计算机网络

第一章_引言

- flow control
  - An allocation problem that occurs at every level is how to keep a fast sender from swamping a slow receiver with data. Feedback from the receiver to the sender is often used.
- congestion control
  - Sometimes the problem is that the network is oversubscribed because too many computers want to send too much traffic, and the network cannot deliver it all. This overloading of the network is called congestion. One strategy is for each computer to reduce its demand when it experiences congestion. It, too, can be used in all layers.
- transmission of technology
  - 1. Broadcast links.
  - 2. Point-to-point links.
- Services
  - A service is a set of primitives (operations) that a layer provides to the layer above it.
- protocol,
  - is a set of rules governing the format and meaning of the packets, or messages that are exchanged by the peer entities within a layer.
- OSI

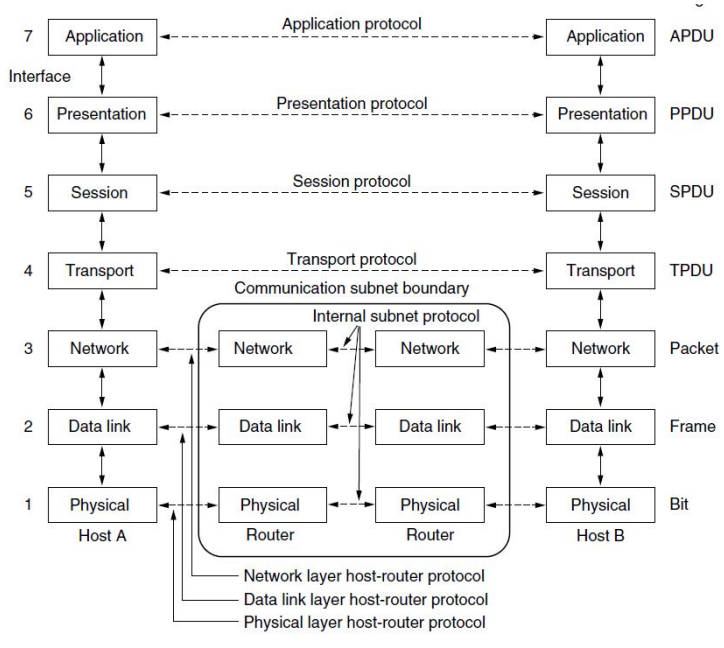


Figure 1-20. The OSI reference model.

  - The transport layer is a true end-to-end layer
  - The difference between layers 1 through 3, which are chained, and layers 4

through 7, which are end-to-end
    - session
        - dialog control
        - token management
        - synchronization
    - Presentation
        - concerned with the syntax and semantics
- TCP/IP

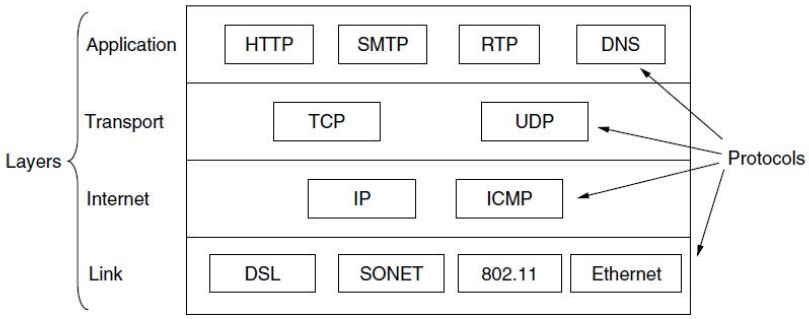


Figure 1-22. The TCP/IP model with some protocols we will study.

- 比较
    - Three concepts are central to the OSI model:
        - 1. Services.
        - 2. Interfaces.
        - 3. Protocols.
    - Both have (inter)network, transport, and application layers
    - OSI先有模型，TCP/IP先有协议

第七章_应用层

- DNS
    - 超过250个顶级域名
        - 两种类型：通用的、国家或地区的
    - An absolute domain name always ends with a period (e.g., eng.cisco.com.)
    - 各组成部分最多63个字符，整个路径最多255个字符
    - 域名资源记录
        - 一条资源记录是一个五元组

Domain_name Time_to_live Class Type Value

        - Class---对于Internet信息总是IN
        - type

Type	Meaning	Value
SOA	Start of authority	Parameters for this zone
A	IPv4 address of a host	32-Bit integer
AAAA	IPv6 address of a host	128-Bit integer
MX	Mail exchange	Priority, domain willing to accept email
NS	Name server	Name of a server for this domain
CNAME	Canonical name	Domain name
PTR	Pointer	Alias for an IP address
SPF	Sender policy framework	Text encoding of mail sending policy
SRV	Service	Host that provides it
TXT	Text	Descriptive ASCII text

    - 递归查询：
        - 从发起方到本地域名服务器
    - 迭代查询：
        - 从本地服务器
- 邮件
    - The ASCII encoding of binary data is called base64 encoding. In this scheme, groups of 24 bits are broken up into four 6-bit units, with each unit being sent as a legal ASCII character. The coding is ‘‘A’’ for 0, ‘‘B’’ for 1, and so on, followed by the 26 lowercase letters, the 10 digits, and finally + and / for 62 and 63, respectively. The == and = sequences indicate that the last group contained only 8 or 16 bits, **3B->4B**
    - SMTP
        - 是一个简单的ASCII协议
        - 不包括认证，FROM可以给出任何发件地址
        - 以明文出现
        - 25号端口
    - IMAP
        - 监听143端口
    - POP3
        - 110端口
- 万维网
    - Tim Berners-Lee
    - URL
        - the protocol
        - the DNS name of the machine
        - the path uniquely indicating the specific page
    - MIME类型

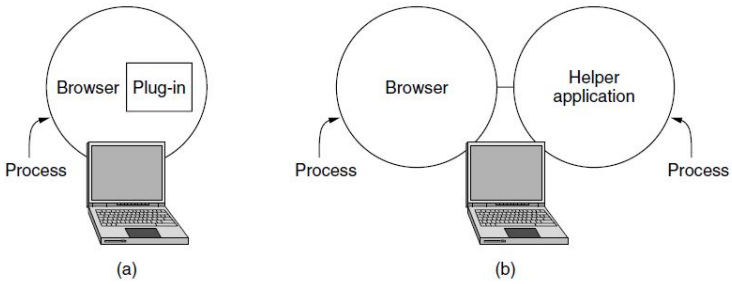


Figure 7-20. (a) A browser plug-in. (b) A helper application.

- cookie
  - 过期时间 expires
    - 退出时丢弃，非持续
    - 提供时间和日期，持续
- 服务器端动态web页面生成
  - CGI 公共网关接口
    - 允许web服务器与后端程序及脚本进行通信。
    - Python、Ruby、Perl
  - 在HTML页面嵌入脚本
    - PHP
  - Java服务器页面（JSP JavaServer Pages）
    - 与PHP相似
  - 活动服务器页面（ASP.NET, Active Server Page）
- 用来生成动态页面的不同技术

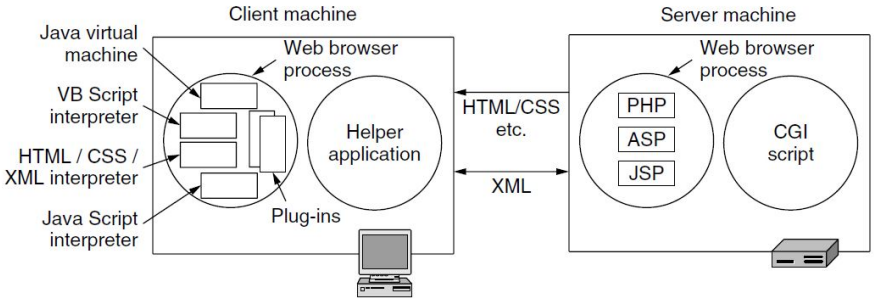


Figure 7-35. Various technologies used to generate dynamic pages.

- HTTP
  - HTTP response messages never have an empty message body. ->false
  - HTTP 1.0 with multiple connections
  - HTTP 1.1 with persistent connections
  - HTTP 1.1 with persistent connections and pipeline requests.
- web cache

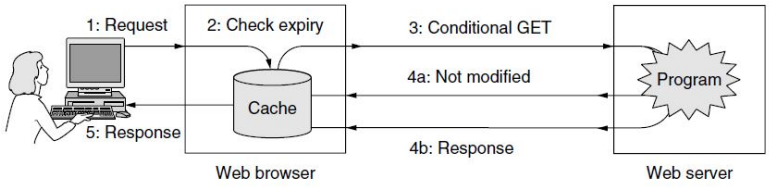


Figure 7-40. HTTP caching.

- 第二章_物理层
- 尼奎斯特公式

The maximum data rate of a channel

– In 1924, Henry Nyquist derived an equation expressing the maximum data rate for a finite bandwidth *noiseless* channel:
Max data rate = 2H $\log_2$ V bits/sec, or
Max symbol rate = 2H
where H is the bandwith, V is the discrete levels

– If we use B to denote the data rate (bits/s), and S to denote the symbol rate (symbol/s),
B <= 2H $\log_2$ V
S <= 2H

- 香农公式

The max data rate of a channel

– In 1943, Claude Shannon extend it to channel *with random noise*
Max data rate = H $\log_2$ (1+S/N) bits/sec
where S/N is the signal to noise ratio

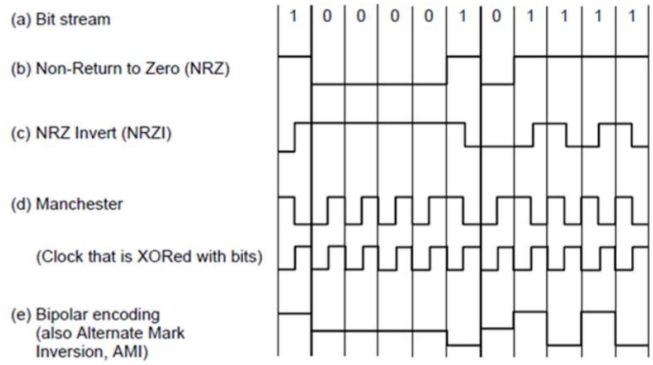
– If we use B to denote the data rate (bits/s)
B <= H $\log_2$ (1+S/N)

– SNR: typically measured in dB, decibel
dB = 10$\log_{10}$ S/N
Ex: S/N=10→10dB, S/N=100→20dB, ...

- GUIDED TRANSMISSION MEDIA 有线数据传输
  - •Magnetic media (磁介质)
  - •Twisted pair (双绞线)
  - •Coaxial cable (同轴电缆)
  - •Power Lines (电力线)
  - •Fiber optics (光纤)
- WIRELESS TRANSMISSION 无线传输
  - •The electromagnetic spectrum (电磁波谱)
    - (spread spectrum, 扩频)
      - –Frequency hopping spread spectrum (military, 802.11, Bluetooth)
      - –Direct sequence spread spectrum (3G mobile phones)
      - –UWB(UltraWideBand)

- •Radio transmission (无线电传输)
  - VLF, LF, MF radio waves follow the curvature（弯曲） of earth.
  - HF, bounce off the ionosphere
- •Microwave transmission (微波传输)
  - The higher the transmission towers are, the further apart they can be.
  - can cause multipath fadingor be absorbed by rain.
  - 管理者
    - ITU, CN, FCC
  - Free frequencies ISM
- •Infrared and millimeter waves (红外线和毫米波)
- •Lightwavetransmission (光波传输）

- DIGITAL MODULATION AND MULTIPLEXING数字调制与多路复用
  - •Baseband Transmission (基带传输)
    - –signal occupies frequencies from zero up to a maximum



    - –NRZ (Non-Return-to-Zero,不归 零 制):
      - the symbols are simply voltage levels.
    - 曼彻斯特
      - 与clock XOR
    - NRZI
      - As a step in the right direction, we can simplify the situation by encoding a 1 as a transition and a 0 as no transition, or vice versa.
      - Used in USB
      - long runs of 0s? -> 4B/5B
  - •PassbandTransmission(通带传输)
    - –signal occupies a band of frequencies around the frequency of the carrier signal
    - 星座图
  - •Frequency Division Multiplexing (频分复用)
    - OFDM正交频分复用
  - •Time Division Multiplexing (时分复用)
  - •Code Division Multiplexing (码分复用)

A = (−1 −1 −1 +1 +1 −1 +1 +1)
B = (−1 −1 +1 −1 +1 +1 +1 −1)
C = (−1 +1 −1 +1 +1 +1 −1 −1)
D = (−1 +1 −1 −1 −1 −1 +1 −1)

(a)

(a) Chip sequences for four stations.
(c) Six examples of transmissions.
(d) Recovery of station C's

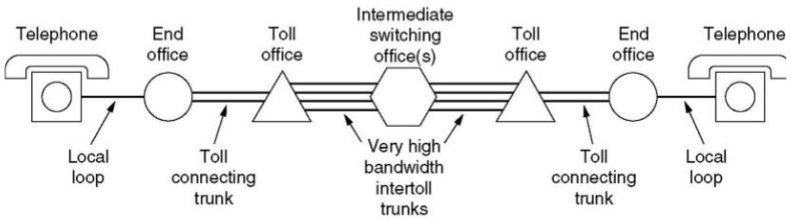
$S_1 = C = (-1\ +1\ -1\ +1\ +1\ +1\ -1\ -1)$
$S_2 = B+C = (-2\ 0\ 0\ 0\ +2\ +2\ 0\ -2)$
$S_3 = A+\bar{B} = (0\ 0\ -2\ +2\ 0\ -2\ 0\ +2)$
$S_4 = A+\bar{B}+C = (-1\ +1\ -3\ +3\ +1\ -1\ -1\ +1)$
$S_5 = A+B+\bar{C}+D = (-4\ 0\ -2\ 0\ +2\ 0\ +2\ -2)$
$S_6 = A+B+\bar{C}+D = (-2\ -2\ 0\ -2\ 0\ -2\ +4\ 0)$

(c)

$S_1 \cdot C = [1+1-1+1+1+1-1-1]/8 = 1$
$S_2 \cdot C = [2+0+0+0+2+2+0+2]/8 = 1$
$S_3 \cdot C = [0+0+2+2+0-2+0-2]/8 = 0$
$S_4 \cdot C = [1+1+3+3+1-1+1-1]/8 = 1$
$S_5 \cdot C = [4+0+2+0+2+0-2+2]/8 = 1$
$S_6 \cdot C = [2-2+0-2+0-2-4+0]/8 = -1$

(d)

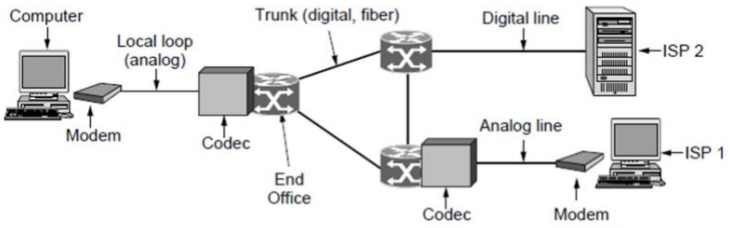
- PUBLIC SWITCHED TELEPHONE SYSTEM 公用电话系统
  - •Structure of the Telephone System
    - five levels
      - end office (本 地局),
      - toll office (长途局),
      - primary office (初级局),
      - sectional office (地区局),
      - regional office (区域局).
    - three major parts
      - –Local loops:
        - the wires between the customers and the switching offices,
        - Analog twisted pairs going to houses and businesses (not open-loop).
      - –Trunks:
        - the long-distance connections between the switching offices,
        - Digital fiber optics connecting the switching offices (used to be coaxial cables).
      - –Switching offices.
        - to switch calls from one trunk to another
    - The call between the subscribers attached to different end offices via toll, primary, sectional, and regional offices.



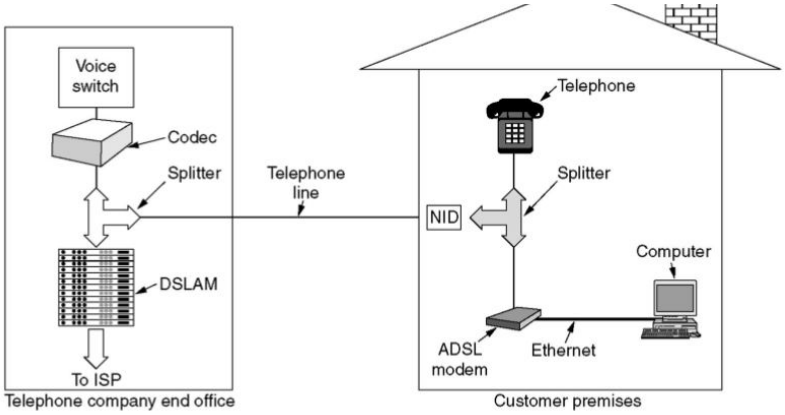
  - •The Politics of Telephones
    - LATA
      - (Local Access and Transport Areas, 本地访问和传输区域).
    - LEC
      - (Local Exchange Carrier, 市话电信局
    - IXC

- InterXchangeCarrier,长话电信局，外部交换

- •The Local Loop: Modems, ADSL, and Wireless
  - modems (modulator demodulator, 调制解调器)
  - codecs(coder decoder, 编码解码器)
  - The data must be converted to analog form for transmission over the local loop.



  - ADSL 非对称数字用户环线
  - typical ADSL equipment configuration
    - –NID (Network Interface Device)
    - –DSLAM (DSL Access Multiplexor)

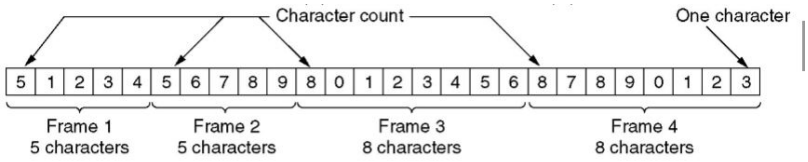


- •Trunks and Multiplexing
  - SONET (Synchronous Optical NETwork, 同步光纤网络)
    - send 810 bytes / 125us 9行90列

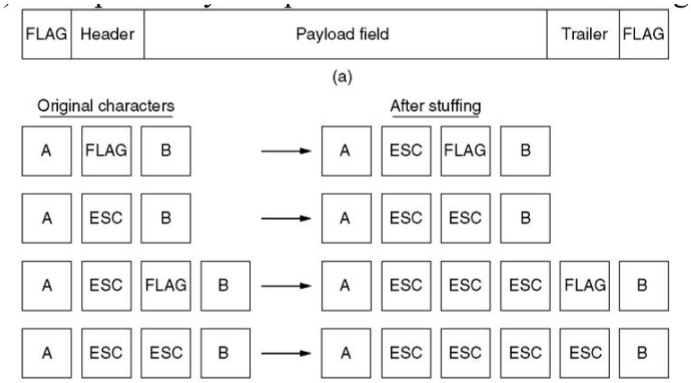
    - SONET frame
      - 前三列：系统管理信息
        - 其中前三行：段的开销
        - 接下来六行：线路的开销
      - 剩下87列=SPE (同步有效载荷信封，其中包含了用户数据)
        - 从任意位置开始（由线路开销第一行指定）
        - 第一列为路径开销
    - Gross data rate 810 B/125us = 6480 b/s = 51.84Mbps
    - Data rate for SPE = 87/90 * 51.84 = 50.112 Mbps
    - User data rate = 86/87 * 50.112 = 49.536 Mbps

  - SDH (Synchronous Digital Hierarchy, 同步数字分级结构)
- •Switching

- 第三章_链路层
- Data Link Layer Design Issues 了解
  - Services provided to the network layer 没有无确认有链接
    - Unacknowledged connectionless service,
      - Ethernet
    - Acknowledged connectionless service,
      - WIFI
    - Acknowledged connection-oriented service.
  - Framing
    - 四种方法
      - Byte count

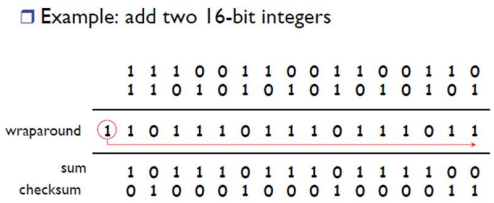


      - Flag bytes with byte stuffing填充



      - Flag bits with bit stuffing
        - Using flag bits with bit stuffing. Ex: flag bits 01111110
        - 遇到5个连续1，补0
      - Physical layer encoding violation.
  - Error control
    - To provide the sender with some feedback
    - To provide timeout timers
    - To number frames
  - Flow control
    - What to do with a sender that systematically wants to transmit frames faster than the receiver can accept them.
    - two approaches
      - –Feedback-based flow control
      - –Rate-based flow control

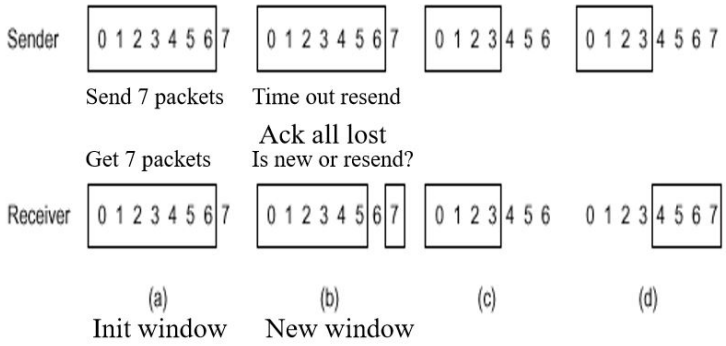
- Error Detection and Correction 重要
- hamming distance
  - # of bit positions in which the 2 codewordsdiffer
  - To detect d errors: you need d+1Hamming distance code.
  - to correct d errors: you need 2d+1 Hamming distance code.
- error correction
  - m message bits, r check bits
  - (m + r + 1) <= 2 ^ r
  - Hamming code for single error:
    - P23
- error detection
  - •Parity
    - 偶校验：若1的个数为偶数，则校验位为0 even
  - •Checksum



  - •CRC
    - 110001 -> x5+ x4+x0.
    - G(X)的阶为r，在frame低端加r个0
    - All single errors can be detected as long as G(X) has more than one term
    - All double errors can be detected as long as G(X) does not divide xk+1 for some k. E(X) = xi+xj = xj(xi-j+1). Ex: x15+x14+1 will not divide xk+1 for any value of k below 32768

- Elementary Data Link Protocols重要
  - •A Utopian Simplex Protocol
    - simplex
    - never damages or loses frames
  - •A Simplex Stop-and-Wait Protocol
    - simplex
    - error free
    - The sender is always ready. The receiver is NOT always ready or the receiver has limited buffer space.
  - •A Simplex Protocol for a Noisy Channel
    - simplex
    - not free of errors
    - receiver not always ready

- Sliding Window Protocols重要
- Sender window increases when ack received
- Piggybacking ACK+DATA IN A FRAME
- 发送方：
  - Window shrinks from trailing edge as frames are sent
  - Window expands from leading edge as acknowledgments are received
- 接收方：
  - Window shrinks from trailing edge as frames are received
  - Window expands from leading edge as acknowledgments are sent
- go back N
  - Send Window Size <= MAX_SEQ (0...N)
  - Receive Window Size=1
  - ACK(n): 确认到序列号n(包含n)的分组均已被正确接收
- Selective Repeat ARQ
  - Send Window Size <= (MAX_SEQ+1)/2
  - Receive Window Size = Send Window Size



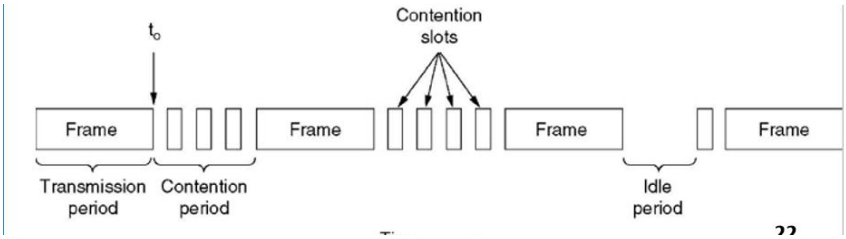
第四章_介质访问控制层

- MULTIPLE ACCESS PROTOCOLS
  - ALOHA
    - pure
      - vulnerable period = 2 frame time
    - slotted
      - vulnerable period = 1 frame time
  - Carrier Sense Multiple Access (CSMA) protocols 载波侦听多路访问协议
    - CSMA without CD (collision detection)
      - Persistent CSMA
        - If the channel is idle, the station transmits a frame.
        - If the channel is busy, the station waits until it becomes idle. Then the station transmits a frame.
        - If a collision occurs, the station waits a random amount of time and starts all over again.
        - The protocol is called 1-persistent because the station transmits with a

probability of 1 whenever it finds the channel idle

- Non-persistent CSMA
  - If the channel is idle, the station transmits a frame.
  - If the channel is in use, the station does not continually sense it. Instead, it waits a random period of timeand then repeats the algorithm.
  - If a collision occurs, the station waits a random amount of timeand starts all over again.
- p-persistent CSMA
  - Applied to slotted channels
  - If the channel is idle, it transmits with a probability p.
  - If the channel is busy, it waits until the next slot and applies the above algorithm.
  - 802.11

- CSMA with CD
  - As soon as stations detect a collision, they stop their transmissions.

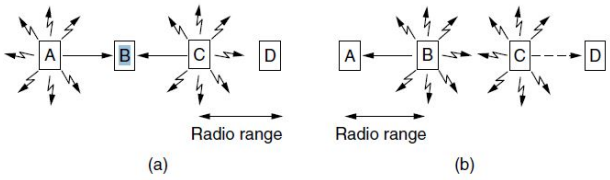


- Collision-free protocols
  - A Bit-Map Protocol

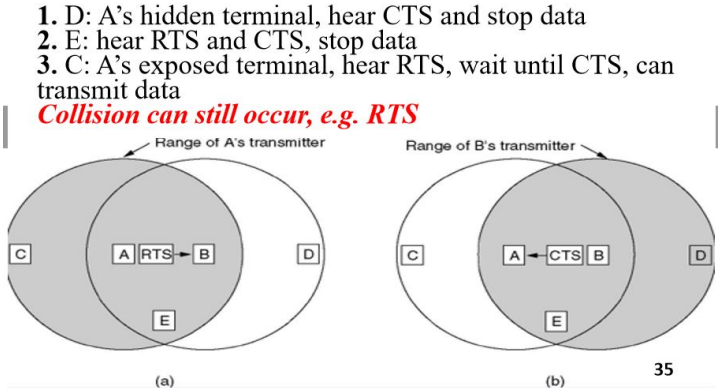
■ The time unit is one contention bit
■ N: contention period
■ d: data frame length
At low load:
■ Delay for low-numbered stations: $0.5N + 1N = 1.5N$
■ Delay for high-numbered stations: $0.5N$
→ The mean delay for all stations is N.
→ The efficiency: $d/(N + d)$.
At high load:
■ the N bit contention period is distributed over N frames, yielding an overhead of only 1 bit per frame,
→ the efficiency: $d/(d + 1)$.
25

  - Token Passing
  - Binary Countdown二进制倒数
    - as soon as a station sees that a high-order bit position that is 0 in its address has been overwritten with a 1, it gives up.

- Limited contention protocols
  - Adaptive Tree Walk自适应树
- Wireless LAN Protocols



- Hidden station problem (隐藏终端问题)
  - 竞争者离得太远而导致站无法检测到潜在竞争者
- Exposed station problem (暴露终端问题)
  - 暴露终端是指在发送接点的覆盖范围内而在接收节点的覆盖范围外的节点。
- The MACA (Multiple Access with Collision Avoidance) protocol. A sending an RTS to B, B responding with a CTS to A.



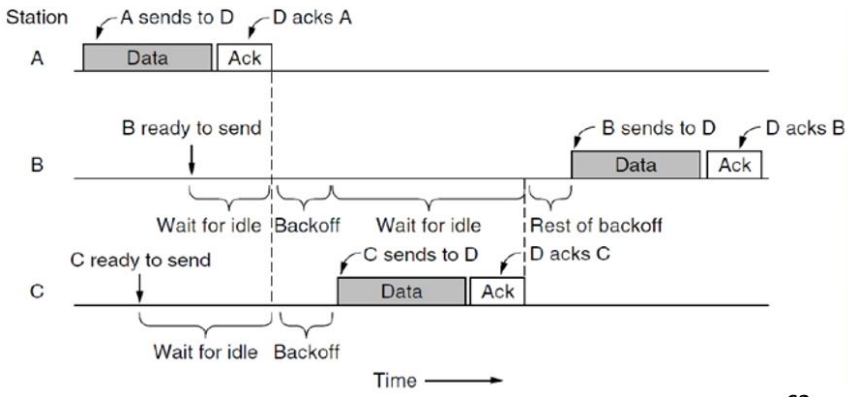
  - 基本思想是发送方刺激接收方输出一个短帧，以便其附近的站能检测到这次传输，从而避免在接下去进行的（较大）数据帧传输中也发送数据。

- IEEE 802.3 (Ethernet)
  - Differential Manchester encoding (used by Token Ring):
    - 在最初信号的时候，即第一个信号时：如果中间位电平从低到高，则表示0；如果中间位电平从高到低，则表示1。
    - 后面的信号（从第二个开始）就看每个信号位开始时有没有跳变来决定：在信号位开始时改变信号极性，表示逻辑"0"；在信号位开始时不改变信号极性，表示辑"1"。
  - MAC Sublayer Protocol

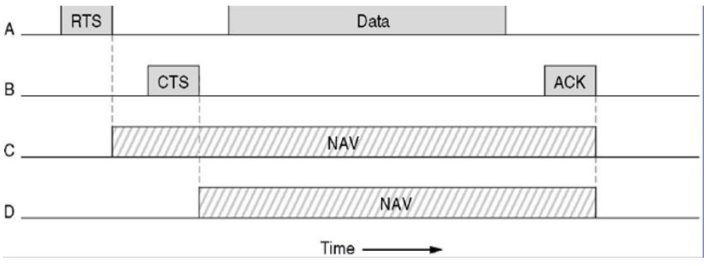
    - Preamble: 8 bytes, containing the bit pattern 10101010, used for synchronization
    - Destination address and source address: 6 bytes each
      - Unicast address
      - Broadcast address
      - Multicast address
    - Type field or length field
    - Data: 1500 bytes (maximum), 46 bytes (minimum)
    - Checksum: 32 bits, CRC

  - Binary Exponential Backoff Algorithm 二进制指数后退
    - After i-thcollisions, each station picks either 0,1,2,..., 2^i-1 at random and waits that number of slot times
    - After 10th collisions, the randomization interval is frozen at a maximum of

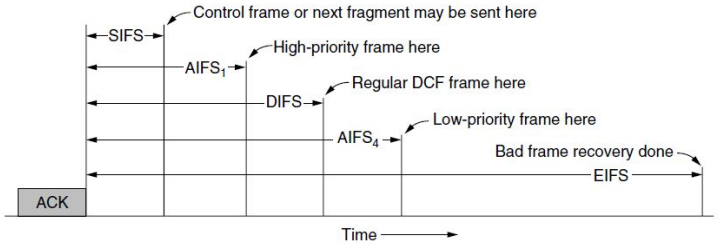
1023 slots.
    - After 16 consecutive collisions, the controller reports failure back to the computer.
  - Fast Ethernet (100-Mbps)
    - IEEE 802.3u or Fast Ethernet (1992~1995)
    - IEEE 802.3z or gigabit Ethernet (1995~1998)
- IEEE 802.11 (WIRELESS LANS)
  - Physical Protocol
    - 802.11: FHSS(Frequency Hopping Spread Spectrum) and Infrared
      - (2.4Mbps)
    - 802.11a: OFDM(Orthogonal Frequency Division Multiplexing) at 5GHz
      - (54Mbps)
    - 802.11b: HR-DSSS(High Rate Direct Sequence Spread Spectrum)
      - (11Mbps)
    - 802.11g: OFDM(Orthogonal Frequency Division Multiplexing) at 2.4 GHz
      - (54Mbps)
    - 802.11n: MIMOOFDM(Multiple-Input Multiple-Output Orthogonal Frequency Division Multiplexing) at multiple frequencies.
      - (600Mbps)
  - CSMA/CA



  - 802.11 takes a conservative approach: Stations hearing RTS cannot send data during NAV, i.e. exposed terminals of A cannot concurrently send data with A (as opposed to MACA)



  - 时间间隔



- SIFS: Short InterFrameSpacing
  - 允许一次对话的各方有有限抓取信道的机会
- PIFS: PCF InterFrameSpacing
- DIFS: DCF InterFrameSpacing
  - 任何站都可以在介质空闲DIFS后尝试抓取信道发送一个新帧
- EIFS: Extended InterFrameSpacing

- Data Link Layer Switching
  - bridge
    - The first 802 bridge is a learning bridge or transparent bidge(透明网桥).
    - Spanning Tree Bridges
      - The bridge with the lowest serial number becomes the root.
  - VLAN
    - IEEE 802.1Q: add a VLAN tag in Ethernet header
  - Which device is in which layer.

Application layer	Application gateway
Transport layer	Transport gateway
Network layer	Router
Data link layer	Bridge, switch
Physical layer	Repeater, hub

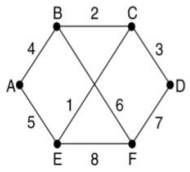
  - 冲突域
    - A collision domainis a section of a network where data packets can collide with one another when being sent on a shared medium or through repeaters, particularly when using early versions of Ethernet. A network collision occurs when more than one device attempts to send a packet on a network segment at the same time.
  - 广播域
    - A broadcast domainis a logical division of a computer network, in which all nodes can reach each other by broadcast *at the data link layer.* A broadcast domain can be within the same LAN segment or it can be bridged to other LAN segments.
  - Some Clarifications
    - The host connected to a single hub belong to

        - –Same collision domain and same broadcast domain
    - •Bridges can
        - –Reduce the collision domain
        - –Increase the broadcast domain
    - •Routers can
        - –Increase the number of broadcast domains
    - switch
        - All stations are in the same broadcast domain H1, H2, … are in the same collision domain

第五章_网络层

- I.Network Layer Design Issues (网络层设计概述)
    - 1.Store-and-Forward Packet Switching (存储转发分组交换)
    - 2.Services Provided to the Transport Layer (为传输层提供的服务)
    - 3.Implementation of Connectionless Service (无连接服务的实现)
    - 4.Implementation of Connection-Oriented Service (面向连接服务的实现)
    - 5.Comparison of Virtual-Circuit and Datagram Subnets (虚电路或数据报子网的比较)
- II.Routing Algorithm(路由算法)
    - Shortest Path Routing (最短路径路由)
    - Distance Vector (DV) Routing (距离向量路由)
        - Delay is used as a metric
        - The algorithm often took too long to converge.
    - Link State (LS) Routing (链路状态路由)
        - Step 1: Learning about the neighbors
        - Step 2: Measuring line cost
        - Steps 3: Building link state packets
        - Steps 4: Distributing the link state packets

Routing Algorithms: LS routing



			Send flags			ACK flags			
Source	Seq.	Age	A	C	F	A	C	F	Data
A	21	60	0	1	1	1	0	0	
F	21	60	1	1	0	0	0	1	
E	21	59	0	1	0	1	0	1	
C	20	60	1	0	1	0	1	0	
D	21	59	1	0	0	0	1	1	

The data structures for router **B**

    - Steps 5: Computing the new routes
    - LS与DV的比较

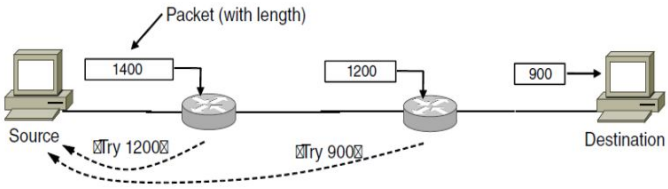
	Link-State	Distance Vector
Summary	"Tell the world about neighbors"	"Tell the neighbors about the world"
# Messages	More (Broadcast)	Less (Exchange)
Message Size	Smaller (Link Costs)	Larger (DVs)
Convergence Speed	Faster	Slower
Robustness	Better	Worse

- Hierarchical Routing (分层路由)

  - How many levels should the hierarchy have?
    - Consider a subnet with 720 routers
      - No hierarchy: every router needs 720 routing table entries.
      - 30 routers/region x 24 regions: every router needs 30 for local entries + 23 for other regions = 53 table entries.
      - 10 routers/region x 9 regions/cluster x 8 clusters: every router needs 10 + 8 + 7 = 25 table entries.
    - ➔Kamount and Kleinrock (1979): The optimal number of levels for an N router subnet is $\ln N$, requiring a total of $e \ln N$ entries per router.

  - Broadcast routing
    - Sink-tree based broadcast
      - optimal
      - Does this optimal broadcasting possible in a network only with DV routing?
        - NO
    - Reverse path forwarding
      - All lines instead of all sink-tree lines, cause some duplicate transmissions
      - work with DV? YES

- III.Internetworking (网络互连)
  - routing
    - –Internet (inter-AS) routing: Exterior gateway protocol
    - –Intranet (intra-AS) routing: Interior gateway protocol
  - fragmentation
    - (a) Transparent fragmentation.
      - 每个网络都感觉不到曾经发生过分段
    - (b) Nontransparent fragmentation
    - path MTU discovery

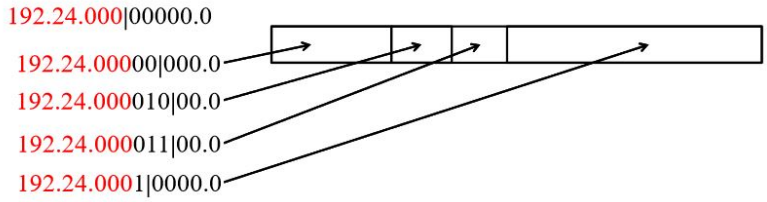


- IV.The Network Layer in the Internet
  - •The IPv4 Protocol
    - Total length:
      - the length of header and data. The maximum is 65,535 bytes.
  - •IP Addresses
    - CIDR

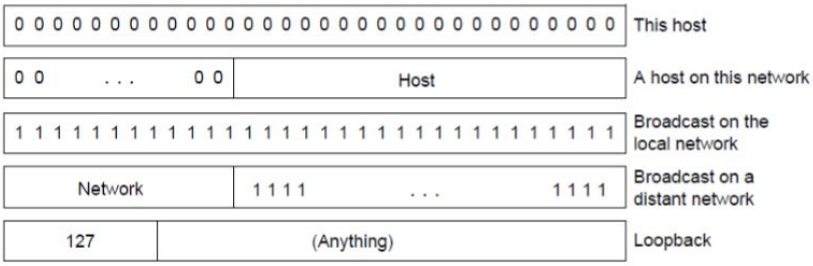
A set of IP address assignments

Divide the prefix 192.24.0.0/19 ($2^{(32-19)}$=8192 IPs) into:

University	First address	Last address	How many	Prefix
Cambridge	194.24.0.0	194.24.7.255	2048	194.24.0.0/21
Edinburgh	194.24.8.0	194.24.11.255	1024	194.24.8.0/22
(Available)	194.24.12.0	194.24.15.255	1024	194.24.12/22
Oxford	194.24.16.0	194.24.31.255	4096	194.24.16.0/20



    - Special IP addresses

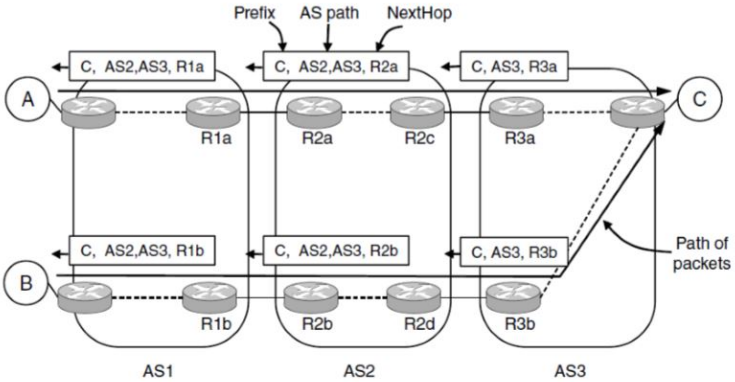


    - NAT网络地址转换
      - –To assign each company a single IP address (or at most, a small number of them) for Internet traffic
      - –Within the company, every computer gets a unique IP address
      - 保留IP地址

- 10.0.0.0/8
  - **10.0.0.0 ~ 10.255.255.255**, 16,777,216 hosts
- 172.16.0.0/12
  - **172.16.0.0 ~ 172.31.255.255**, 1,048,576 hosts
- 192.168.0.0/16
  - **192.168.0.0 ~ 192.168.255.255**, 65,536 hosts

- •The IPv6 Protocol
- •Internet Control Protocols
  - Internet Control Message Protocol (ICMP)
    - When something unexpected occurs during packet processing at a router, the event is reported to the sender by the ICMP.
    - traceroute
  - The Address Resolution Protocol (ARP)
    - IP -> link address
  - DHCP
    - –used to dynamically obtain IP address.
    - –Link address -> IP
  - MPLS (MultiProtocol Label Switching
- •OSPF –The Interior Gateway Routing Protocol
  - A routing algorithm within an AS is called an interior gateway protocol
  - –Distance vector (DV) protocol
    - RIP(Routing Information Protocol)
  - –Link state (LS) protocol
    - OSPF(Open Shortest Path First in 1990), IS-IS
  - OSPF “advanced” features
    - security
    - multiple same-cost paths allowed
    - For each link, multiple cost metrics for different TOS (Type of Service
    - multicast support
    - hierarchical
  - 结构
    - •two-level hierarchy:local area, backbone（0号）
      - –Link-state advertisements only in area
      - –each node has detailed area topology; only know direction (shortest path) to nets in other areas.
    - •area border routers:“summarize” distances to nets in own area, advertise to other Area Border routers.
      - 必须是骨干区域的一部分
    - •backbone routers:run OSPF routing limited to backbone.
      - 骨干区域内的路由器

- •boundary routers:connect to other AS’ s
- RIP
  - distance metric: # of hops (max = 15 hops)
  - each advertisement: list of up to 25 destination subnets within AS
- •BGP –The Exterior Gateway Routing Protocol
  - fundamentally a distance vector protocol
  - Each BGP router keeps track of the path used
  - Each BGP router tells its neighbors the exact path it is using



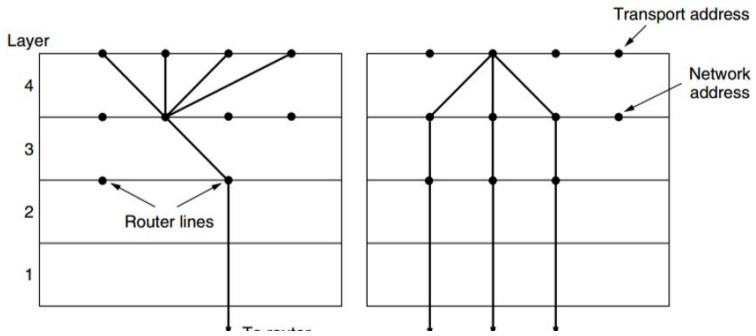
第六章_传输层

- •The Transport Service (传输服务)
  - The transport entity (传输实体):
    - the hardware and/or software within the transport layer that does the work. Its positions:
      - –in the OS kernel, in a separate user process, in a library package bound to network applications,
      - –on the network interface card (NIC).
  - BerkeleySockets

Primitive	Meaning
SOCKET	Create a new communication endpoint
BIND	Associate a local address with a socket
LISTEN	Announce willingness to accept connections; give queue size
ACCEPT	Passively establish an incoming connection
CONNECT	Actively attempt to establish a connection
SEND	Send some data over the connection
RECEIVE	Receive some data from the connection
CLOSE	Release the connection

  - •The C/S Application
    - –Server: SOCKET/BIND/LISTEN/ACCEPT
    - –Client: SOCKET/CONNECT
  - –C/S: SEND/RECEIVE
  - –C/S: CLOSE(symmetric)

- •Elements of Transport Protocols (传输协议的若干问题)
- •Addressing
  - NSAP->ip 32 bits
  - TSAP ->port 16bits
  - portmapper: service name -> TSAP address
- •Connection Establishment
- •Connection Release
  - Asymmetric release:
    - When one part hangs up, the connection is broken.
  - Symmetric release
    - to treat the connection as two separate unidirectional connections and require each one to be released separately.
- •Error Control and Flow Control
- •Multiplexing
  - 多路复用，逆向多路复用（SCTP）



- •Crash Recovery

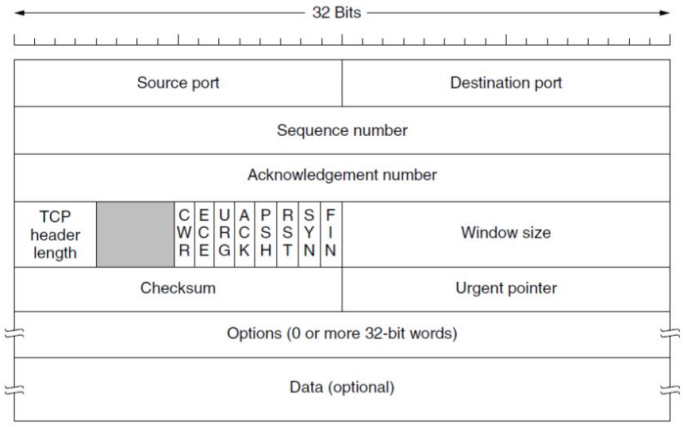
- •Congestion Control（拥塞控制）
  - power = load / delay
  - AIMD
- •The Internet Transport Protocols (Internet传输协议): UDP
  - •UDP (User Datagram protocol)
    - 8-byte header
    - 没有流量控制，拥塞控制和重传
  - •RPC (Remote Procedure Call)
  - •RTP (Real-time Transport Protocol)
    - 应用于多媒体
    - 没有流量控制，没有错误控制，没有确认，也没有重传
    - 有时间戳
  - RTCP
    - Does not transport any data
- •The Internet Transport Protocols (Internet传输协议): TCP
  - reliable end-to-end byte stream

- well-known ports

23 for Telnet, 69 for TFTP, 79 for Finger, 119 for NNTP

Port	Protocol	Use
20, 21	FTP	File transfer
22	SSH	Remote login, replacement for Telnet
25	SMTP	Email
80	HTTP	World Wide Web
110	POP-3	Remote email access
143	IMAP	Remote email access
443	HTTPS	Secure Web (HTTP over SSL/TLS)
543	RTSP	Media player control
631	IPP	Printer sharing

- daemon
  - inetd&xinetd
- All TCP connections are full duplex and point-point.
- Each segment, including the TCP header, must fit in the 65515 (=65535-20) byte IP payload.
- TCP header



  - TCP header length:
    - how many 32-bit words are contained in the TCP header.
  - SYN for CONNECTION REQUEST,
  - SYN+ACK for CONNECTION ACCEPTED.
- Nagle’ s algorithm
  - When data come into the sender one byte at a time, just send the first byte and buffer all the rest until the outstanding byte is acknowledged. Then send all the buffered characters in one TCP segment and start buffering again until they are acknowledged
- Clark’ s solution
  - 禁止receiver发送只有1个字节的窗口更新段
- 计时器管理
  - Retransmission timer

– SRTT (Smoothed Round-Trip Time) (α=7/8)

$$SRTT = \alpha\, SRTT + (1-\alpha)\, R$$

– RTTVAR(Round-Trip Time VARiation) (β=3/4)

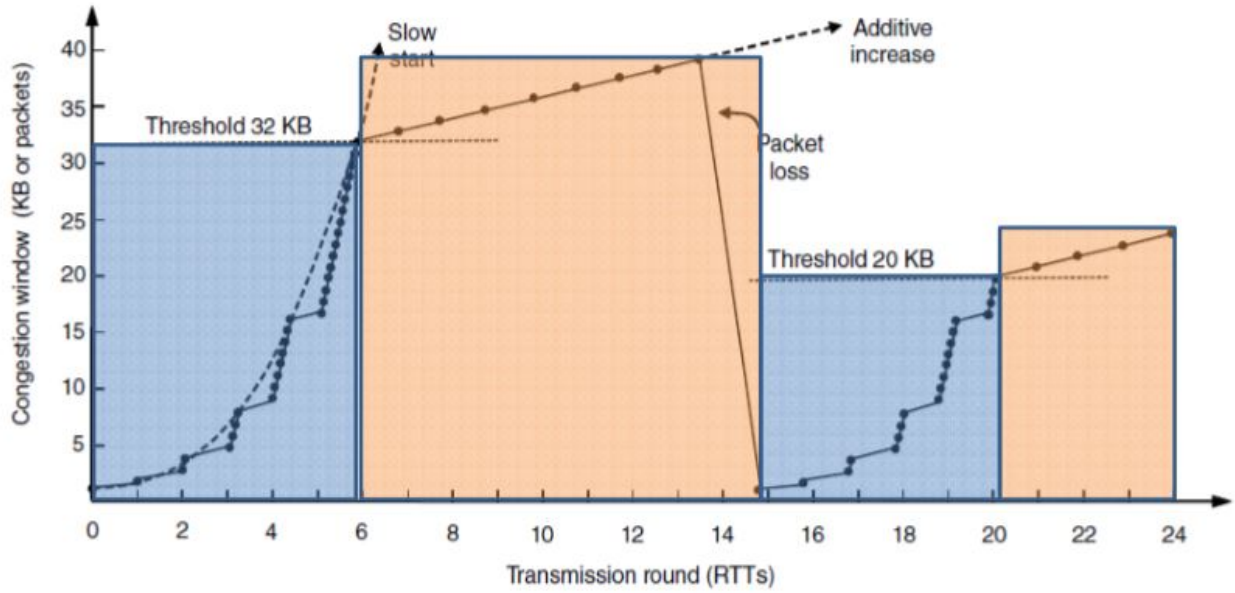
$$RTTVAR = \beta\, RTTVAR + (1-\beta)\,|\,SRTT - R\,|$$

– RTO (Retransmission TimeOut)

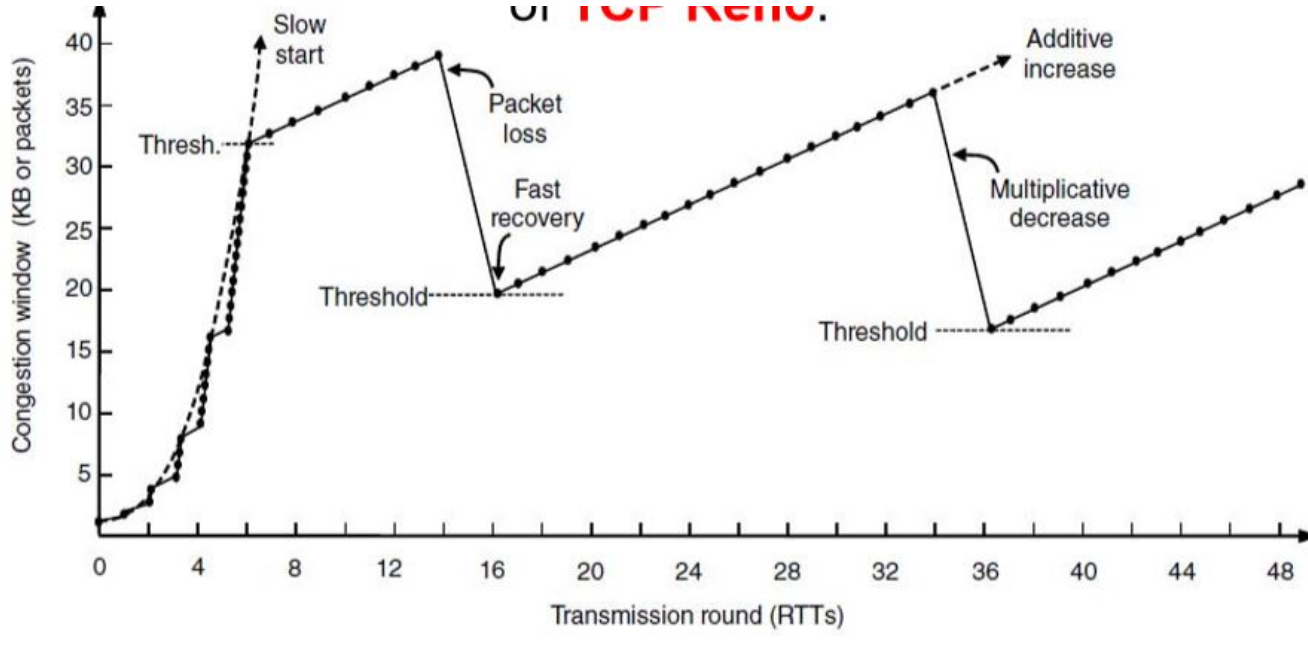
$$RTO = SRTT + 4\, RTTVAR$$

- Persistence timer
  - prevent the deadlock:
- Keepalive timer
  - to check whether the other side is still there.

- TCP Tahoe



- TCP Reno



- After 3 dup ACKs:

        - –cwndis cut in half
    - after timeout event:
        - cwndinstead set to 1 MSS;
- slow-start -> cwnd指数增长
- congestion-avoidance -> cwnd线性增长
- sending rate

- **cwnd** is measured in bytes (or number of pkts)
- Send rate? Implicitly indicated by **cwnd**:
  Rate = **cwnd**/RTT
- # ACKs per RTT = cwnd/MSS
- Linear increase: +MSS per RTT
  → +MSS*(MSS/cwnd) per ACK
- Exponential increase: double Rate
  → +cwnd per RTT
  → +MSS per ACK

- Average throughout: 0.75 W/RTT