Internet Protocols and Networks.

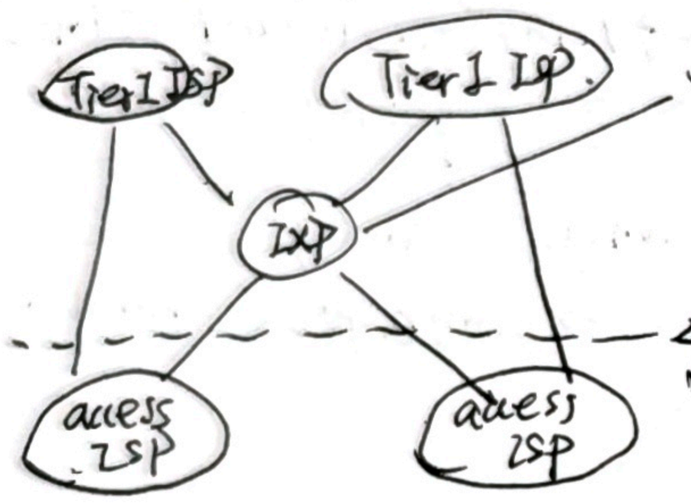
<Block 1>

A. Introduction to IP.

1. The network core (转发)

store-and-forward $\frac{L\text{ bit}}{R\text{ bits/s}}$ → end-end delay (1 mbits = 10^6 bits)

2. network of network.



Content provider network / Content Distribution Networks (CDNs) push content close to users

core

Internet Service Providers (ISP)

Internet Exchange point (IXP)

edge

3. loss ← packet arrival rate to link exceeds output link capacity

$d_{nodal} = d_{processing} + d_{queue} + d_{transmission} + d_{propagation}$

↳ determine output link.

Delay / loss } limit → throughput { instantaneous / averages } rate (bps) at which bits transferred between sender/receiver

bottleneck link: link on end-end path that constrains end-end throughput.

网络核心一般高速率链路，bottleneck link 四周在接入网.

4. 协议栈

Why { explicit structure allows identification, relationship of complex system's pieces.

modularization eases maintenance, updating of system.

Layer

	message	Application
		Presentation
		Session
port	segment	Transport
IP	datagram	Network
MAC	frame	Data Link
	Layer 1	physical

TCP/IP (4层): Transmission Control Protocol/Internet Protocol

ISO/OSI (7层): International Standards Office/ Open Systems Interconnection.

Traslates data from an application (encode)

Take care of a connection between two hosts for the lifetime of that connection (Authentication + authorisation)

认证+授权.

B. Transport layer (I)

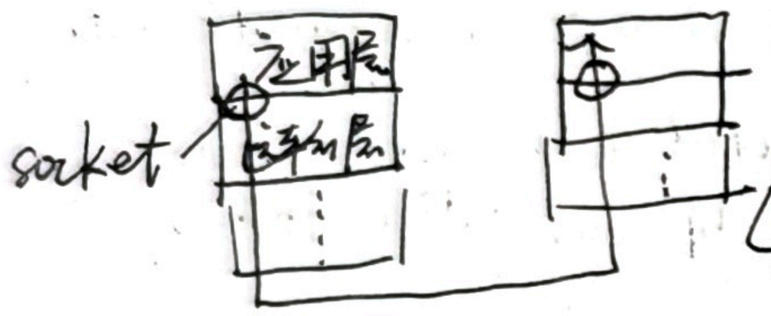
1. transport-layer services.

传输层 { 应用层 → logical communication between processes
网络层 → logical communication between hosts.

{ send. break app messages into segments pass to network layer.
recieve. reassemble segments into messages, pass to app layer

? reliable, in-order TCP, UDP;

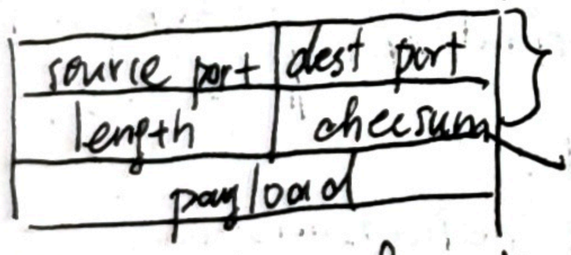
2. Multiplexing / demultiplexing



Mux: Combining several streams of data into a single stream.
Demux: A stream of data is separated out into its individual components.
传输层提供服务的一部分，涉及 header 的操作.

UDP socket = U<destination IP, destination port #>.
TCP socket = T<destination IP, destination port, source IP, source port>

3. Connectionless transport: UDP (User Datagram Protocol)



32 bit = 4 Byte.
detect "errors" in transmitted segment.

4. Principles of reliable data transfer

reliable data transfer protocol → (rdt) 接口

rdt 1.0 在 bit errors, 在 lost.

ACK/NAKs
error detection
sequence number
countdown
timer
Acko, 1,

rdt 2.0 在 bit errors
rdt 2.1:
rdt 2.2 → 只有 ACK, 无 NAK.
rdt 3.0 在 bit errors 和丢包

bit 出错
ack error
丢包
bit 出错 / 丢包 ← duplicate ⇒ stop and wait

sequence number 接收方在上层调用时期望收到任何分组都不以此序号为(或比此序号)到顺序
currently sending

<Block 2>

A. Transport layer (II)

1. pipelined protocols

利用率 → 改善 utilization $\frac{L/R}{RTT+L/R}$.

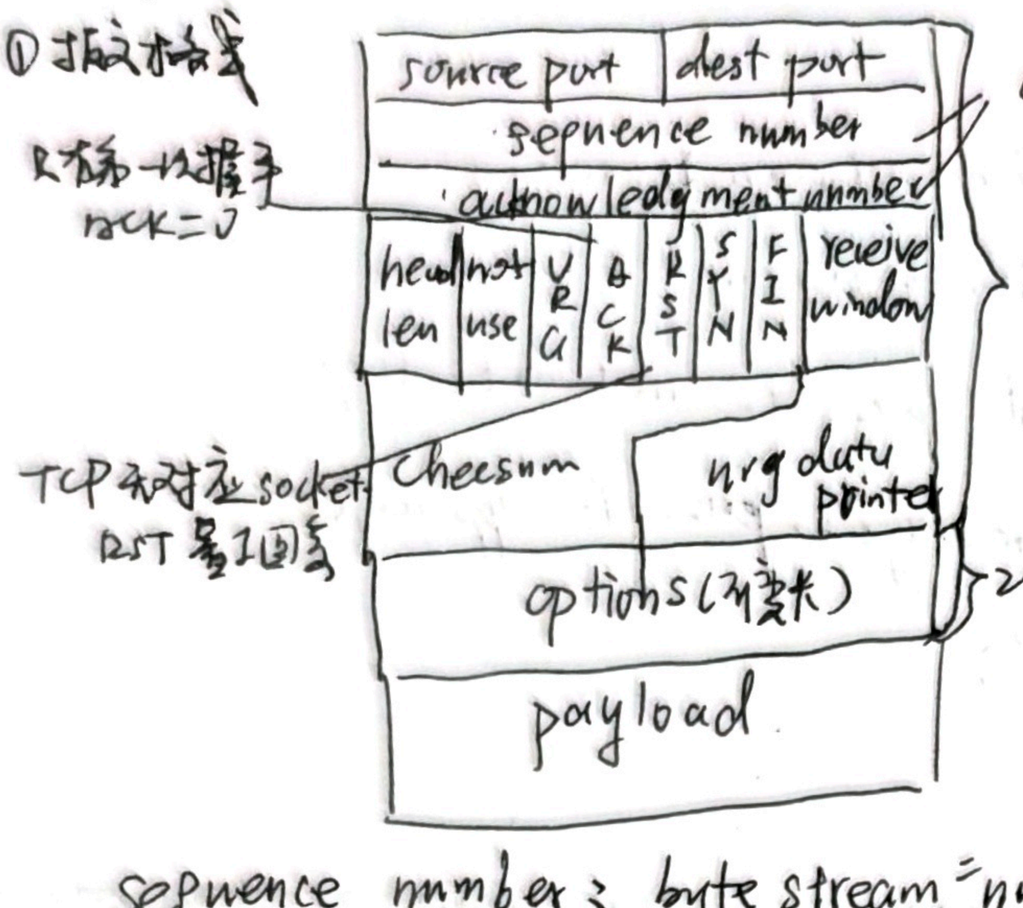
Go-Back-N (GBN) { 累积确认 cumulative ACK { 管道化
失序报文 drop
只有一个定时器 } 出错方

Selective Repeat (SR) { 每个报文都有 ACK, 已确认
自己超时重发, 失序(窗口内)
报文缓存并发 ACK
基本原则, 窗口的为 N 个已确认 ACK } 也可在接收序直接 deliver

维护 window
GBN 发送方 window
SR 都有 window, SR 中 window 不一定一致, 会动态改变窗口
发和序号空间, 一般
窗口长 ≤ $\frac{\text{序号空间}}{2}$

2. TCP

① 报文格式



seq. ack number 是TCP最重要的部分.

sequence number: byte stream "number" of first byte in segment's data

~~acko~~ acknowledgement number: next byte expected from other side

(TCP只确认流中第一个丢失的字节为止, 即 cumulative ACK)

如到来序报文段的处理方法因初用而异.

② RTT (round trip time) and timeout.

EstimatedRTT = $(1-\alpha)$ EstimatedRTT + $\alpha \cdot$ SampleRTT ($\alpha = 0.1$左右)

DevRTT = $(1-\beta)$ DevRTT + β | SampleRTT − EstimatedRTT | ($\beta = 0.25$) → 为防 retransmissions.

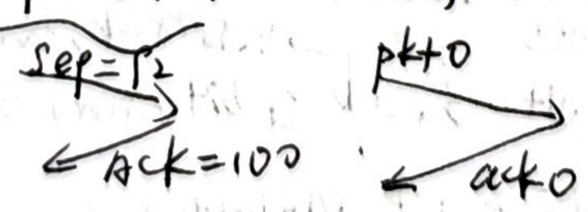
TimeoutRTT = EstimatedRTT + 4 DevRTT.

Tmeout RTT { too short: early timeout, unnecessary retransmissions
too long: slow reaction to segment loss.

TCP fast retransmit: triple duplicate Acks (看连续的 duplicate)

注意. TCP中, ACK序号指的是想要接收的报文段序号而不是已经接收的序号

(ACK根据大小字在PPT中含义不同, ACK n 下字段, ack 1则向加1报文段或1名组)



③ TCP flow control

recieve window → rwnd

reciever { 保证接收 buffer 不 overflow. LastByteRcvd − LastByteRead ≤ RcvBuffer.
接收窗口 rwnd = RcvBuffer − [LastByteRcvd − LastByteRead] → 已使用缓存.

Sender { LastByteSend − LastByteAcked ≤ rwnd

MSS (MTU) = IP header + TCP header + MSS (MTU)

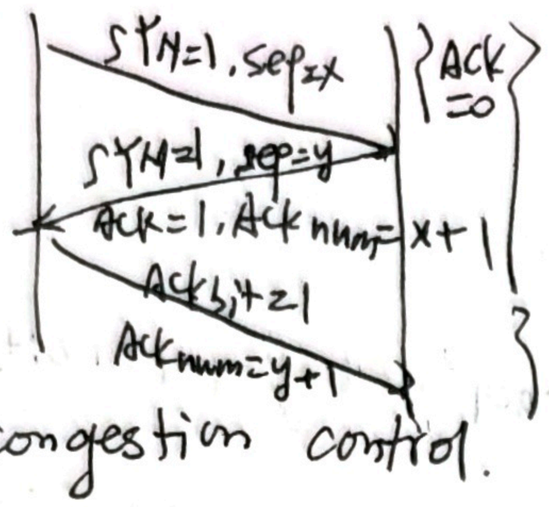
Maximum Segment Size, 一般为 1460 Byte — Maximum Transmission unit.

接收方认为自己的接收缓冲大小, rwnd中, 发送方保证发送方不至data 导致 rwnd 为0

Nagle's Algorithm { Cause: when an application generates data very slowly, Overhead of this is huge
发送方发送多个数据
发送方在buffer积攒数据, 直到 { 收到TCP发送ACK
buffer 达到 MSS (Data长度)

④ Connection Management

建立



断开

3. TCP congestion control.

Flow control = controlled by receiver limit window for connection. Stops receiver getting too much traffic

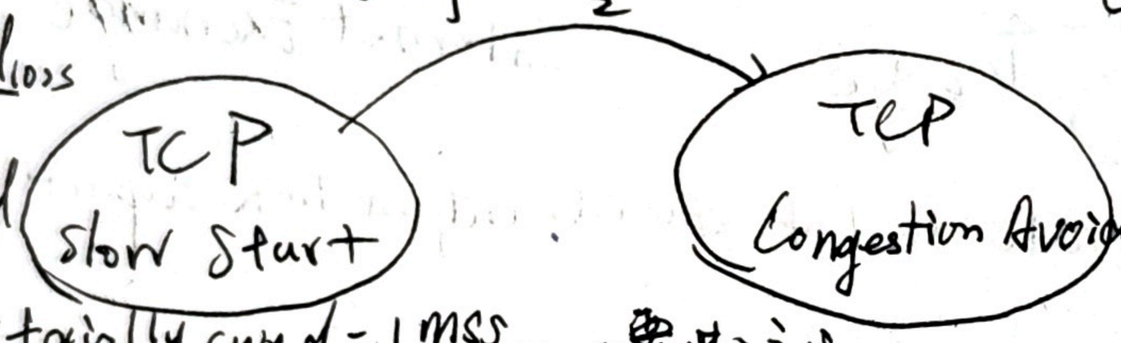
Congestion control = reacts to network itself being overloaded. Stops network getting too much traffic

Sender = LastByteSent − LastByteAcked ≤ min{cwnd, rwnd}

↳ congestion window.

ssthresh = $\frac{1}{2}$ cwnd loss 发生时的 cwnd

比之前 $\frac{ssthresh}{2}$

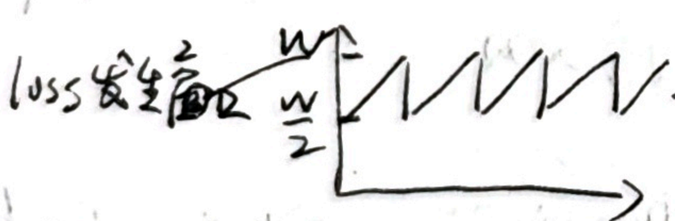


initially cwnd = 1 MSS → 要进入时没有.

initial rate is slow but ramps up exponentially fast.

TCP flavor { TCP RENO: { 3 duplicate Ack → 快速恢复, cwnd = $\frac{cwnd}{2}$, grow linearly; timeout → 快启动.
TCP Tahoe: 无论何种丢包都快启动. (设置 cwnd 为 1).

Additive Increase Multiplicative Decrease (AIMD).



avg TCP ~~throughput~~ throughput = $\frac{3}{4}\frac{W}{RTT}$ bps.

TCP具有 fairness.

B. Network layer. Data plane.

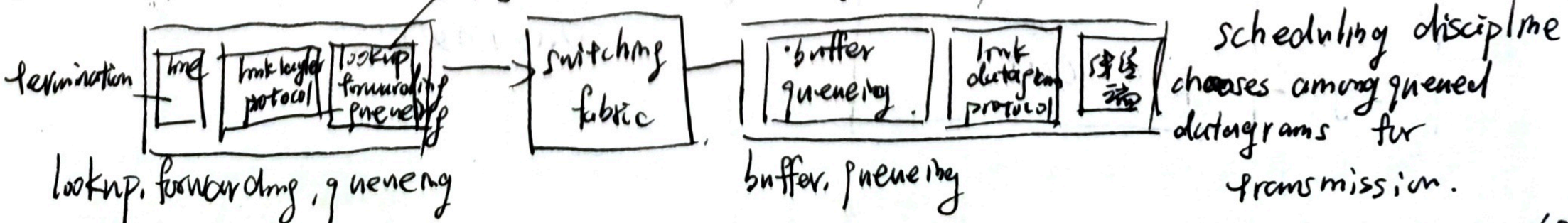
1. Overview.

两个重要功能 { forwarding: move packets from router's input to appropriate router output.
routing: determine route taken by packets from source to destination.

Network service model { guaranteed delivery
guaranteed minimum bandwidth to flow
In order
guaranteed delivery with { 上限的时延 / less than some delay.

2. router.

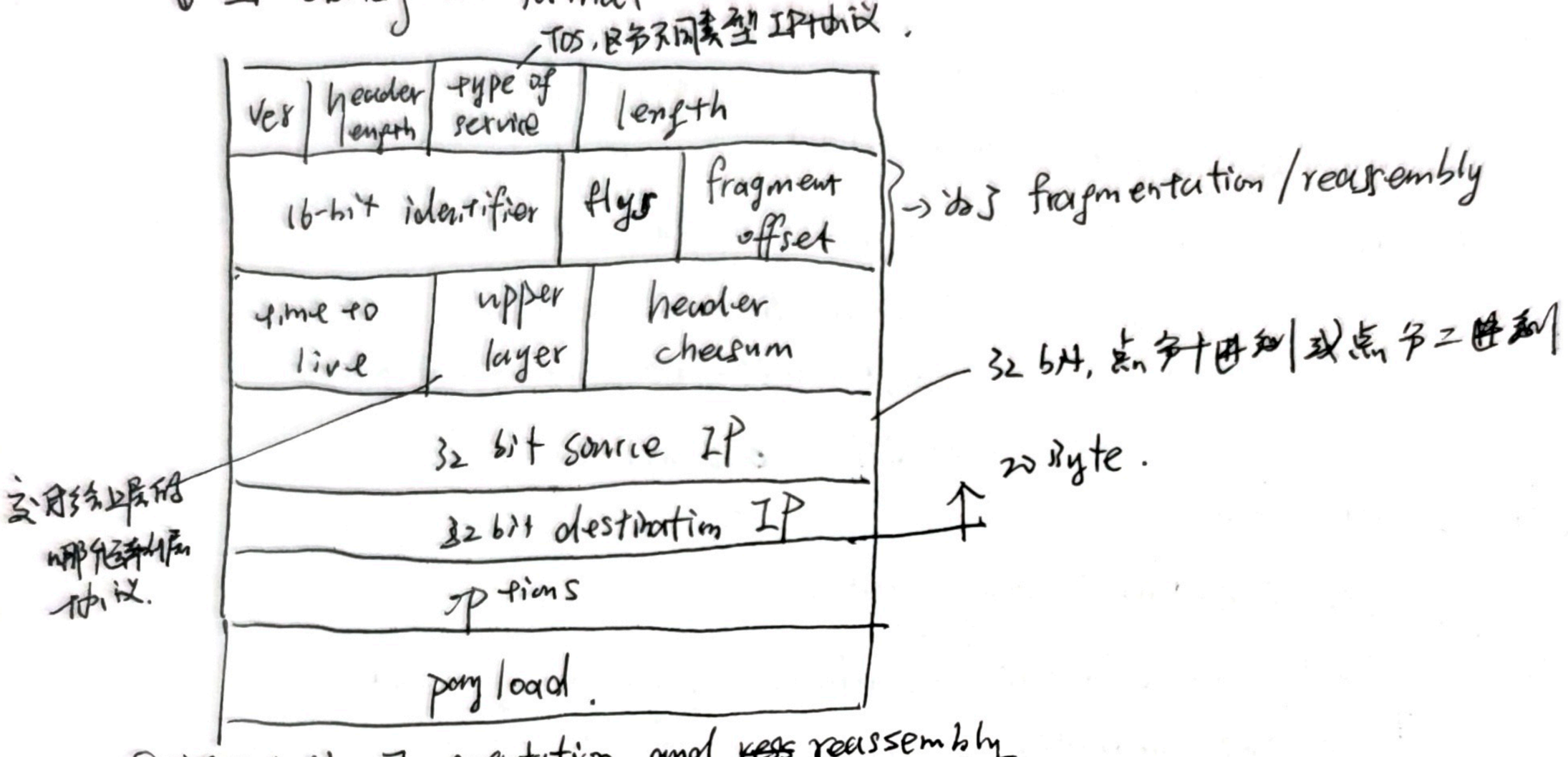
{ destination-based forwarding
generalized forwarding



3、转发表 → Longest prefix matching

4、IPv4.

① IP datagram format.

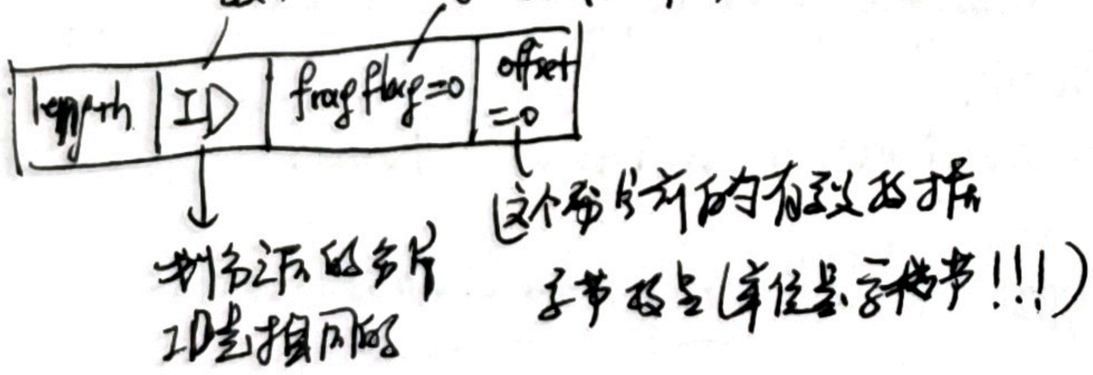


→ 为了 fragmentation / reassembly

② 数据分片，Fragmentation and reassembly

因有 MTU 的存在，并且不同链路有不同的 MTU，发送方和路上的路由器都有对 datagram 进行分片，分片的数据报只有 final destination 进行 reassembled.

IP header bits used to identify, order related fragments.



TCP首部为20，UDP首部为8. → 有时用来计算校载荷.

IP中首部一般20Byte

下层IP中校载荷+首部≤MTU

一般为1500 bytes

③ 子网.

IP address is identifier for host, router interface, associated with each interface.

classful IP addressing. Class A/8, Class B/16, Class C/24 网络部分长度，subnet mask

Subnet: can physically reach each orther without intervening router

<Subnetting>.

What: spliting a block to smaller blocks, each subnetwork having its own subnetwork address.

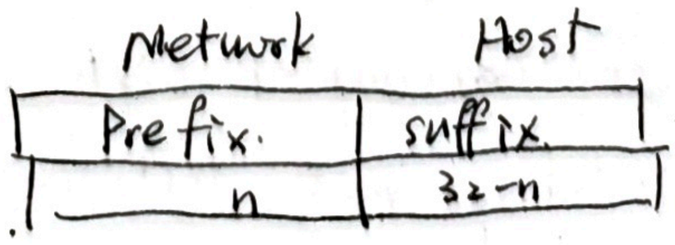
Why: want divide into individual networks, sell, 一个很长的一大段IP，无法全部利用.

a private site router is used to subnetting

划分方法以及背后知识.

也主观叫认为CIDR

Classless addressing - CIDR → 无类别域间路由.



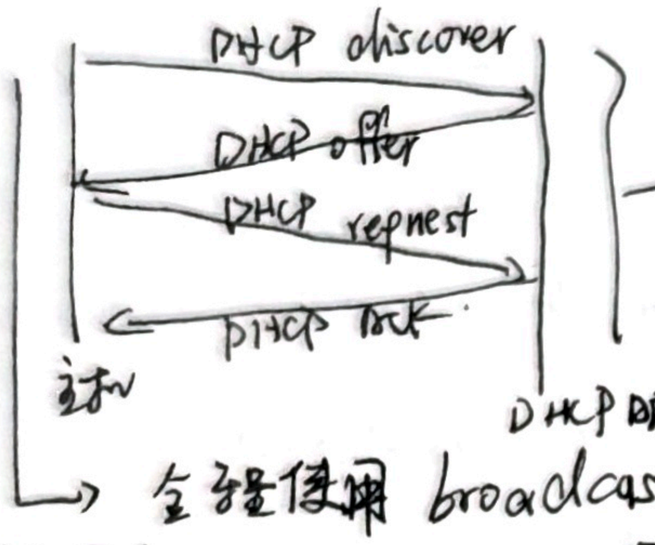
IP → Network address, 除了网掩码，所有主机号全0，表示本地子网 /n.

network address 是子网中第一个可用IP

Broadcast address，主机号全部全1，不可用为主机IP.

两种子网划分方法（已用子网掩码）：
1. 将一块网络划分为几个子网，使得每个子网内主机数相同，靠指派确定子网ID在入分组为，划分出子网的掩码数值相同。
2. Variable length subnet mask (VLSM)，每个子网内的主机数量不同，从大的子网开始，优先确定用户ID的数量（位数），然后从大到小子网设计，一定不能重叠，划分的子网的掩码长度为的不同，因为优先确认的是主机数量。

5. DHCP = Dynamic Host Configuration Protocol.
goal = allow host to dynamically obtain its IP address from network server when it joins network.



→ 都是 message.

DHCP不仅可以获得主机IP，
区别以：first-hop router IP.
IP, name for DNS sever.
network mask.

TCP只有点到点通信，所以 DHCP报文
是在UDP协议之上的

→ 全程使用 broadcast.

6. 使用分层寻址将路由表项目先多个网络的网户被动的地址聚合
hierarchical addressing → route aggregation (地址聚合，路由聚合).
hierarchical addressing allows efficient advertisement of routing information.

<Block 3>.

8.5.7. Network Address Translation. (NAT). } IPv4 address are running out
IPv4 address are unfairly allocated.

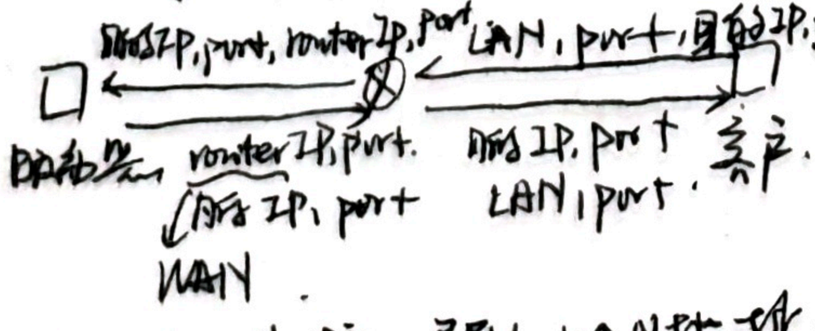
使用 NAT router 隐藏子网中的 hosts，使得整个子网从外部看起来仅有1个IP地址.

NAT router 中有一个

NAT translation table.	
WAN side address	LAN side address
138.76.29.7:5001	10.0.0.1:3345

IP address = port
32 bit. 16 bit.

不需要从ISP申请，可以更改主机IP，可以更改ISP和主机IP不变.

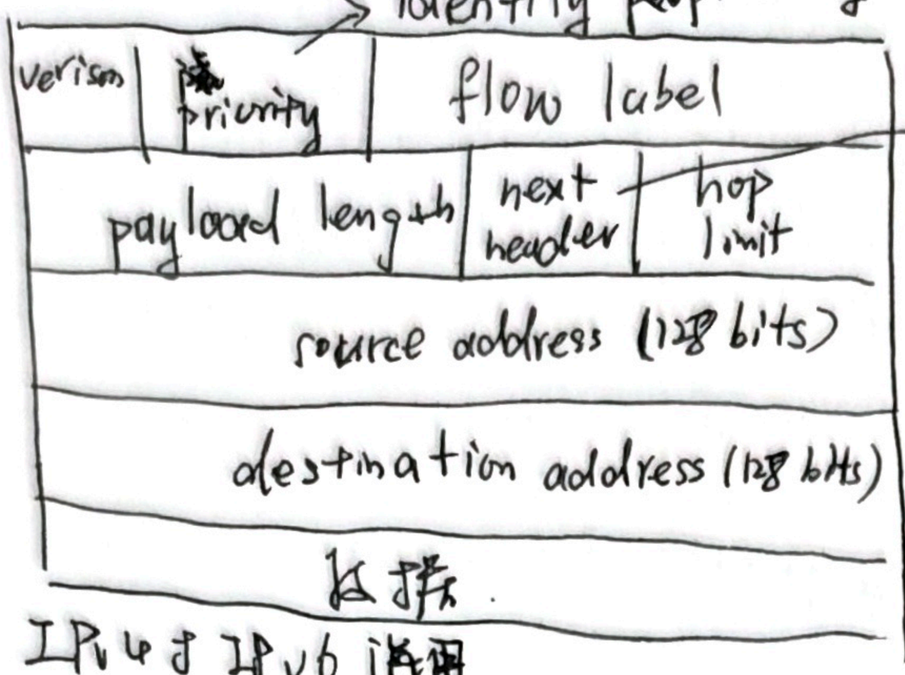


路由器的对外地址都为WAN地址，对于主机来说，可以把 NAT router当作透明的.
对于外部来说，它表现不到客户的存在，只知道在与 router (NAT) 通信.
Controversial: router 在第3层用到第4层信息，违反了网络已被IP的解决，violate end-to-end argument

8. IPv6.
动机：IPv4 用完了.
additional motivation: header format helps processing/forwarding
40 byte 首部，无 fragmentation 和

① IPv6 datagram format.

→ identify priority among datagrams in flow



→ 需要交付给哪个运输层协议.

首部长 40 Bytes.

② IPv4 与 IPv6 选择.

tunneling: IPv6 datagram carried as payload in IPv4 datagram among IPv4 routers

7. Generalized Forward and SDN.

SDN → 控制 → Data plan: match-action can do many different things to packets, much more flexible

logically centralized routing controller → Control plane: a single centralised control point, programmable, fast

↓ 多条 flow table.

OpenFlow ← a specific SDN protocol.

Rule	Action	States

Rule: 多层匹配 Link, Network, Transport 层信息, IP protocol 不为空.

Action: forward, encapsulate and forward to controller; drop, send to normal processing pipeline; modify fields

States: packets + byte counters

B. Network layer (Control) <简>.

1. per-router control / logically centralized control

2. Routing algorithm → global / decentralized

Autonomous System
↳ a group of routers under a single administrative authority

static / dynamic.

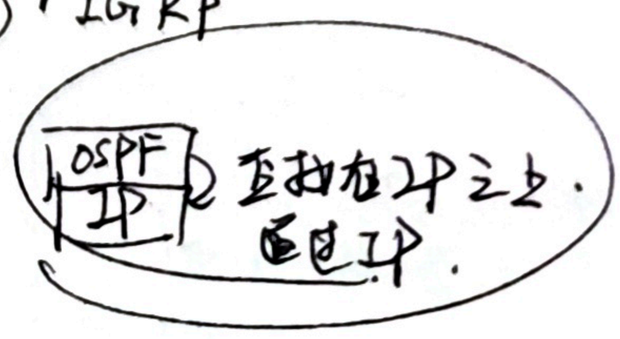
Dijkstra → oscillation possible.

DV → $D_x(y) = \min_v \{ c(x,v) + D_v(y) \}$, const to infinity → poisoned.

3. Intra-AS OSPF.
↳ autonomous system. IGP (Interior gateway protocols) → RIP / OSPF (LS-IS) / IGRP

OSPF → LS algorithm, floods OSPF LS advertisements in entire AS

↳ security, multiple path; multiple costbroad, hierarchical OSPF.



互相在IP之上.
在IP.

4. BGP.

Boraber Gateway Protocol.

基于TCP之上

prefix + attributes = A route　　eBGP. iBGP.

↳ (AS-PATH, NEXT-HOP)

BGP route selection { local preference value attribute = policy decision.
shortest AS-PATH.
closest NEXT-HOP router = hot potato routing
additional criteria.

5. SDN control plane.

Why SDN { easier network management
table-based forwarding allows "programming" routers.
open implementation of control plane

network-control app
SDN controller } control plane.

↕ openflow → 也行在TCP之上

data plane.

6. ICMP. internet control message protocol. → Traceroute → { UDP. unreachable port
↕ name & address
router, RTTs record.

在IP数据报之上 → 必须处理ICMP.

communication network level information.

内容 = type, code, plus 8 Byte of IP datagram causing error

↳ 错误报告发送方确定引发该差错的 datagram.

C. Network security basics.

1. firewall

isolates organization's internal net from larger Internet, allowing some packets to pass, blocking others.

对fw → attack, illegal modification/access of internal data.

① Stateless packet filtering

Access Control Lists; table of rules.

② Stateful packet filtering

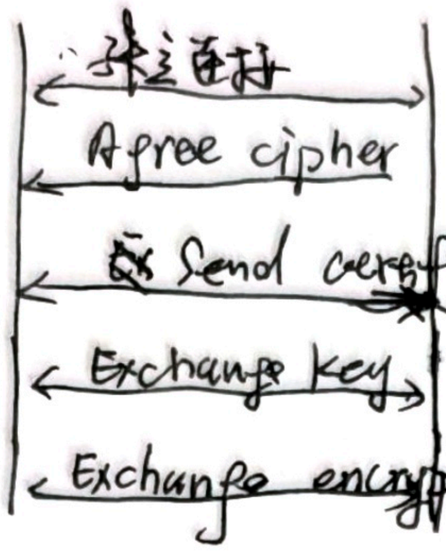
track status of every TCP connection (SYN, FIN, SYN ACK timeout inactive connection).

③ Application gateways.

require all telnet users to telnet through gateway, relays data between connection

2. HTTPs security → TLS (Transport Layer Security) { correct website / seenre connection.

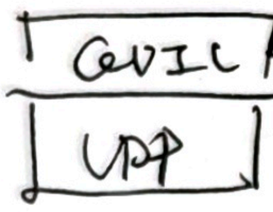
TLS



send certificate + public key
exchange key
exchange encrypted data.

Asymmetric encryption case verified public key for seenre exchange)
symmetric encryption (for fast exchange of data).

3. 快速UDP互联网连接 QUIC.



应用层.

more efficient.

4. Network ethics question

informed consent, consent but ~~dead~~ not informed consent

{ some details of uses my be accessed by orthers.
↓ human subject. (harm to orthers).
agree and understand.

<Block 4>. Data Link layer.

1. Services.

framing, reliable delivery between adjacent nodes.
flow control, error detection, error correction.

<where?> → 网络接口控制器(NIC) 网络接口卡 network interface card.

软硬件结合.

2. Error detection.

CRC 如何检错.

{ Parity checking
Internet checksum.
CRC.

3. multiple access links, MAC protocol

partition channel. { TDMA, time division multiple access.
FDMA, frequency

<Random access protocol>.

① Slotted ALOHA.

retransmits frame in each subsequent slot with prob p until success.

full rate, 简单, decentralized.

② Pure ALOHA

不对时隙, 效率减半.

③ CSMA (carrier sense multiple access).

collision依然发生 → propagation delay

④ CSMA/CD (collision detection)

发送时基带也在发送, 立刻停止, 发送 jamming signal.

停止之后, NIC执行 binary backoff time.

第 m 次碰撞, 随机等待时间 大二 Random {0, 1, 2, 3, ... $2^m - 1$}.

以单位 (512 bits times).

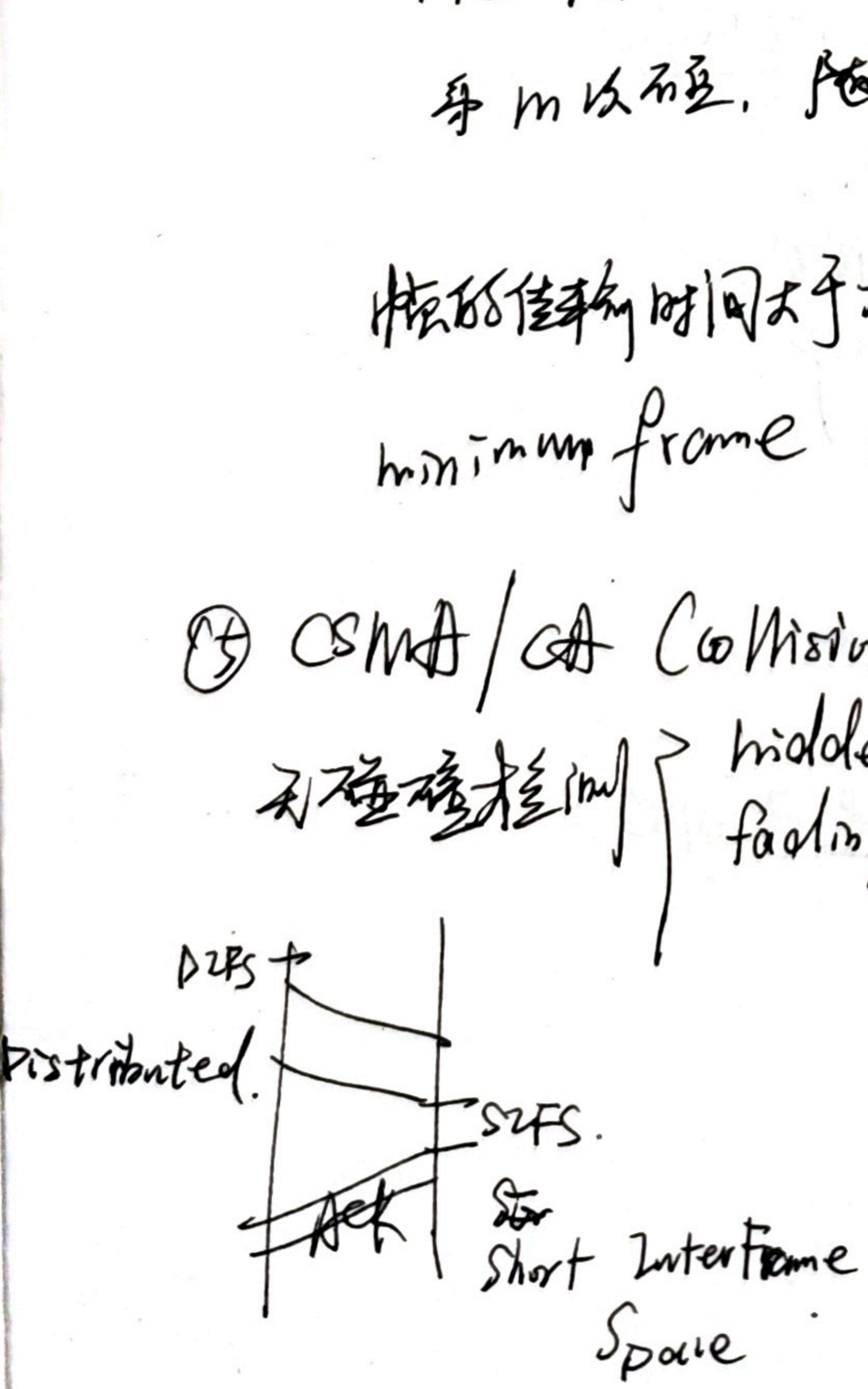
帧的传输时间大于检测时间, 才能检测到 collision.

$$\text{minimum frame size} = \underbrace{RTT} \cdot R$$

$2\frac{L}{V}$ collision window

⑤ CSMA/CA (Collision Avoidance).

无线链路检测 { hidden station/terminal.
fading weak ~~recei~~ received singals.



RTS, CTS → 解决隐藏终端问题.

Request to Send Clear to Send.

都发于播发区.

4. MAC address.

48 bit, 8 Byte, 十六进制.

MAC 对应一个 NIC (adapter)

MAC → never changes, flat address, portability.

IP → can changes, hierarchical.

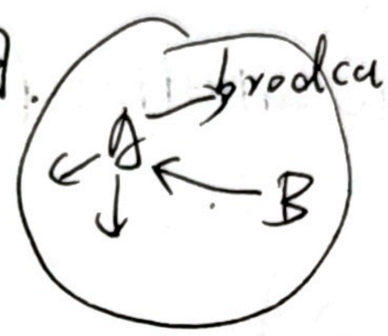
<ARP, Address resolution protocol>.

ARP table. <IP, MAC, TTL>.

* query.

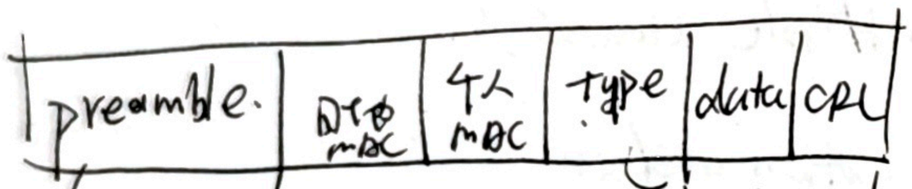
局域网.

broadcast ARP query packet. (B's IP)



IP → ARP → MAC.

5. Ethernet.



synchronize receiver, sender clock rates.

connectionless, unreliable/unslotted CSMA/CD with binary backoff.

6. Ethernet switch. (transparent)

self-learning ← Switch forwarding table. | MAC | 接口 | TTL |

filtering/forwarding { 无 MAC (源网段), 注洪.
有 MAC (源自己), 转发.
有 MAC (丢弃), drop.

7. wireless network.

infrastructure mode / adhoc mode.

{ 802.11 frame.
地址 1, 2, 3
接收站 发送站 路由器接口

802.11和802.3通信, 对于交换机来说, 好像透明的.
路由器

只有 host 才能够知道 MAC 存在

8. Virtual Local Area Network

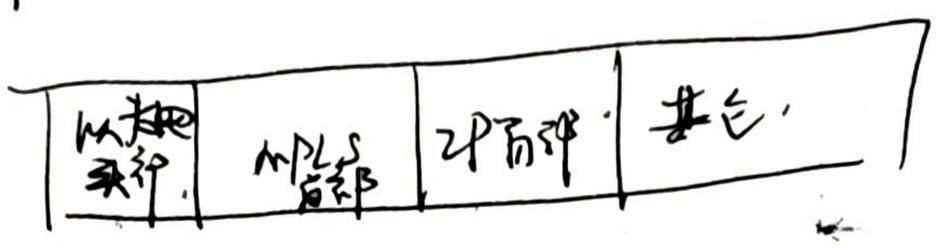
① port-based VLAN 不同 VLAN 互通 router 实现

② trunk port. 802.1q 多 VLAN.

f. MPLS (Multiprotocol lable switching).

使用固定 label 进行高速IP forward.

MPLS 一IP 平台.



MPLS router

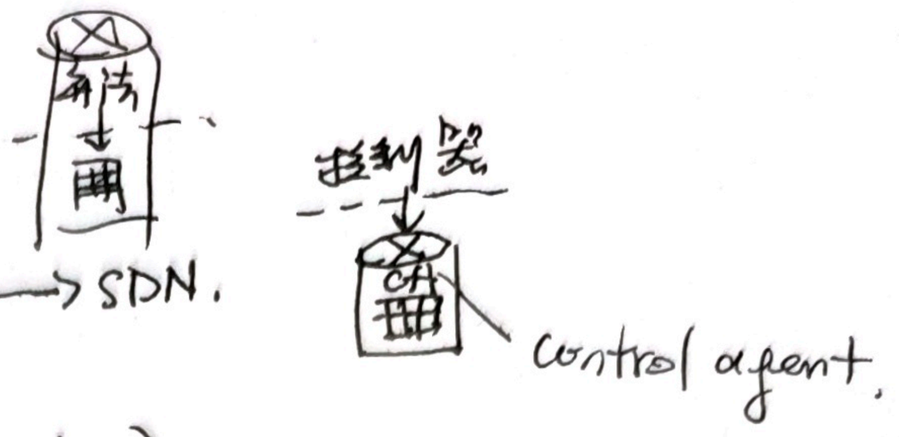
传统IP，只能用IP路由

MPLS → 可和网络IP一起.

MPLS forwarding table.

<网络层·控制层面>.

一、概述



二、路由选择算法.

- 分类
  - 集中式路由选择算法(Link State)
  - 分散式路由选择算法(Distance-Vector)
  - 静态、动态路由选择算法
  - 负载敏感、负载迟钝路由选择算法(OSPF, BGP).

1) 链路状态路由选择算法(LS).

所有网络状态已知 → 每个节点产生最短路径集.

Dijkstra 算法,给定一初始节点,找到它到所有节点的最短路径.

- Input: 初始化邻节点的状态 ($D(v)=c(u,v)$, $p(v)=u$)
- Loop: 加入最小 $D(w)$ 的 w 进子集
- 使用 $D(v)=\min\{D(v), D(w)+c(w,v)\}$ 更新节点.
- until: N' 为全集.

(u 为初始节点, $D(w)$ 为到 w 的最短开销, $p(w)$ 为到 w 的前一个节点, $c(u,v)$ 为 u,v 开销)

→ 使用该算法后,得到 u 路由器的转发表.

时间复杂度 $n(n+1)/2$.

缺点:同步状态下,拥塞都会振荡 → 异步 → 链路面临时间随机性.

2) 距离矢量路由选择算法(DV).

迭代、异步、自终止、分布式.

动态规划 Bellman-Ford equation: $d_x(y)=\min_v\{c(x,v)+d_v(y)\}$.

($d_x(y)$: $x\to y$; $d_v(y)$: $v\to y$; $c(x,v)$: $x\to v$ 邻居点.)

DV算法
- Input: 初始化每个点的路由选择表,并向邻节点发送.
- Loop: wait(链路状态变化或收到某台直连发来)
  $D_x(y)=\min_v\{c(x,v)+D_v(y)\}$
- If: 节点的 routing table 变化(对于节点 x)
  向邻节点发送 D_x.

(注意发送的条件,发送的内容也很重要,要让每个节点收的都是本事为起点的成本量,当然每个节点收的都不一样)

按打性测试法开辟了每一步有用于存放路径和全程开销表,本书未提及.

缺点:链路开销动态变化时容易成 routing loop,可以使用毒性逆转处理,但当节点多起来时并不实用.

LS与DV比较：
- 计算复杂度（收敛时间）：LS $O(N^2)$ > DV（无环时较慢）
- 健壮性：LS各点生成独立的表，DV但一处错误波及整个网络
- 报文复杂性：LS 发送 $O(|N||E|)$ 个报文（全广播）
  DV只需邻节点交换报文

三、Internet中自治系统内部路由选择（OSPF）

互联网络相接大方法使用同一种协议，因特网是 ISP 的网络，ISP有自己的算法。

一般一个ISP一个AS，也包含多个AS（Autonomous System，AS）

一个AS由自行决定路由算法 → 自治系统内部路由选择协议（intra AS ...）
- OSPF
- IS-IS
- ...

开放最短路优先（OSPF）：链路状态协议
- 洪泛链路状态信息
- Dijkstra：路由器向AS其他 router 广播到相邻的距离信息，未变化也周期性广播

优势：无环，快，n->v多条路径，单播多播支持，支持AS层次分割：使用主机域对一个AS负责。

四、ISP之间的路由选择（BGP）→ 四种的IP任播（anycast）

AS间路由选择协议（inter-AS）→ Internet 也是网关协议，BGP（Border gateway protocal）

在BGP中，为每个前缀路由到一个CIDR前缀，即转发表（为，下一）→ CIDR → 接口

1）通告BGP路由信息

AS路由器 →
- 网关（gateway router）
- 内部（internal router）

BGP Connection
- eBGP：若不属于相连AS
- iBGP：相邻对应

gateway router 向其 eBGP 收到，再以 iBGP 对每个AS。

2）确定最好的路由

路由（route）
- 前缀
- BGP attribute
  - AS-PATH → 已通告的AS列表
  - NEXT-HOP → 里潮性播出程，路由当前AS最近的AS的 gateway router 的IP地址
    → 指它的表

选择算法：

a）热土豆路由选择（hot potato routing）

只关注AS（当前）内部相邻最优

b）路由器选择算法

自顶而下的选择条目：
- 本地偏好属性 → 路由选择策略，AS间routing主导作用
- 最短AS路径
- 热土豆
- BGP标识符

五、SDN控制平面.

逻辑上集中，物理上分布.

路由选择 进入控制 ... 东西协议 —北向API→ SDN网络控制应用程序.

控制平面. → SDN控制器

- 对网络控制应用程序层的抽象
- 网络范围 状态 管理层
- 通信层(OpenFlow)

SDN控制器 → 南向API

SDN控制器 特征:
- 基于协议的军发
- 各种控制技术平面
- 可编程的网络.

六、ICMP：因特网控制报文协议.

ICMP，被主机和路由器用来彼此沟通网络层的信息. —困惑→ 差错报告.

ICMP在IP的有效载荷中，有类型字段和编码字段

类型和编码字段 之后即为首部校验和.

Traceroute程序使用ICMP实现，逐次增大TTL字段，获得随后产生的警告ICMP

到达目的主机时，因内容是不为其主端口，获得ICMP.

<链路层和局域网>

从主机到另一主机通信 → 路由选择协议 → 规划路径 → 链路层协议 → 协行每段路径.

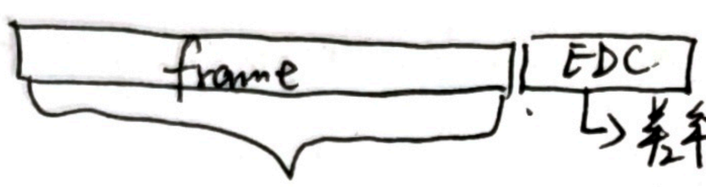
实现 → 硬件(适配器) → 网络适配器 或 网络接口控制器(NIC)
软件

实现功能 → 成帧 / 链路接入 / 可靠交付 / 差错检测和校正.

→ 每协议栈中软硬件交接的地方.

一. 差错检测和纠正技术 (比特级).

增强检测:
(发送方)



↳ 差错检测和纠正比特
(Error-Detection and -Correction, EDC).

检测方也可能查不到差错, 即 未检出比特差错 (undeteced bit error).
差检测和纠错能力当 前向纠错 (Forward Error Correction, FEC).

1) 奇偶校验.

parity bit → 不可检测偶数个bit-error

two-dimensional parity → 单个bit错, 检测并纠正
1)→j行j列, 分别做校验. 两个bit错, 检测不可纠正.

2) 校验和.

d bit → k bit (×n) 序列 → 全加n (回卷加n) → 反码 → Checksum.

3) 循环冗余检测. (Cyclic Redundancy Check)

→ 全为1, 出现0即错误.

互联网 循环冗余检测 (CRC) 编码, 也称多项式编码.

发送方设置CRC —— $R = ((D \cdot 2^r))_G$

R: r bit余数; D: d bit; 2^r: 左移r位补0; G: 生成多项式, r+1比特

所有运算都是 ⊕ 无 XOR.

接收方收到 D R, 按 DR 的 (d+r) bit, 使用 G 去除.

则 $\frac{d+r}{G}$ 余数 = 0 → 无错
other → 差错.

三、多路访问链路和协议（MAC协议）

broadcast link → 多路访问问题（multiple access problem）

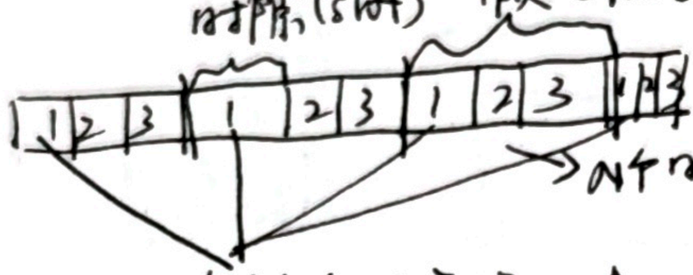
局域网 → 地理上集中在一块区域的网络.

两个或多个结点同时传输 → 碰撞（collide）→ 多路访问协议 → 三类 { channel partitioning / random access / taking-turns.

channel partitioning → 时分 { 单节点发送R吞吐量. / M节点平均 R/M 吞吐量 / 去中心化，易实现 / 时隙(slot) 帧(time frame)

1）信道划分协议.

时分多路复用（TDM）



每个节点公平限制 R/M bps，并等待传本轮 后 Hz，公平划分帧率.

频分多路复用（FDM）

最大速率被限制在 R/M.

码分多址（Code Division Multiple Access, CDMA）→ 应用于无线通信 / 每队 Code 同时传本轮.

2）随机接入协议.

一个节点总以信道全部速率发送，碰撞后等待随机时延

a）时隙ALOHA.

发生 collide，以 p 概率在每个时隙中重传，成功为止.

效率 $Np(1-p)^{N-1}$ $\xrightarrow{\text{limit}}$ 0.37.

b）~~ALOHA~~

相比于ALOHA，取消时隙概念，信号随时开传 collide 后，P概率立刻重发，(1-P)概率等待一个帧传输时间，再P概率决定.

效率 = $Np(1-p)^{2(N-1)}$ $\xrightarrow{\text{limit}}$ $\frac{1}{2e}$ ≈ 0.37.

c）载波侦听多路访问（CSMA）.

Carrier Sense Multiple Access → { 别的节点未传本轮（当前未听见收到） / 自己才传本轮.

受主干到达时延（channel propagation delay）影响.

d）具有碰撞检测的载波侦听多路访问（CSMA/CD）

Collision detection → 发生碰撞立刻停止，停止一个随机时间量.

二进制指数后退 { 经过n次碰撞后，K = Random{0, 1, 2, ..., 2^n-1} / 随机时间量 = K·(512 bit 传输时间)

效率 = $\frac{1}{1+5\,d_{prop}/d_{trans}}$

3) 轮流协议.

a) 轮询协议(polling protocal), 主节点轮询每个节点并指定最大帧

b) 令牌传递协议 (token-passing protocal), 传递令牌, 节点持有令牌才能指定帧

不超过令牌允许最大帧.

主节点损坏? 令牌丢失? 都会问题.

三、交换局域网.

1) 链路层寻址和ARP.

a) MAC地址

Medium Access Control, 介质访问控制 链路层地址 { LAN address / physical address / MAC address } 等价.

正如IP地址对于接口, MAC地址对应路由器和主机的NIC.

(链路层交换机无MAC地址).

MAC地址为6字节, 用十六进制对, 一共六对表示.

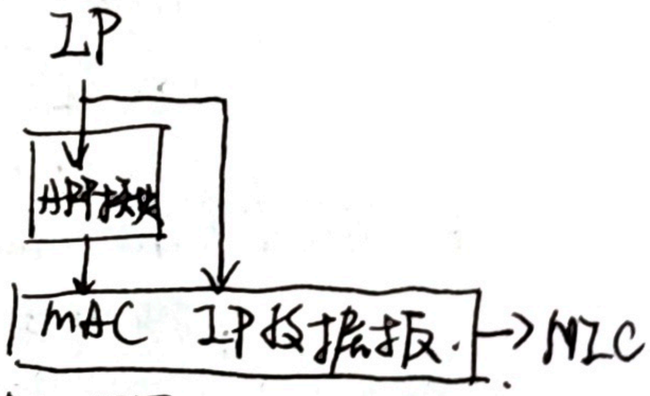
MAC具有扁平结构, 不变, IP有层次结构, 可变.

对于局域网上任意MAC broadcast address = FF-FF-FF-FF-FF-FF.

b) 地址解析协议.

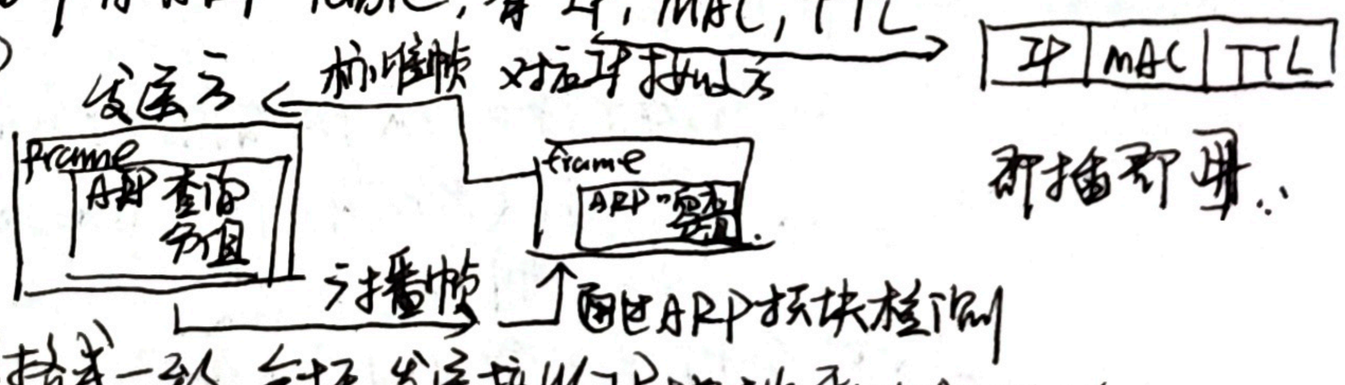
IP ←Address Resolution Protocal, ARP→ MAC.

ARP只由同一子网上的主机和路由器接口解析IP.



每台主机或router的内存中有ARP table, 有IP, MAC, TTL

获取table方法



ARP查询分组和响应分组格式一致, 都有 发送者的IP地址和MAC地址.

ARP分组封装在链路层帧中, 含MAC, IP. 是跨越 网络层和链路层的协议.

c) 发送数据到子网以外.

主机向另一子网发送数据报, MAC地址并强调的主机对应的 MAC地址, ARP只在同一个子网中使用, 对于该子网, 并没有MAC对应的IP, 会丢弃.

主机 →帧→ 路由器接口 →转发→ 另一接口 →成帧→ 另一主机

ARP

路由器接口MAC

另一主机MAC.

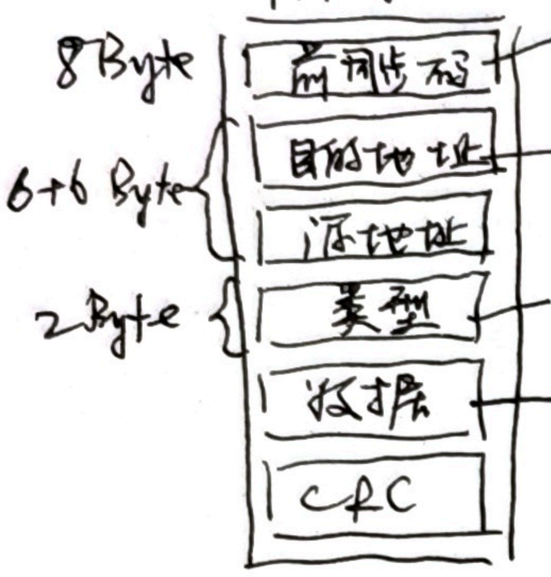
2) 以太网.

交换以太网, 无碰撞, 有储转发分组交换机.

a) 以太网帧结构.

主机A 以太网帧 → 转发 → 帧 → IP数据报 → 主机B.

以太网帧结构：



前7Byte为10101010，最后一个为10101011，用于时钟同步，以标志帧传输开始

MAC，广播地址或一个NIC的MAC地址.

支持的网络层protocol，一般为IP

MTU=1000Byte，范围46~1500字节，少于补齐，大于加首部标识字段区分，多于分片.

以太网提供无连接服务(面向网络层).

(b) 现代以太网或许已经需要MAC协议，但是帧结构保持不变.
(因为交换机以存储转发.

5) 链路层交换机.

对于子网中的主机和路由器，交换机是透明的.

a) 转发和过滤

借助于交换机表(switch table)来完成交换机表记录的：

地址(MAC)	通往该地址的接口	表项记录时间

到一个目的MAC的到达交换机，3种情况：

1. 表中无MAC地址，向该机之外的所有接口缓存广播该帧.
2. 查询到收出接口与该帧入的是个接口，丢弃该帧进行过滤.
3. 有对应的某个接口，向某个接口的缓存转发.

b) 自学习(self-learning).

即插即用设备(plug-and-play device).

1. 全表
2. 对每个接口收帧，有储存(①MAC地址 ②接口 ③时间)
3. 一段时间(老化期 aging time)后，交换机没有收到以MAC1为源地址的帧，删除这个表项.

交换机任何接口能用时会自动接收，全双工.

c) 性质.

1. 消除冲突，交换机不会在同一个网段上转发好什么缓存帧.
2. 异质媒介连接.
3. 管理，一个NIC(异常会话帧，断开异常NIC连接.

d) 交换机与路由器.

交换机
- 优：即插即用，过滤转发速率快.
- 缺：防止广播帧loop，限制拓扑结构.
  - 部署大，处理慢.
  - 对广播风暴不提供保护.

→ 用小网络，不需要工作配置.

路由器
- 优：分层次，有已知路径也不循环，可扁平不限制在生成树拓扑
  对第二层广播风暴提供防火墙保护
- 缺：非即插即用，处理时间长.

→ 大网络(非即用).
要降低.

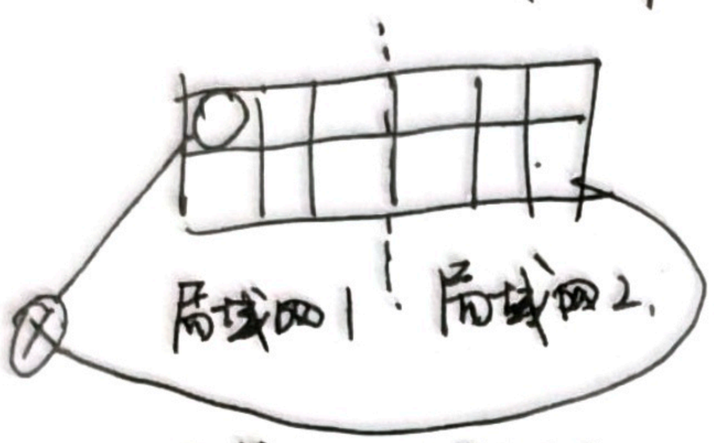
应用层
↑端口号字段
运输层
↑协议字段.
网络层
↑类型字段
链路层.

4) 虚拟局域网

交换机的缺点：
- 缺乏流量隔离.
- 无效使用（没有充分利用端口容量）
- 管理用户必须改变布局.

→ 支持虚拟局域网 (Virtual Local Network, VLAN) 的交换机可解决.

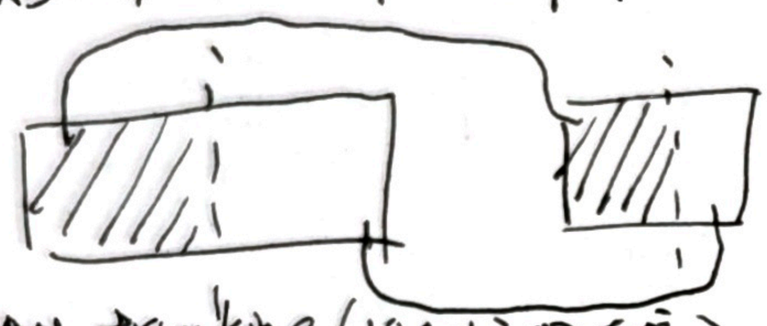
VLAN交换机允许一个单一的物理局域网基础设施定义多个虚拟局域网.



局域网1,2只能有各自的内部任意两台主机互连的帧,
发到局域网任意两台主机并与不属于1,2的端口用路由器相连.

对于需要划分多个端口，如何操作呢.

1.



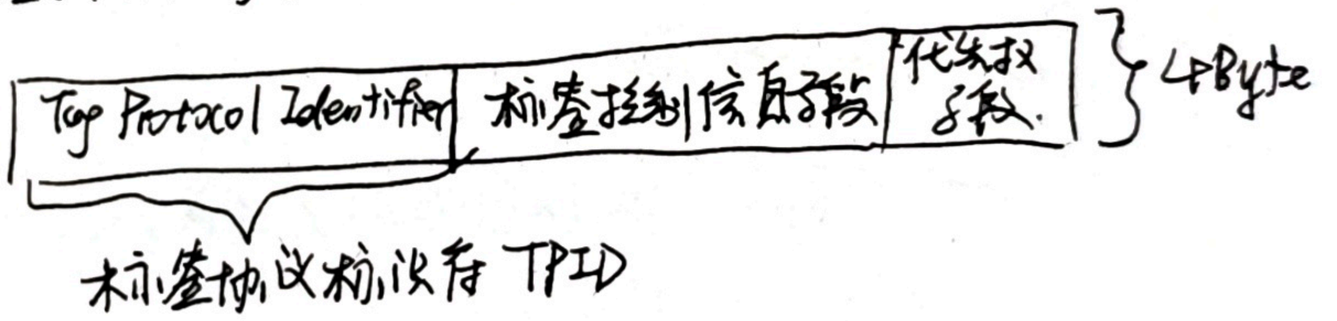
新增一个交换机，相同局域网互联
不具备扩展性.

2. VLAN trunking (VLAN干线连接).

开辟一特殊端口，属于所有的交换机所有的 VLAN.
如何判断该交换机的帧属于哪个VLAN？

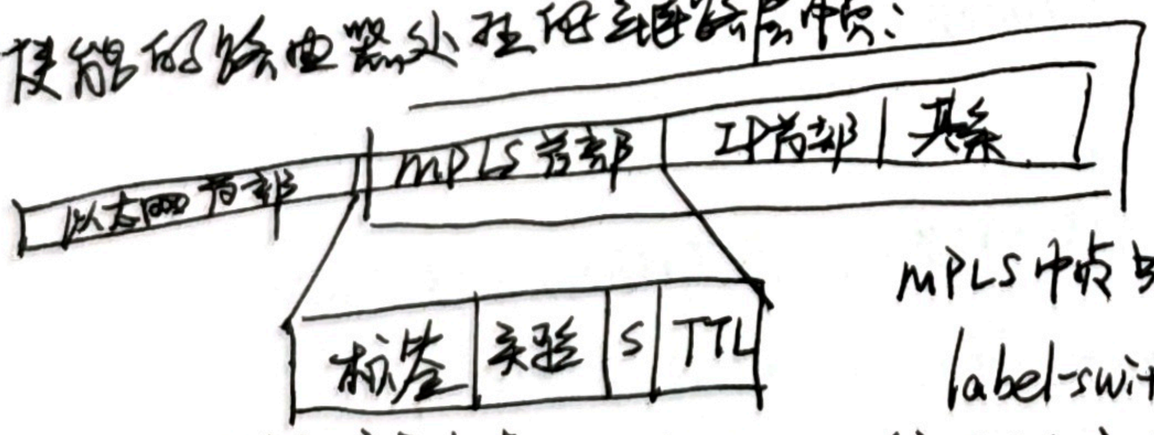
以太网帧进行扩展 → 802.1Q. 用于跨越VLAN的帧.

在标准以太网帧中添加 4 Byte的 VLAN tag.



五、多协议标签交换(Multiprotocol Label Switching, MPLS)，为IP提供面向连接的分组转发技术.

MPLS使能的路由器处理的链路层帧：



帧有改变.

MPLS帧只能在MPLS使能路由器中使用.
label-switched router 标签交换路由器.

MPLS router 在转发表中查找 MPLS tag，然后转发，无须提取IP再匹配查找.

基于标签执行交换，不涉及IP地址

将路由本身视为MPLS网络.

转发速度增加.
提供多条路径到目的网 → 流量工程
(traffic engineering)

Virtual Private Network. VPN.

五. 数据中心网络.

Internet ··· 边界路由器 (border router) ← 数据中心 ← 数据中心网络 → 用于连接内部主机.

主机 → 刀片 (blade). 机架顶部交换机 (Top of Rack, TOR).

1. 负载均衡.

负载均衡器 load balancer
- 基于目的立前地号和目的工件的决策.
- 功能
  - 向主机发送请求，均衡主机负载.
  - 类似NAT功能，转换工件，提供安全.

2. 多级的交换机.

数据中心网络应用 路由器和交换机等级结构(Chierarchy of router and switch)

解决问题：不同机架两台主机间最大速率受限.

作法：① 高速率交换机和路由器.

② 相关服务和数据整合的接口，以减少采购后.

③ 增强相邻交换机层次的连接.

3. 云服务趋势

① 成本降低. ④ 集中式SDN控制和管理. ⑦ 虚拟化

② 物理结构 ⑤ 硬件模块化和互操作.

〈无线网络和移动网络〉.

一、概述.

· wireless host, wireless communication link

base station 基站 ⟶ 接入点 (access point, AP), 链路层中往

主机与基站的关联 → 基础设施模式(infrastructure mode).

未关联 → 自组织网络 (ad hoc network)

主机在接切换基站 → 切换 (handoff)

二、无线链路和网络特征.

1. 路径损耗 (path loss)
2. 其他干扰
3. 多径传播 (multipath propagation)

多路访问
- 隐藏终端问题 (hidden terminal problem)
- 信号强度衰减 (fading)

⇒ CDMA, 同时接收多路径, 由编码恢复.

三、Wi-Fi：802.11无线局域网.

a) 体系结构.

基本构件 → 基本服务集 (Basic Service Set, BSS)

部署此的无线局域网为基础设施无线局域网 (infrastructure wireless LAN.)

b) 信道与关联.

各个无线站点在能发送/接收网络层数据之前，必须与一个AP相关联.

AP { 服务标识符 Service Set Identifier, SSID
信道号(共11个).

每个时间期性发送(广播)信标帧(包含SSID和MAC)，通过扫描信道发现AP，选择关联.

passive scanning: AP发送信标帧.
主站向AP发送关联请求帧
AP向主站发送关联响应帧.

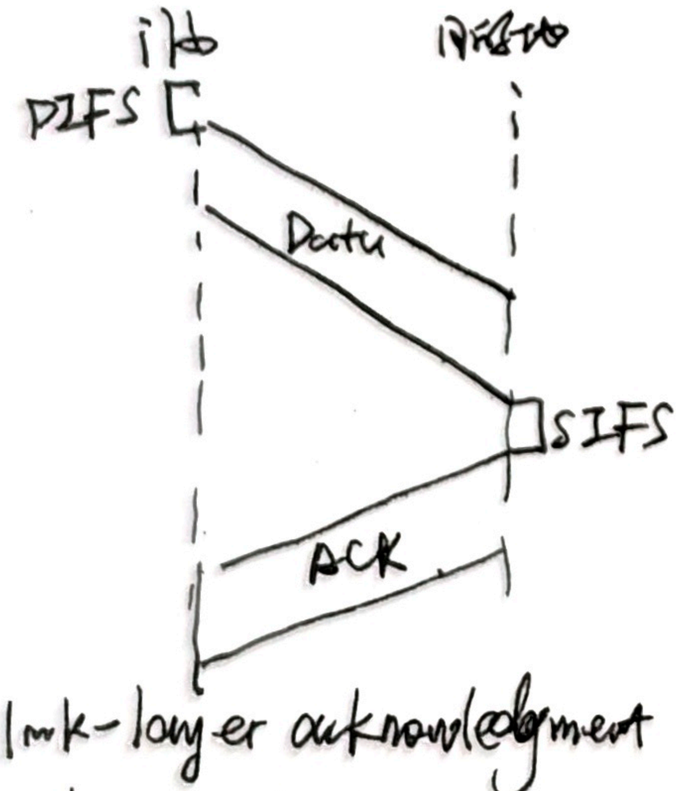
active scanning: 站点主动探索，广播请求帧
AP发送探测响应帧
主站向AP发送关联请求帧
AP向站点发送关联响应帧.

c) 802.11 MAC协议.

随机接入协议，采碰撞避免的CSMA (CSMA with collision avoidance, CSMA/CA)
比特差错率较高，使用链路层确认/重传ARQ (Automatic Repeat request)

并未实现碰撞检测的原因:
1. 物理硬件代价大(具有同时发送和检测的能力).
2. 电子隐藏终端和暴露问题无法检测所有碰撞.

1) 碰撞避免.



link-layer acknowledgment

分布式帧间间隔 (Distributed Inter-frame space)
短帧间间隔 (short Inter-frame space).

① 站点监听到信道空闲，等待 DIFS，发送.

② 监听不空闲，选择随机回退值，在空闲时递减计数，信道忙时回退值不变.

③ 回退值计数值>0，发送并等待确认.

④ 收到ACK，发送另一帧，再发回退值.
未ACK，选择更大范围回退值.

若站点相互隐藏，无法检测.
非常重要的回退值，可能碰撞.

2) 处理隐藏终端问题: RTS/CTS.

Request to Send, Clear to Send 主动侦听帧.

RTS. CTS 都发给播发送

可以设置门限，仅将RTS/CTS用于长序列发送的分组.

d) IEEE 802.11帧.

① 具有4个MAC地址字段.(自组织时才使用第4个).

地址1是接收帧的无线站点(即MAC, eAP或站点)

地址2是(发)MAC地址. (同一个AP或LAN的一般不经过路由器).

地址3是路由器接口MAC地址(可以是目的也可以是源)

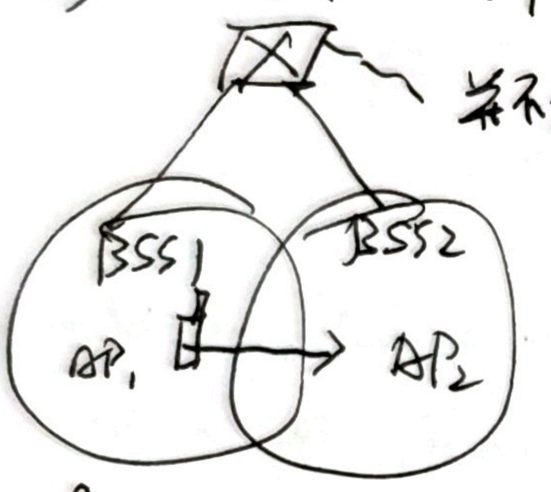
AP可以接收更以太网中其他LAN帧并转发.

② 具有序号字段,B名新传输帧中成未确定值.

③ 持续期字段(传输时间和确认传输后确认时间).

④ 帧*控制字段有类型和子类型用于区分关联,RTS,CTS,ACK和数据帧.

e) 在扩用子网工中的移动性.



若不经路由器,则从BSS1到BSS2属于同一子网,IP到下不会变换.

信号表弱,收到其它信标帧 → 直接建立新关联(先解除).

对于交换机,接口改变,利用自学习,让主机换BSS2之后使AP2

向交换机广播主机MAC,交换机更新交换表.

f) 特色.

速率适应,自适应调制技术.

功率管理,睡眠和唤醒.