哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY
立足航天，服务国防，面向国民经济主战场

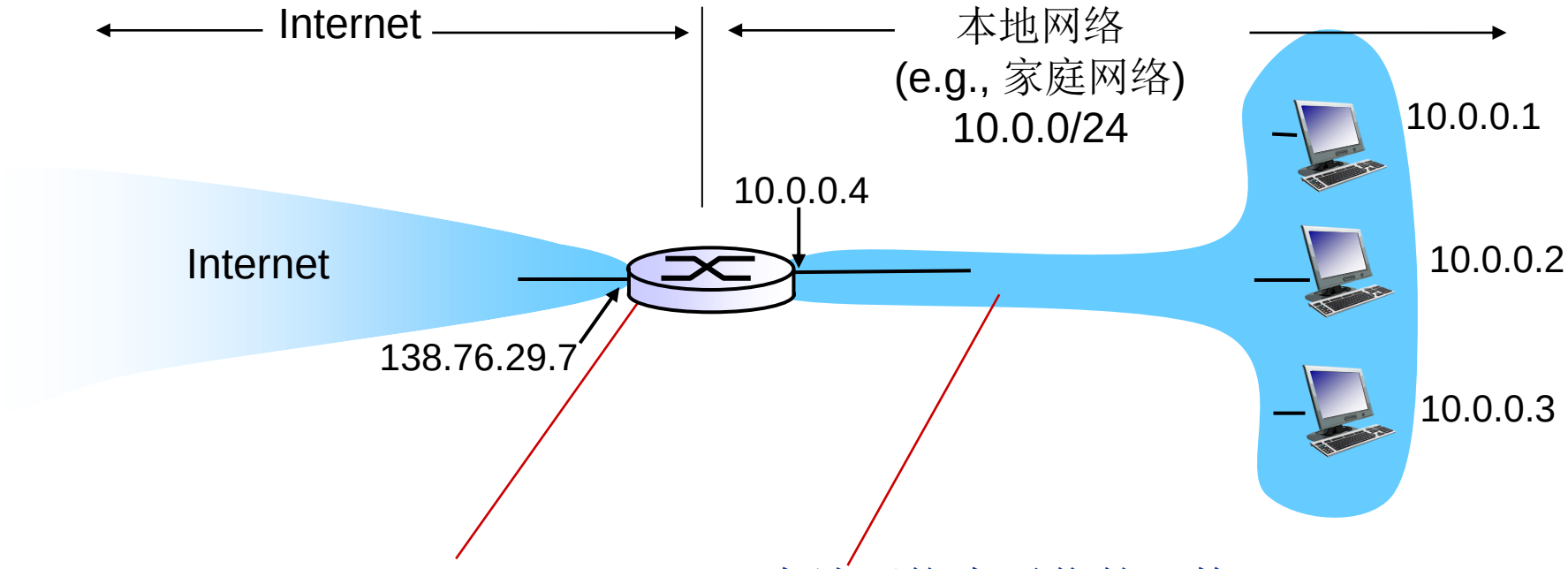
计算机网络之探赜索隐

主讲人：李全龙

本讲主题

网络地址转换(NAT)

网络地址转换(NAT)



所有离开本地网络去往Internet的数据报的源IP地址需替换为相同的NAT IP地址：138.76.29.7以及不同的端口号

本地网络内通信的IP数据报的源与目的IP地址均在子网10.0.0/24内

网络地址转换(NAT)

动机:

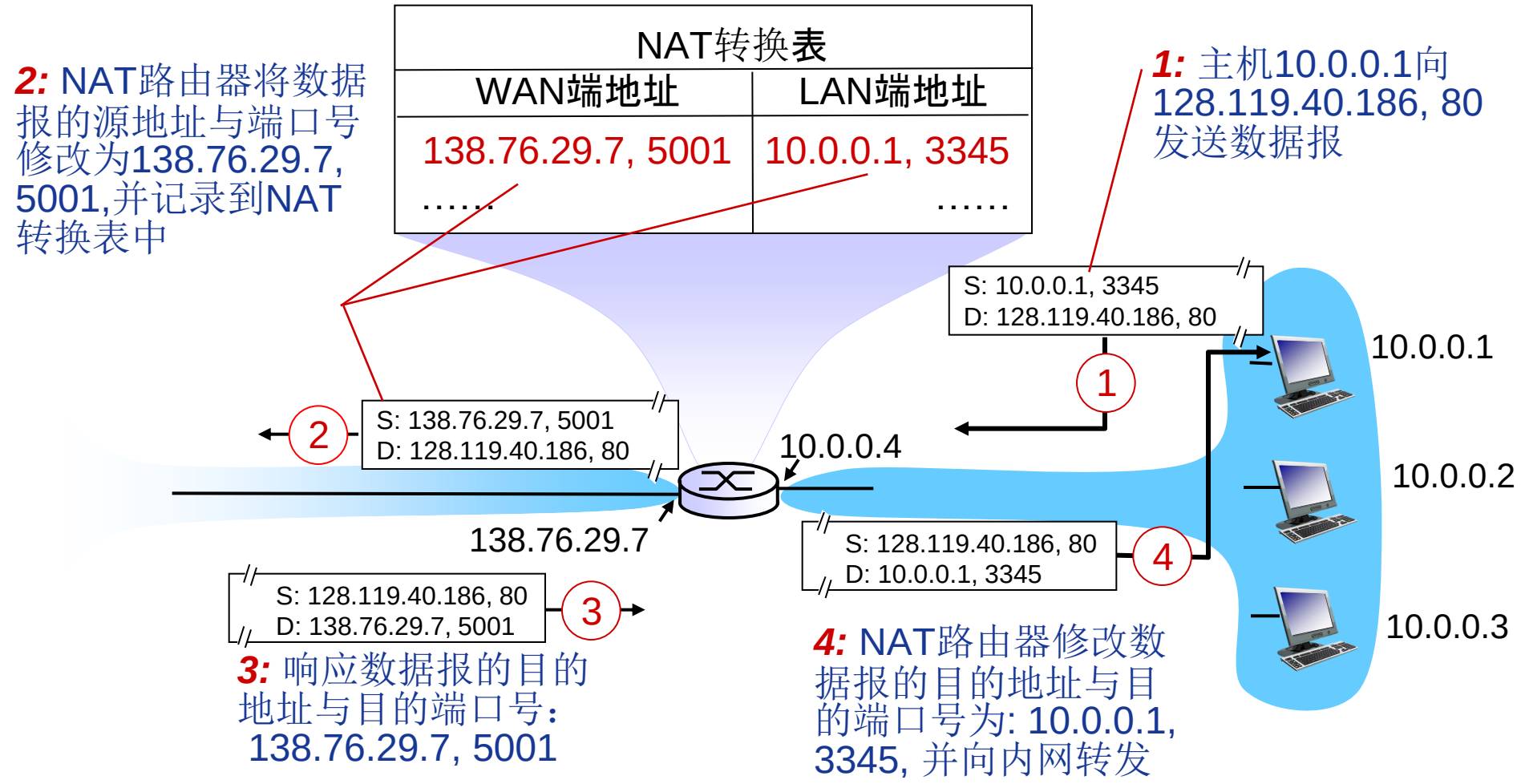
- 只需/能从ISP申请一个IP地址
  - IPv4地址耗尽
- 本地网络设备IP地址的变更，无需通告外界网络
- 变更ISP时，无需修改内部网络设备IP地址
- 内部网络设备对外界网络不可见，即不可直接寻址(安全)

网络地址转换(NAT)

实现:

- 替换
  - 利用(**NAT IP地址,新端口号**)替换每个外出IP数据报的(**源IP地址,源端口号**)
- 记录
  - 将每对(**NAT IP地址, 新端口号**) 与(**源IP地址, 源端口号**)的替换信息存储到NAT转换表中
- 替换
  - 根据NAT转换表，利用(**源IP地址, 源端口号**)替换每个进入内网IP数据报的(**目的IP地址,目的端口号**)，即(NAT IP地址, 新端口号)

网络地址转换(NAT)

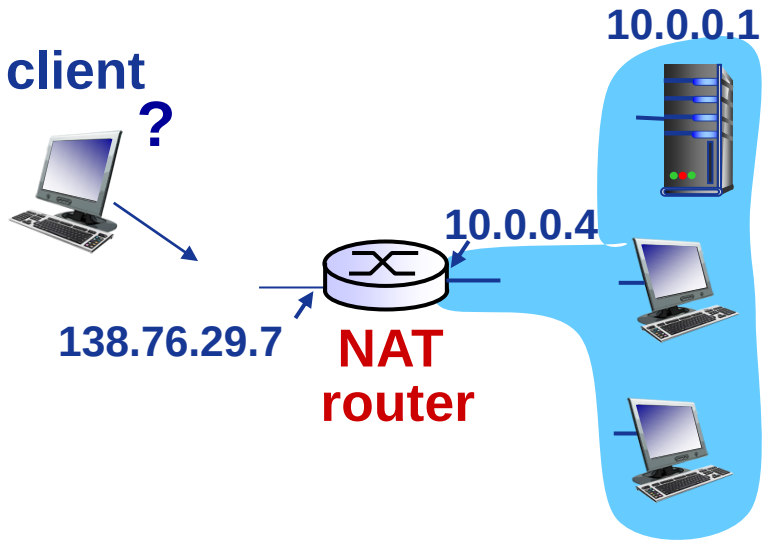

NAT转换表
WAN端地址
LAN端地址
138.76.29.7, 5001
10.0.0.1, 3345
……
……
2: NAT路由器将数据报的源地址与端口号修改为138.76.29.7, 5001,并记录到NAT转换表中
1: 主机10.0.0.1向128.119.40.186, 80发送数据报
S: 10.0.0.1, 3345
D: 128.119.40.186, 80
1
10.0.0.1
10.0.0.2
10.0.0.3
10.0.0.4
2
S: 138.76.29.7, 5001
D: 128.119.40.186, 80
138.76.29.7
S: 128.119.40.186, 80
D: 10.0.0.1, 3345
4
S: 128.119.40.186, 80
D: 138.76.29.7, 5001
3
3: 响应数据报的目的地址与目的端口号：138.76.29.7, 5001
4: NAT路由器修改数据报的目的地址与目的端口号为: 10.0.0.1, 3345, 并向内网转发

网络地址转换(NAT)

- 16-bit端口号字段:
  - 可以同时支持60,000多并行连接!
- NAT主要争议:
  - 路由器应该只处理第3层功能
  - 违背端到端通信原则
    - 应用开发者必须考虑到NAT的存在，e.g., P2P应用
  - 地址短缺问题应该由IPv6来解决

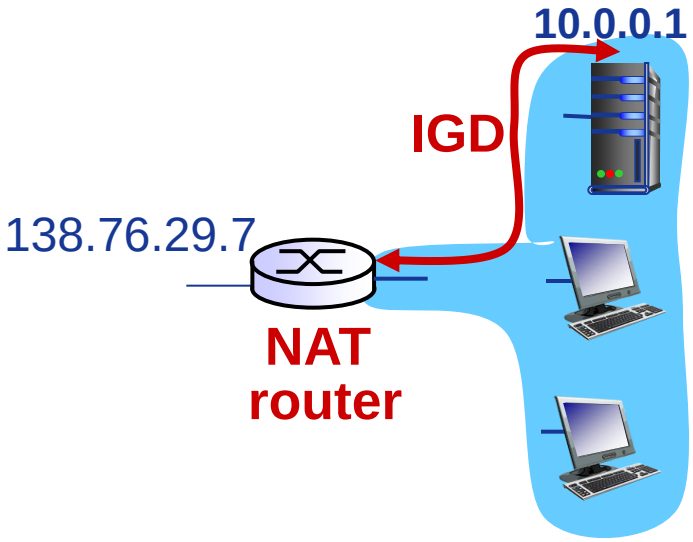
NAT穿透问题

- 客户期望连接内网地址为10.0.0.1的服务器
  - 客户不能直接利用地址10.0.0.1直接访问服务器
  - 对外唯一可见的地址是NAT地址: 138.76.29.7
- 解决方案1: 静态配置NAT，将特定端口的连接请求转发给服务器
  - e.g., (138.76.29.7, 2500) 总是转发给(10.0.0.1, 25000)



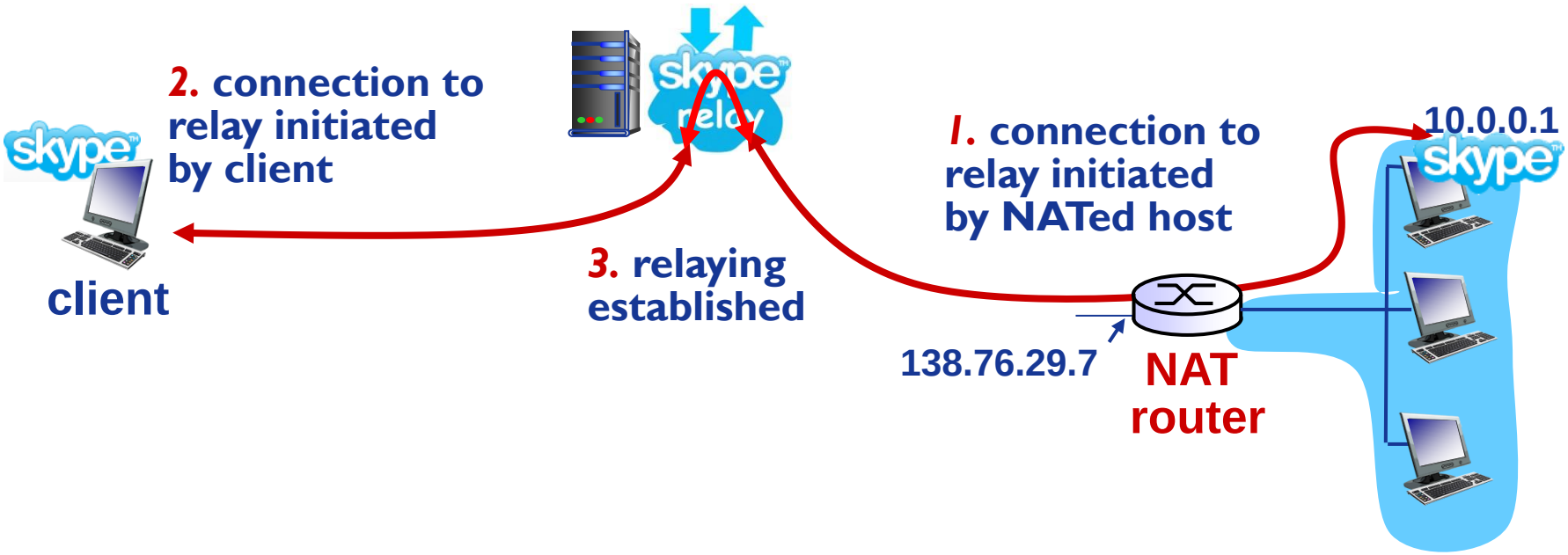
NAT穿透问题

- 解决方案2: 利用UPnP (Universal Plug and Play) 互联网网关设备协议 (IGD-Internet Gateway Device) 自动配置:
  - 学习到NAT公共IP地址 (138.76.29.7)
  - 在NAT转换表中，增删端口映射



NAT穿透问题

- 解决方案3: 中继(如Skype)
  - NAT内部的客户与中继服务器建立连接
  - 外部客户也与中继服务器建立连接
  - 中继服务器桥接两个连接的分组



哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场

谢 谢！

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场

计算机网络之探赜索隐

主讲人：李全龙

本讲主题

互联网控制报文协议(ICMP)

互联网控制报文协议(ICMP)

- 互联网控制报文协议 ICMP (Internet Control Message Protocol)支持主机或路由器：
  - 差错(或异常)报告
  - 网络探询
- 两类ICMP 报文:
  - 差错报告报文(5种)
    - 目的不可达
    - 源抑制(Source Quench)
    - 超时/超期
    - 参数问题
    - 重定向 (Redirect)
  - 网络探询报文(2组)
    - 回声(Echo)请求与应答报文(Reply)
    - 时间戳请求与应答报文

ICMP报文

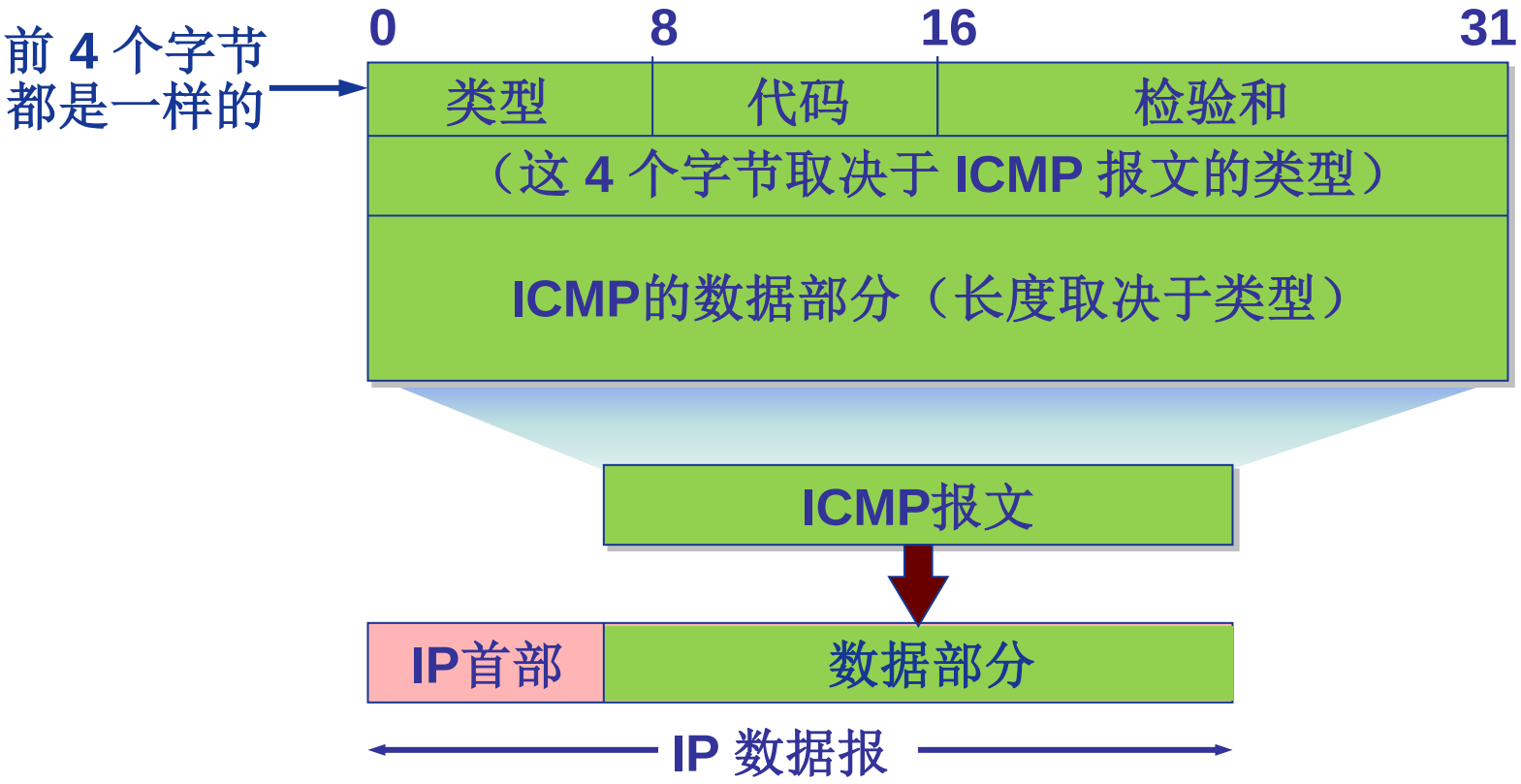
类型(Type)	编码(Code)	description
0	0	回声应答 (ping)
3	0	目的网络不可达
3	1	目的主机不可达
3	2	目的协议不可达
3	3	目的端口不可达
3	6	目的网络未知
3	7	目的主机未知
4	0	源抑制(拥塞控制-未用)
8	0	回声请求(ping)
9	0	路由通告
10	0	路由发现
11	0	TTL超期
12	0	IP首部错误

例外情况

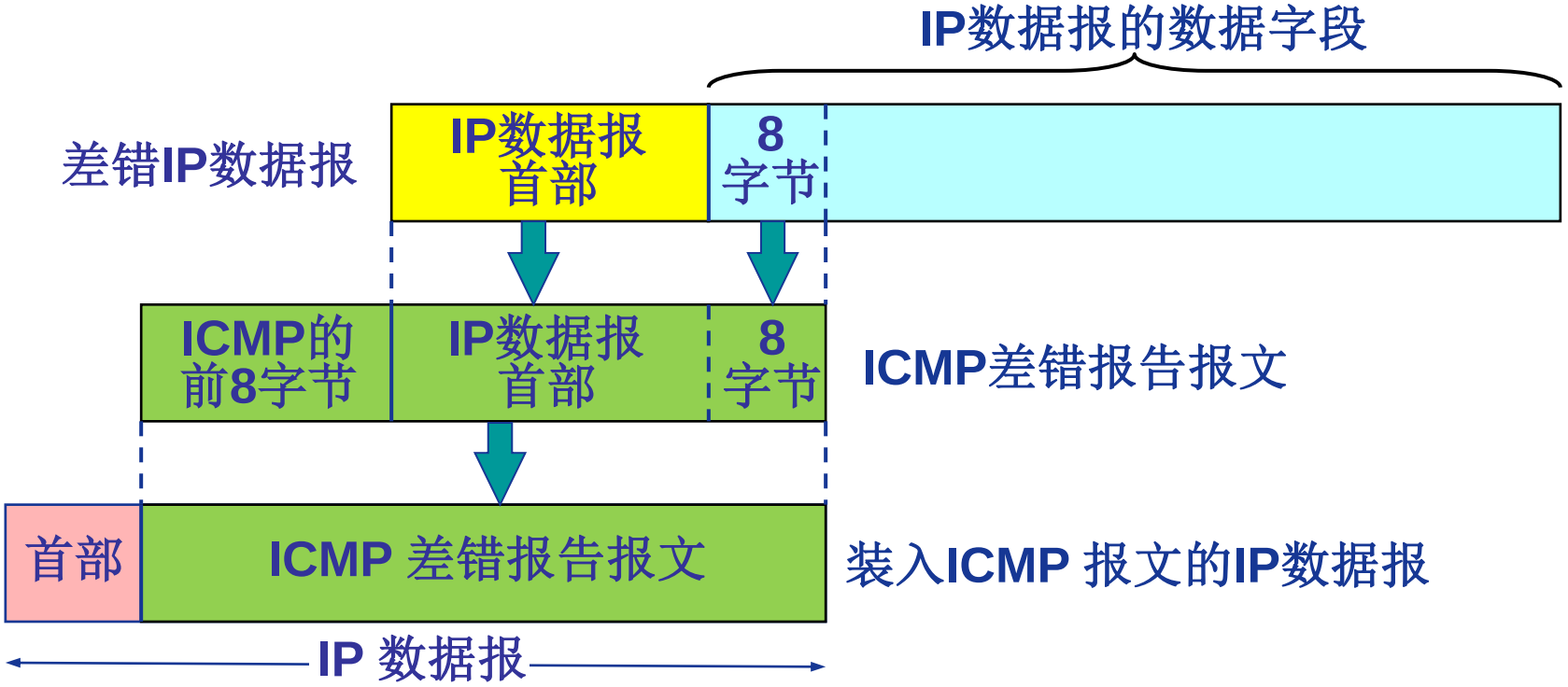
- 几种不发送 ICMP差错报告报文的特殊情况：
  - 对ICMP差错报告报文不再发送 ICMP差错报告报文
  - 除第1个IP数据报分片外，对所有后续分片均不发送ICMP差错报告报文
  - 对所有多播IP数据报均不发送 ICMP差错报告报文
  - 对具有特殊地址（如127.0.0.0 或 0.0.0.0）的IP数据报不发送 ICMP 差错报告报文
- 几种 ICMP 报文已不再使用
  - 信息请求与应答报文
  - 子网掩码请求和应答报文
  - 路由器询问和通告报文

ICMP报文的格式

- ICMP报文封装到IP数据报中传输



ICMP差错报告报文数据封装

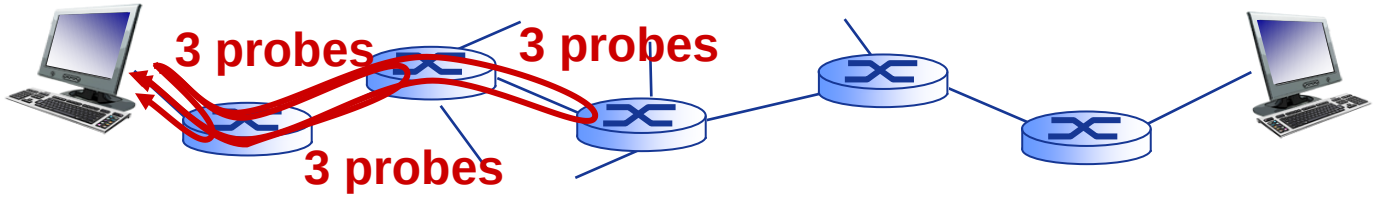


ICMP的应用举例：Traceroute

- 源主机向目的主机发送一系列UDP数据报
  - 第1组IP数据报TTL =1
  - 第2组IP数据报TTL=2, etc.
  - 目的端口号为不可能使用的端口号
- 当第*n*组数据报(TTL=*n*)到达第*n*个路由器时：
  - 路由器丢弃数据报
  - 向源主机发送ICMP报文(type=11, code=0)
  - ICMP报文携带路由器名称和IP地址信息
- 当ICMP报文返回到源主机时，记录RTT

停止准则：

- **UDP数据报最终到达目的主机**
- **目的主机返回“目的端口不可达”ICMP报文 (type=3, code=3)**
- **源主机停止**



哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场

谢 谢！

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场

计算机网络之探赜索隐

主讲人：李全龙

本讲主题

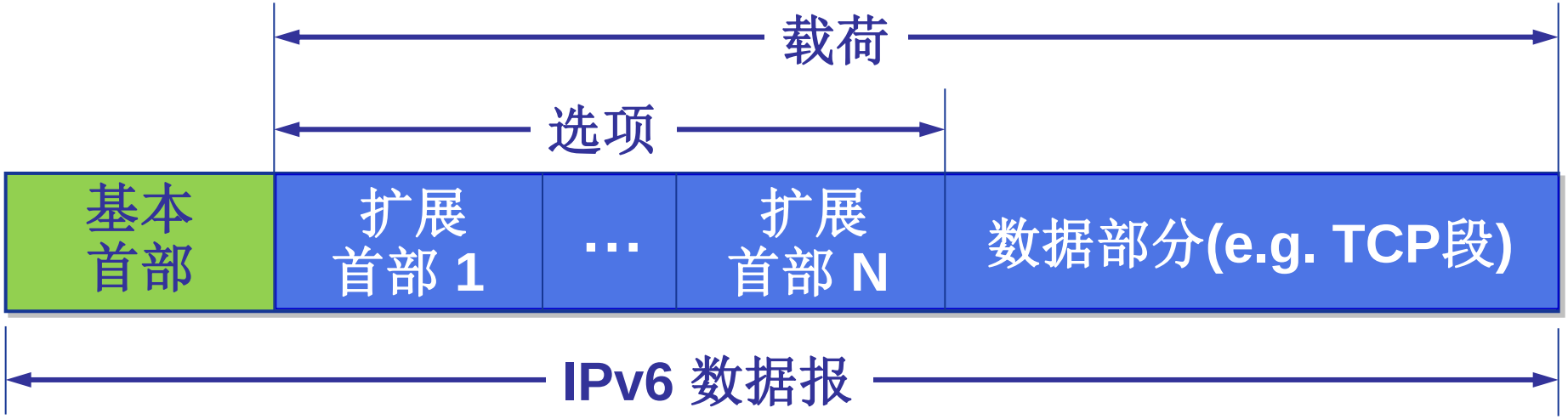
IPv6简介

IPv6：动机

- 最初动机: 32位IPv4地址空间已分配殆尽
- 其他动机: 改进首部格式
  - 快速处理/转发数据报
  - 支持QoS

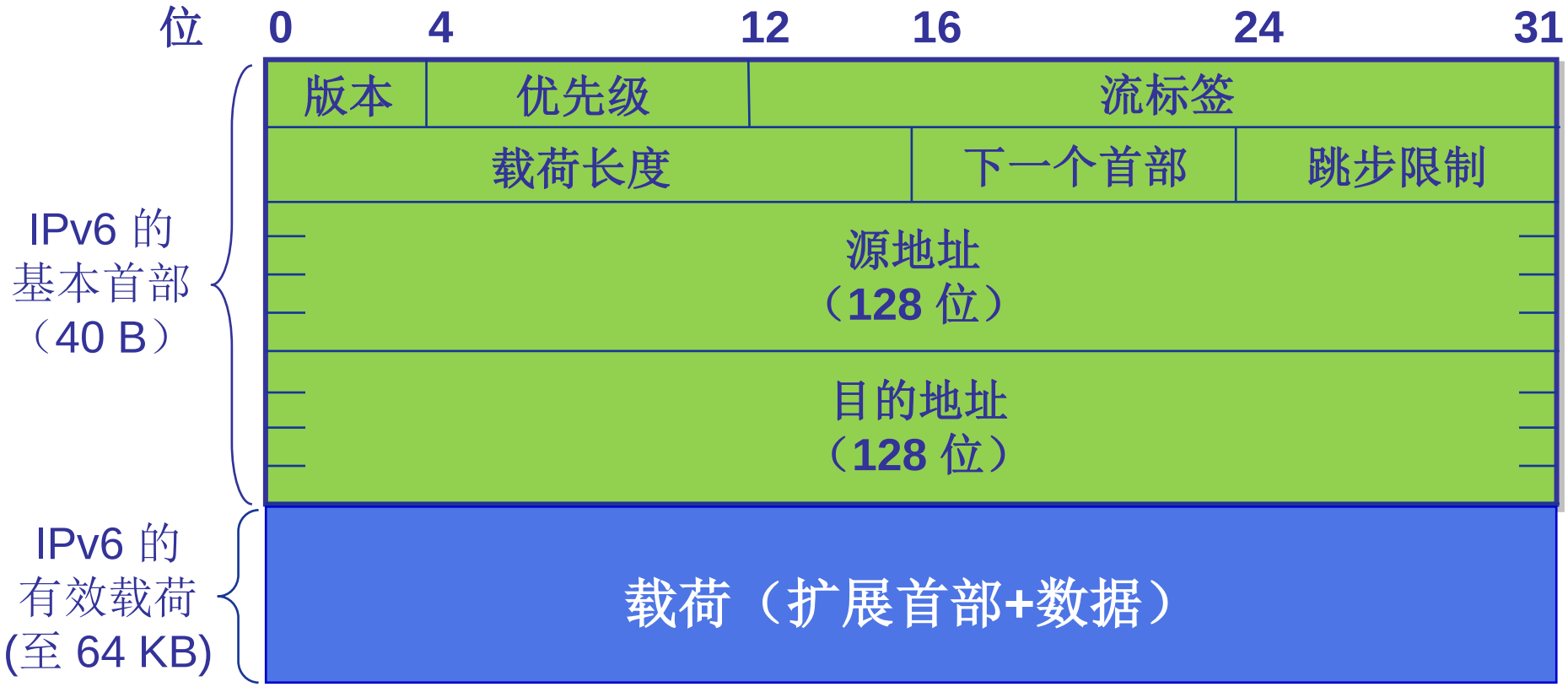
IPv6数据报格式:

- 固定长度的40字节基本首部
- 不允许分片



IPv6数据报格式

优先级(priority): 标识数据报的优先级
流标签(flow Label): 标识同一“流”中的数据报
下一个首部(next header): 标识下一个选项首部或上层协议首部(如TCP首部)



其他改变 vs IPv4

- 校验和(checksum): 彻底移除，以减少每跳处理时间
- 选项(options): 允许，但是从基本首部移出，定义多个选项首部，通过“下一个首部”字段指示
- ICMPv6: 新版ICMP
  - 附加报文类型，e.g. “Packet Too Big”
  - 多播组管理功能

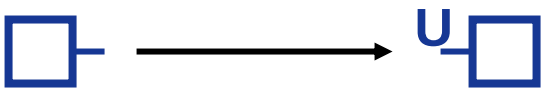
IPv6地址表示形式

- 一般形式: 1080:0:FF:0:8:800:200C:417A
- 压缩形式: FF01:0:0:0:0:0:0:43
  压缩→FF01::43
- IPv4-嵌入形式: 0:0:0:0:0:FFFF:13.1.68.3
  或 ::FFFF:13.1.68.3
- 地址前缀: 2002:43c:476b::/48
  (注: IPv6不再使用掩码!)
- URLs: http://[3FFE::1:800:200C:417A]:8000

IPv6基本地址类型

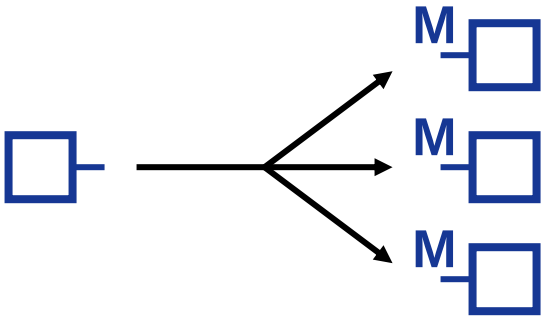
单播(unicast):

一对一通信



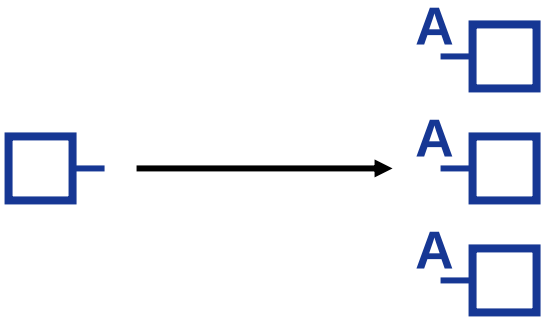
多播(multicast):

一对多通信



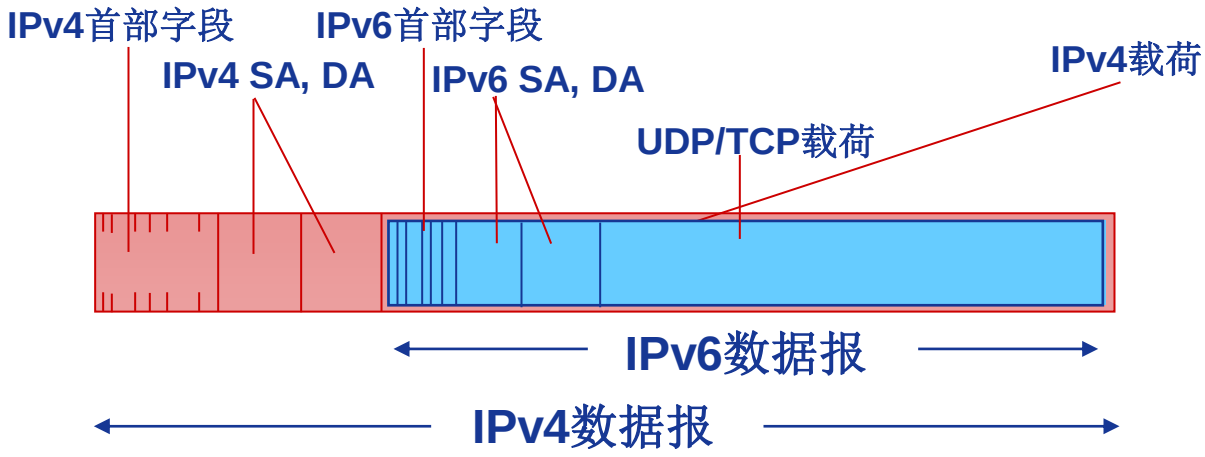
任意播(anycast):

一对一组之一（最近一个）通信

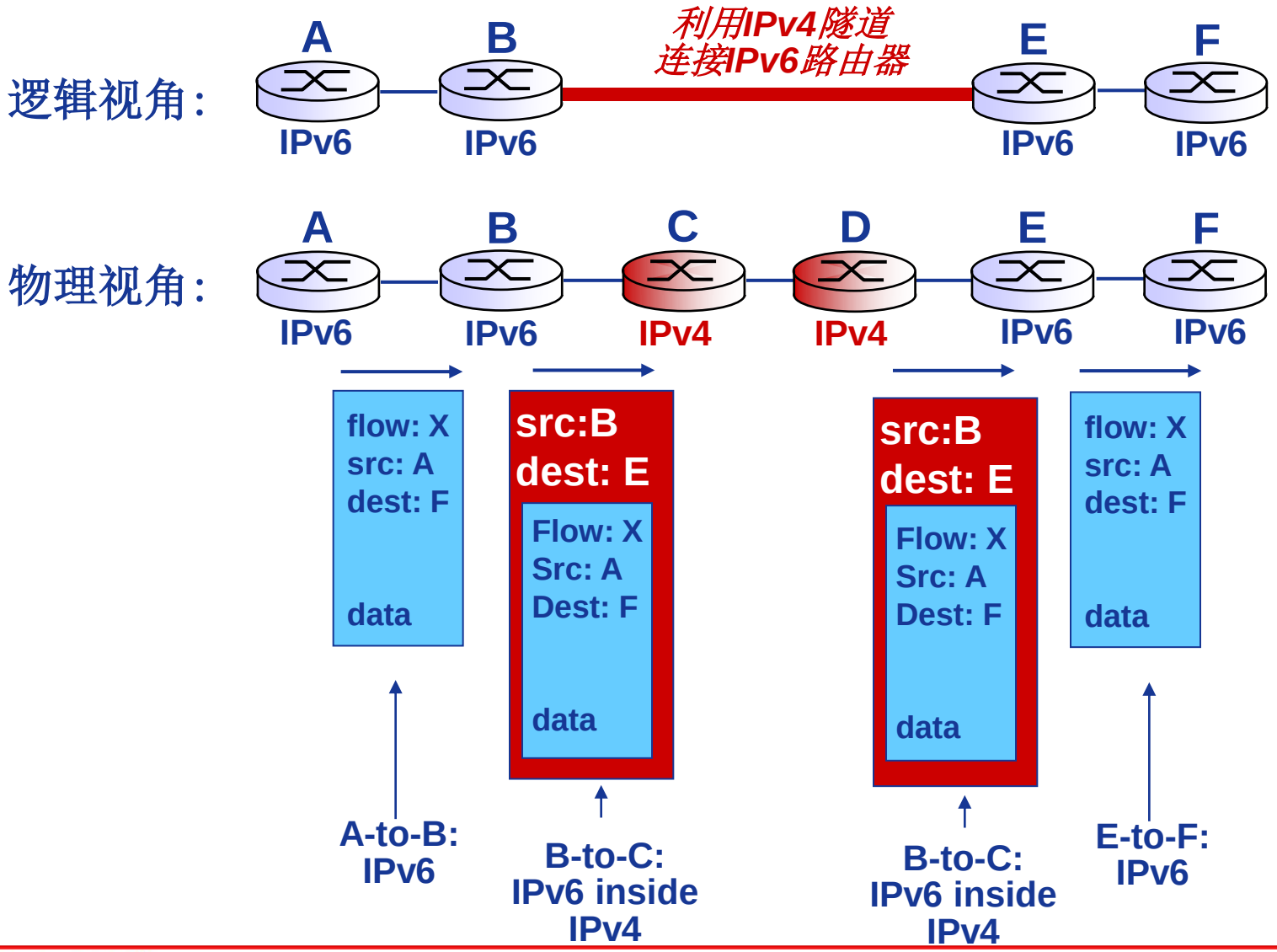


IPv4向IPv6过渡

- 不可能在某个时刻所有路由器同时被更新为IPv6
  - 不会有 “标志性的日期”
  - IPv4和IPv6路由器共存的网络如何运行？
- 隧道(tunneling): IPv6数据报作为IPv4数据报的载荷进行封装，穿越IPv4网络



隧道（tunneling）



哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场

谢 谢！

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY
立足航天，服务国防，面向国民经济主战场

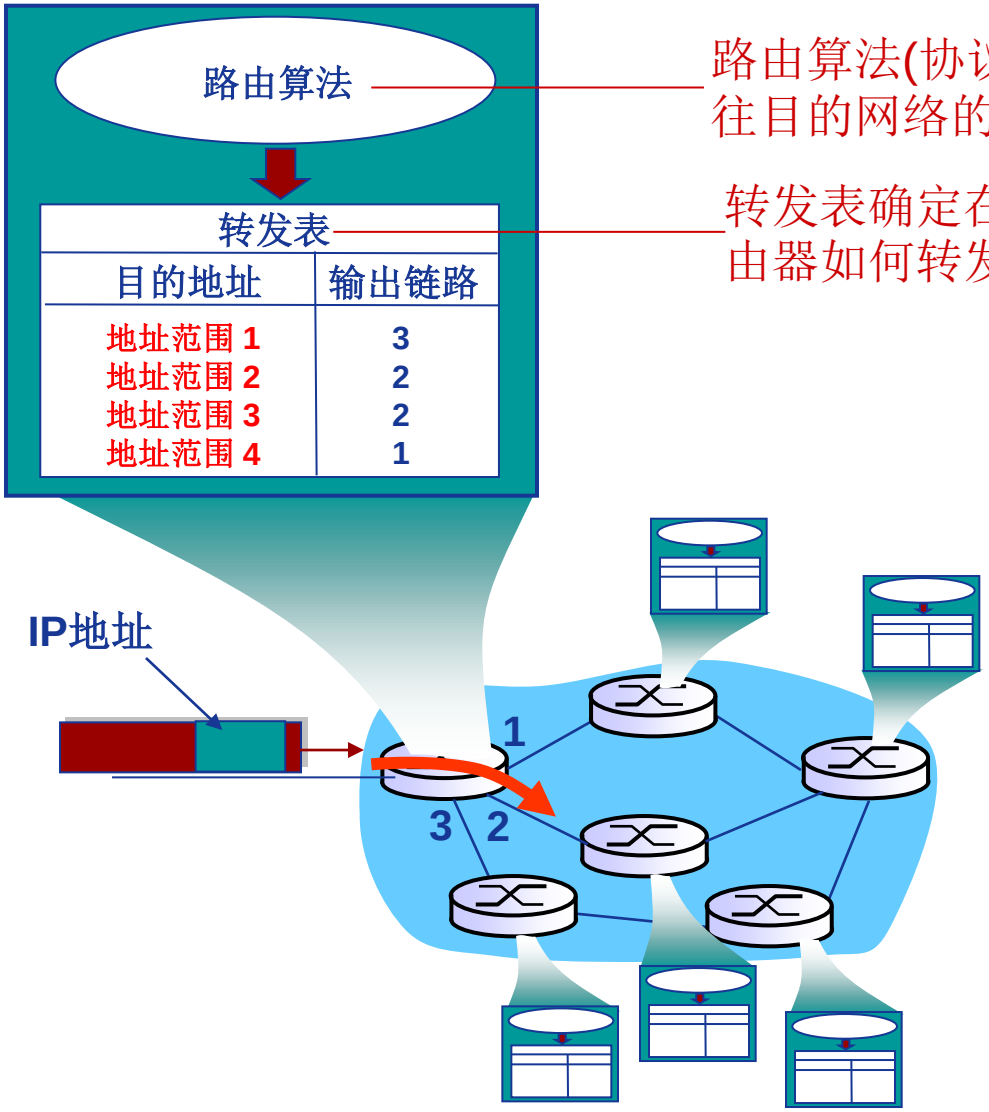
计算机网络之探赜索隐

主讲人：李全龙

本讲主题

路由算法

路由与转发



路由算法(协议)确定去往目的网络的最佳路径

转发表确定在本路由器如何转发分组

网络抽象：图

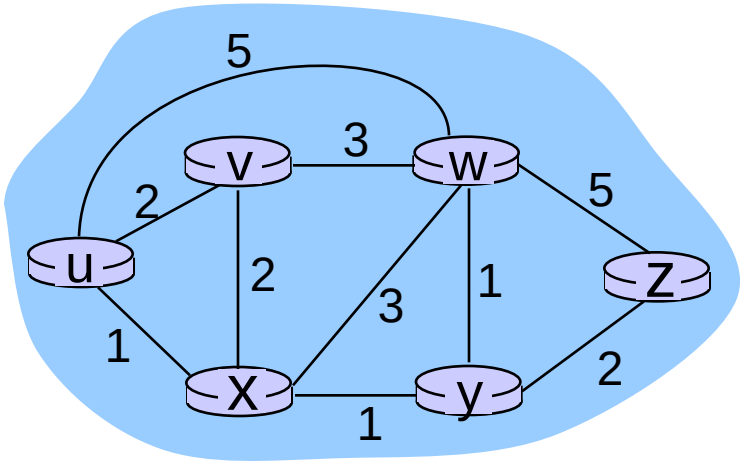


图: G = (N, E)

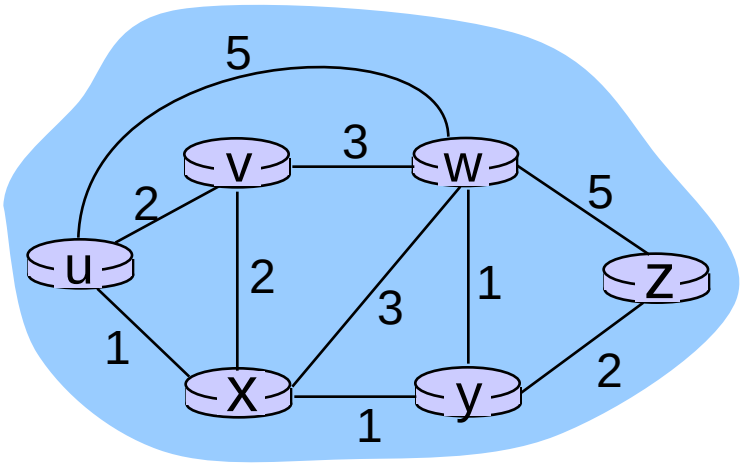
N = 路由器集合= { u, v, w, x, y, z }

E = 链路集合 ={ (u,v), (u,x), (v,x), (v,w), (x,w), (x,y), (w,y), (w,z), (y,z) }

附注: 图的抽象在网络领域应用很广泛

E.g.:P2P，其中，N是 peers集合，而E是TCP连接集合

图抽象：费用(Costs)



c(x, x') = 链路(x, x')的费用
e.g., c(w, z) = 5

每段链路的费用可以总是1，

或者是，

带宽的倒数、拥塞程度等

路径费用：$(x_1, x_2, x_3, \ldots, x_p) = c(x_1,x_2) + c(x_2,x_3) + \ldots + c(x_{p-1},x_p)$

关键问题: 源到目的（如u到z）的最小费用路径是什么？
路由算法: 寻找最小费用路径的算法

路由算法分类

静态路由 vs 动态路由？

静态路由：

- 手工配置
- 路由更新慢
- 优先级高

动态路由：

- 路由更新快
  - 定期更新
  - 及时响应链路费用或网络拓扑变化

全局信息 vs 分散信息？

全局信息：

- 所有路由器掌握完整的网络拓扑和链路费用信息
- **E.g. 链路状态(LS)路由算法**

分散(decentralized)信息：

- 路由器只掌握物理相连的邻居以及链路费用
- 邻居间信息交换、运算的迭代过程
- **E.g. 距离向量(DV)路由算法**

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场

谢 谢！

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY
立足航天，服务国防，面向国民经济主战场

计算机网络之探赜索隐

主讲人：李全龙

本讲主题

链路状态路由算法

网络抽象：图

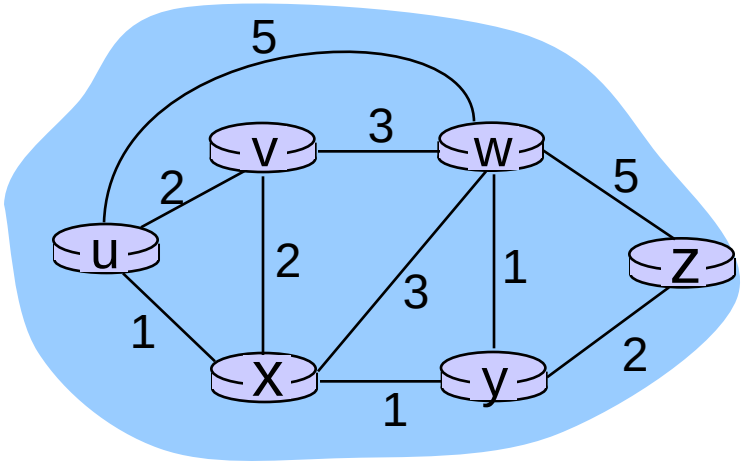


图: G = (N, E)

N = 路由器集合= { u, v, w, x, y, z }

E = 链路集合 ={ (u,v), (u,x), (v,x), (v,w), (x,w), (x,y), (w,y), (w,z), (y,z) }

链路状态路由算法

Dijkstra 算法

- 所有结点(路由器)掌握网络拓扑和链路费用
  - 通过“链路状态广播”
  - 所有结点拥有相同信息
- 计算从一个结点(“源”)到达所有其他结点的最短路径
  - 获得该结点的转发表
- 迭代: k次迭代后，得到到达k个目的结点的最短路径

符号:

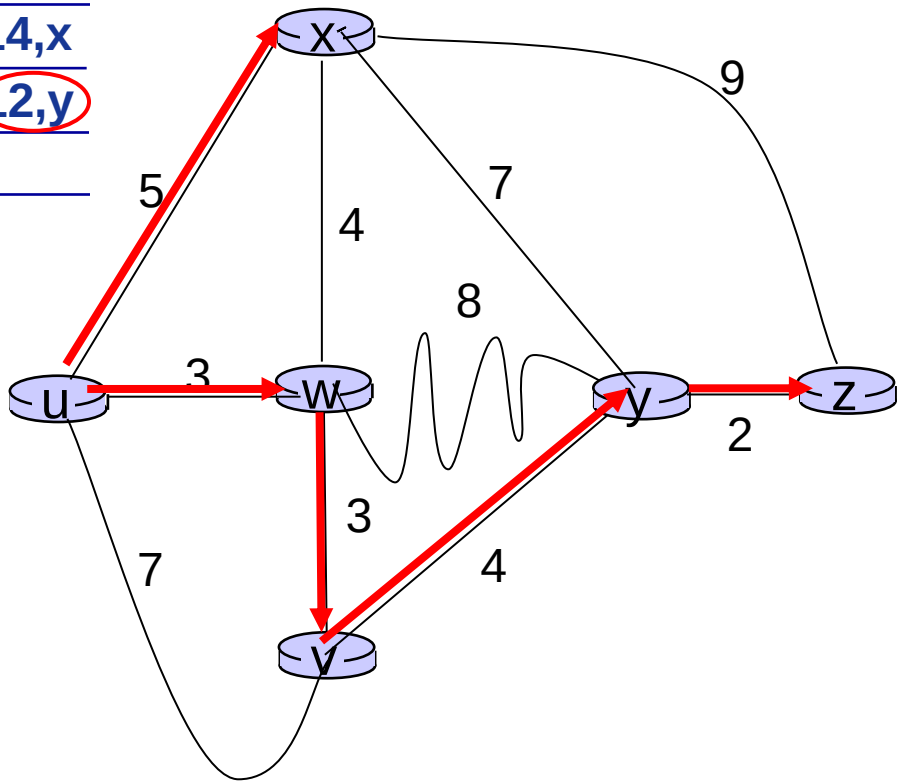
- c(x,y): 结点x到结点y链路费用；如果x和y不直接相连，则=∞
- D(v): 从源到目的v的当前路径费用值
- p(v): 沿从源到v的当前路径，v的前序结点
- N’: 已经找到最小费用路径的结点集合

Dijkstra 算法

```
1  初始化:
2    N' = {u}
3    for 所有结点v
4      if v毗邻u
5          then D(v) = c(u,v)
6      else D(v) = ∞
7
8   Loop
9     找出不在 N'中的w ， 满足D(w)最小
10    将w加入N'
11    更新w的所有不在N'中的邻居v的D(v) :
12       D(v) = min( D(v), D(w) + c(w,v) )
13    /*到达v的新费用或者是原先到达v的费用，或者是
14       已知的到达w的最短路径费用加上w到v的费用 */
15  until 所有结点在N'中
```

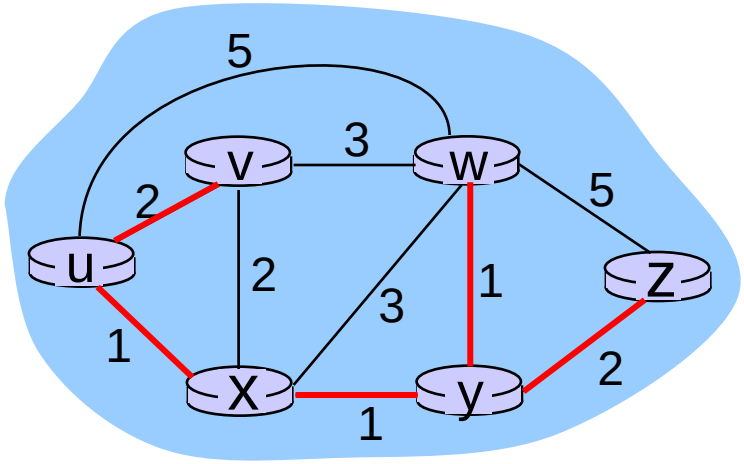
Dijkstra 算法:例1

Step	N'	D(v) p(v)	D(w) p(w)	D(x) p(x)	D(y) p(y)	D(z) p(z)
0	u	7,u	3,u	5,u	∞	∞
1	uw	6,w		5,u	11,w	∞
2	uwx	6,w			11,w	14,x
3	uwxv				10,v	14,x
4	uwxvy					12,y
5	uwxvyz					



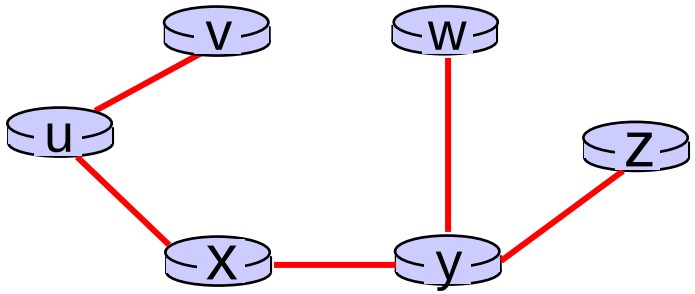
Dijkstra 算法:例2

Step	N'	D(v),p(v)	D(w),p(w)	D(x),p(x)	D(y),p(y)	D(z),p(z)
0	u	2,u	5,u	1,u	∞	∞
1	ux	2,u	4,x		2,x	∞
2	uxy	2,u	3,y			4,y
3	uxyv		3,y			4,y
4	uxyvw					4,y
5	uxyvwz					



Dijkstra 算法:例2

u的最终最短路径树:



u的最终转发表:

目的	链路
v	(u,v)
x	(u,x)
y	(u,x)
w	(u,x)
z	(u,x)

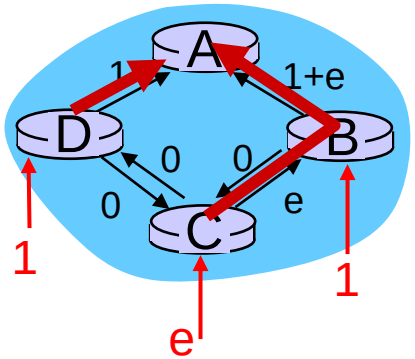
Dijkstra 算法:讨论

算法复杂性: n个结点

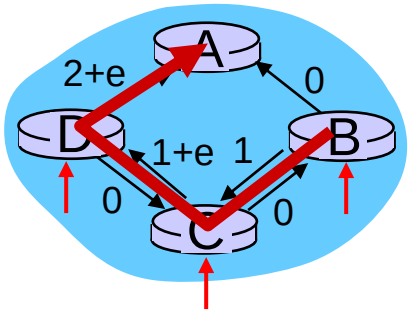
- 每次迭代: 需要检测所有不在集合N'中的结点w
- n(n+1)/2次比较: $O(n^2)$
- 更高效的实现: O(nlogn)

存在震荡(oscillations)可能:

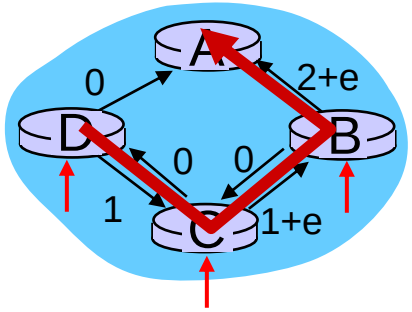
- e.g., 假设链路费用是该链路承载的通信量:



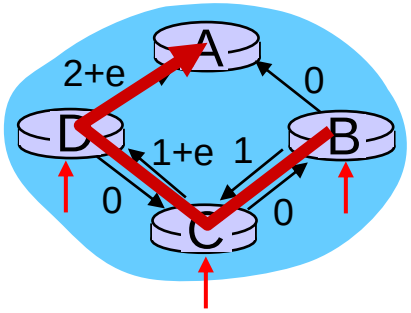
初始



基于这些费用
计算新的路由…



基于这些费用
计算新的路由…



基于这些费用
计算新的路由…

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场

谢 谢！

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY
立足航天，服务国防，面向国民经济主战场

计算机网络之探赜索隐

主讲人：李全龙

本讲主题

距离向量路由算法(1)

距离向量(Distance Vector)路由算法

Bellman-Ford方程(动态规划)

令：

$d_x(y)$:=从x到y最短路径的费用（距离）

则：

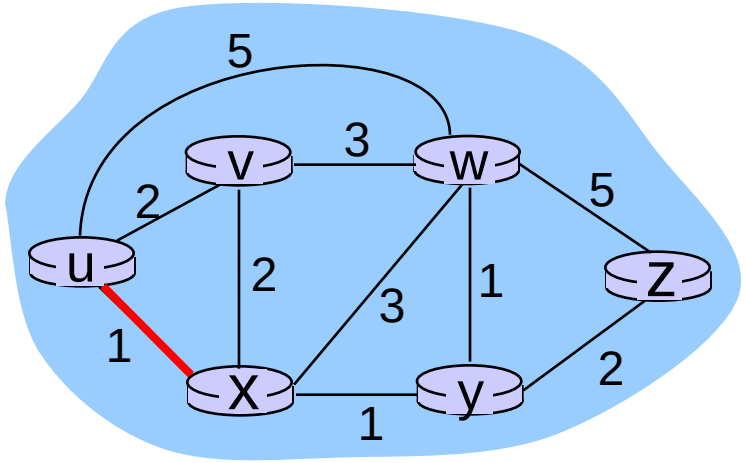
$$d_x(y) = \min_v \{c(x,v) + d_v(y)\}$$

在x的所有邻居v中取最小值

x到邻居v的费用

从邻居v到达目的y的费用（距离）

Bellman-Ford 举例



显然: $d_v(z) = 5$, $d_x(z) = 3$, $d_w(z) = 3$

根据B-F方程:

$$d_u(z) = \min \{ c(u,v) + d_v(z), \; c(u,x) + d_x(z), \; c(u,w) + d_w(z) \}$$
$$= \min \{2 + 5, \; 1 + 3, \; 5 + 3\} = 4$$

重点：结点获得最短路径的下一跳，该信息用于转发表中！

距离向量路由算法

- $D_x(y)$ = 从结点x到结点y的最小费用估计
  - x维护距离向量(DV): $\mathbf{D}_x = [D_x(y): y \in N]$
- 结点x:
  - 已知到达每个邻居的费用: c(x,v)
  - 维护其所有邻居的距离向量: $\mathbf{D}_v = [D_v(y): y \in N]$

核心思想:

- 每个结点不定时地将其自身的DV估计发送给其邻居
- 当x接收到邻居的新的DV估计时，即依据B-F更新其自身的距离向量估计：

$$D_x(y) \leftarrow \min_v\{c(x,v) + D_v(y)\} \quad \text{for each node } y \in N$$

- $D_x(y)$将最终收敛于实际的最小费用 $d_x(y)$

距离向量路由算法

异步迭代:

- 引发每次局部迭代的因素
  - 局部链路费用改变
  - 来自邻居的DV更新

分布式:

- 每个结点只当DV变化时才通告给邻居
  - 邻居在必要时（其DV更新后发生改变）再通告它们的邻居

每个结点:

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场

谢 谢！

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY
立足航天，服务国防，面向国民经济主战场

计算机网络之探赜索隐

主讲人：李全龙

本讲主题

距离向量路由算法（2）

距离向量路由算法：举例

node x table

from \ cost to	x	y	z
x	0	2	7
y	∞	∞	∞
z	∞	∞	∞

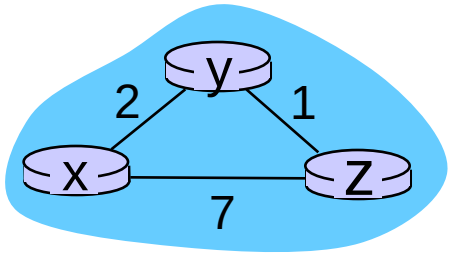
from \ cost to	x	y	z
x	0	2	3
y	2	0	1
z	7	1	0

$Dx(y) = min\{c(x,y) + Dy(y), c(x,z) + Dz(y)\}$
$= min\{2+0 , 7+1\} = 2$

$D_x(z) = \min\{c(x,y) + D_y(z), c(x,z) + D_z(z)\}$
$= \min\{2+1 , 7+0\} = 3$

node y table

from \ cost to	x	y	z
x	∞	∞	∞
y	2	0	1
z	∞	∞	∞

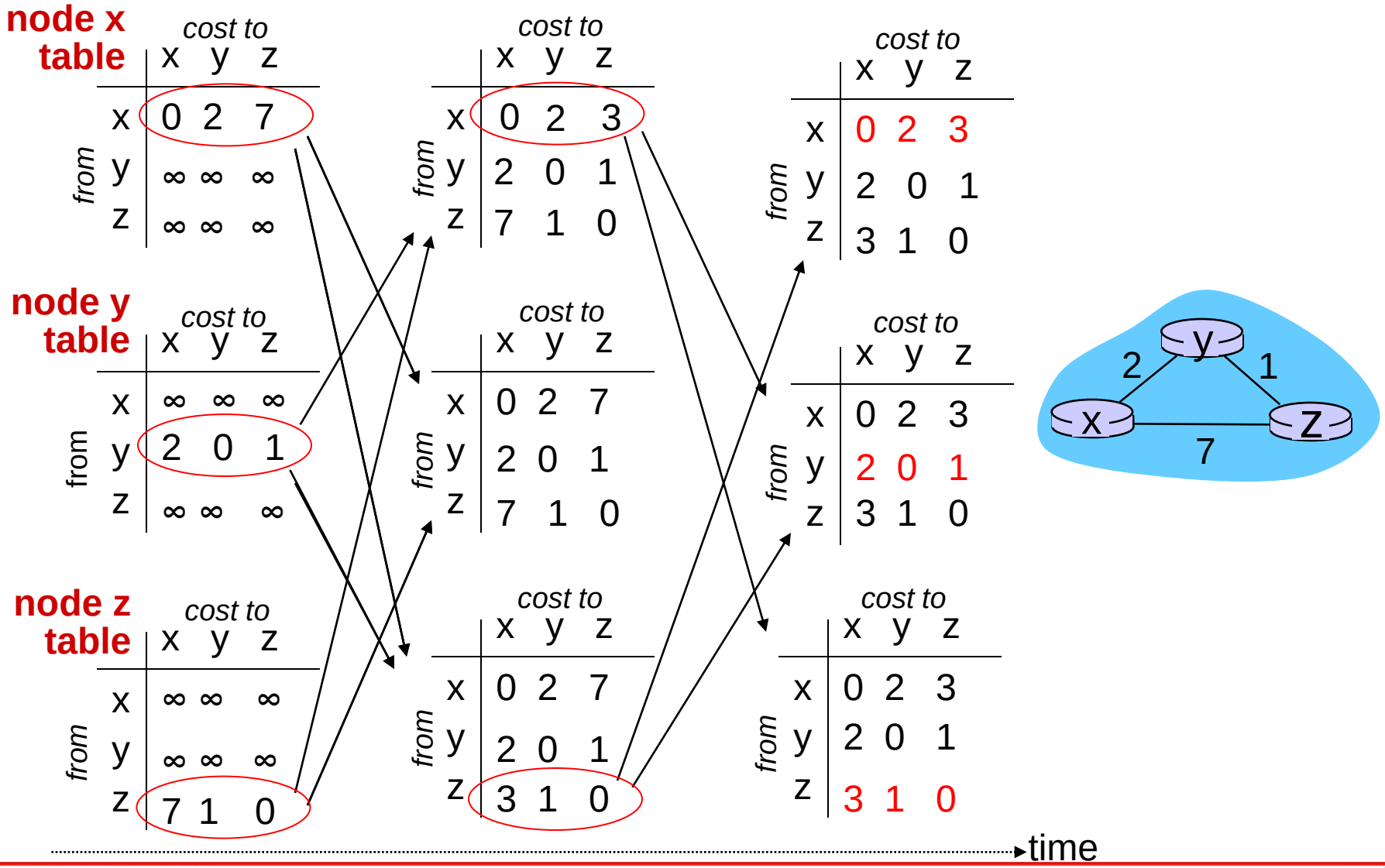


node z table

from \ cost to	x	y	z
x	∞	∞	∞
y	∞	∞	∞
z	7	1	0

time

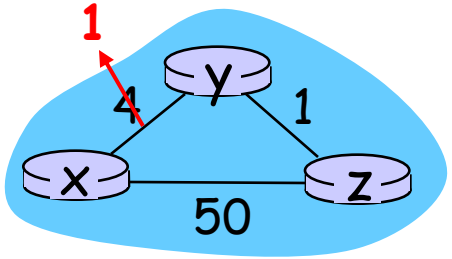
距离向量路由算法：举例


node x table
node y table
node z table
cost to
x y z
from
x 0 2 7
y ∞ ∞ ∞
z ∞ ∞ ∞
x 0 2 3
y 2 0 1
z 7 1 0
x 0 2 3
y 2 0 1
z 3 1 0
x ∞ ∞ ∞
y 2 0 1
z ∞ ∞ ∞
x 0 2 7
y 2 0 1
z 7 1 0
x 0 2 3
y 2 0 1
z 3 1 0
x ∞ ∞ ∞
y ∞ ∞ ∞
z 7 1 0
x 0 2 7
y 2 0 1
z 3 1 0
x 0 2 3
y 2 0 1
z 3 1 0
y
x
z
2
1
7
time

距离向量DV: 链路费用变化

链路费用变化:

- 结点检测本地链路费用变化
- 更新路由信息，重新计算距离向量
- 如果DV改变，通告所有邻居



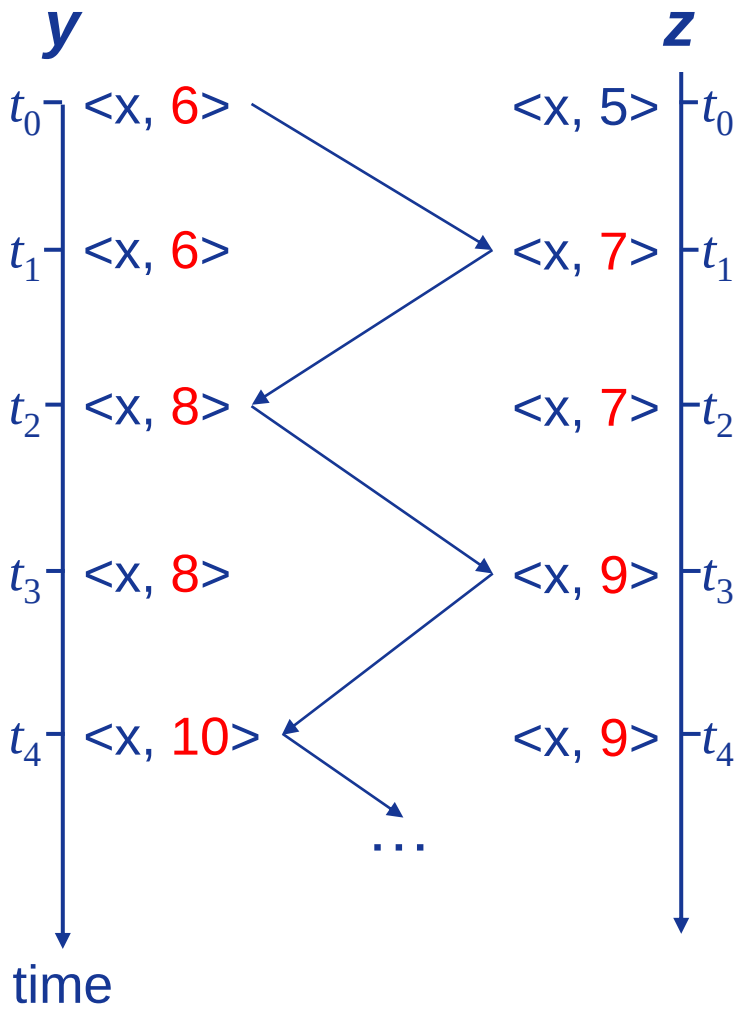
t_0 : *y*检测到链路费用改变，更新DV，通告其邻居.

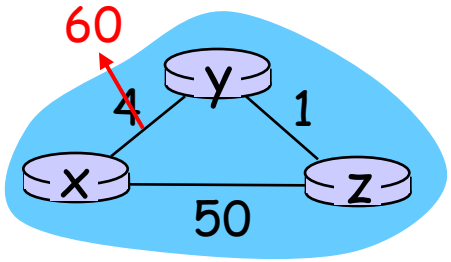
t_1 : *z*收到*y*的DV更新，更新其距离向量表，计算到达*x*的最新最小费用，更新其DV，并发送给其所有邻居.

t_2 : *y*收到*z*的DV更新，更新其距离向量表，重新计算*y*的DV，未发生改变，不再向*z*发送*DV.*

“☺好消息传播快！”　　“坏消息会怎么样呢？”

距离向量DV: 无穷计数问题





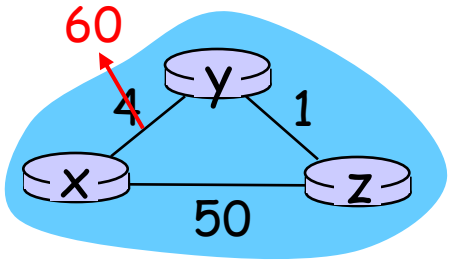
坏消息传播慢！

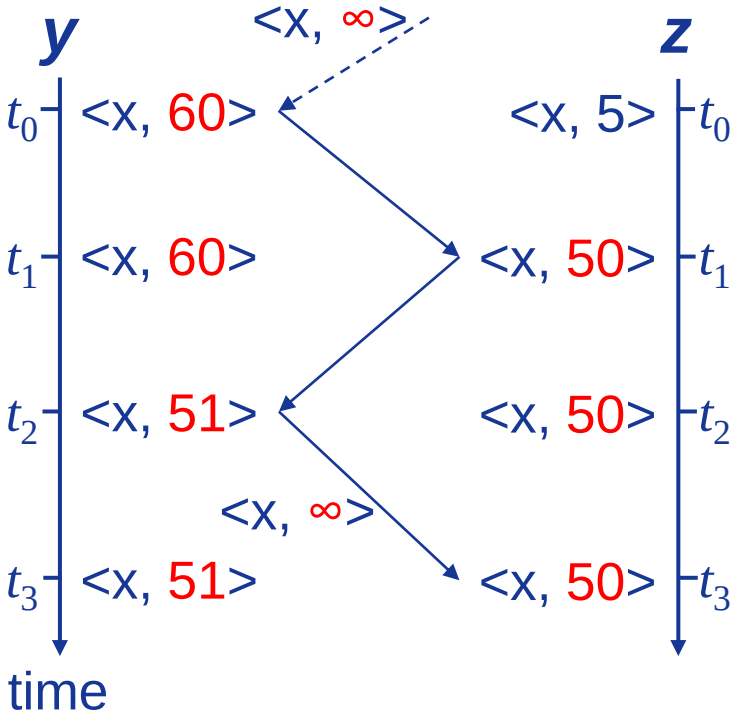
— “无穷计数(count to infinity)”问题！

距离向量DV: 无穷计数问题

毒性逆转(poisoned reverse):

- 如果一个结点(e.g. Z)到达某目的(e.g.X)的最小费用路径是通过某个邻居(e.g.Y)，则：
  - 通告给该邻居结点到达该目的的距离为无穷大



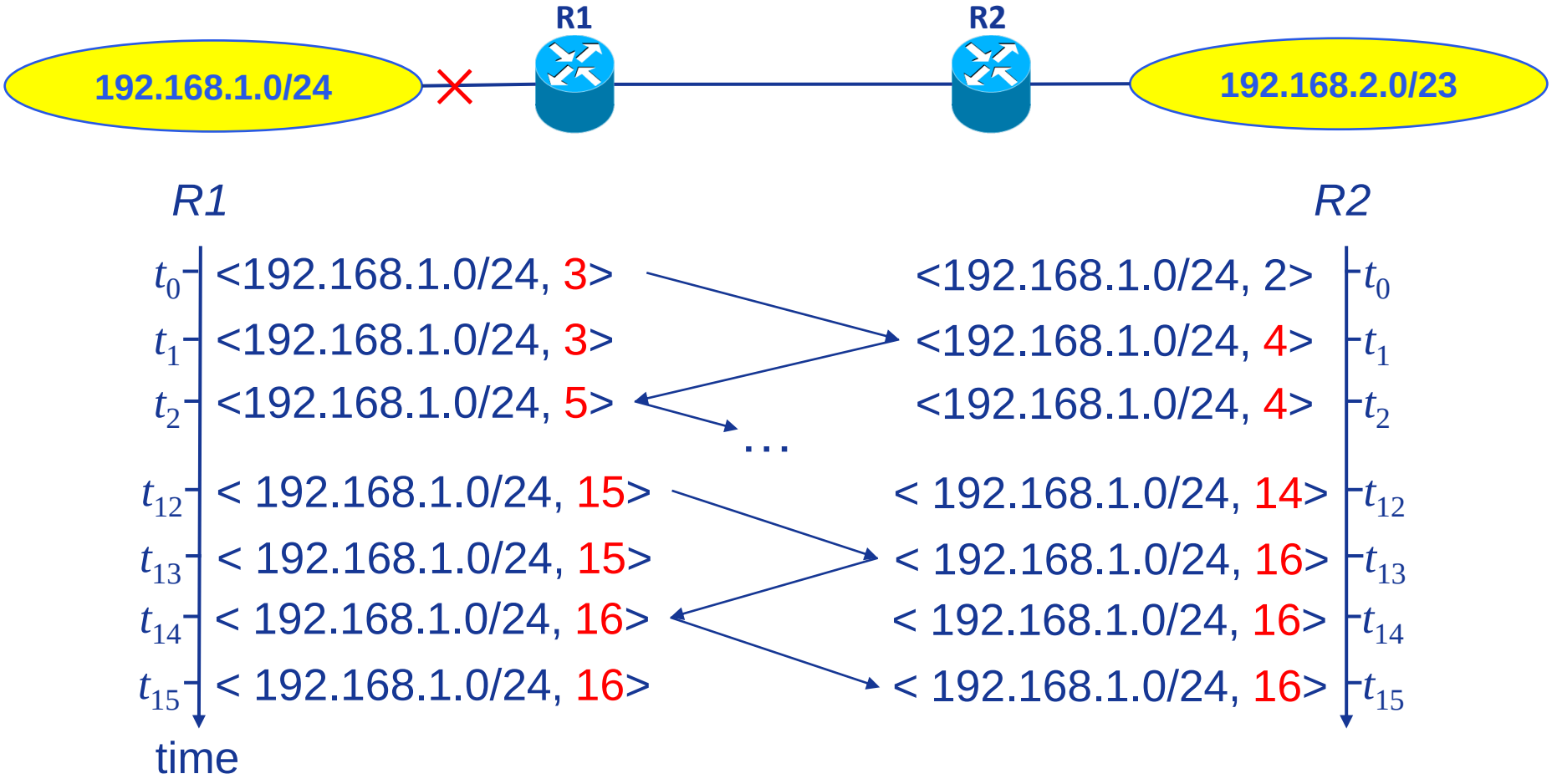


毒性逆转能否彻底解决无穷计数问题？

距离向量DV: 无穷计数问题

定义最大度量(maximum metric):

- 定义一个最大的有效费用值，如15跳步，16跳步表示∞



哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场

谢 谢！

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场

计算机网络之探赜索隐

主讲人：李全龙

本讲主题

层次路由

层次路由

将任意规模网络抽象为一个图计算路由-过于理想化

- 标识所有路由器
- “扁平”网络

——在实际网络（尤其是大规模网络）中，不可行！

网络规模：考虑6亿目的结点的网络

- 路由表几乎无法存储！
- 路由计算过程的信息（e.g. 链路状态分组、DV）交换量巨大，会淹没链路！

管理自治：

- 每个网络的管理可能都期望自主控制其网内的路由
- 互联网(internet) = 网络之网络(network of networks)

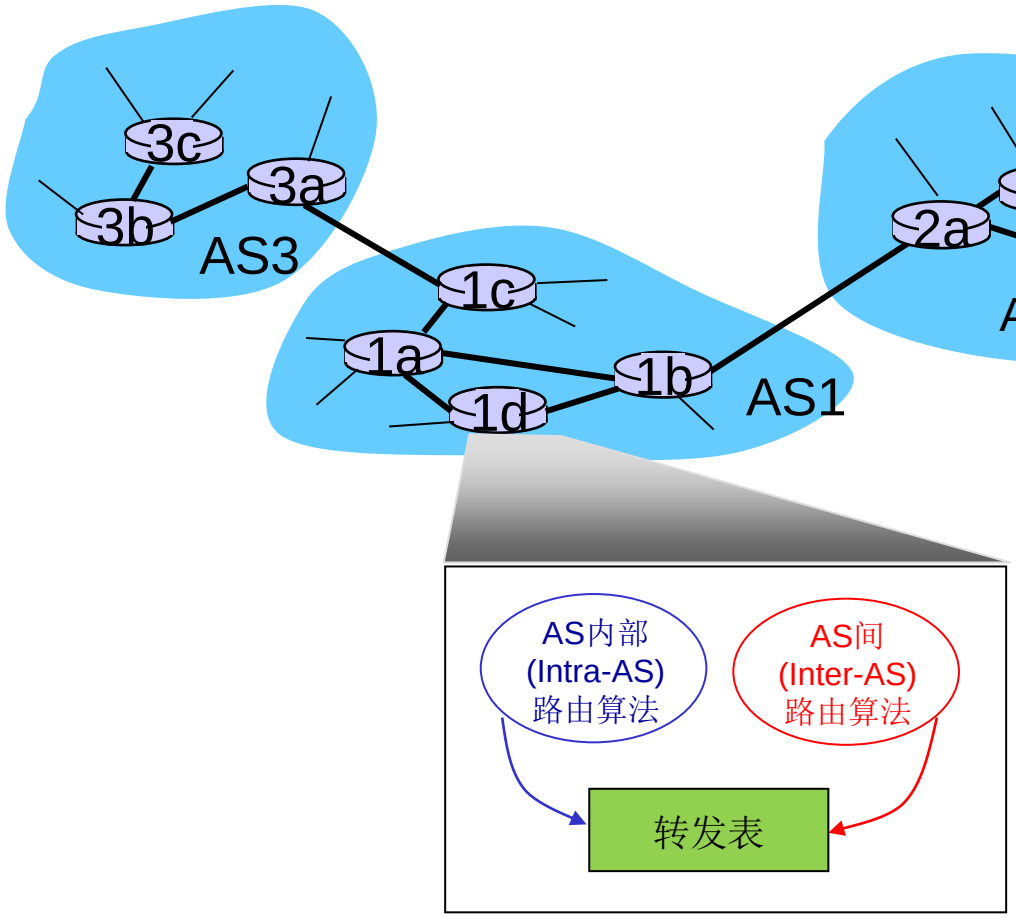
层次路由

- 聚合路由器为一个区域：自治系统AS (autonomous systems)
- 同一AS内的路由器运行相同的路由协议(算法)
  - 自治系统内部路由协议(“intra-AS” routing protocol)
  - 不同自治系统内的路由器可以运行不同的AS内部路由协议

网关路由器(gateway router):

- 位于AS“边缘”
- 通过链路连接其他AS的网关路由器

互连的AS



- 转发表由AS内部路由算法与AS间路由算法共同配置
  - AS内部路由算法设置AS内部目的网络路由入口(entries)
  - AS内部路由算法与AS间路由算法共同设置AS外部目的网络路由入口

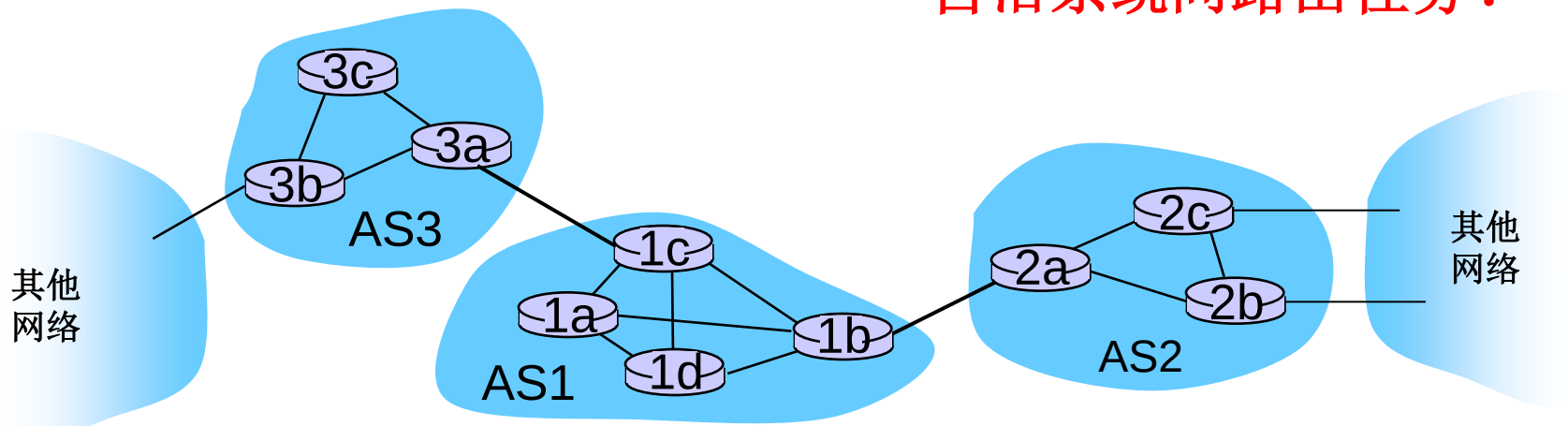
自治系统间(Inter-AS)路由任务

- 假设AS1内某路由器收到一个目的地址在AS1之外的数据报:
  - 路由器应该将该数据报转发给哪个网关路由器呢？

AS1必须:

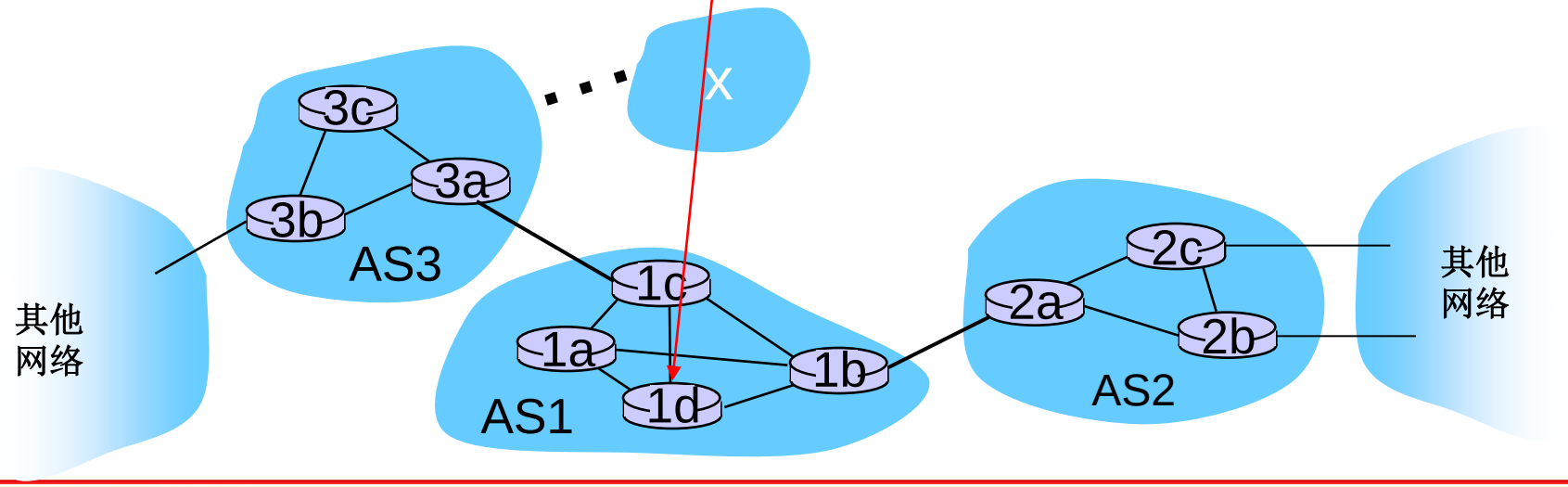
1. 学习到哪些目的网络可以通过AS2到达，哪些可以通过AS3到达
2. 将这些网络可达性信息传播给AS1内部路由器

自治系统间路由任务！



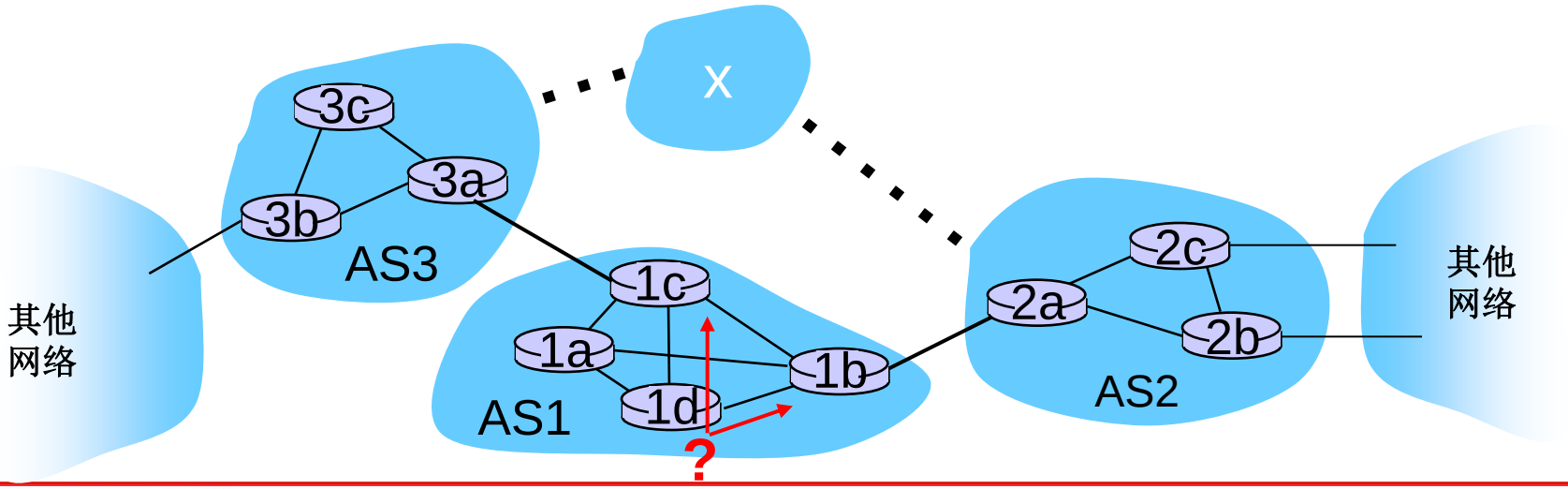
例: 路由器1d的转发表设置

- 假设AS1学习到(通过AS间路由协议)：子网*x*可以通过AS3 (网关 1c)到达，但不能通过AS2到达
  - AS间路由协议向所有内部路由器传播该可达性信息
- 路由器1d：利用AS内部路由信息，确定其到达1c的最小费用路径接口*I*
  - 在转发表中增加入口：(*x*, *I*)



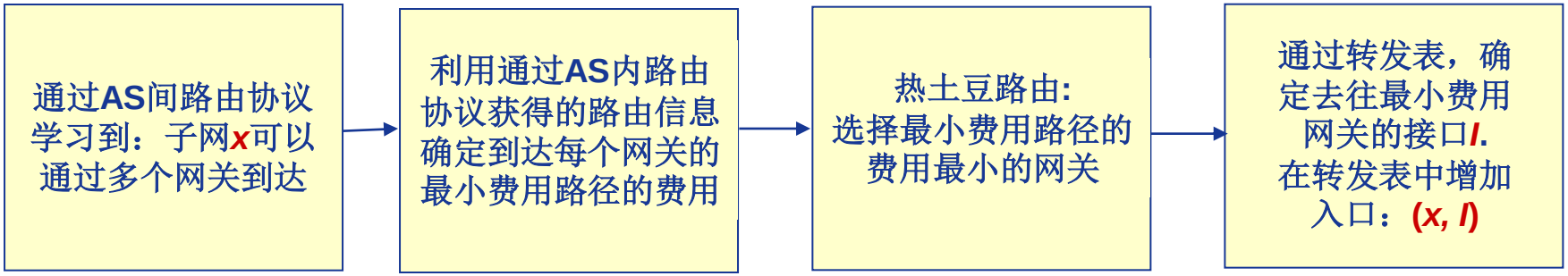
例: 在多AS间选择

- 假设AS1通过AS间路由协议学习到：子网*x*通过AS3和AS2均可到达
- 为了配置转发表，路由器1d必须确定应该将去往子网*x*的数据报转发给哪个网关？
  - 这个任务也是由AS间路由协议完成!



例: 在多AS间选择

- 假设AS1通过AS间路由协议学习到：子网*x*通过AS3和AS2均可到达
- 为了配置转发表，路由器1d必须确定应该将去往子网*x*的数据报转发给哪个网关？
  - 这个任务也是由AS间路由协议完成!
- **热土豆路由: 将分组发送给最近的网关路由器.**



哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场

谢 谢！

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场

计算机网络之探赜索隐

主讲人：李全龙

本讲主题

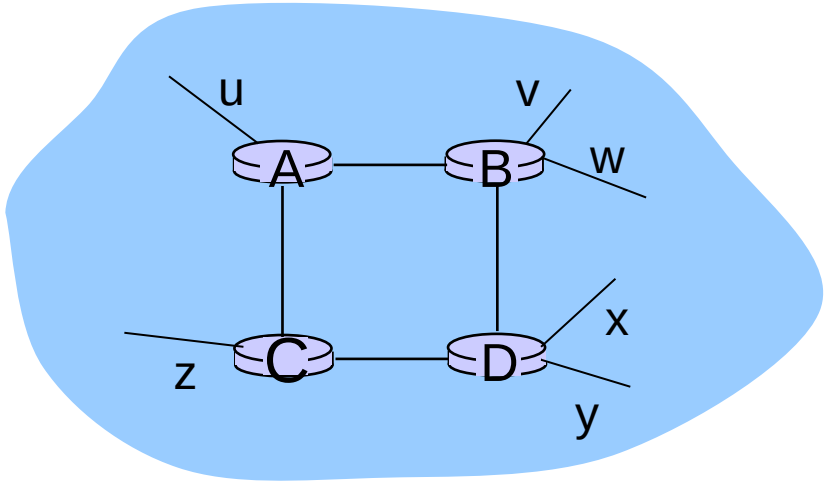
RIP协议简介

AS内部路由

- Internet采用层次路由
- AS内部路由协议也称为内部网络协议IGP (interior gateway protocols)
- 最常见的AS内部路由协议:
  - 路由信息协议：RIP(Routing Information Protocol)
  - 开放最短路径优先：OSPF(Open Shortest Path First)
  - 内部网关路由协议：IGRP(Interior Gateway Routing Protocol)
    - Cisco私有协议

RIP

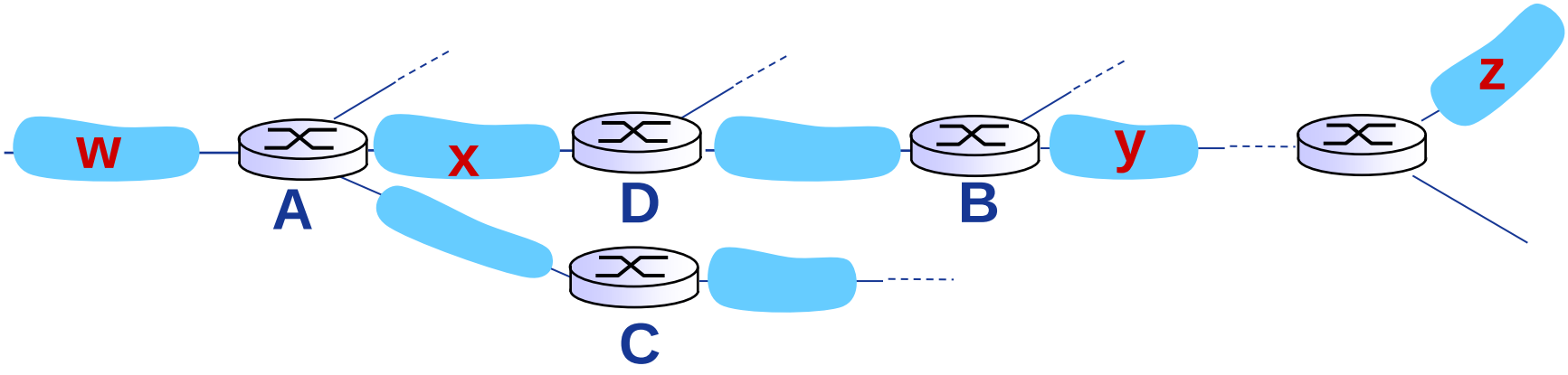
- 早于1982年随BSD-UNIX操作系统发布
- 距离向量路由算法
  - 距离度量：跳步数 (max = 15 hops), 每条链路1个跳步
  - 每隔30秒，邻居之间交换一次DV，成为通告(advertisement)
  - 每次通告：最多25个目的子网(IP地址形式)



从路由器A到目的子网:

subnet	hops
u	1
v	2
w	2
x	3
y	3
z	2

RIP：举例



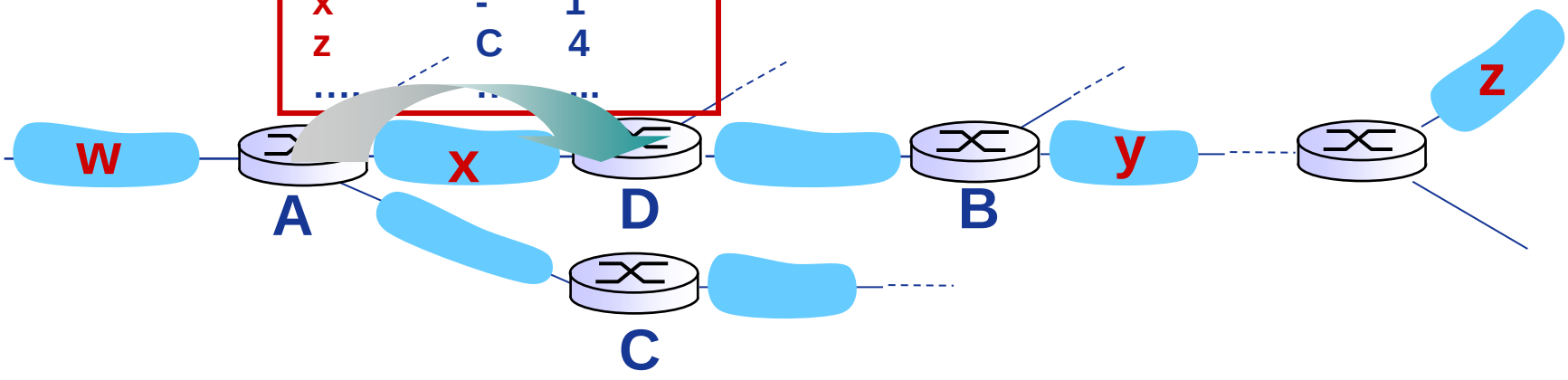
路由器D的路由表

destination subnet	next router	# hops to dest
w	A	2
y	B	2
z	B	7
x	--	1
….	….	….

RIP：举例

A-to-D advertisement

dest	next	hops
w	-	1
x	-	1
z	C	4
....	...	...



路由器D的路由表

destination subnet	next router	# hops to dest
w	A	2
y	B	2
z	B → A	7 → 5
x	--	1
....	....	....

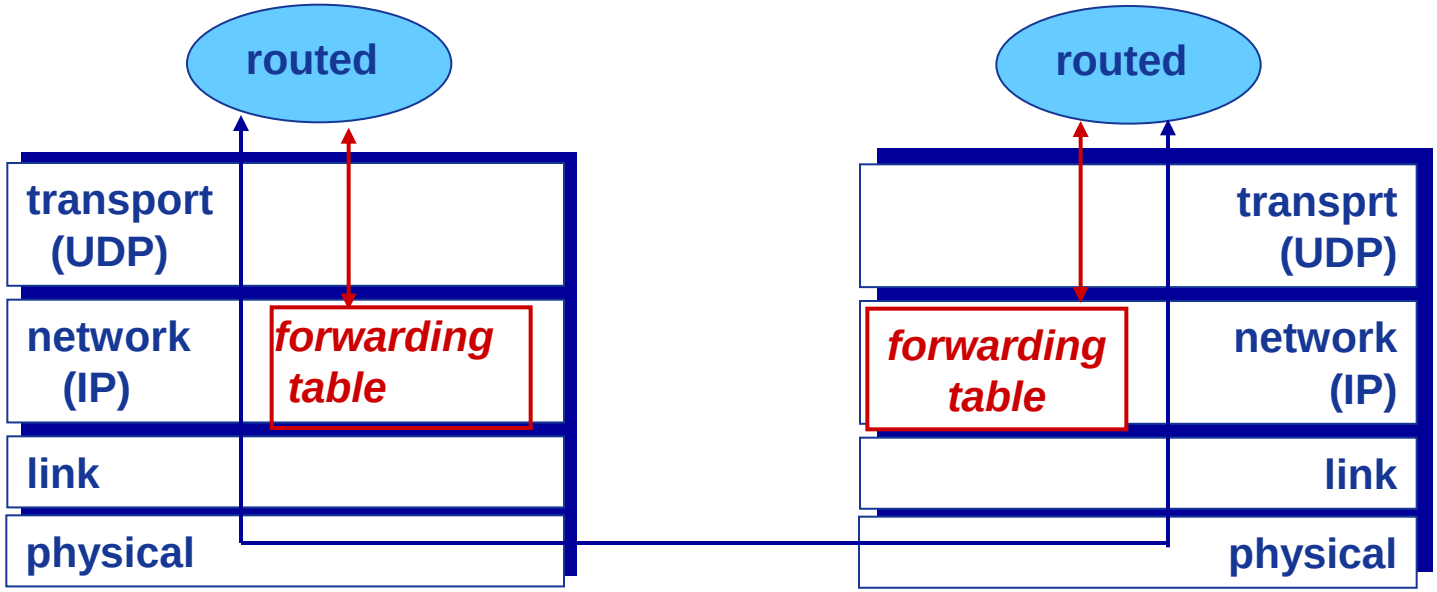
RIP: 链路失效、恢复

如果180秒没有收到通告→邻居/链路失效

- 经过该邻居的路由不可用
  - 重新计算路由
- 向邻居发送新的通告
- 邻居再依次向外发送通告（如果转发表改变）
- 链路失效信息能否快速传播到全网？
  - 可能发生无穷计数问题
- 毒性逆转技术用于预防乒乓(ping-pong)环路
  (另外：无穷大距离 = 16 hops)

RIP路由表的处理

- RIP路由表是利用一个称作route-d (daemon)的应用层进程进行管理
  - 应用进程实现
- 通告报文周期性地通过UDP数据报发送



哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场

谢 谢！

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY
立足航天，服务国防，面向国民经济主战场

计算机网络之探赜索隐

主讲人：李全龙

本讲主题

OSPF协议简介

OSPF (Open Shortest Path First)

- “开放”：公众可用
- 采用链路状态路由算法
  - LS分组扩散（通告）
  - 每个路由器构造完整的网络(AS)拓扑图
  - 利用Dijkstra算法计算路由
- OSPF通告中每个入口对应一个邻居
- OSPF通告在整个AS范围泛洪
  - OSPF报文直接封装到IP数据报中
- 与OSPF极其相似的一个路由协议:*IS-IS路由协议*

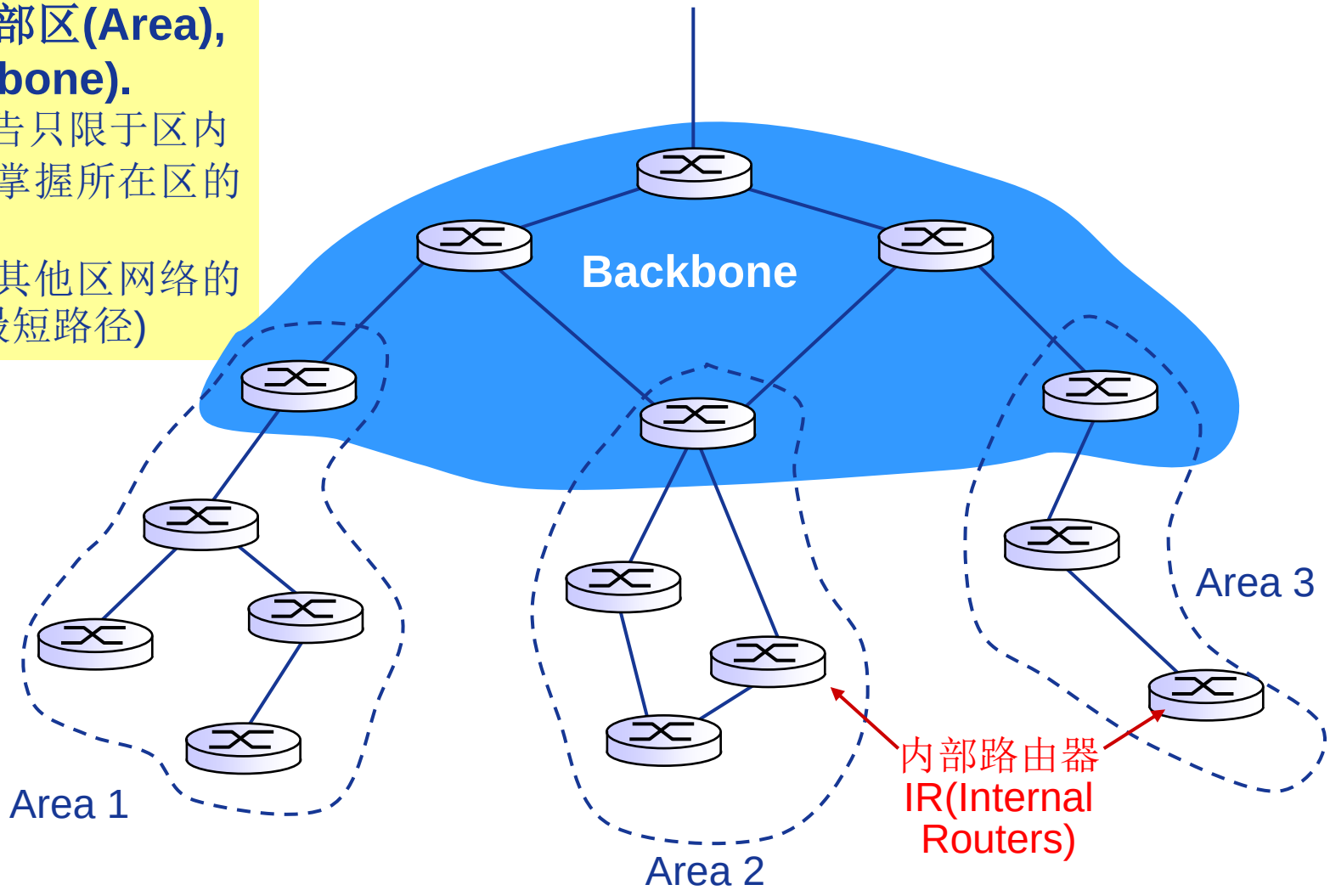
OSPF优点(RIP不具备)

- 安全(security): 所有OSPF报文可以被认证 (预防恶意入侵)
- 允许使用多条相同费用的路径 (RIP只能选一条)
- 对于每条链路，可以针对不同的TOS设置多个不同的费用度量 (e.g., 卫星链路可以针对“尽力”(best effort) ToS设置“低”费用；针对实时ToS设置“高”费用)
- 集成单播路由与多播路由:
  - 多播OSPF协议(MOSPF) 与OSPF利用相同的网络拓扑数据
- OSPF支持对大规模AS分层(hierarchical)

分层的OSPF

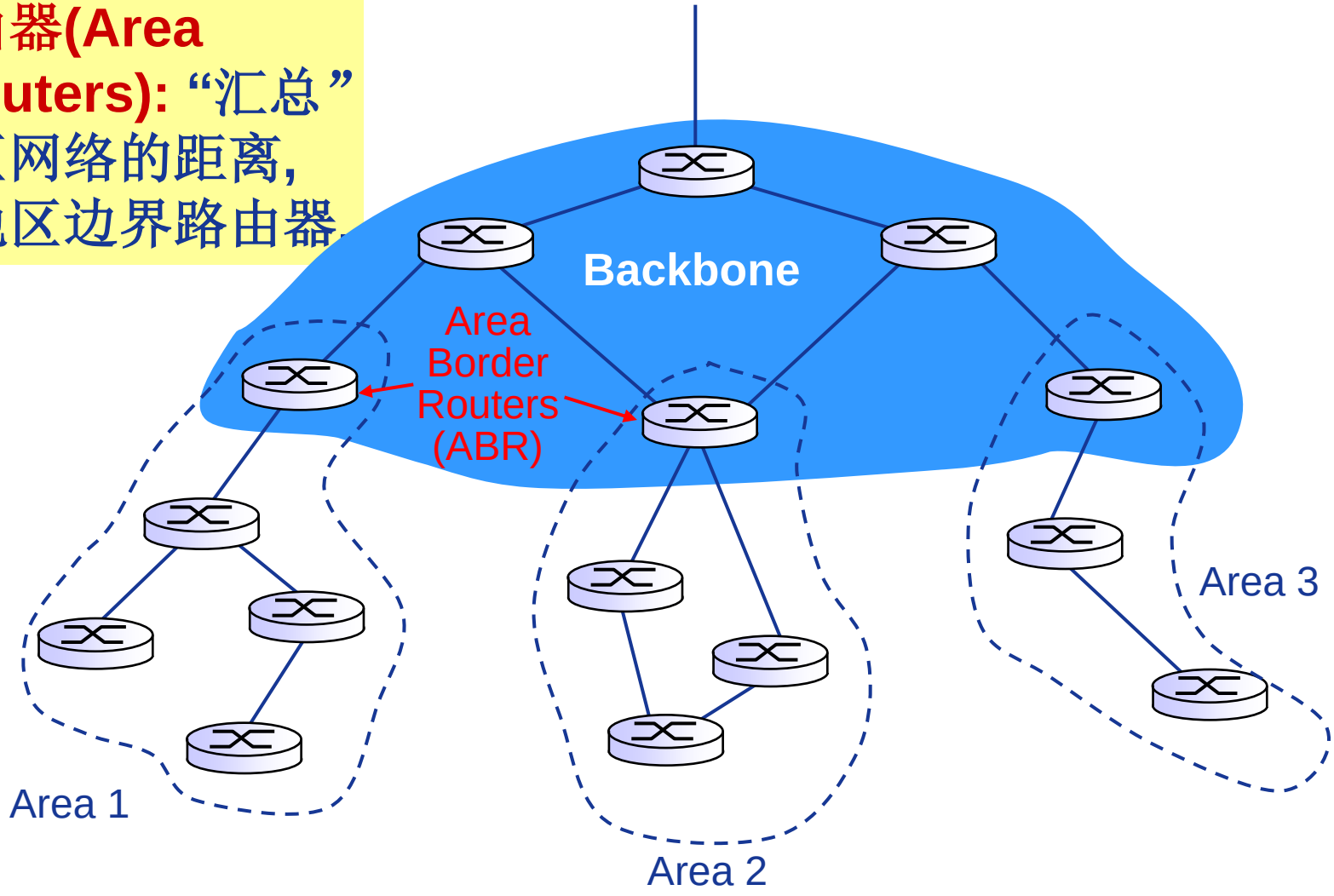
两级分层: 局部区(Area), 主干区(Backbone).

- 链路状态通告只限于区内
- 每个路由器掌握所在区的详细拓扑
- 只知道去往其他区网络的“方向”(最短路径)



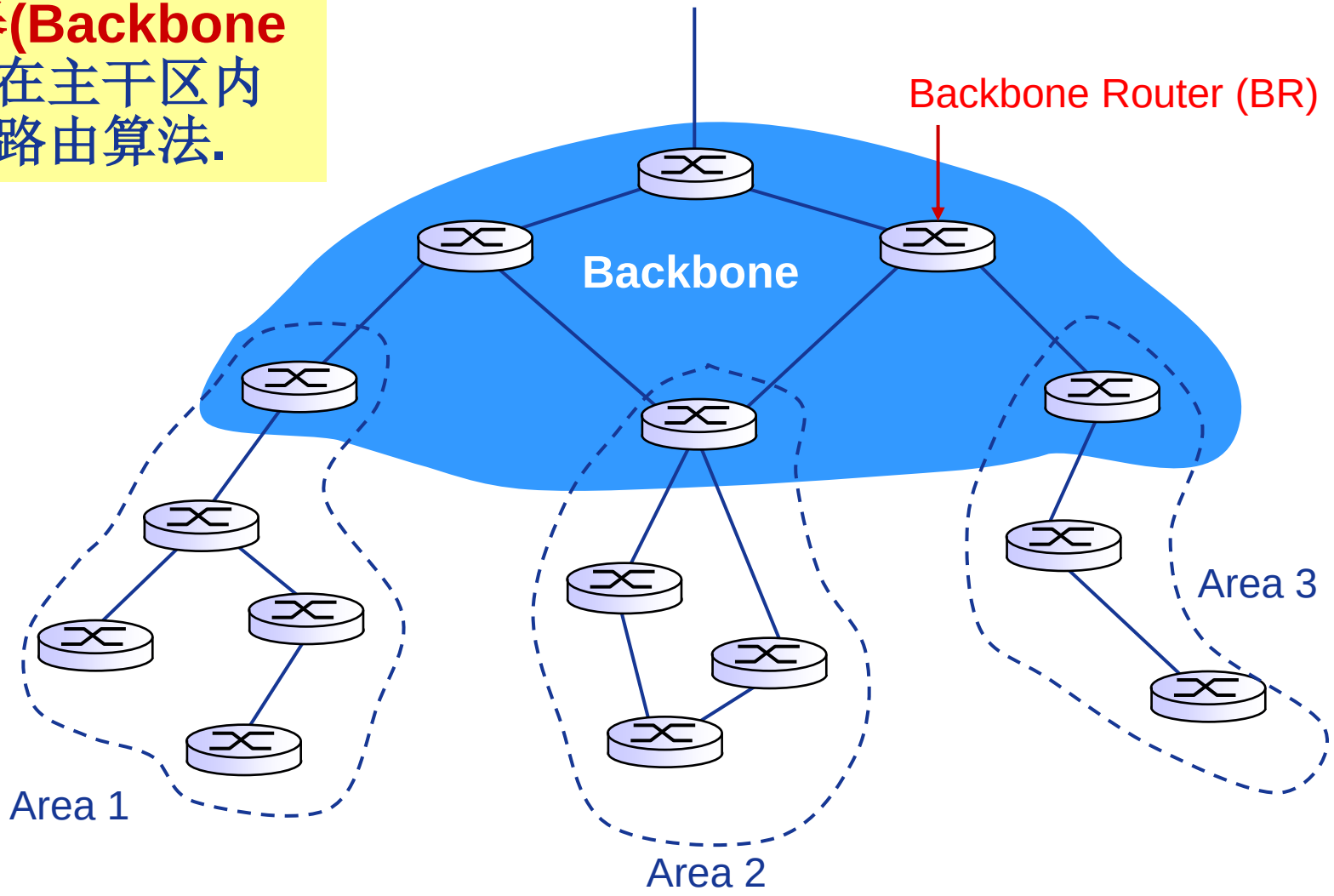
分层的OSPF

区边界路由器(Area Border Routers): “汇总”到达所在区网络的距离，通告给其他区边界路由器。



分层的OSPF

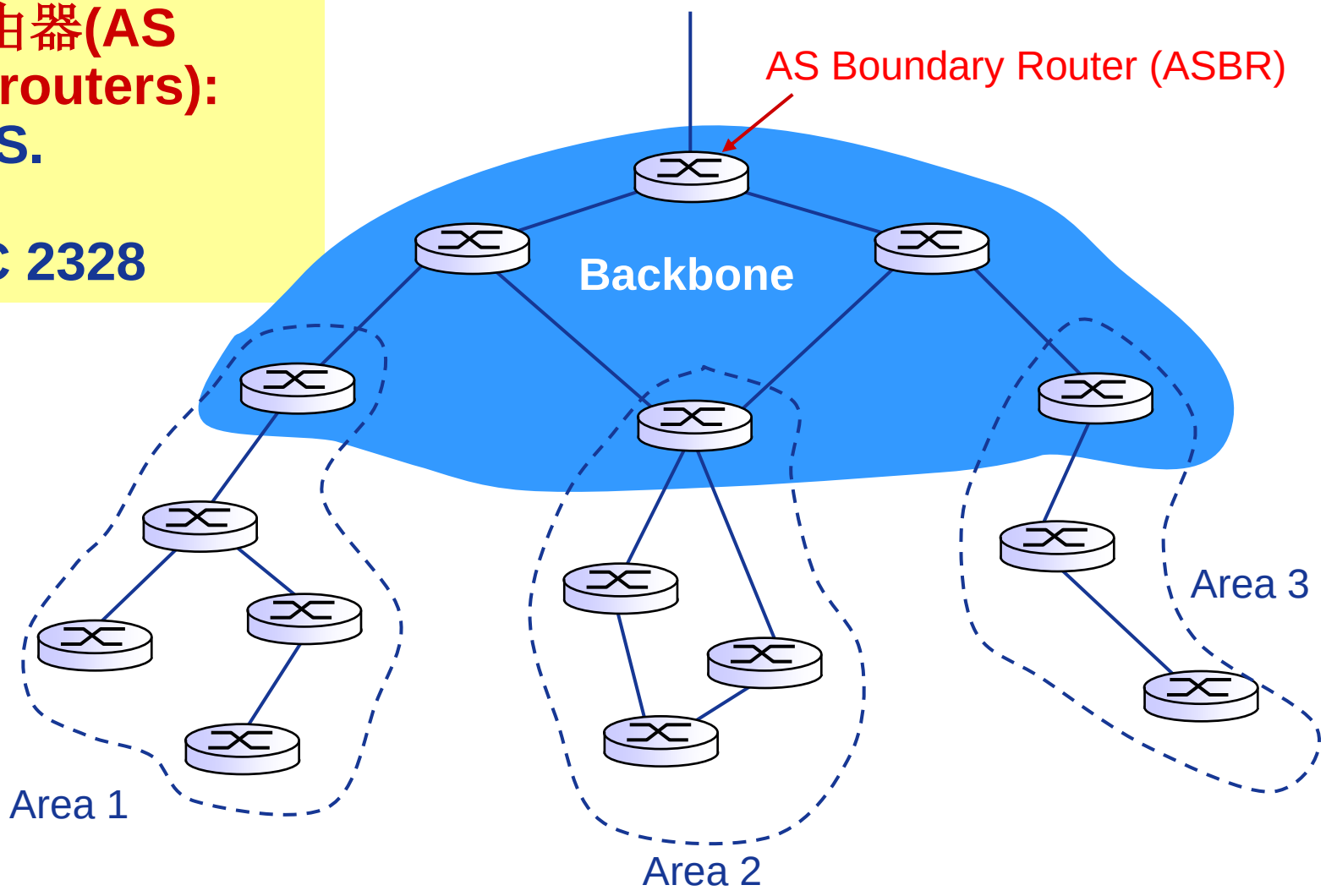
主干路由器(Backbone Routers): 在主干区内运行**OSPF**路由算法.



分层的OSPF

AS边界路由器(AS boundary routers):
连接其他AS.

参考：RFC 2328



哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场

谢 谢！

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY
立足航天，服务国防，面向国民经济主战场

计算机网络之探赜索隐

主讲人：李全龙

本讲主题

BGP协议简介（1）

Internet AS间路由协议: BGP

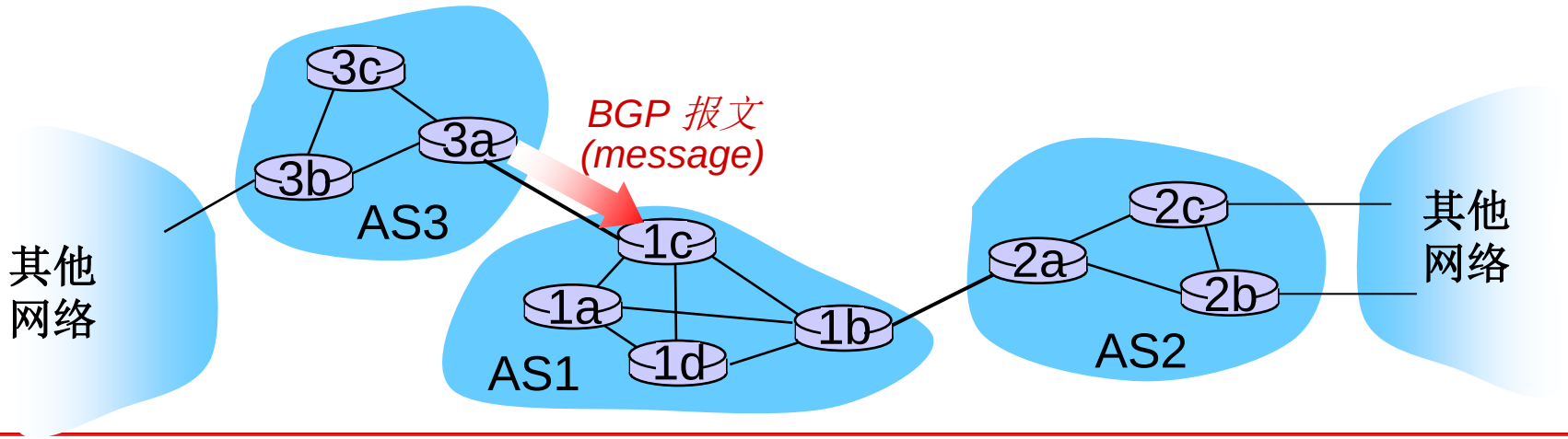
- 边界网关协议BGP (Border Gateway Protocol): 事实上的标准域间路由协议
  - 将Internet “粘合”为一个整体的关键
- BGP为每个AS提供了一种手段:
  - eBGP: 从邻居AS获取子网可达性信息.
  - iBGP: 向所有AS内部路由器传播子网可达性信息.
  - 基于可达性信息与策略，确定到达其他网络的“好”路径.
- 容许子网向Internet其余部分通告它的存在：**“我在这儿！”**

BGP基础

- BGP会话(session): 两个BGP路由器 (“Peers”) 交换BGP报文:
  - 通告去往不同目的前缀（prefix）的路径 (“路径向量(path vector)”协议)
  - 报文交换基于半永久的TCP连接
- BGP报文:
  - OPEN: 与peer建立TCP连接，并认证发送方
  - UPDATE: 通告新路径 (或撤销原路径)
  - KEEPALIVE: 在无UPDATE时，保活连接；也用于对OPEN请求的确认
  - NOTIFICATION: 报告先前报文的差错；也被用于关闭连接

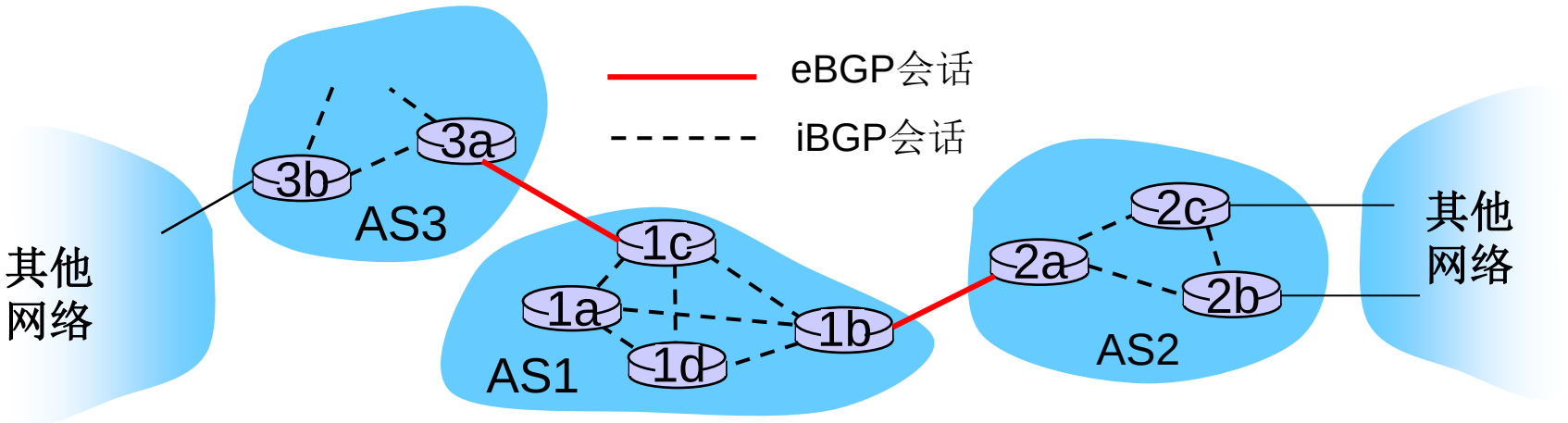
BGP基础

- 当AS3通告一个前缀给AS1时:
  - AS3承诺可以将数据报转发给该子网
  - AS3在通告中会聚合网络前缀



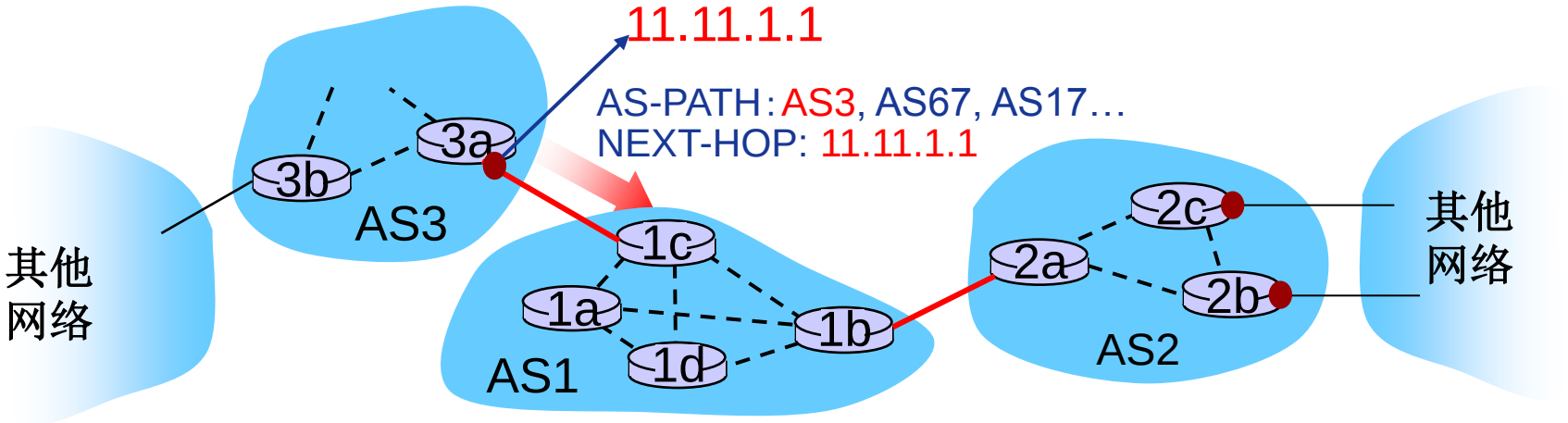
BGP基础: 分发路径信息

- 在3a与1c之间, AS3利用eBGP会话向AS1发送前缀可达性信息.
  - 1c则可以利用iBGP向AS1内的所有路由器分发新的前缀可达性信息
  - 1b可以（也可能不）进一步通过1b-到-2a的eBGP会话，向AS2通告新的可达性信息
- 当路由器获得新的前缀可达性时，即在其转发表中增加关于该前缀的入口（路由项）.



路径属性与BGP路由（route）

- 通告的前缀信息包括BGP属性
  - 前缀+属性= “路由”
- 两个重要属性:
  - AS-PATH(AS路径): 包含前缀通告所经过的AS序列: e.g., AS 67, AS 17
  - NEXT-HOP(下一跳): 开始一个AS-PATH的路由器接口，指向下一跳AS.
    - 可能从当前AS到下一跳AS存在多条链路



哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场

谢 谢！

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY
立足航天，服务国防，面向国民经济主战场

计算机网络之探赜索隐

主讲人：李全龙

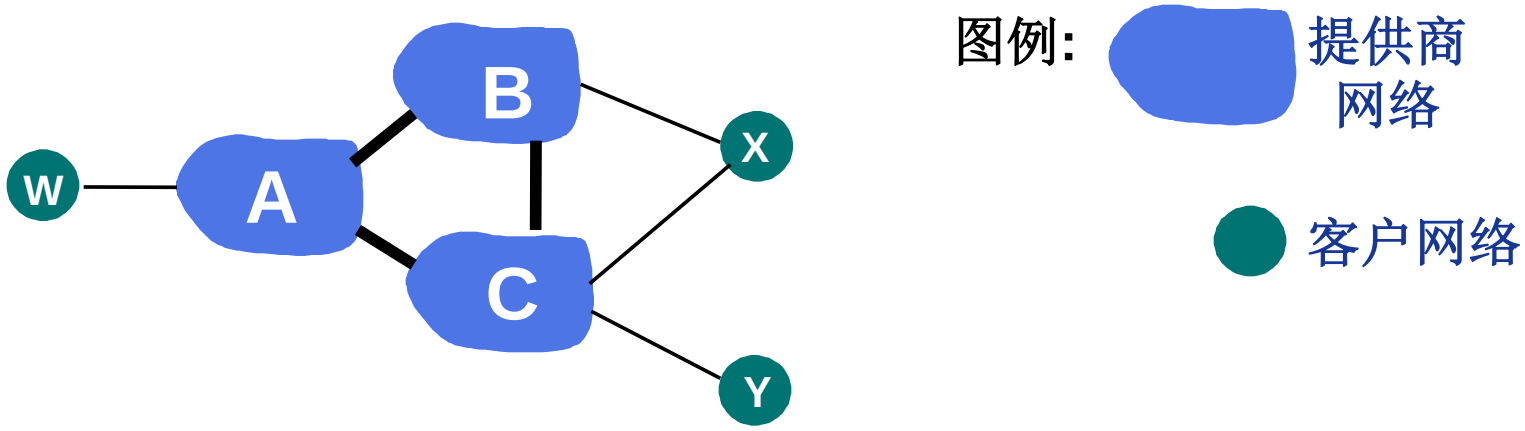
本讲主题

BGP协议简介（2）

BGP路由选择

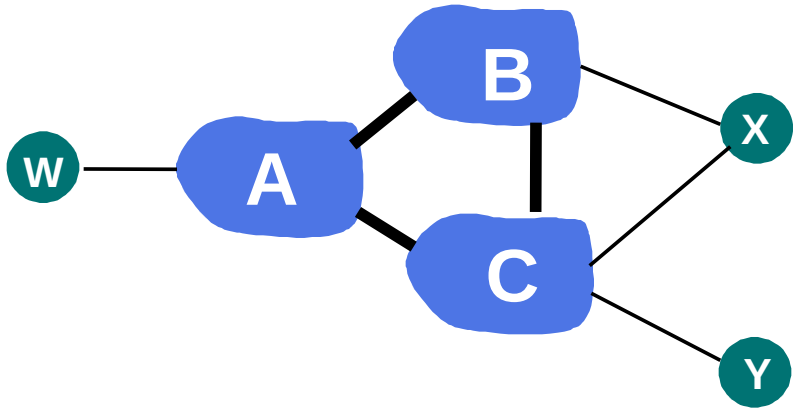
- 网关路由器收到路由通告后，利用其输入策略(import policy)决策接受/拒绝该路由
  - e.g., 从不将流量路由到AS x
  - 基于策略(policy-based) 路由
- 路由器可能获知到达某目的AS的多条路由，基于以下准则选择：
  1. 本地偏好(preference)值属性: 策略决策(policy decision)
  2. 最短AS-PATH
  3. 最近NEXT-HOP路由器: 热土豆路由(hot potato routing)
  4. 附加准则

BGP路由选择策略



- A,B,C是提供商网络/AS(provider network/AS)
- X,W,Y是客户网络(customer network/AS)
- W,Y是桩网络(stub network/AS): 只与一个其他AS相连
- X是双宿网络(dual-homed network/AS): 连接两个其他AS
  - X不期望经过他路由B到C的流量
  - … 因此，X不会向B通告任何一条到达C的路由

BGP路由选择策略



图例：

提供商网络

客户网络

- A向B通告一条路径：AW
- B向X通告路径：BAW
- B是否应该向C通告路径BAW呢?
  - 绝不! B路由CBAW的流量没有任何“收益”，因为W和C均不是B的客户。
  - B期望强制C通过A向W路由流量
  - B期望只路由去往/来自其客户的流量!

为什么采用不同的AS内与AS间路由协议?

策略(policy):

- inter-AS: 期望能够管理控制流量如何被路由，谁路由经过其网络等.
- intra-AS: 单一管理，无需策略决策

规模(scale):

- 层次路由节省路由表大小，减少路由更新流量
- 适应大规模互联网

性能(performance):

- intra-AS: 侧重性能
- inter-AS: 策略主导

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

立足航天，服务国防，面向国民经济主战场

谢 谢！