






网络层

 Courses	 计算机网络
<input checked="" type="checkbox"/> Done	<input checked="" type="checkbox"/>
 Status	Done

网络层服务

- 交换/路由：确定分组从源到目的地所采取的端到端路径的网络范围处理过程
- 转发：将分组从一个输入链路接口转移到适当的输出链路接口的路由器本地动作
- 建立连接（虚电路）

分组交换网络，基本思想

虚电路网络

携带的是VC number而不是地址信息

信令协议：建立、维持、拆除虚电路

数据报网络

无连接的通信模式

每个路由器维护一个forwarding table

对比总结

特性	虚电路网络	数据报网络
连接建立	需要预先建立连接	无需建立连接
数据包传输路径	固定路径	动态路径
资源分配	预留资源	无资源预留

特性	虚电路网络	数据报网络
传输顺序	保证有序传输	不保证顺序
控制复杂度	维护连接状态和路径信息，复杂度高	无需维护连接状态，简单
适用场景	语音通信、视频会议等实时和稳定传输	互联网、电子邮件、文件传输等

分组交换网络中路由

性能评估指标：

- 跳数最少
- 花费最少
- 延时最少（队列长度）
- 吞吐量最大

路由信息的更新方式：

- 信息源
 - 本地信息
 - 相邻的交换机
 - 网络中所有的交换机
- 更新时间
 - 周期性
 - 重大改变时
 - 人工维护

路由算法

集中式路由

源到目的地单条路径，最小代价算法，网络拓扑发生重大变化时更新

分布式路由

洪泛

所有path：不需要网络信息，但网络中会存在过多副本， cycle problem： hop count 解决

随机路由

随机一个path（随机or循环）：不需要网络信息，通常不是最佳

自适应路由

路由决策随着网络条件的变化而变化，帮助拥塞控制

最小代价路由算法及其性能分析

- Bellman-Ford：一跳—两跳—三跳...（分布式（距离向量DV）、局部信息）

```
w(i, j) = __inf__ if i and j not directly connected
w(i, j) = e(i, j) if i and j is directly connected
# h:最大跳数
# Lh(n):从源点s到点n在h跳内的最小代价
# initialization
L1(n)=w(s, n)
# update
L(h+1)(n)=min(j)[L(h)(j) + w(j, n)]
```

- Dijkstra：集中式（链路状态LS）、全局信息

一二三代互联网路由算法之间的对比和改进



- 第一代：BF算法
 - 输出队列长度用于定义链路开销
- 第二代：dj算法
 - 测量的延迟用于定义链路开销

- 第三代
 - 处理抖动震荡问题
 - 应用链接利用率来表示链接的状态
 - 基于先前值和新利用率的调平
 - 使用跃点规范化指标计算链路开销

链路代价的计算

应用**链路利用率**来表示链接的状态

$$\rho = 2(T_s - T) / (T_s - 2T)$$

T : *current measured delay*

T_s : *mean packet length(600 bit)/transmission rate of the link*

基于先前值和新利用率的调级

Leveling: 老化算法

$$U_n = \alpha \times \rho_n + (1-\alpha) \times U_{n-1}$$

自治系统与路由方式

IGP与EGP

- IGP (内部网关协议) : AS内部路由器间信息传递
 - 例如RIP (使用DV) 、 OSPF (使用LS) 、 IGRP
- EGP (外部网关协议) : **可达性**信息
 - 例如BGP

内部路由协议

距离向量协议RIP (Routing Information Protocol)

距离指标: 跳数

每隔30s更新一次

如果180s未收到, 则认为已经丢失连接

over **UDP**

开放最短路优先OSPF

- 收敛速度快，无跳数限制
- 配置复杂

通过洪泛将更新状态信息（LSA）传输到整个AS

over **IP**

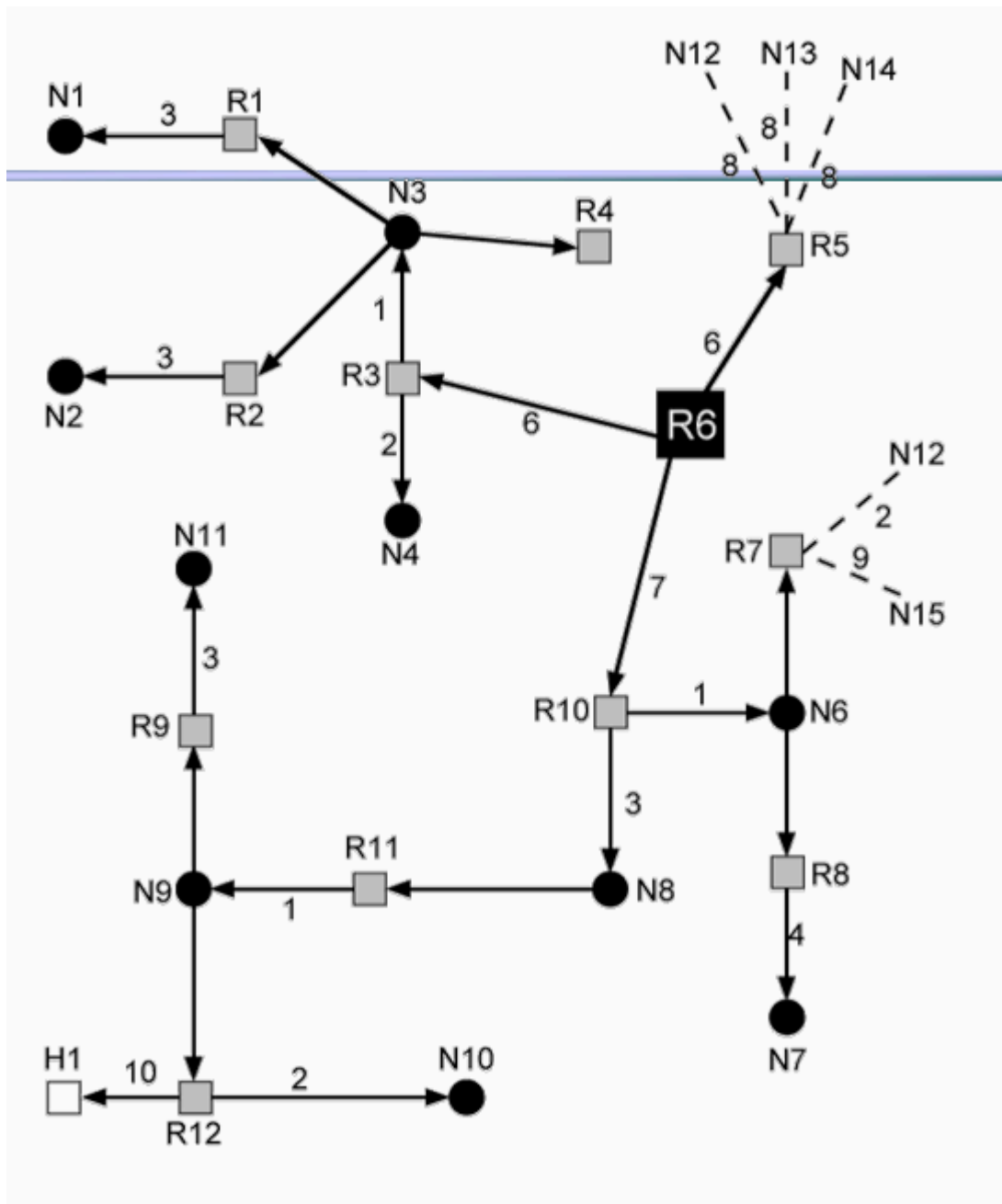
dijkstra算法

路由表：（of R6）

Destination	Next Hop	Distance	Destination	Next Hop	Distance
N1	R3	10	N11	R10	14
N2	R3		H1		21
N3	R3		R5		6
N4	R3		R7		8
N6	R10		N12		10
N7	R10		N13		14
N8	R10		N14		14
N9	R10	11	N15	R10	17
N10	R10	13			

Destination	Next Hop	Distance	Destination	Next Hop	Distance
N1	R3	10	N11	R10	14
N2	R3	10	H1	R10	21
N3	R3	7	R5	R5	6
N4	R3	8	R7	R10	8
N6	R10	8	N12	R10	10
N7	R10	12	N13	R5	14
N8	R10	10	N14	R5	14
N9	R10	11	N15	R10	17
N10	R10	13			

路由结构图：



SPF Operation: 将网络、主机和BGP路由器当做目标，每个路由器计算其SPF树，显示到其他所有目标的最低成本路径，仅路由数据包中使用的下一跳

边: router—router router—network

节点: router&network

图: 有向图

边界网关协议BGP

作为一种AS间路由选择协议。BGP中，分组到达的是CIDR化的前缀。 <IP prefix, 路由器接口号>

- BGP的功能
 - 从邻居AS获得前缀的可达性信息
 - 确定到该前缀的“最好的”路由
- 基本报文类型和工作方式
 - 网关路由器：位于AS边缘的路由器
 - **eBGP**(外部BGP)：跨越两个AS的BGP
 - **iBGP**(内部BGP)：相同AS中两台路由器之间的BGP会话

IP协议

IP基本原理

- 分组交换：将数据分成小块
- 无连接协议：没有握手
- 不可靠传输：尽力而为

异构网络环境下internet协议的工作过程

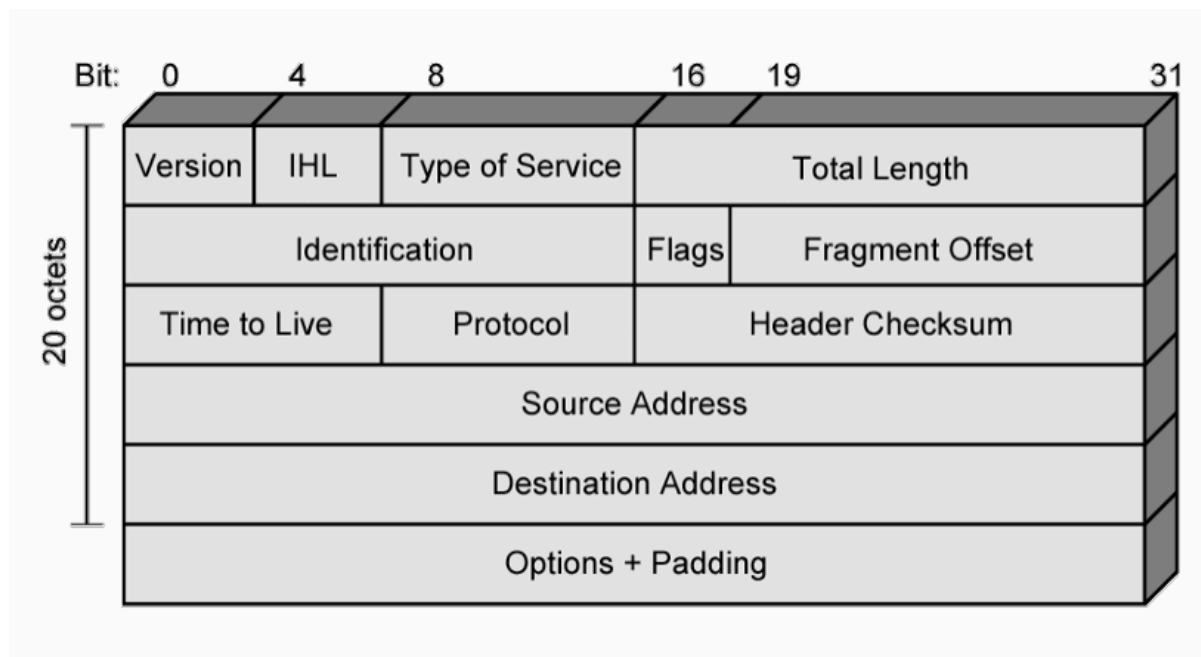
- 数据封装与传输
- 路由与转发
- 分片与重组
- 终端递送

协议内容

- Routing：主机和路由器维护一个**路由表**，指示应将数据报发送给哪个路由器
- Datagram Lifetime: 数据报有TTL部分—Hop count
- Fragmentation and Re-assembly（分片和重组）：
 - 为何？分组的长度超过后面的网络的最长限制长度
 - **MTU**（最大传输单元）：一个数据帧中可以传输的最大数据量（以字节为单位），不包括数据链路层的头部信息，典型的MTU为1500个字节
- Error Control：如果丢弃数据包，路由器应尝试通知源—ICMP

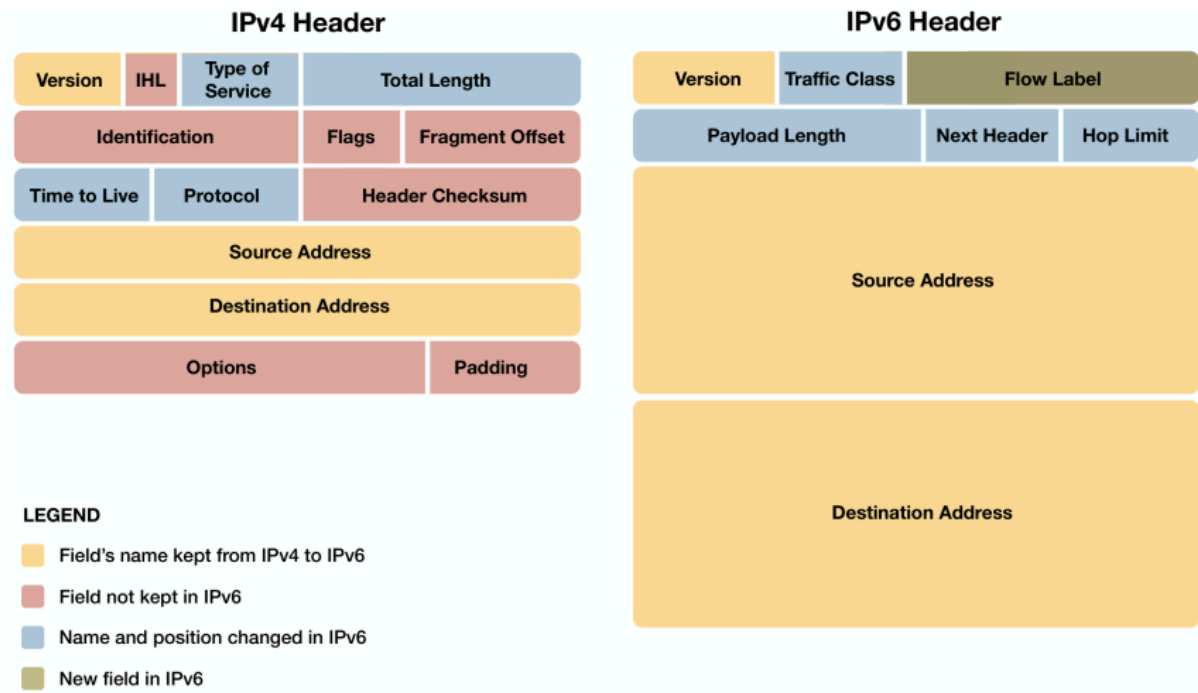
- Flow Control: 当缓冲区满时

IPv4首部格式



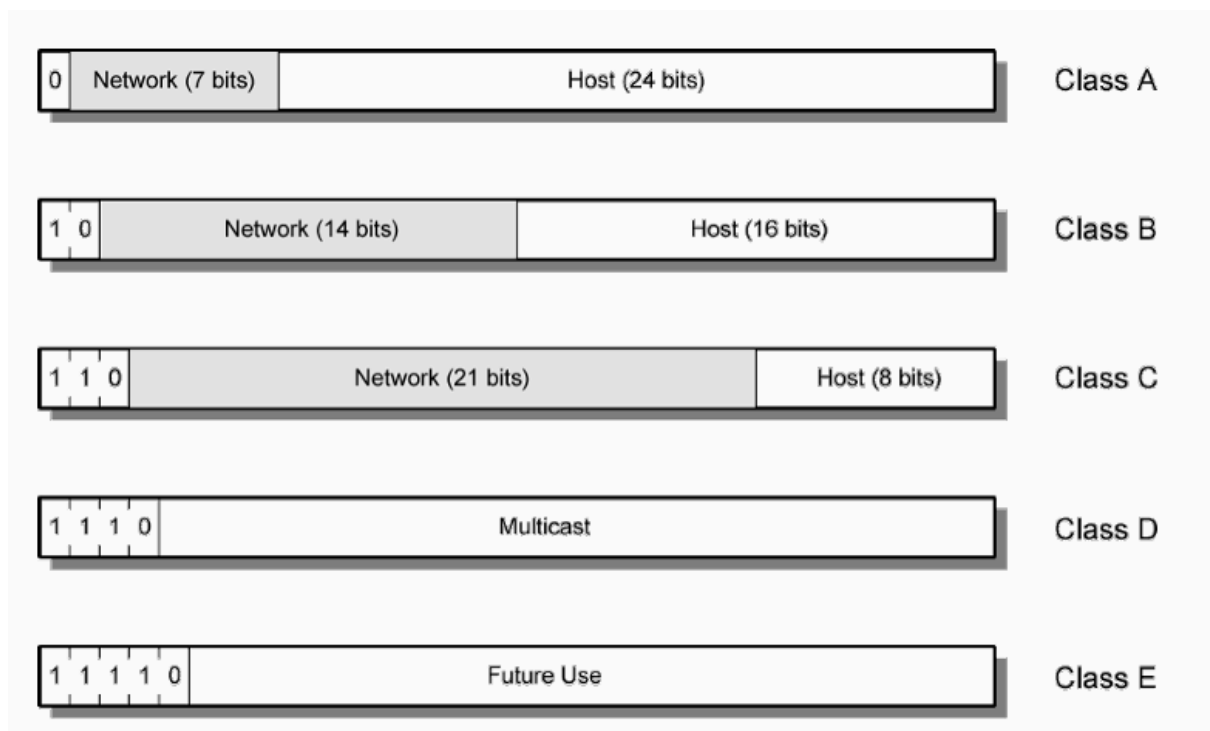
- IHL (包头长度)：指明IPv4协议包头长度包括多少个32位（1bit代表4bytes），最小值为5
- total length：共16位，指明数据报长度为多少个字节，包括包头和数据部分，最小值20
- Identification：用于唯一标识数据报。在数据报需要分片时，所有片段都具有相同的标识字段值，以便接收端重新组装。
- Flags：是否还有更多的分片
- Offset：表示当前分片在原始数据报中的相对位置，多少个8字节

到IPv6的变化



IP地址

点分十进制表示法



CIDR表达

A.B.C.D/n, n是IP前缀

子网划分：

ISP's block	<u>11001000 00010111 00010000 00000000</u>	200.23.16.0/20
Organization 0	<u>11001000 00010111 00010000 00000000</u>	200.23.16.0/23
Organization 1	<u>11001000 00010111 00010010 00000000</u>	200.23.18.0/23
Organization 2	<u>11001000 00010111 00010100 00000000</u>	200.23.20.0/23
...
Organization 7	<u>11001000 00010111 00011110 00000000</u>	200.23.30.0/23

IPv6和IPv4的比较

IPv6没有ARP。

IPv6没有boardcast

NAT原理及优缺点

在NAT中，路由器会将本地网络中的私有IP地址转换为一个或多个公共IP地址，用于与外部网络通信。

NAT转换表：包含端口号及其IP地址

- 优点：
 - 节省IP地址
 - 增强网络安全
 - 便于管理
- 缺点
 - 影响性能
 - 破坏端到端通信
 - 复杂性：额外配置

ARP地址解析原理和流程

ARP只为在同一个子网上的主机和路由器接口解析IP地址

查询ARP报文在广播帧中发送，相应ARP报文在标准帧中发送

DHCP动态地址获取过程

DHCP：动态主机配置协议

- DHCP服务器发现：客户生成包含DHCP发现报文（UDP、port=67）的IP数据报，使用广播地址，源ip：0.0.0.0
- DHCP服务器提供：服务器收到发现报文时，用DHCP提供报文响应，使用广播地址，报文包含：事务ID、向客户推荐的IP地址，网络掩码和IP地址租用期
- DHCP请求：客户从服务器中选择一个，发送DHCP请求报文回显配置参数
- DHCP ACK：服务器用DHCP ACK报文响应

ICMP

在主机和路由器之间传输错误和控制信息

Ping

echo request & **echo reply** 可达性

回显请求—回显应答

可用于：计算往返时间RTT、计算到目标的跳数（use TTL）

Traceroute

源主机：发送具有不可达UDP端口号的UDP报文，TTL为1/2/3/4。。。为每个数据报启动定时器，以计算RTT

路由器：发送ICMP警告报文

IP组播

one-to-many

- 组播地址
 - D类地址，即224.0.0.0~239.255.255.255 **(1110开头)**
- 组播模型
 - 发送者将数据发送到一个特定的组播地址
 - 加入该组播组的接收者会接收该组播地址的流量

- multicast tree:路由器只会沿着生成树转发
- 组播组管理：IGMP
 - 主机通过发送IGMP消息告诉本地路由器它希望加入或离开某个组播组
 - 路由器通过IGMP查询（**Query**）消息**定期**检查组成员是否依然存在

组播路由机制

生成一棵从源到所有接收者的最短路径树，使用Dijkstra算法（同OSPF）、

Shared-tree

Source-Based tree

- 反向路径转发机制（Reverse Path Forwarding）：
 - 只去转发来自上游接口的Mcast报文。
 - 确保组播报文在正确的接口到达，只有这些组播报文才会被路由器沿着组播分发树进行转发，如果报文在错误的接口到达，路由器将丢弃这些报文。
 - Pruning（修剪）：转发树中包含不在多播组内的成员，自下而上发送Prune msgs

Shared-Tree & Source-Based tree

- Shared-Tree:
 - 所有组播流量通过一个共享的核心节点（Rendezvous Point, RP）进行传输
 - Steiner Tree连接所有路由器与附加组成员的最小成本树
 - 由于其计算复杂度未在实践中使用
- Source-Based tree
 - 为每个发送者构建单独的组播树