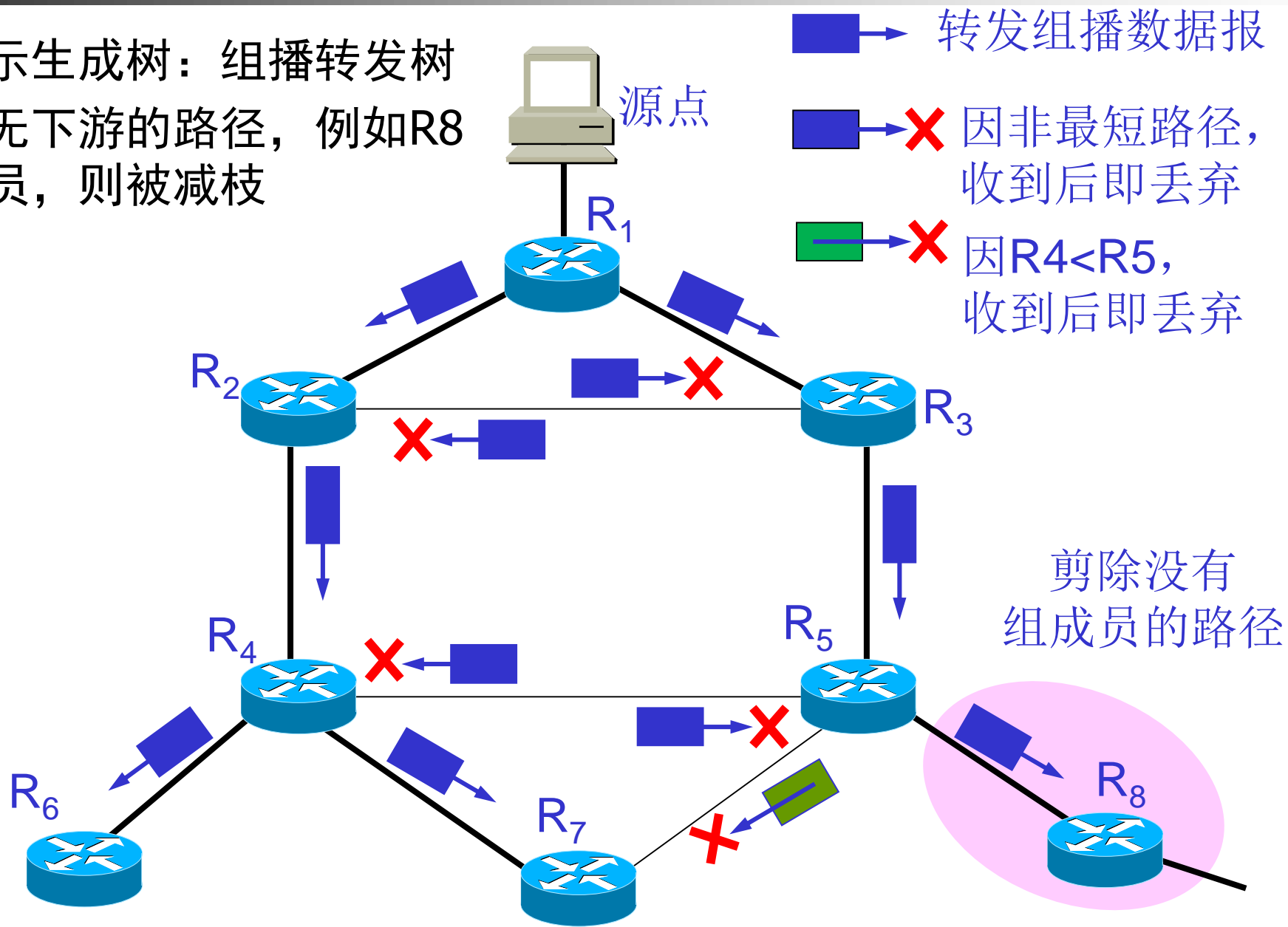
转发组播数据报的方法

(1) 基于生成树

- 路由器转发组播数据报使用泛洪（即广播）。为避免兜圈子，采用反向路径转发RPF (Reverse Path Forwarding)策略
- RPF要点：
 - 路由器收到组播数据报时，先检查是否从源点经最短路径传送来的
 - 可根据IP分组头部中的TTL，或者路由表中的代价
 - 若是，就向除进入方向之外的其他方向转发；
 - 否则就丢弃；
 - 如果存在几条最短路径，则只选择IP地址最小的路由器

反向路径转发RPB和剪枝

粗线表示生成树：组播转发树
要剪除无下游的路径，例如R8
无组成员，则被减枝



转发组播数据报的方法

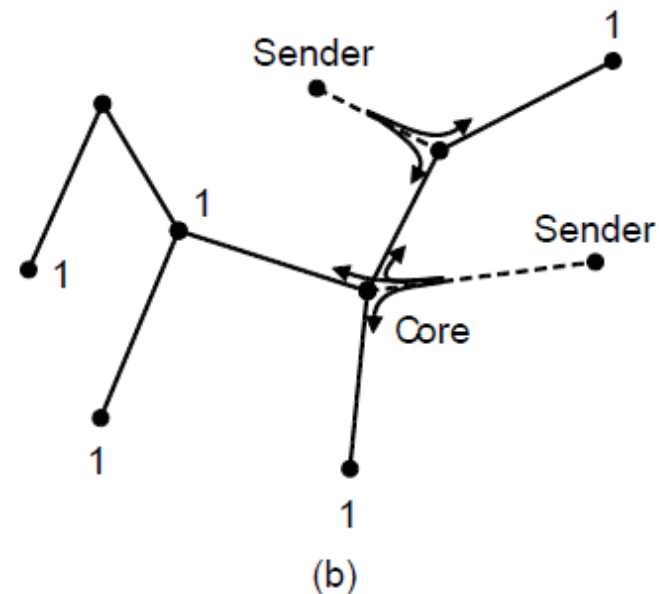
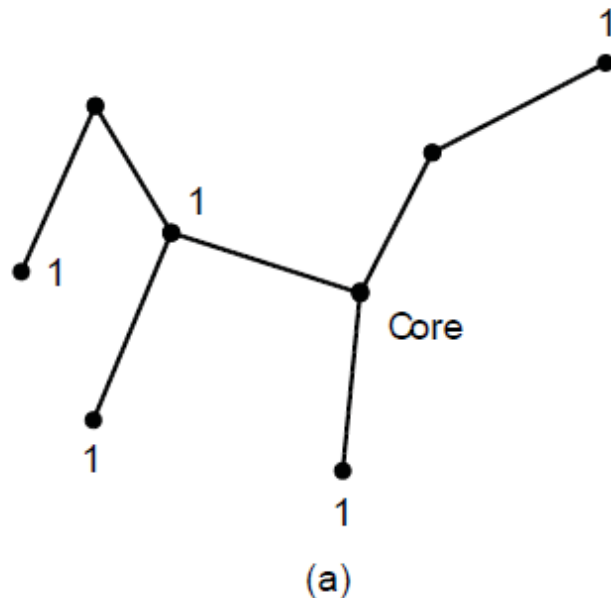
- 基于核心树：计算某个组的单棵生成树，方法是所有的路由器同意某一路由器作为核心，各路由器发送数据分组以建立这棵树

(a) 组1的核心树：成员先发送组播分组给核心

(b) 由核心发送组播分组给组1的各成员

若发送者距离核心较近，基于核心树的分发是最优的；

若发送者距离核心较远，将以发送者作为核心（生成树算法）是最优的





几种组播路由选择协议

- 距离向量组播路由协议DVMRP (Distance Vector Multicast Routing Protocol)
 - 基于反向路径转发和剪除算法
- 组播MOSPF (Multicast Extensions to OSPF)
 - 基于反向路径转发和剪除算法
 - 每个路由器为每个发送者构造一棵剪除后的生成树
- 协议独立组播PIM(Protocol Independent Multicast)
 - 基于核心树，每个路由器为每个组保存一棵树
- 基于核心的转发树CBT (Core Based Tree)



组播路由选择算法比较

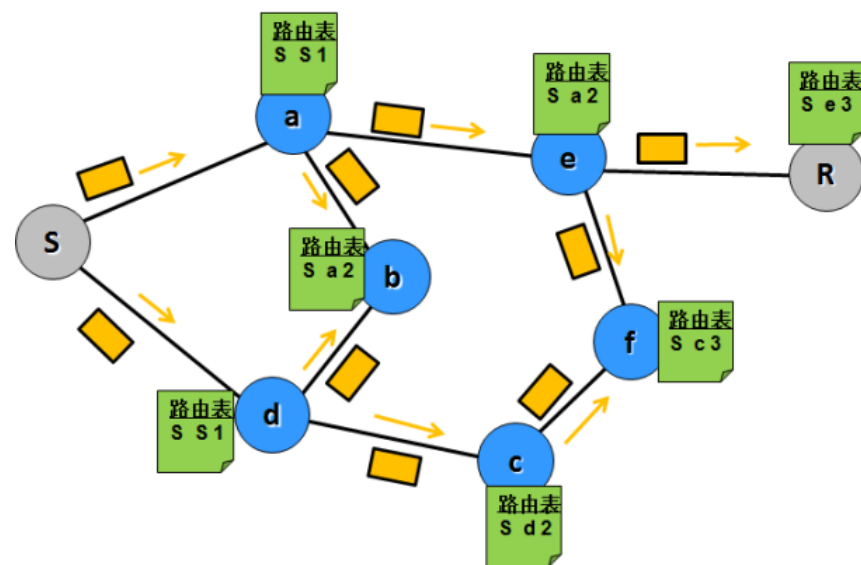
- 性能与网络规模及组播的密度相关
- 组播密度大：基于生成树（广播+剪枝）
- 组播密度小：基于核心树
- “基于生成树”与“基于核心树”的比较
 - 若组播组总数是 n ，每个组的结点数是 m
 - 基于生成树，每个路由器维护 nm 棵生成树
 - 在 m 个节点中，以每个节点作为根都有一个修剪的生成树； n 个组相当于有 n 个图；因此共有 nm 棵树。
 - 其优点是，可以建立一个以组播服务器（可位于网络任何位置）为根的组播树
 - 基于核心树，每个路由器维护 n 棵核心树；若发送者距离核心较近，则基于核心树的分发是最优的；否则，需要经过多跳转给核心之后再转发

练习题

IP组播树的构造采用一种称为反向路径转发（RPF）技术。在如图所示的网络中，假设S为组播源，R为组成员。各路由器给出的路由表部分信息为：

目的地、下一跳、距离值。

当S以泛洪方式发出一个组播数据分组，节点b和f将收到该分组的多个拷贝。试问b和f是如何转发收到的分组？结合图例说明RPF技术。

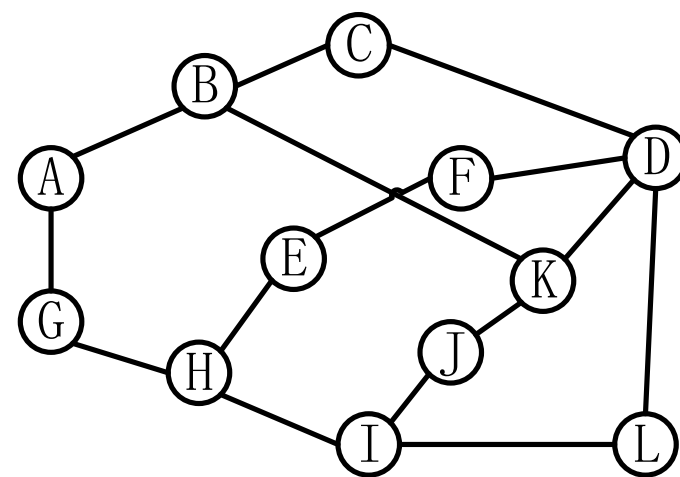


练习题

- 在如图所示的网络中，组播成员分布在路由器A、B、C、D、E、F、I、K上，采用反向路径转发（RPF）技术。请给出路由器C的组播生成树。

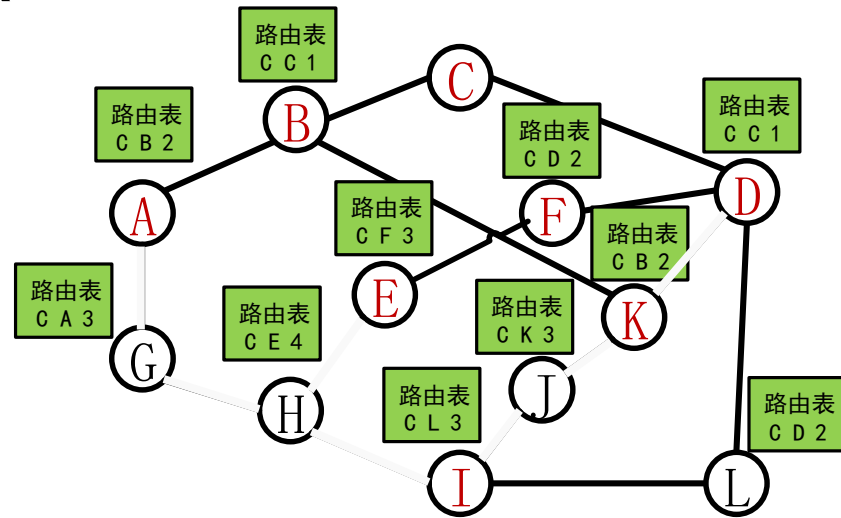
提示：

- 先写出各路由器的路由表
- 再基于RPF给出组播生成树



练习题（解答）

- 组播成员：A、B、C、D、E、F、I、K
- 路由表：以C为源
- 以C为根的最小生成树
- 剪枝





再议路由问题

■ 路由与转发：

- 路由器执行路由协议与算法、形成路由表及转发表
 - 最小距离路由、距离矢量路由
- 路由器接收分组、查表选择转发端口

■ IP组播及路由

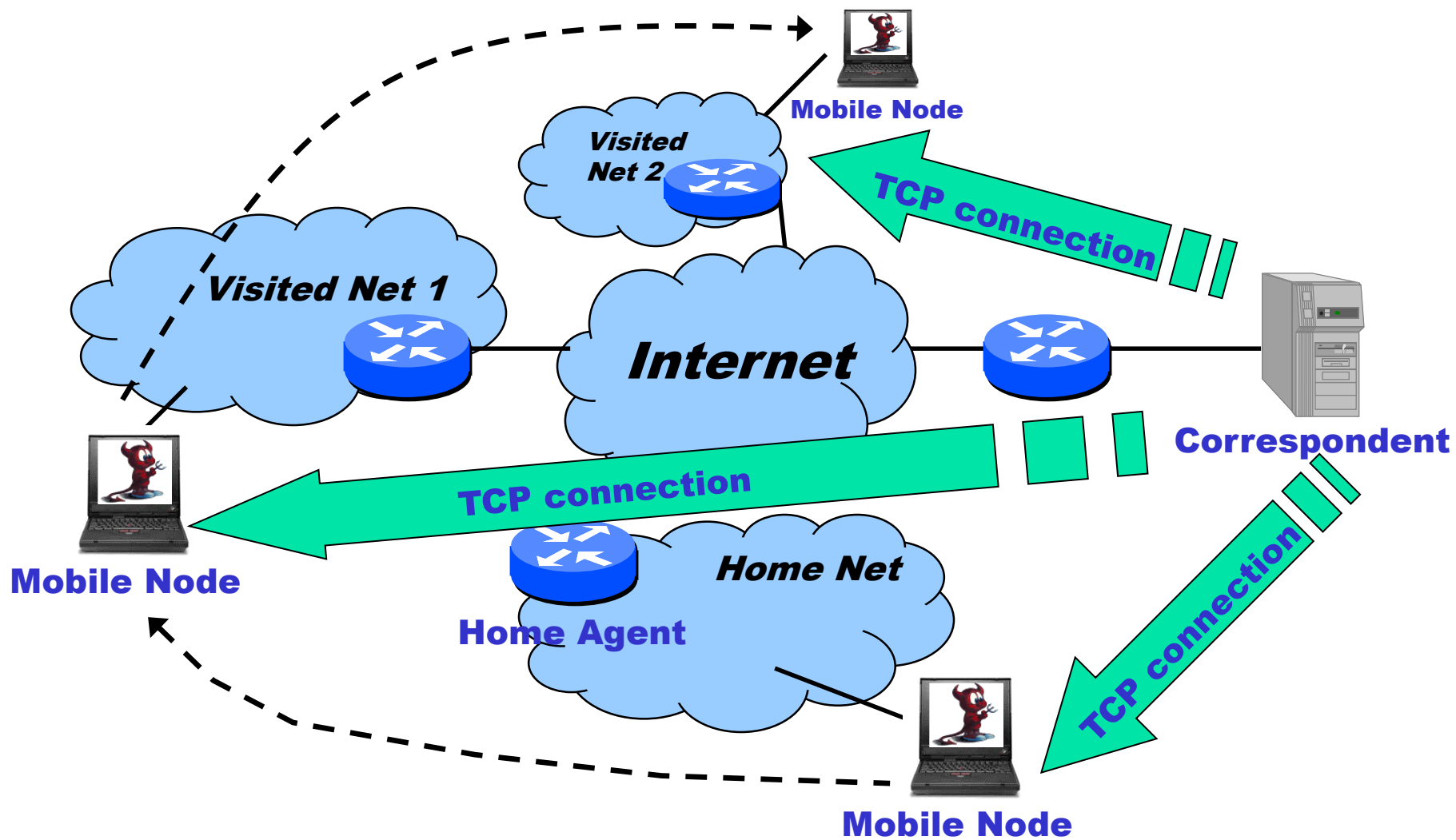
■ 移动主机路由

- 网络中的路由器不移动，结点移动，采用移动IP

■ 自组织网络路由

- 网络中的所有节点都移动
 - MANET(Mobile AdHoc)
 - WSM((Wireless Sensor Networks)

移动主机路由





使用两个IP地址及其问题

IP 地址与应用相关

- IP分组交换是根据分组的目的IP地址进行路由
- IP地址既标识主机又标识主机上的应用程序
- 当改变主机与网络连接点时，根据IP路由机制，要求改变IP地址；而改变IP地址，则导致应用程序与网络之间的连接中断，因为基于UDP/TCP的进程间通信，网络进程标识使用套接字（IP，端口）
- 若采用DHCP动态IP地址分配，则在移动切换过程中需要改变IP，则出现通信中断——再连接的问题

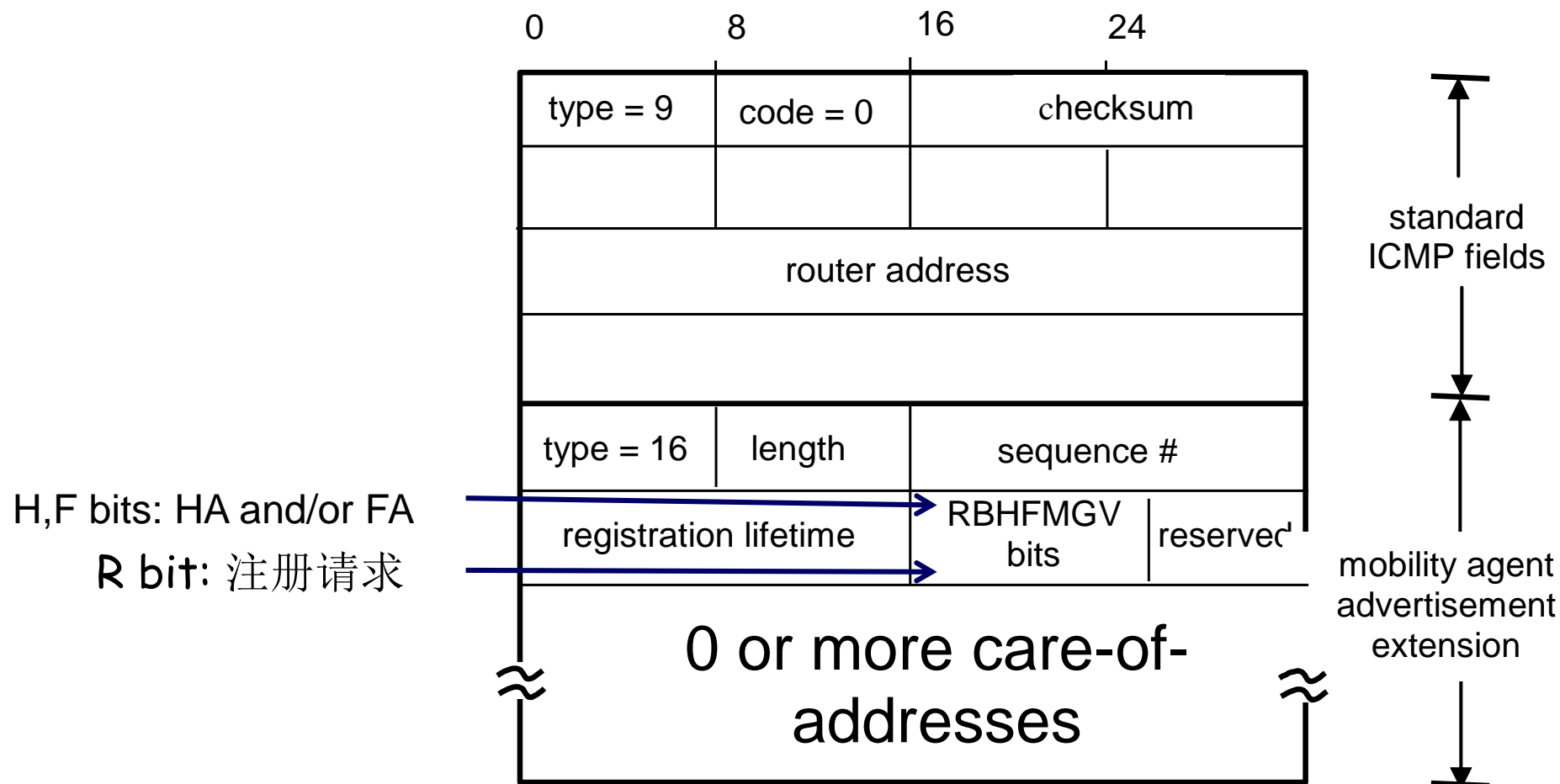


移动IP

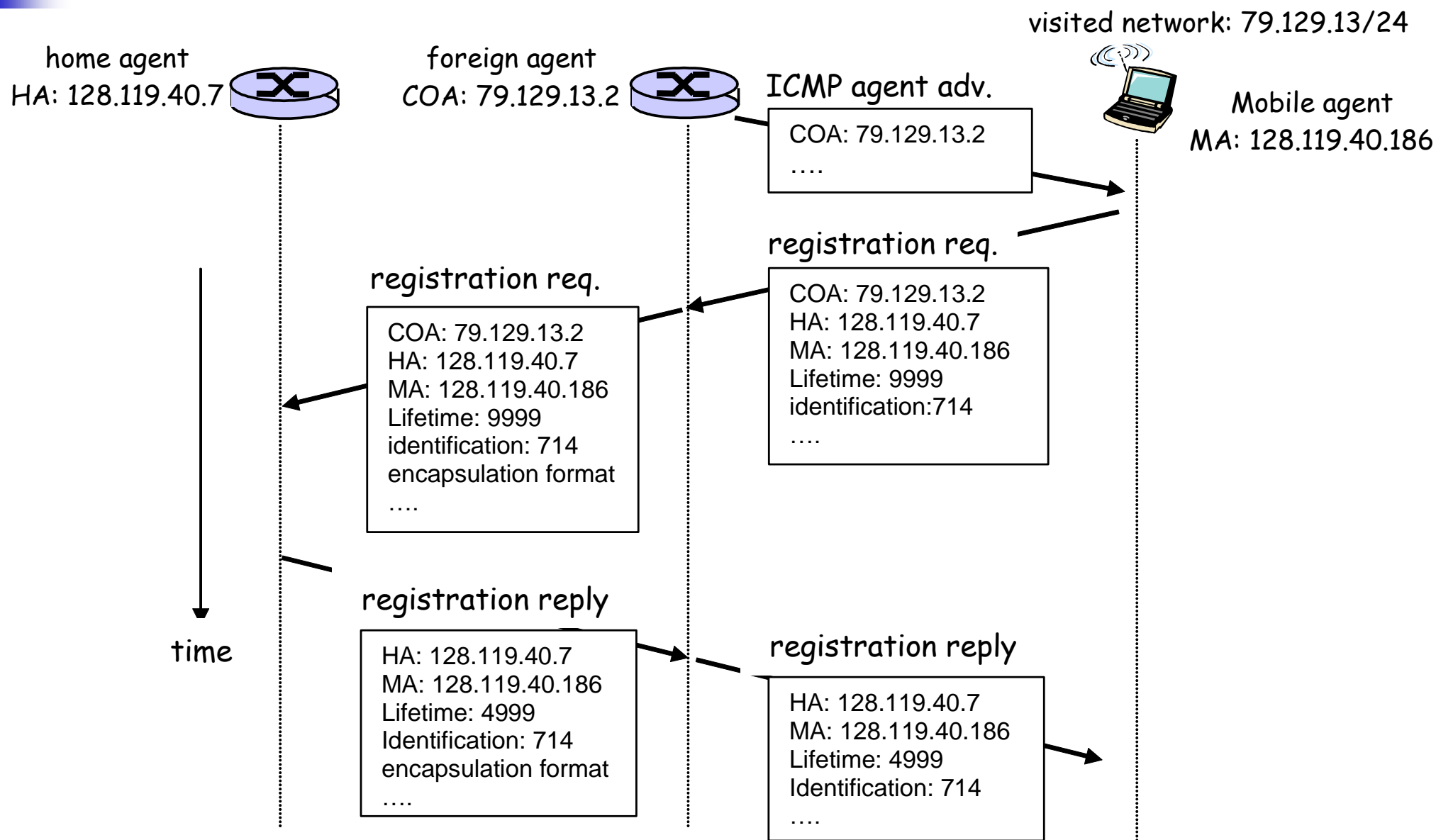
- RFC 3344中的主要部件：
 - 归属代理HA（Home Agent），外部代理FA（Foreign Agent），FA注册，关联地址CoA（care-of-addresses），封装
- 移动节点MN（Mobile Node）根据HA/FA的代理通告，获得其当前的位置
 - 可以采用链路层技术或网络层技术
- 当MN移动并改变网络连接点时，获得CoA
- MN向HA注册，建立（CoA，MNIP）地址绑定关系
- 建立隧道
 - 目的地址为MN的IP分组，由HA经过隧道传送给FA或MN

移动IP：代理发现

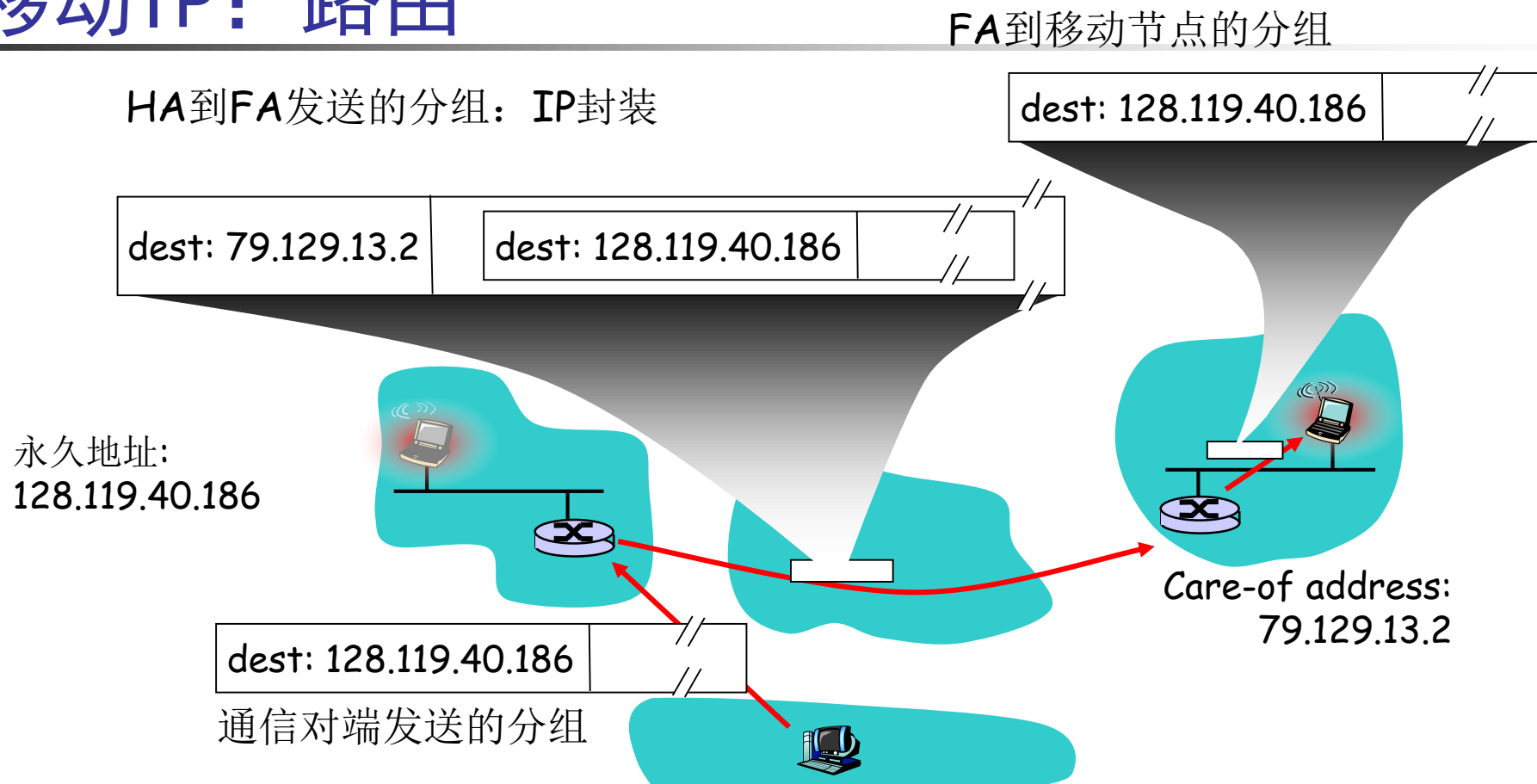
- 代理通告：FA/HA广播ICMP消息（typefield = 9），通告移动业务



移动IP: 注册过程



移动IP：路由



- 间接路由：由HA捕获到MN的分组并封装IP（目的地址为COA）；存在“三角路由”问题
- 直接路由：由HA通知通信对端，由其封装；通信对端可有多并要支持移动IP
- 问：HA一定要在路由器上吗？HA可否为普通主机？



再议路由问题

■ 路由与转发：

- 路由器执行路由协议与算法、形成路由表及转发表
 - 最小距离路由、距离矢量路由
- 路由器接收分组、查表选择转发端口

■ IP组播及路由

■ 移动主机路由

- 网络中的路由器不移动，结点移动，采用移动IP

■ 自组织网络路由

- 网络中的所有节点都移动
 - MANET(Mobile AdHoc)
 - WSM((Wireless Sensor Networks)



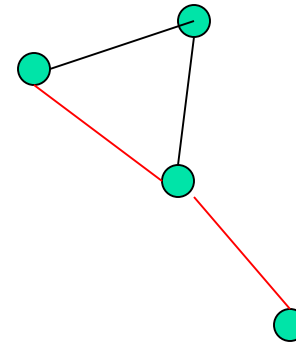
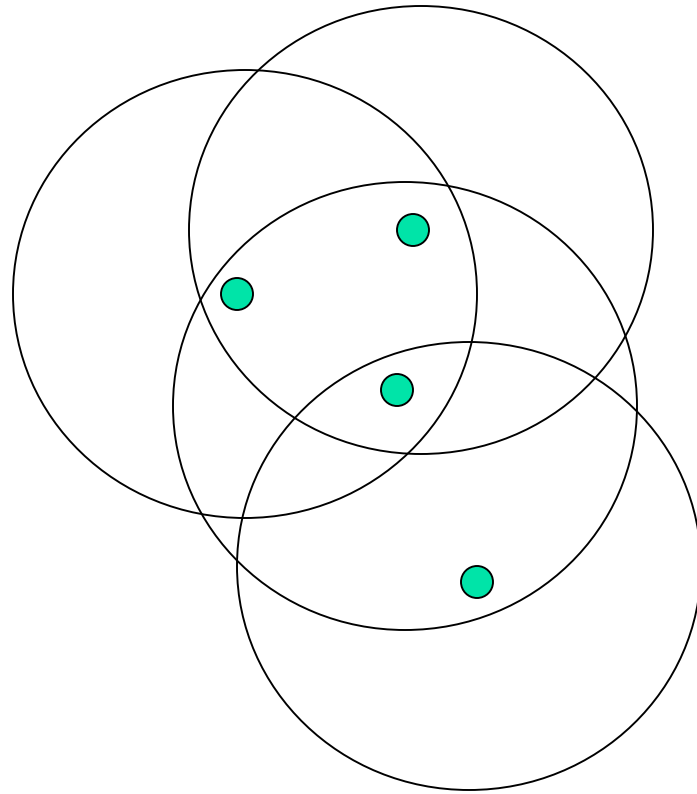
自组织（Ad Hoc）网络

■ 何谓Ad Hoc

- Ad hoc is a Latin phrase which means “for this [purpose]”.
- 由主机、移动节点组成的无线网络
- 无固定基础设施
- 网络拓扑结构可变：网络中的所有节点均可以移动
- 节点之间的路由为多跳：为便于节点移动，要求设备小型化、低功耗，因而，节点的通信能力有限，为支持更远距离的节点之间的通信，需要多跳

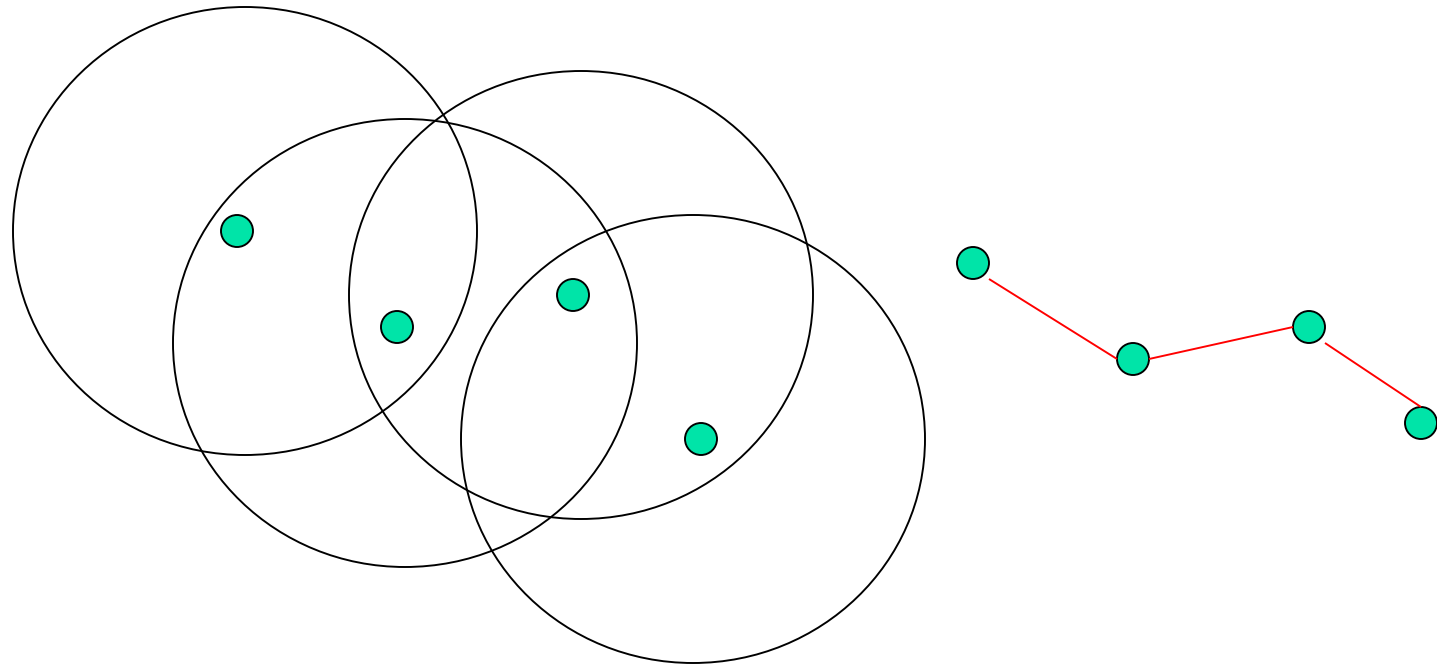
Ad Hoc Networks

- 分组传输需要经过多跳（经过多个节点的中继）才能到达目的节点



Ad Hoc Networks (MANET)

- 移动导致网络拓扑改变，要求路由变化



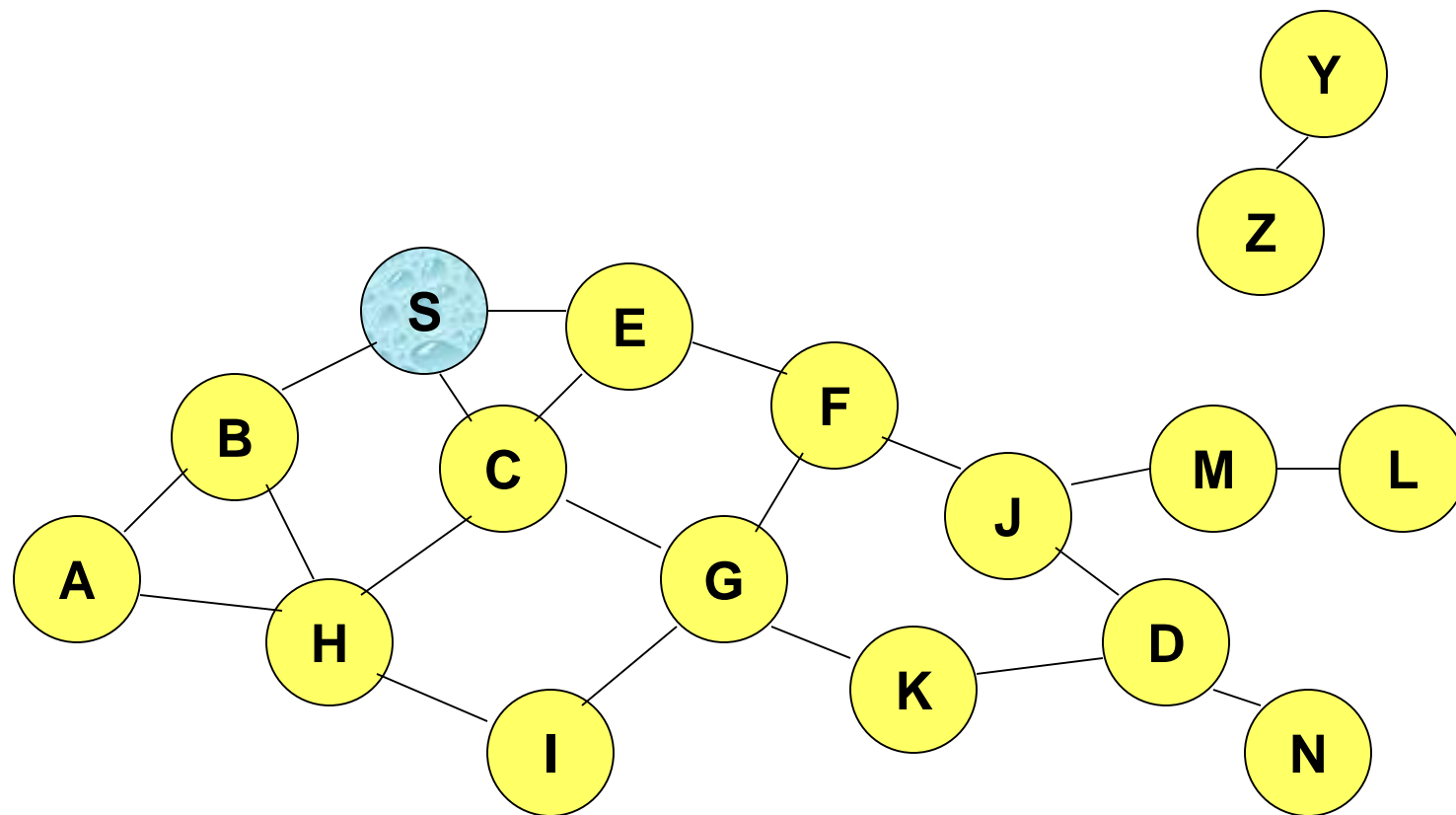
- 如何路由？
 - 泛洪或广播？
 - 最短路径或距离矢量路由？



AODV路由协议

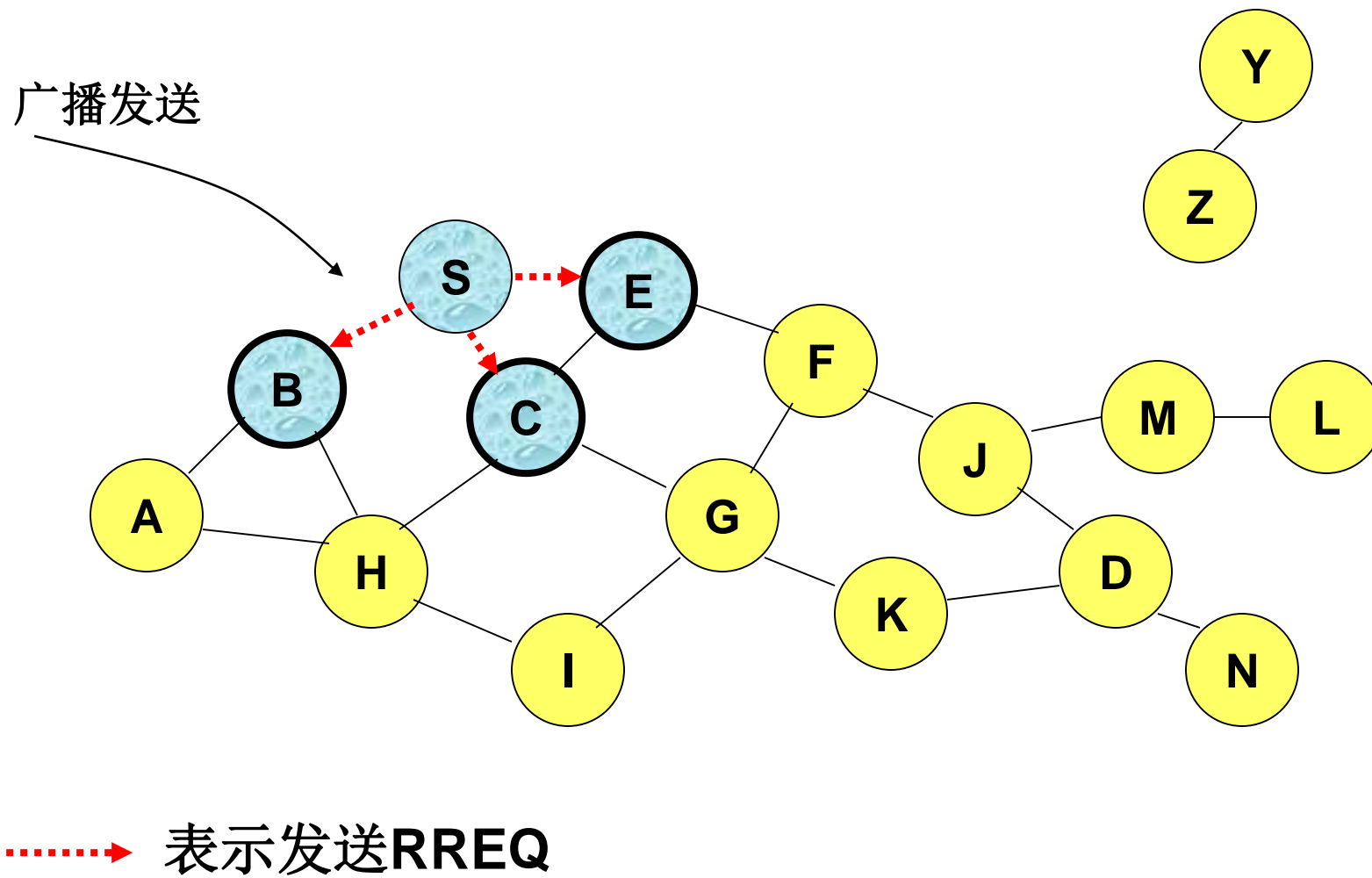
- AODV: Ad Hoc On-Demand Distance Vector
- 当节点S要发送分组给节点D，但没有到节点D的路径，则启动路由发现过程 **route discovery**
- 源节点S泛洪**Route Request (RREQ)** 分组
- 其他节点收到RREQ后再广播，同时建立它到源节点的反向路径
 - AODV假设链路是双向对称的
- 当目的节点D收到RREQ时，发送**Route Replay (RREP)** 作为响应
- RREP沿着RREQ转发过程中所建立的相反路径转发

AODV中的路由请求

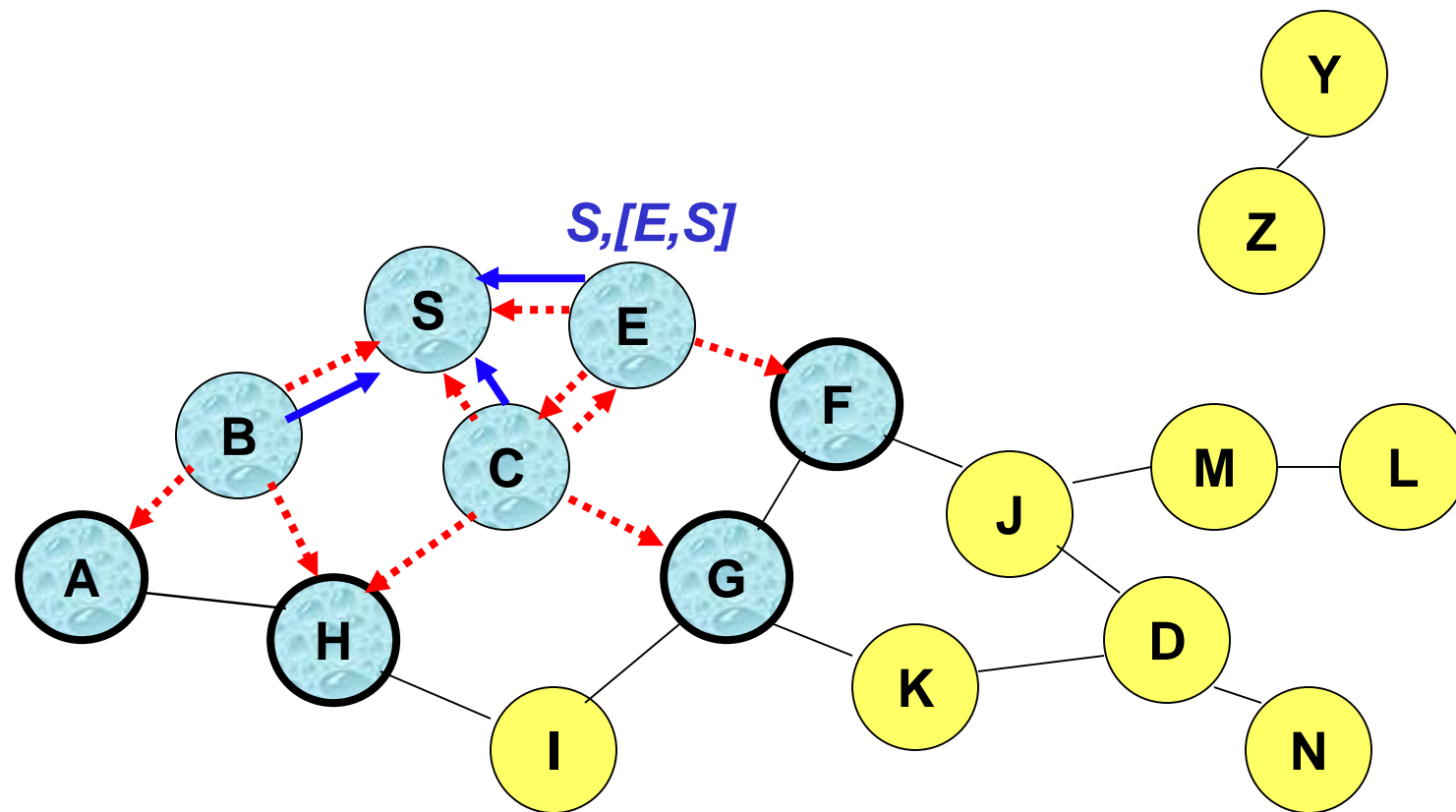


代表接收从S到D的RREQ分组的节点

AODV中的路由请求



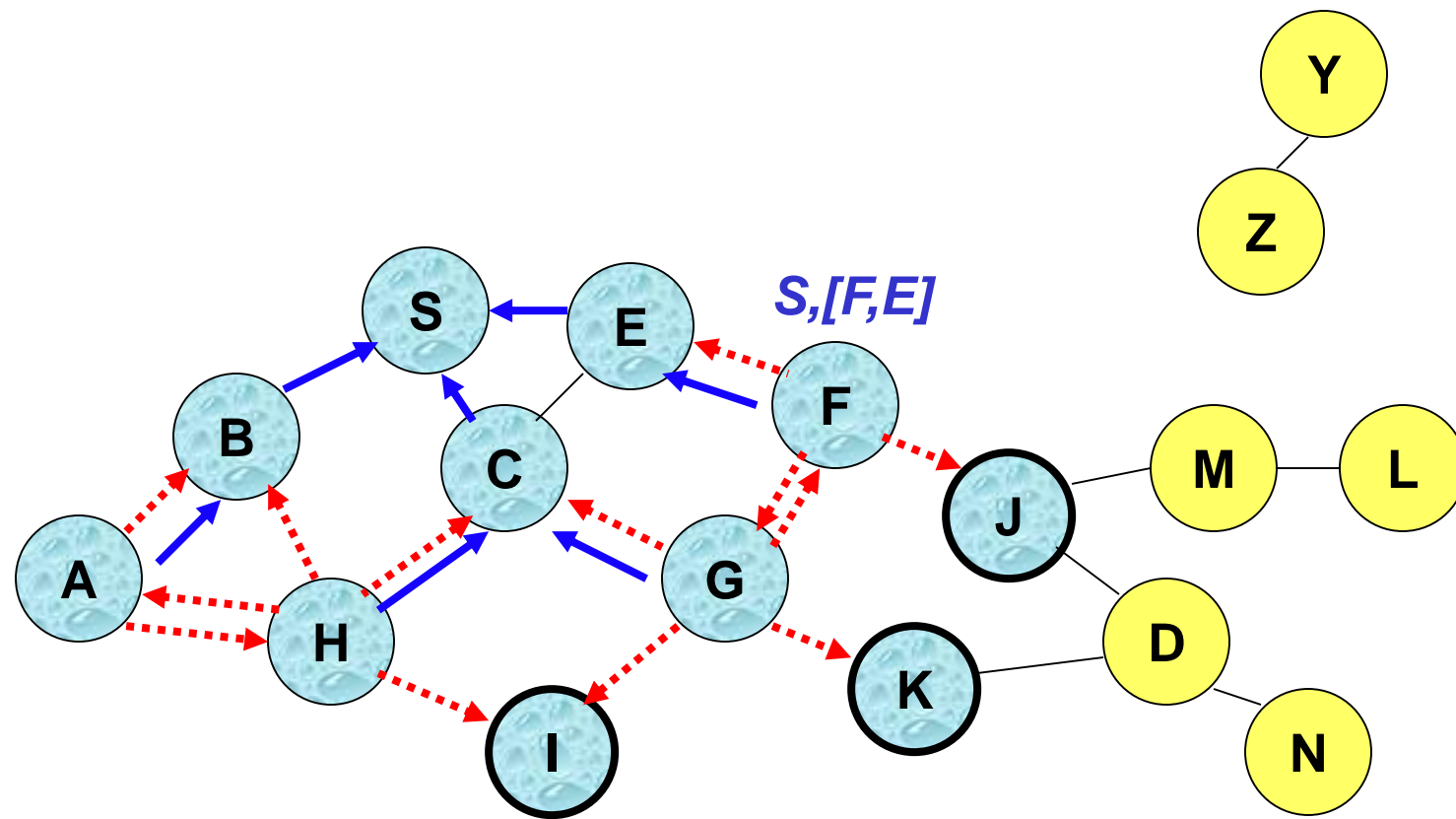
AODV中的路由请求



← 表示反向路径的链路

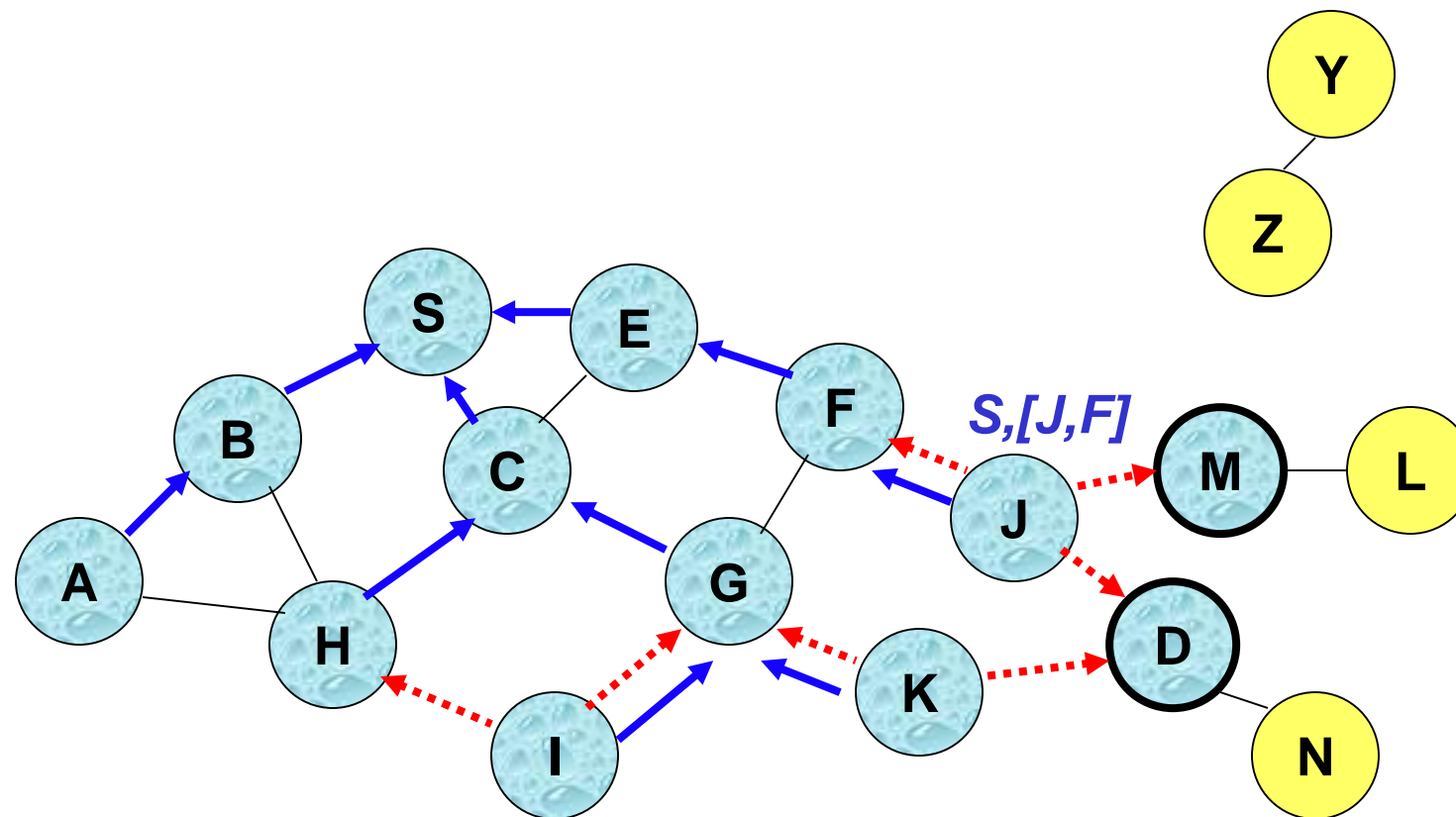
$S,[E,S]$ 表示到目的节点S经过转发链路[E,S]

AODV中建立的反向路径

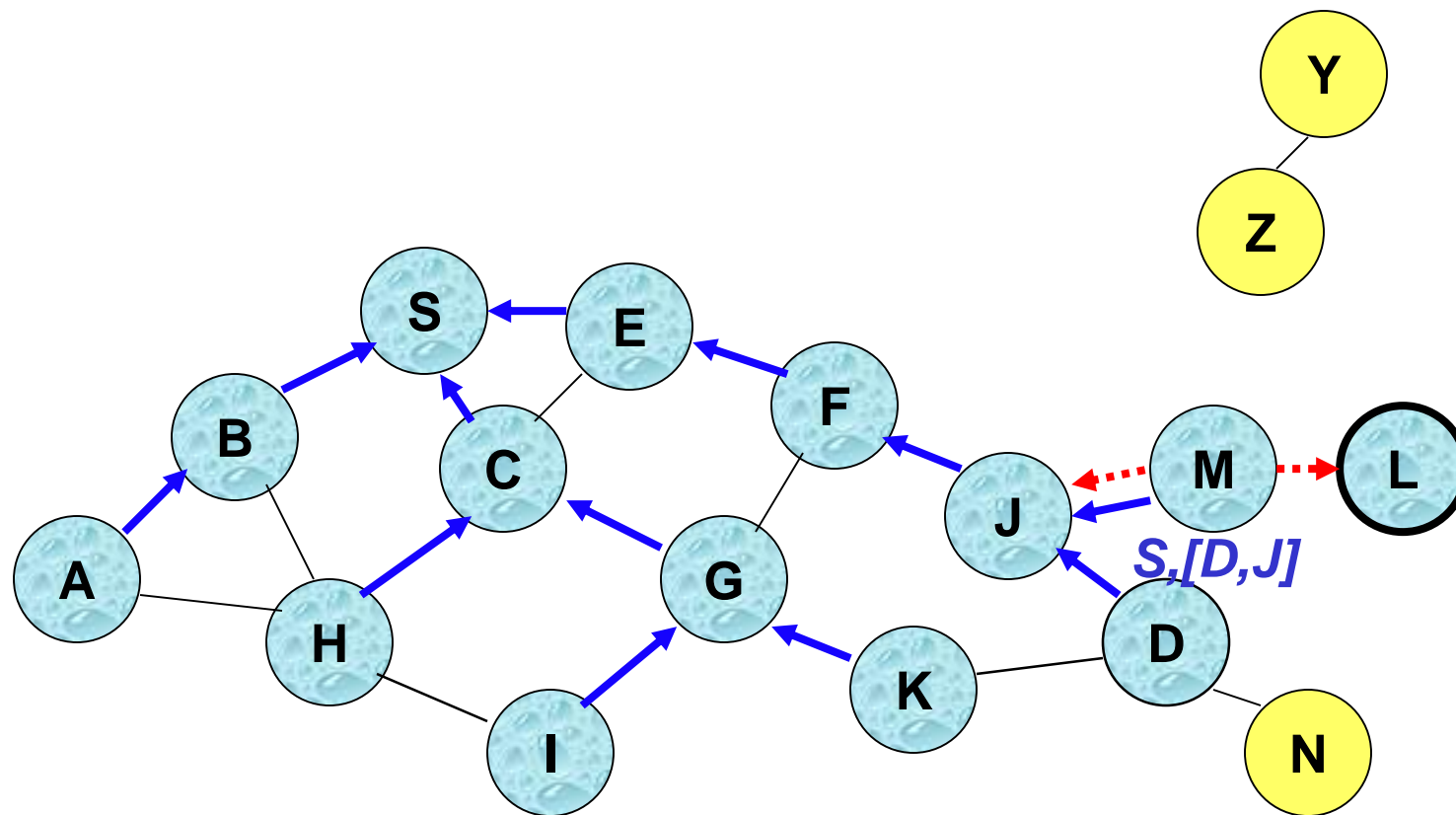


- 节点**C**收到从**G**和**H**的**RREQ**也不再转发，因它已经转发过

AODV中建立的反向路径

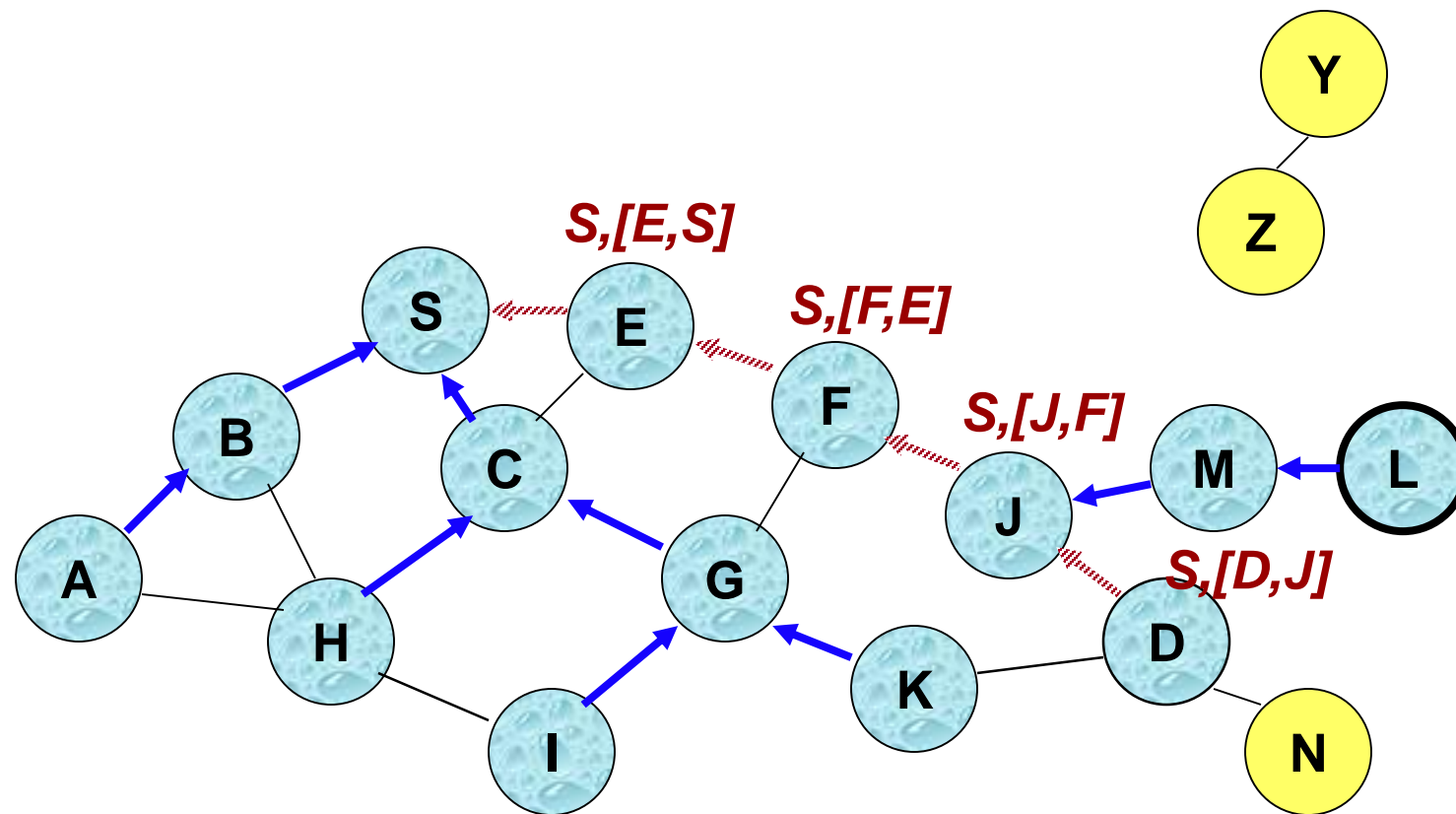


AODV中建立的反向路径



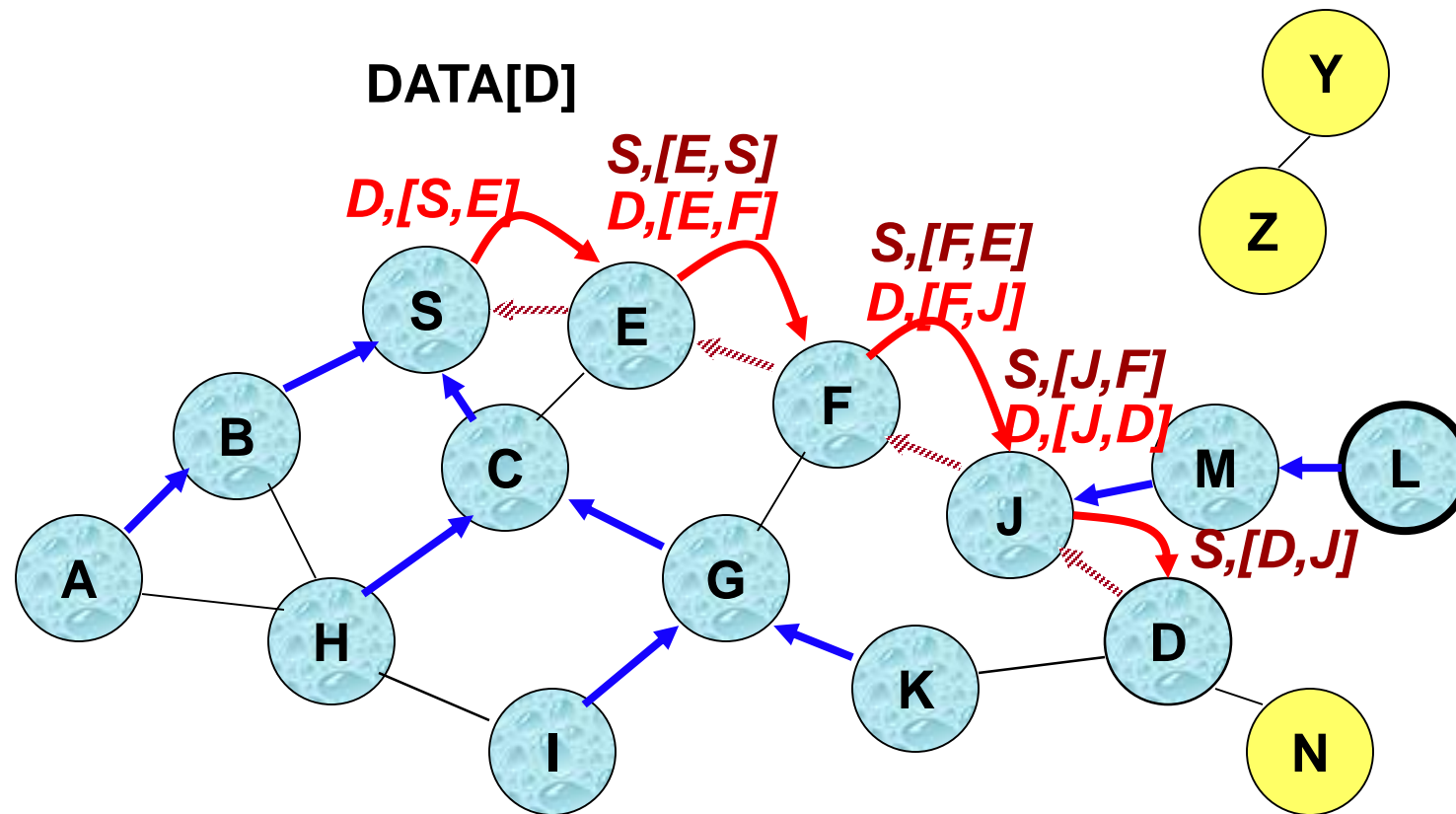
- 节点D不再转发RREQ, 因 D为RREQ 的目的节点

AODV的路由响应



 表示RREP选择的链路

AODV中建立的前向路径



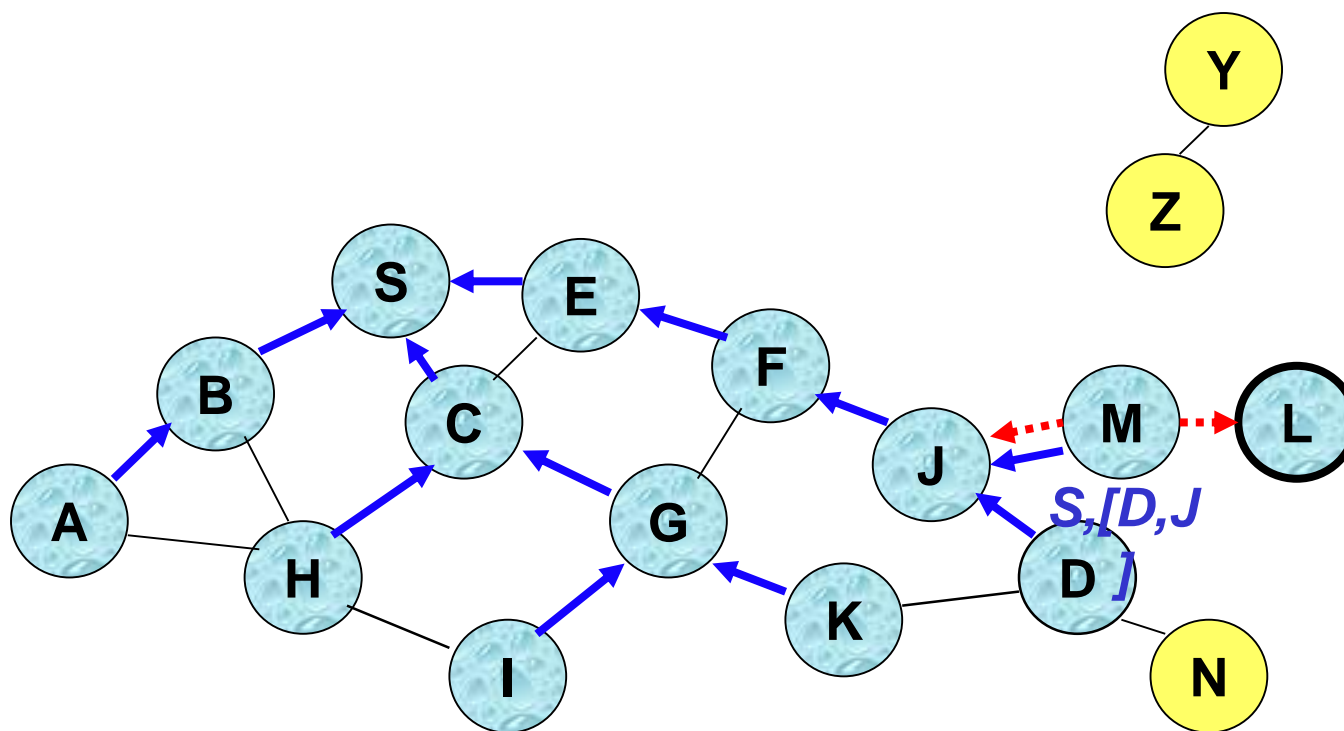
表示RREP选择的链路

当RREP沿反向路径传输时，则建立了D到S的前向链路
表示前向路径上的链路

在DATA中不含路由信息，各节点根据路由表转发分组

降低AODV 路由发现的开销

- 在大型网络中，AODV RREQ泛洪，开销很大，例如：
 - S到Z：泛洪后无法建立路由
 - S到C：1次广播即建立路由，但也需要等待泛洪结束
- 减少开销的方法：设置AODV RREQ中IP分组中的TTL，由1依次递增

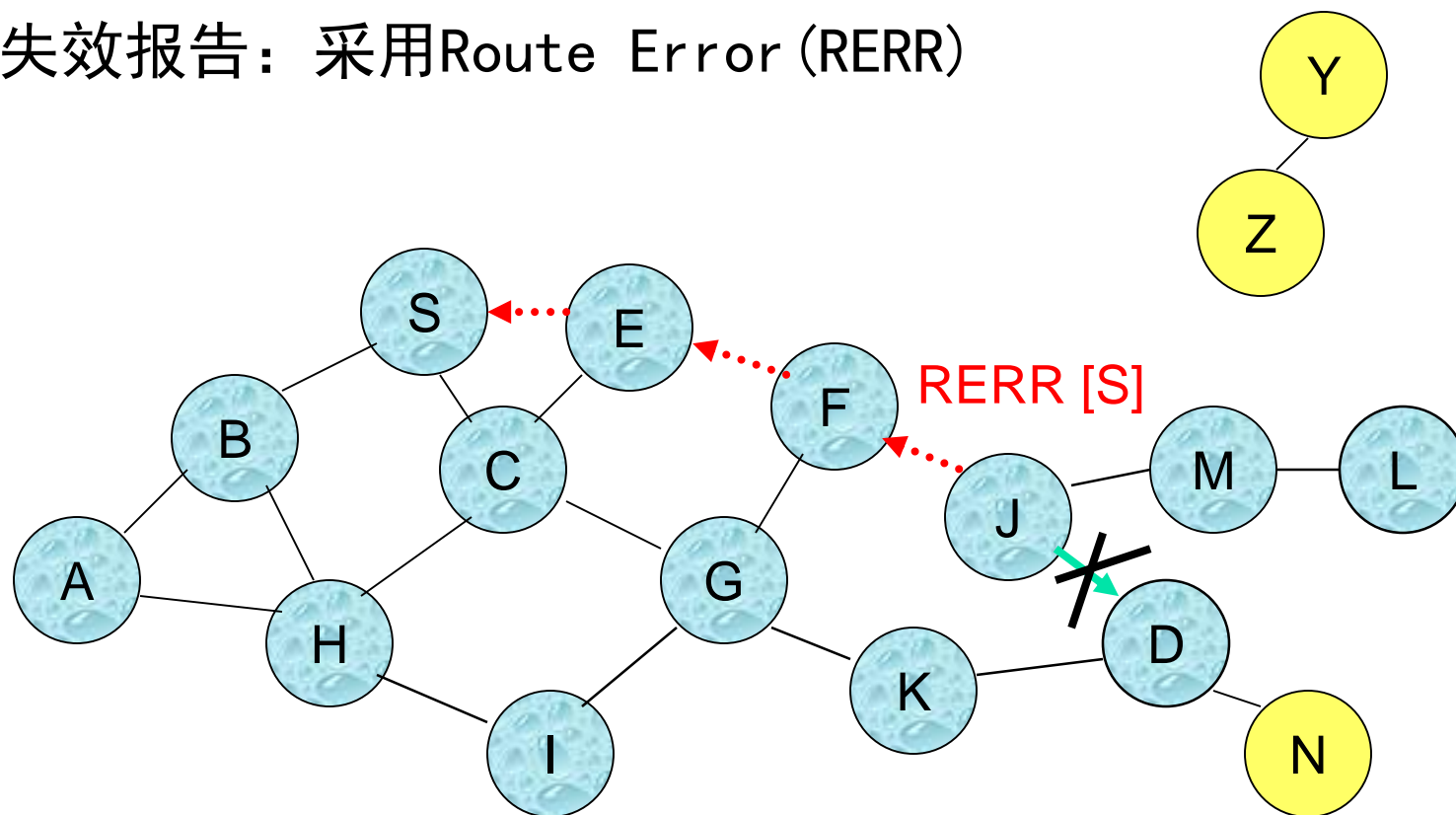


AODV协议中的细节

- 目的地序列号DSN：用于用于确定路由的新旧
 - 在RREQ、RREP、RERR中含有DSN (*destination sequence numbers*)
 - 若节点S到D发送一个新的RREQ，则增大DSN
 - 中间节点有到目的D的路由
 - 若DSN比RREQ中的小，则不发送RREP
 - 若DSN比RREQ中的大，则可以发送RREP
- 超时定时器：
 - 在广播RREQ过程中建立的D到S的**反向路径**路由表项是临时的，若超时则删除；超时时间应足够长，以保证可返回RREP
 - S到D的一条**前向路径**，在*active_route_timeout*内没有发送数据，则删除，以降低保存路由表的内存开销

AODV协议中的细节

链路失效报告：采用Route Error (RERR)



- 当J在J-D链路上转发分组失败时，则DSN+1并向S发送RERR
- 收到RERR的节点，转发RERR并删除路由表项
- S收到RERR则发送RREQ（其中的DSN加1）启动到节点D的路由发现过程



AODV: 小结

- 在分组头中不含路由信息，降低了头部开销
- 节点维护一张路由表，每一路由表项设置一个定时器，表示路由是否过期
- 一般情况下，每个节点为每个目的节点维持下一跳地址
- 即使网络拓扑不变，不用的路由也会因过期而被删除



本章小结

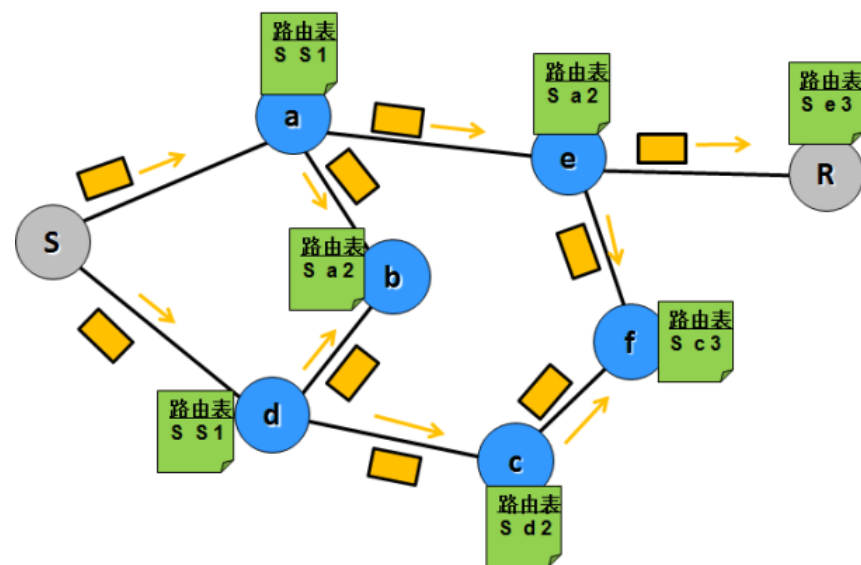
- 网络层的功能：为传输层提供与路由器的数量、类型和拓扑结构无关的服务；统一编址；网络拥塞控制、保证服务器质量、网络互联
- 路由器：网络层互联设备，实施路由与转发
- 网络互联
 - 地址分配：IP地址，三种编址方式；DHCP、NAT与专用地址
 - 分组传送：ARP，分段与重组
 - 隧道技术
- 路由：距离矢量路由算法与RIP；最小距离路由算法与OSPF
- 网络控制：超时、差错恢复、状态报告、拥塞控制
- IP组播及路由：IGMP、组播路由
- 移动主机路由：代理（HA、FA）通告、注册、隧道
- 自组织网络路由

练习题

IP组播树的构造采用一种称为反向路径转发（RPF）技术。在如图所示的网络中，假设S为组播源，R为组成员。各路由器给出的路由表部分信息为：

目的地、下一跳、距离值。

当S以泛洪方式发出一个组播数据分组，节点b和f将收到该分组的多个拷贝。试问b和f是如何转发收到的分组？结合图例说明RPF技术。



练习题

在如图所示的网络中，组播成员分布在路由器A、B、C、D、E、F、I、K上，采用反向路径转发（RPF）技术。请给出路由器C的组播生成树。

提示：

- 先写出各路由器的路由表
- 再基于RPF给出组播生成树

