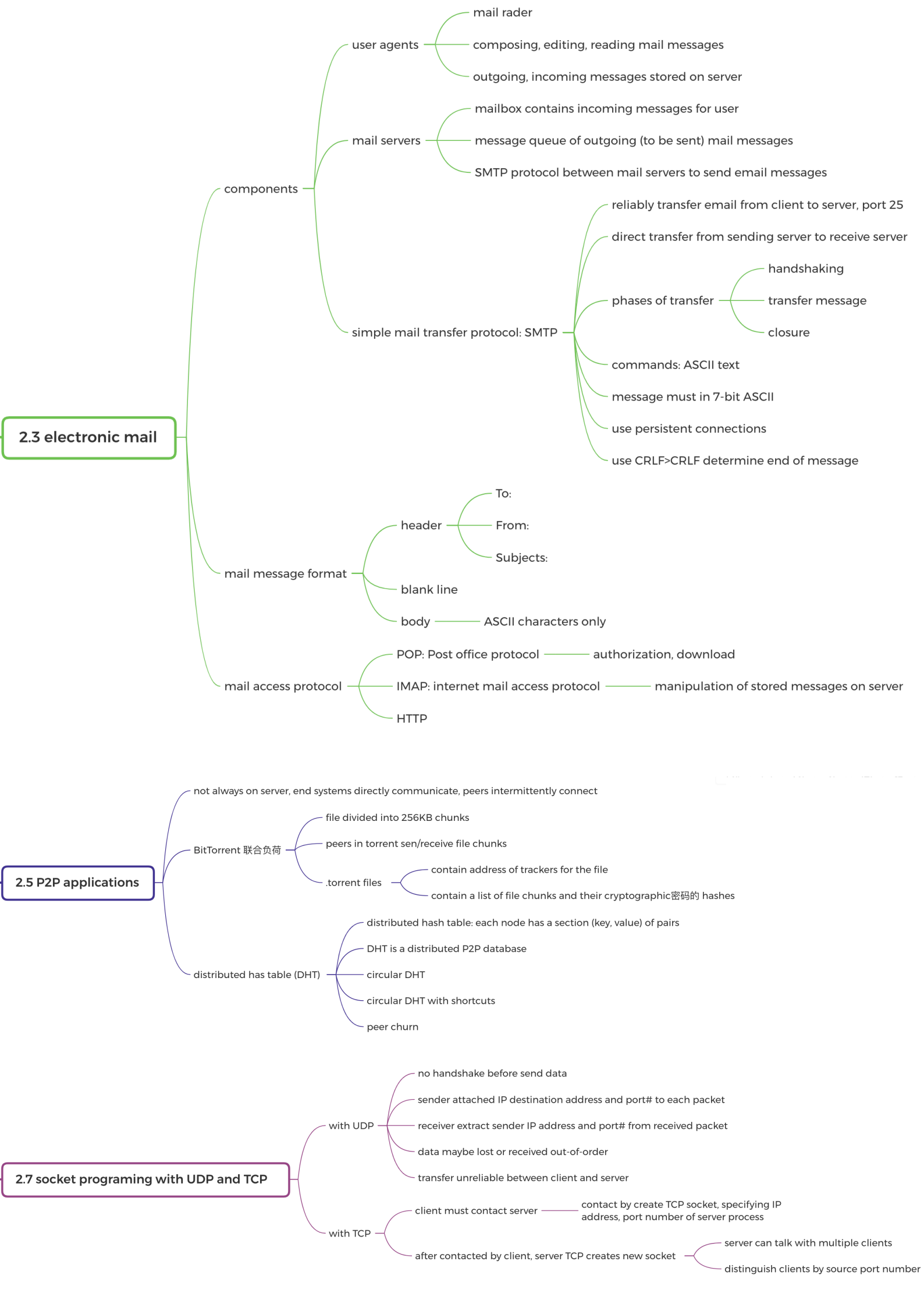
2.1 principles of network applications

- Interprocess Communication (IPC)
  - processes talk to each other through IPC
  - On a single machine: share memory
- Sockets
  - process sends/receives messages to/from its socket
  - analogous相似 to door
- addressing processes
  - process must have identifier to receive messages
  - host device has unique 32- bit IP address
  - identifier includes both IP address and port numbers
- Client-server architecture
  - server
    - exports request/response interface
    - long-lived process that waits for requests
    - Upon receiving request, carries it out
    - always-on host
    - Permanent IP address
    - Static port conventions (http: 80, email: 25, ssh:22)
  - clients
    - Short-lived process that makes requests
    - “User-side” of application
    - Initiates the communication
    - not communicate directly with each other
    - May have dynamic IP addresses
- P2P -- peer to peer
  - no always-on server
  - peers directly communicate
  - Symmetric相称的 responsibility
  - peers request service from other peers, provide service in return to other peers
  - Speed: parallelism, less contention
  - Reliability: redundancy, fault tolerance
  - Geographic distribution
  - Privacy protection
  - problems of decentralized control
    - State uncertainty: no shared memory or clock
    - Action uncertainty: mutually conflicting decisions
- App-layer protocol
  - types of messages exchanged
  - message syntax
  - message semantics
  - rules
  - open protocols
    - defined in RFCs
    - allows for interoperability 互用性
  - proprietary protocols 私有的
- transport service app need
  - data intergrity
    - some apps require 100% reliable data transfer
    - some apps can tolerate some loss
  - timing
    - some apps require low delay
  - throughput
    - some apps require minimum amount of throughput
  - security
    - encryption 编密码
    - data integrity
- Internet transport protocols services
  - TCP
    - reliable transport
    - flow control
    - congestion control
    - no provide timing, minimum throughput guarantee, security
    - connection-oriented: setup required
  - UDP
    - unreliable data transfer
    - no reliability, flow control, congestion control, timing, throughput guarantee, security, or-connection setup
    - faster -- less overhand

- web page consists of objects and base HTML-file
  - object can be HTML file, JPEG image, Java applet, audio file,...
  - base HTML-file which includes several referenced objects
  - each object is addressable by a URL
  - www.someschool.edu/someDept/pic.gif
    - www.someschool.edu -- host name
    - someDept/pic.gif -- path name
- Uniform Resource Locator (URL) — protocol://host-name[:port]/directory-path/resource
- client: browser that requests, receives, "displays" Web objects
- server: Web server sends objects in response to requests
- use TCP:
  - client initiates TCP connection to server, port 80
  - server accepts TCP connection from client
  - HTTP messages exchanged between browser and Web server
  - TCP connection closed
- HTTP request message
  - ASCII
  - request line (GET, POST, HEAD commands)
  - carriage return: \r\n
- HTTP response message
  - status line (protocol status code status phrase)
  - data, e.g., requested HTML file
  - status code
    - 200 OK
    - 301 Moved Permanently
    - 400 Bad Request
    - 404 Not Found
    - 505 HTTP Version Not Supported
    - 451 Unavailable for Legal Reasons
    - 429 Too Many Requests
  - request types in HTTP/1.0:
    - Get: Request page
    - POST: Uploads user response to a form
    - HEAD: asks server to leave requested object out of response
  - request types in HTTP/1.1:
    - GET, POST, HEAD
    - PUT: uploads file in entity body to path specified in URL field
    - DELETE: deletes file specified in the URL field
    - TRACE, OPTIONS, CONNECT, PATCH: For persistent connections
- HTTP is all text — But not the most efficient — sending '12345678' as a string is 8 bytes, as integer is 4 bytes
- Cookies
  - Components
    - cookie header line: HTTP response message
    - cookie header line: next HTTP request message
    - cookie file kept on user's host, managed by user's browser
    - back-end后端 database at Web site
  - Darkside
    - learn a lot about you
    - lots of personal details
    - could turn off cookies, but can't do anything on internet
- Page load time (PLT) is metric衡量标准 — From click until user see page
- depend on
  - page content/structure
  - protocols involved
  - network bandwidth and RTT

2.2 Web and HTTP

- HTTP: Hypertext Transport Protocol
  - Performance of HTTP
    - goals
      - for user
        - fast download — By improve HTTP
        - high availability — By caching and replication
      - for provider
        - happy user
        - cost effevtive — By exploit economies of scale 利用规模经济
      - for network — avoid overload
    - improve PLT (page load time)
      - reduce content size for transfer — smaller images, compression
      - change HTTP to make better use of available bandwidth — persistent connections and pipelining
      - change HTTP to avoid repeated transfers of same content
      - move content closer to client
    - non-persistent HTTP
      - one object sent per TCP connection
      - downloading multiple objects required multiple connections
      - response time
        - RTT: time for a small packet travel from client to server and back
        - = 2RTT + file transmission time
      - HTTP/1.0: Multiple TCP slow-start phases
    - persistent HTTP
      - persistent without pipelining 流水线
        - one RTT per object
        - 发完一个收到 再发
      - persistent with pipelining
        - as little as one RTT for total objects
        - 发完一个 立刻发下一个
    - improving HTTP performance: caching
      - proxy server 代理服务器
        - goal: satisfy client request without involving origin server
        - user access web via cache
        - browser send all HTTP request to cache
        - cache act as both server and client
        - advantage use caching
          - reduce response time for client request
          - reduce traffic on institution's access link
          - let slow content provider effectively deliver content
          - install web cache cheaper than increase access link speed
      - conditional GET
        - if cache has up to date, not send object
        - cache specify date of cashes copy in HTTP request: "If-modified-since: <date>"
        - server response contains no object if cache up to date: "HTTP/1.0 304 Not Modified"
      - replication 复制
        - replicate popular web site
          - spread load on servers
          - place content closer to clients
          - helps when content not cacheable
        - problem
          - want direct client
            - balance load across server replicas
            - pair clients with nearby servers
          - expensive
        - solution — DNS returns different addresses based on client's geo location, server load
      - CDN
        - combination of caching and replication
        - pull: direct result of clients' requests
        - push: expectation of high access rate
  - HTTP is insecure
    - use TLS (transport layer security)
    - provides authentication鉴定 and bidirectional encryption双向加密
- HTTP/2
  - better content structure
  - Improvment
    - servers can push content and reduce overhead of additional request cycle
    - fully multiplexed
    - prioritization of the order for sending
    - data compression of HTTP headers


2.3 electronic mail
components
user agents
mail rader
composing, editing, reading mail messages
outgoing, incoming messages stored on server
mail servers
mailbox contains incoming messages for user
message queue of outgoing (to be sent) mail messages
SMTP protocol between mail servers to send email messages
simple mail transfer protocol: SMTP
reliably transfer email from client to server, port 25
direct transfer from sending server to receive server
phases of transfer
handshaking
transfer message
closure
commands: ASCII text
message must in 7-bit ASCII
use persistent connections
use CRLF>CRLF determine end of message
mail message format
header
To:
From:
Subjects:
blank line
body
ASCII characters only
mail access protocol
POP: Post office protocol
authorization, download
IMAP: internet mail access protocol
manipulation of stored messages on server
HTTP
2.5 P2P applications
not always on server, end systems directly communicate, peers intermittently connect
BitTorrent 联合负荷
file divided into 256KB chunks
peers in torrent sen/receive file chunks
.torrent files
contain address of trackers for the file
contain a list of file chunks and their cryptographic密码的 hashes
distributed has table (DHT)
distributed hash table: each node has a section (key, value) of pairs
DHT is a distributed P2P database
circular DHT
circular DHT with shortcuts
peer churn
2.7 socket programing with UDP and TCP
with UDP
no handshake before send data
sender attached IP destination address and port# to each packet
receiver extract sender IP address and port# from received packet
data maybe lost or received out-of-order
transfer unreliable between client and server
with TCP
client must contact server
contact by create TCP socket, specifying IP address, port number of server process
after contacted by client, server TCP creates new socket
server can talk with multiple clients
distinguish clients by source port number

2.4 DNS: domain name system

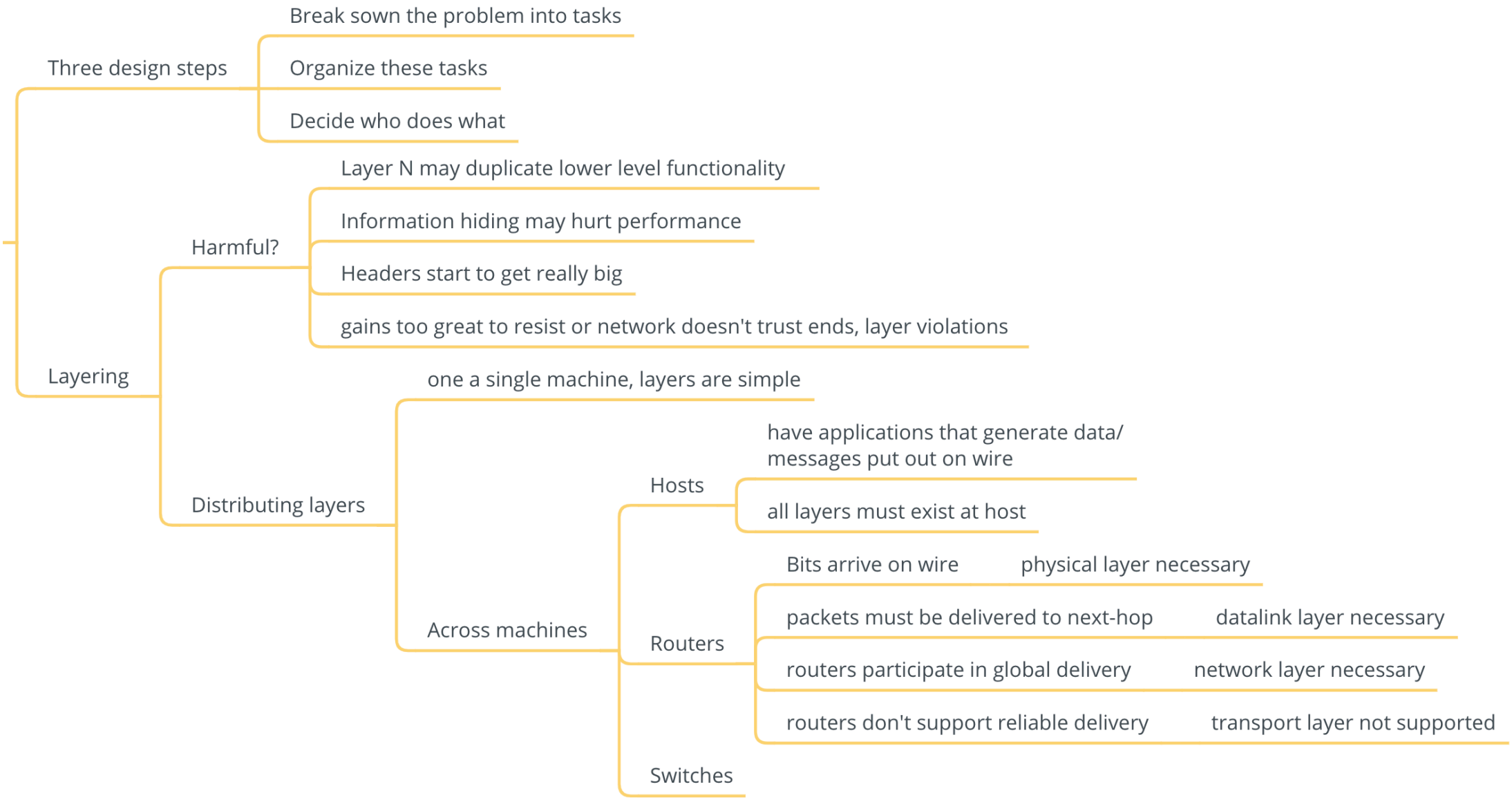
- increase reliable and matainess
- DNS services
  - hostname to IP address translation
  - host aliasing别名
  - mail server aliasing
  - load distribution
- goals
  - no naming conflict
  - scalable
  - distributed, autonomous administration自治管理
  - highly available
  - lookups should be fast
- hierarchy 等级制度
  - hierarchical namespace 分级命名空间
    - top level domains at top
    - domains are sub-trees
    - name is leaf-to-root path
    - depth of tree is arbitrary (limit 128)
    - name collision avoided (一般不会重名)
  - hierarchically administered 分级管理
  - hierarchy of servers 服务器层次结构
    - top: root servers
      - location hardwired into other servers
      - 13 root servers
    - next level: top-level domain (TLD) servers
      - .com, .edu, .uk, .fr, .ca, .jp
    - bottom level: authoritative DNS servers
      - store name-to-address mapping
      - organization's own DNS
      - maintain by corresponding administrative suthority
    - local DNS name server
    - Each server stores a small subset of DNS database
    - each server know other servers responsible for other portions 每个服务器知道其他层级的服务器
- example
  - robot.cs.washington.edu
  - local. authoritative. top level domain. root
- resource records
  - format: (name, value, type, ttl)
  - type
    - A: name is hostname, value is IP address
    - NS: name is domain, value is domain, value is hostname
    - CNAME: name is alias name, value is canonical公认的 name
    - MX: value is name of mail-server associated关联 with name
- protocol, messages
  - query and reply messages, both with same message format
  - message header: identification 2bytes and flags 2bytes
- DNS provide indirection间接
  - address can change underneath
  - name can map multiple IP address
  - multiple names for same address

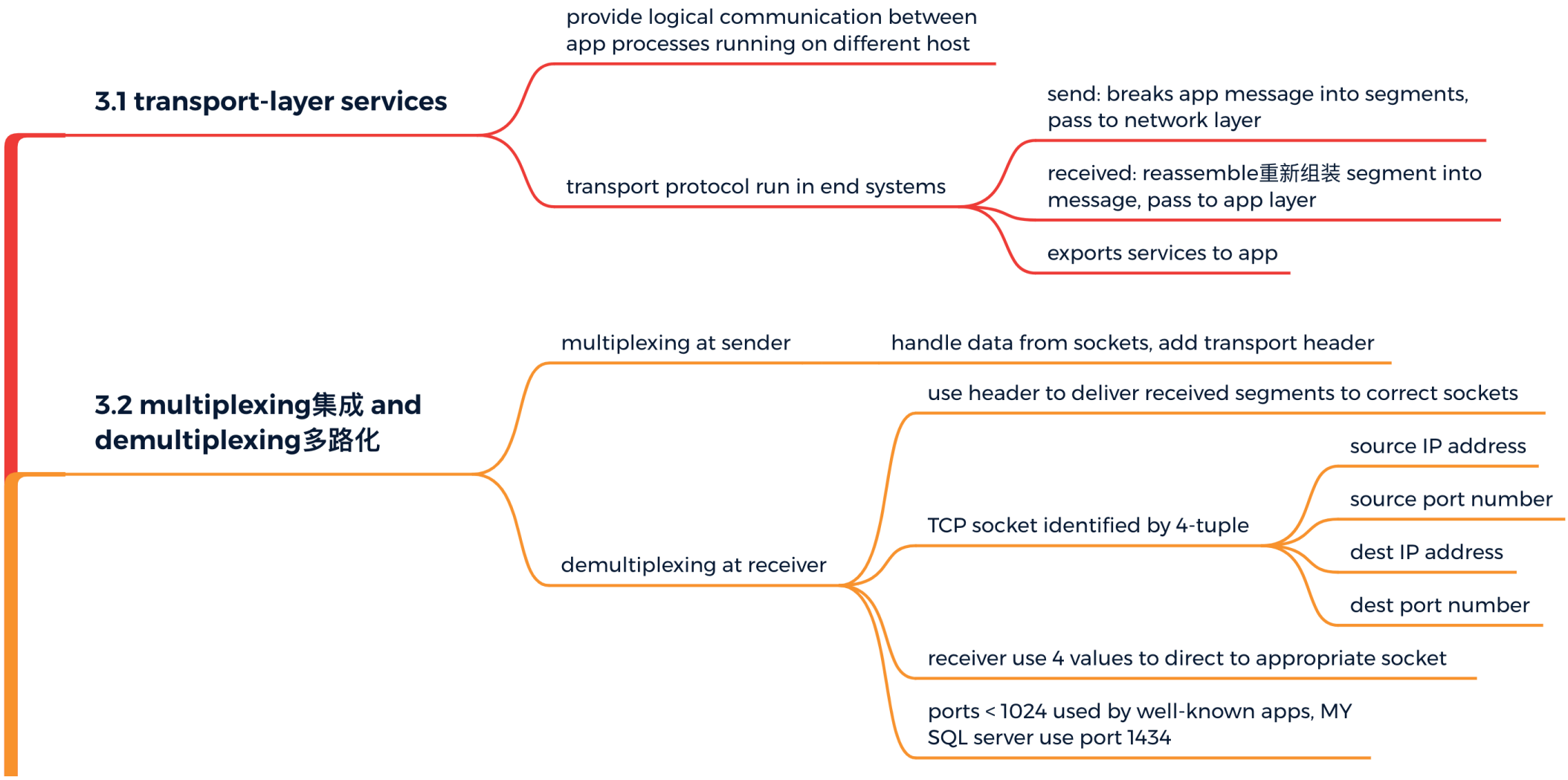
1.4 delay, loss, throughput in networks

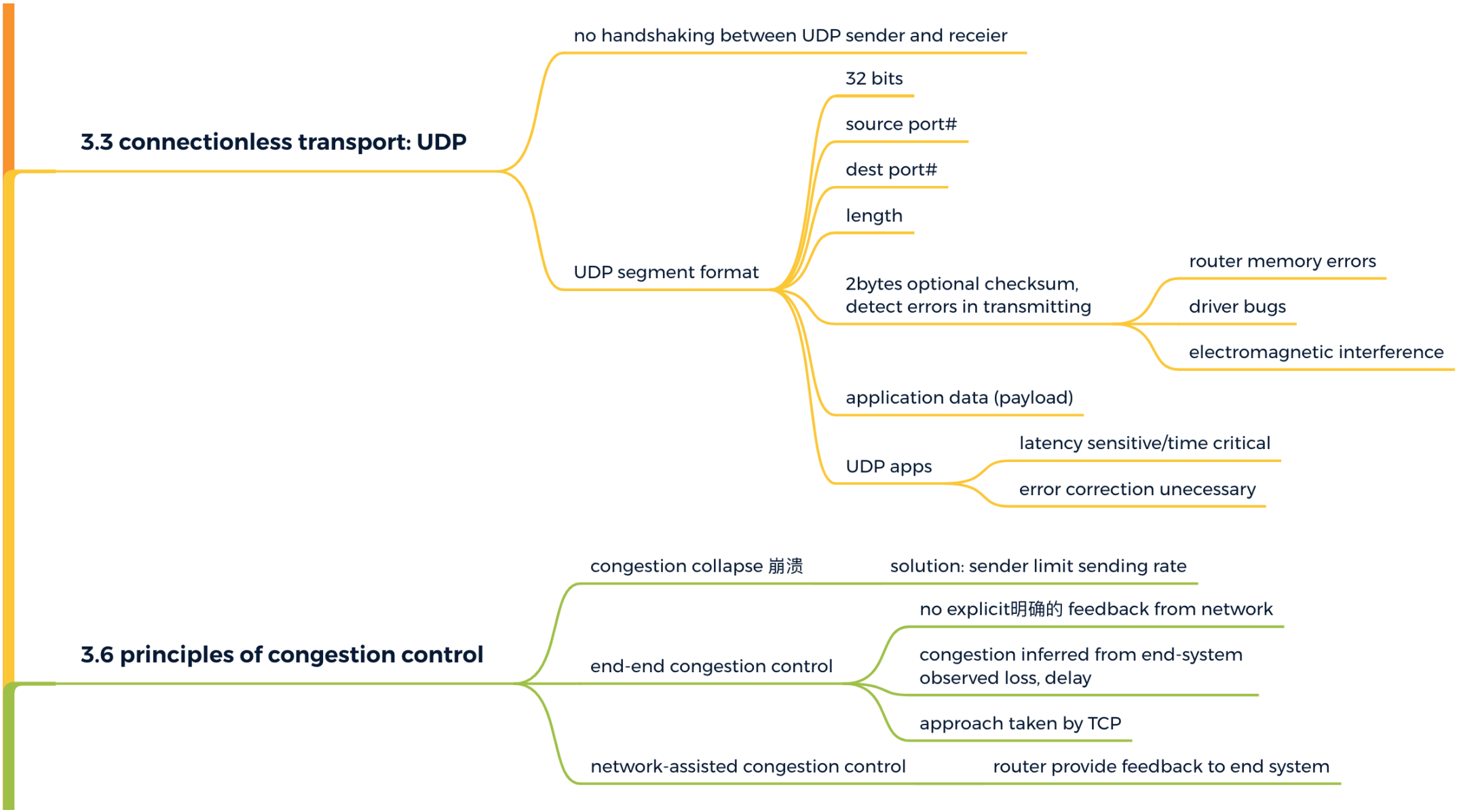
When packet arrival rate to link exceeds output link capacity, delay/ loss occur

- d nodal
  - d proc: nodal processing 节点处理
    - check bit errors
    - determine output link
    - typically < msec
  - d queue: queueing delay 排队延迟
    - time waiting at output link for transmission
    - depends on congestion level of router
    - aL bits arrive to queue /sec
    - R bits leave the router / sec
    - aL/R: traffic intensity
      - aL/R > 1 -- drop, average delay infinite
      - aL/R <= 1 -- fine
      - aL/R = 1, average delay = 0, transmit immediately
  - d trans: transmission delay 传输延迟
    - d trans = L/R
    - L: packet length (bits)
    - R: link bandwidth (bps)
    - depend on bandwidth
  - d prop: propagation delay 传播延迟
    - d prop = d/s
    - d: length of physical link
    - s: propagation speed in medium (2 * 10^8 m/sec)
    - depend on speed
- real internet delays and routes
  - traceroute 跟踪路由
    - provide delay measurement from source to router along end-to-end internet path toward destination
    - sends 3 packets will reach router i on path towards destination
    - router i will return packets to sender
    - sender times interval between transmission and reply
  - end-end delay = sum of all d nodal along the path
- Packet loss
  - queue preceding link in buffer has finite capacity
  - packet arriving to full queue will be drop -- lost
  - lost packet may be retransmitted
- Throughput
  - rate at which bits transferred between sender/receiver
  - unit: bits/time
  - instantaneous: rate at given point in time
  - average: rate over longer period of time
  - depend on the smaller one
  - Rc or Rs often bottleneck 阻碍， 瓶颈
- Links
  - Bandwidth: 'width' of the link — Number of bits sent per unit of time (bps)
  - Propagation delay: 'length' of the link — Propagation time to travel along the link (seconds)
  - Bandwidth-Delay Product: 'volume' of the link — Amount of data in flight (bps * s = bits)

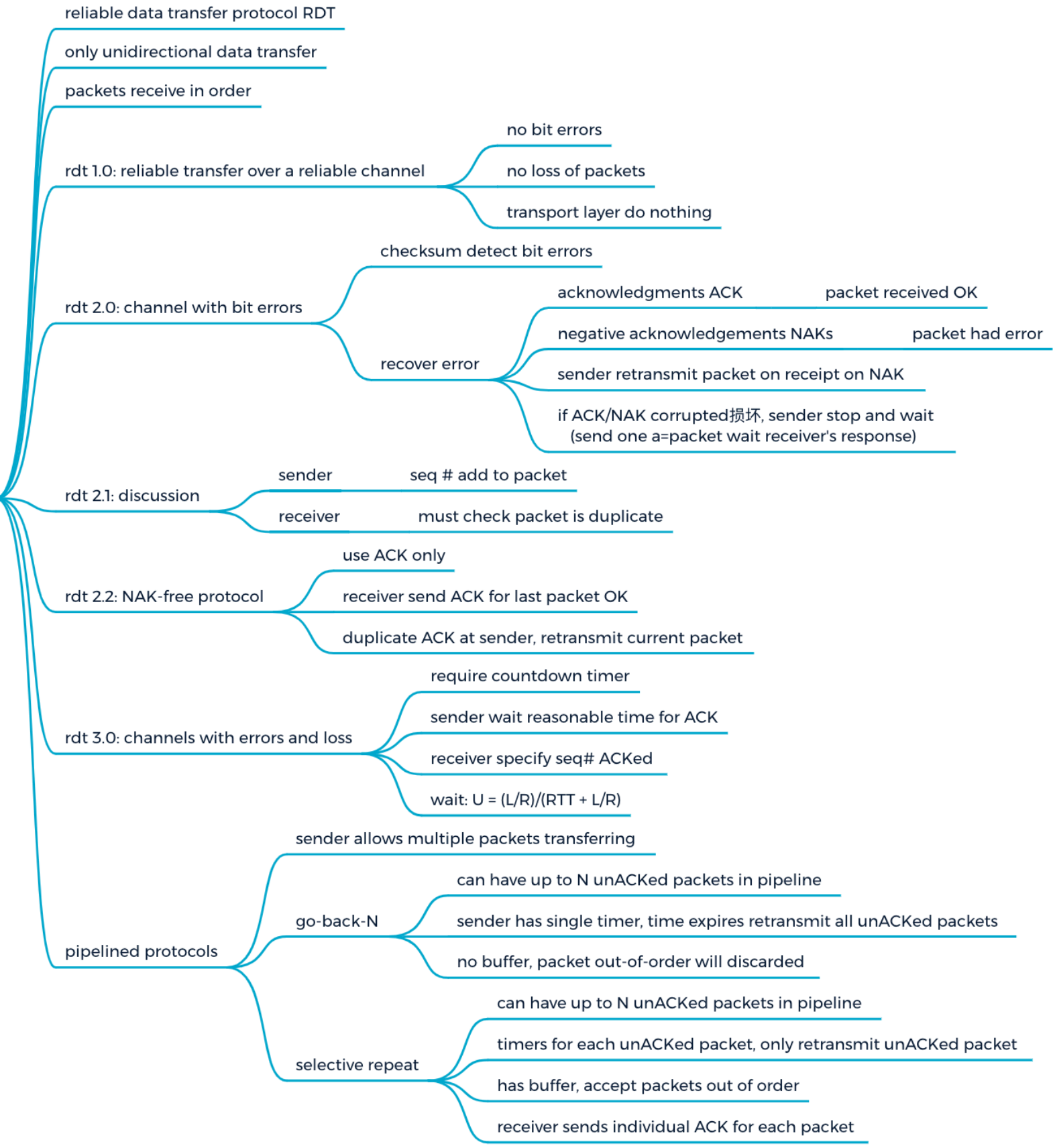
1.5 protocol layers, service models

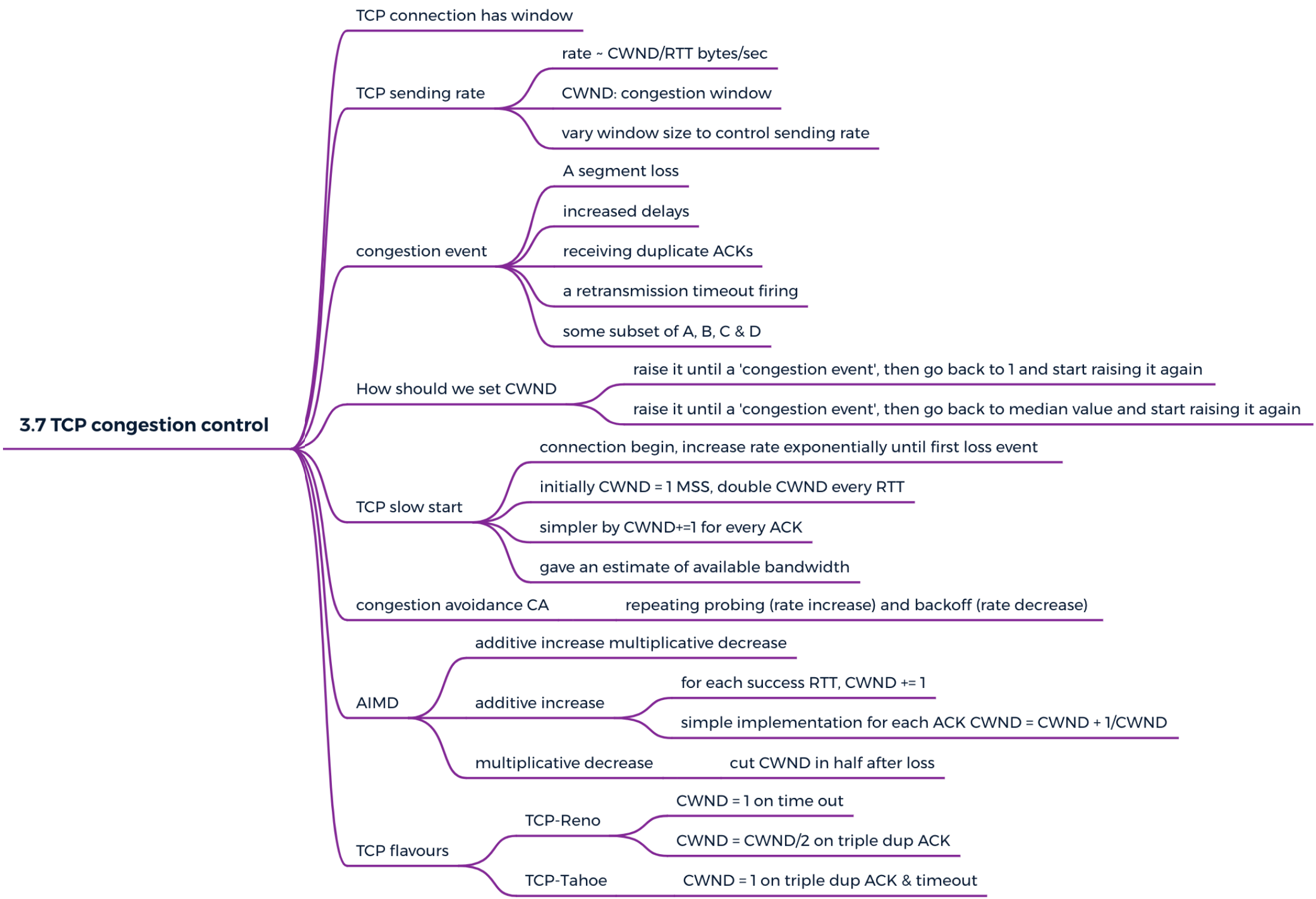


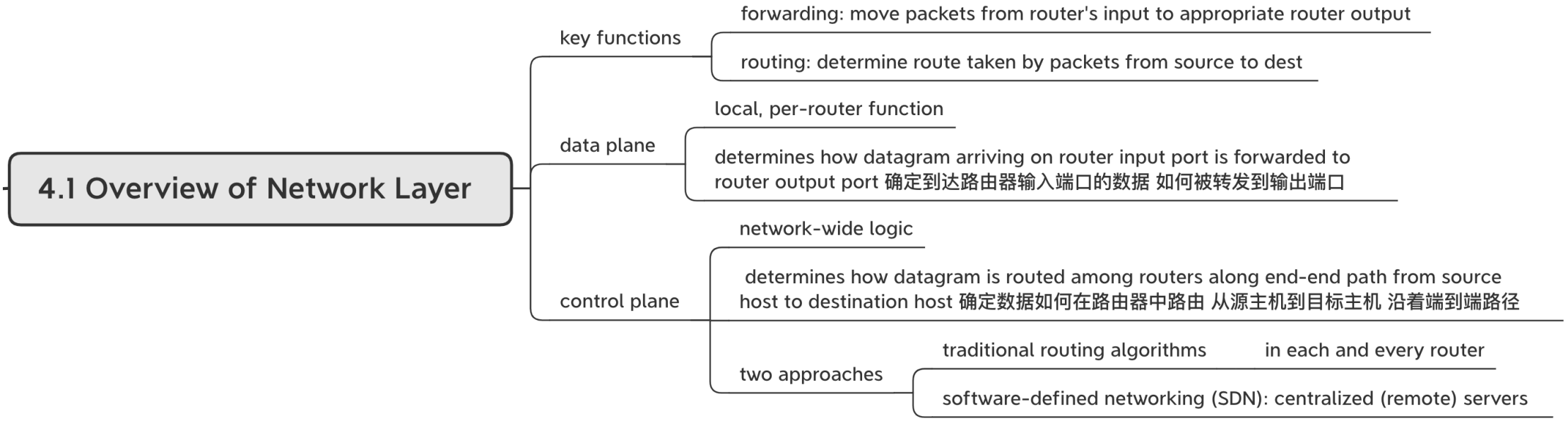

3.1 transport-layer services
provide logical communication between app processes running on different host
transport protocol run in end systems
send: breaks app message into segments, pass to network layer
received: reassemble重新组装 segment into message, pass to app layer
exports services to app
3.2 multiplexing集成 and demultiplexing多路化
multiplexing at sender
handle data from sockets, add transport header
demultiplexing at receiver
use header to deliver received segments to correct sockets
TCP socket identified by 4-tuple
source IP address
source port number
dest IP address
dest port number
receiver use 4 values to direct to appropriate socket
ports < 1024 used by well-known apps, MY SQL server use port 1434


3.3 connectionless transport: UDP
no handshaking between UDP sender and receier
UDP segment format
32 bits
source port#
dest port#
length
2bytes optional checksum,
detect errors in transmitting
router memory errors
driver bugs
electromagnetic interference
application data (payload)
UDP apps
latency sensitive/time critical
error correction unecessary
3.6 principles of congestion control
congestion collapse 崩溃
solution: sender limit sending rate
end-end congestion control
no explicit明确的 feedback from network
congestion inferred from end-system
observed loss, delay
approach taken by TCP
network-assisted congestion control
router provide feedback to end system

3.4 principles of reliable data transfer




3.7 TCP congestion control
TCP connection has window
TCP sending rate
rate ~ CWND/RTT bytes/sec
CWND: congestion window
vary window size to control sending rate
congestion event
A segment loss
increased delays
receiving duplicate ACKs
a retransmission timeout firing
some subset of A, B, C & D
How should we set CWND
raise it until a 'congestion event', then go back to 1 and start raising it again
raise it until a 'congestion event', then go back to median value and start raising it again
TCP slow start
connection begin, increase rate exponentially until first loss event
initially CWND = 1 MSS, double CWND every RTT
simpler by CWND+=1 for every ACK
gave an estimate of available bandwidth
congestion avoidance CA
repeating probing (rate increase) and backoff (rate decrease)
AIMD
additive increase multiplicative decrease
additive increase
for each success RTT, CWND += 1
simple implementation for each ACK CWND = CWND + 1/CWND
multiplicative decrease
cut CWND in half after loss
TCP flavours
TCP-Reno
CWND = 1 on time out
CWND = CWND/2 on triple dup ACK
TCP-Tahoe
CWND = 1 on triple dup ACK & timeout


4.1 Overview of Network Layer
key functions
forwarding: move packets from router's input to appropriate router output
routing: determine route taken by packets from source to dest
data plane
local, per-router function
determines how datagram arriving on router input port is forwarded to router output port 确定到达路由器输入端口的数据 如何被转发到输出端口
control plane
network-wide logic
determines how datagram is routed among routers along end-end path from source host to destination host 确定数据如何在路由器中路由 从源主机到目标主机 沿着端到端路径
two approaches
traditional routing algorithms
in each and every router
software-defined networking (SDN): centralized (remote) servers

4.3 IP: Internet Protocol

- IP packet structure
  - 32-bit word
  - reading packet correctly
    - 4-bit version number, typically '4' (means IPv4)
    - 4-bit header length, typically '5' (32-bit words)
    - 16-bit total length,how many bytes int packet, max size 65535 (2^16 - 1)
  - how to handle packet — 8-bit Protocol
  - preventing loops — 8-bit (TTL) Time-to-live, initially 255, -1 for each hop, stop at 0
  - Header corruption — 16-bit checksum, not correct, router discard packet
- fragmentation
  - large IP datagram fragmented 分裂 to several datagrams
  - "reassembled" only at final destination
  - use IP header identify and order related fragments
- IPv4 addressing
  - IP address: 32-bit
  - subnetting子网划分 — 3 levels of hierarchy
    - netid
    - subnetid
    - hostid
  - CIDR: class interDomain routing — a.b.c.d/x
  - longest prefix matching — use longest address prefix matches
  - DHCP: Dynamic Host Configuration Protocol
    - address allocation for hosts
    - offer message: IP address, length of lease, subnet mask, DNS servers, default gateway
- NAT: network address translation
  - router
    - outgoing datagrams: replace (source IP address, port #) to (NAT IP address, new port #)
    - remember in NAT translation table: every (source IP address, port #) to (NAT IP address, new port #) translation pair
    - income datagrams: replace (NAT IP address, new port #) in dest fields (source IP address, port #) stored in NAT table
  - advantage & disadvantage
- IPv6
  - fixed-length 40 byte header
  - no fragmentation allowed
  - dataformat
    - priority
    - flow label
    - next header: identify upper layer protocol for data
    - NO checksum
  - tunneling — transition IPv4 to IPv6

5 Network Layer: Control Plane

5.2 Routing protocols

- Internet routing
  - intra-domain 域内: routes within domain
  - inter-domain 域间; routes between domain
  - routing algorithm: find least coat path
  - link cost
- LS: link state
  - each node maintain its local 'LS'
  - each node floods its local link state
  - each node learn the entire network topology
  - routers transmit LSA on links (link state advertisement)
- distance vector
  - DV: shortest path to neighbor
  - Each router has DV and exchange it with neighbors and select best one
  - each node notify others only when DV changes

5.6 ICMP: The Internet Control Message Protocol

- hosts & routers to communicate network level infromation
  - error reporting
  - echo request/reply
- ICMP messages carried in IP datagrams
- message format
  - type
  - code plus IP header
  - first 8 bytes of IP datagram payload causing error

6.1 Introduction, services

- focus within a subset
- hosts and routers: nodes
- communication channels: links
  - wired links
  - wireless links
  - LANs
- layer-2 packet: frame, encapsulates datagram
- datalink layer transfer datagram from one node to physically adjacent node over a link
- link layer services
  - framing, link access
  - reliable delivery between adjacent nodes
  - flow control
  - error detection
  - error correction
  - half-duplex and full-duplex
- each host implement link layer
- framing
  - physical layer talks in terms of bits
  - start of frame
    - preamble: seven bytes with pattern 10101010
    - Start of Frame Delimiter (SFD): 10101011
  - end of frame — Absence of transition in Manchester encoded signal
  - in Ethernet — inter frame gap is 12 bytes of idle state

6.2 Error detection, correction

- EDC
  - Error detection and Correction bits
  - not 100% reliable
    - protocol may miss some error
    - larger EDC field yields better detection and correstion
- error coding
- add check bits to the message bits
- checksum: 16-bit words
- CRC cyclic redundancy check
  - view data bits D as a binary number
  - choose r+1 bit pattern G
  - choose r CRC bits
    - <D, R> exactly divisible by G (modulo 2)
    - receiver knows G,divide and check remainder
    - if remainder non-zero, ERROR detected
  - R = remainder[D * 2^r / G]

6.3 multiple access protocols

- types of links
  - point-to-point
  - broadcast (shared wire or medium)
    - single shared broadcast channel
    - if two or more simultaneous transmission will collision
- MAC protocols
  - determines how nodes share channel
  - determine when node can transmit
- ideal MAC
  - channel of rate R
  - when one node transmit, rate=R
  - when M nodes transmit, average rate = R/M
- board classes
  - channel partitioning
    - divide channel into smaller
    - allocate piece to node for exclusive use
    - TDMA: time division multiple access
      - each station gets fixed length slot
      - length = packet trans time
    - FDMA: frequency division multiple access
      - divided into frequency bands
      - each station gets fixed frequency band
  - random access
    - channel not divided, allow collision
    - 'Slotted ALOHA'
      - max efficiency = 37%
    - 'Pure (unslotted) ALOHA'
      - max efficiency = 18% (even worse)
    - CSMA: carrier sense multiple access
      - listen before transmit
      - channel idle, transmit
      - channel busy, defer transmit
      - CSMA/CD: collision detection
        - easy in wired LANs
        - difficult in wireless LANs
      - min packet size
        - Ethernet min size = 64 bytes
        - 64 bytes = 2 * 6-bytes address, 2-bytes type, 4-bytes CRC, 46 bytes data
  - taking turns
    - node take turns, more to send can take longer turns
    - polling轮流
    - token passing

6.4 Switched LANs

- IP address 32-bit
  - dynamically
  - hierarchical name space of 32 bits
  - not portable, can not change
  - e.g. 12.178.66.9
- MAC address 48-bit
  - hard-coded in read-only memory
  - flat name space 48 bits
  - portable 便携, depend on host
  - e.g. 00-0E-9B-6E-49-76
- addressing LANs
  - each adapter适配器 on LAN has unique LAN address
  - ARP: address resolution protocol
    - APR table: each IP node on LAN has
      - <IP address; MAC address; TTL>
      - TTL: time to live, typically 20 mins
    - same LAN
    - routing to another LAN
- Ethernet
  - cheaper than LANs and ATM
  - speed: 10 Mbps - 10 Gbps
  - bus: popular through mid 90s
    - all nodes in same collision domain
    - CSMA/CD
  - star: prevails today
    - active switch in centre
    - no sharing, no CSMA/CD
  - Ethernet frame struct
    - preamble 序文： 7 bytes with pattern 10101010
    - SFD: 1 byte: 10101011
    - Dest MAC 6 bytes
    - Source MAC 6 bytes
    - Type/Length 2 Bytes
    - Payload 46-1500 Bytes
    - FCS/CRC 4 Byte
    - Inter Frame Gap: 12 bytes of idle state
      - 0.96 ms for 100 Mbit/s Ethernet
      - 0.096 ms for Gbps Ethernet
  - no handshaking between sending and receiving NICs
  - unreliable: no sending ACKs or NACKs when receiver receiving NIC
- switches
  - link-layer device: takes an active role
    - store, forward Ethernet frames
    - examine incoming frame's MAC address
    - selectively forward frame to outgoing links
  - transparent
    - hosts unaware of switched 主机不知道交换机的存在
  - plug-and-play, self-learning
    - no need be configured
    - switch learns which hosts can be reached
    - record sender/location pair in switch table
  - multiple simultaneous transmissions 多个同时传输
    - host direct connect to switch
    - switches buffer packets
    - no collisions, full duplex
    - transmit 'A to B' and 'C to D' can simultaneously, no collisions
  - interconnecting switches
    - connect switches together

7 Wireless Networks

7.1 Introduction

- Wireless 101
- elements of wireless network
  - wireless host
    - smartphone, laptop
    - may be stationary
  - base station
    - connect to wired network
    - relay 转播： send packets between wireless network and wireless hosts
  - wireless link
    - used to connect mobile移动台 to base station
    - various data rates
    - various transmission distance
  - infrastructure mode 基础建设模式
    - base station connects mobiles into wired network
    - handoff
  - ad hoc mode 临时模式
    - no base station
    - nodes can only transmit to others within link coverage
    - nodes organize themselves into a network
    - route among nodes

7.2 Wireless links, characteristics

- decreased signal strength
- interference from other sources
- multipath propagation
  - effect
    - Signals bounce off surface and interfere
    - self-interference 自干扰
- path loss/ path attenuation 衰弱
  - FSPL: free space path loss
- SNR: signal-to-noise ratio
  - larger SNR easier to extract from noise

7.3 IEEE 802.11 wireless LANs (Wi-Fi)

- wireless host communicate with base station
  - base station = access point
- BSS: (base service set) in infrastructure mode contains
  - wireless hosts
  - access point: base station
  - ad hoc mode: hosts only
- 802.11b
  - 2.4GHz - 2.485GHz divided into 11 channels
  - AP admin choose frequency for AP
  - host must associate with an AP
  - nearby channels overlap
  - passive/active scanning
  - multiple access
    - avoid collisions
      - using small reservation packets
    - CSMA sense before transmit
    - no collision detection
    - different receivers hear different signals
    - only care if receiver can hear the sender clearly