计算机网络

1. Introduction

- Networks: exchange something from one site to the others
- Why computer networks?【s1 p20】
    - Communication, resource sharing, collaboration.

Connectivity

- What do we need to connect two computers?【s1 p27】
    - computers
    - links
- Building blocks【s1 p28】
    - Nodes
        - Hosts: work for the users
        - Switches: work for the network
    - Direct links
        - point to point
        - **multiple access** (常用)
- Device at the center
    - Hub（集线器）, a box
    - Switch（交换机）, a computer
    - Router（路由器）
- Switched networks【s1 p30】
    - suitable for scaling up
    - circuit switched (tel.), packet switched (computer)
    - can be defined recursively
- Network scales【s1 p32】
    - SAN, LAN, WAN, MAN, Internet

Networking resource sharing

- Fundamental resource sharing concept: **multiplexing**（复用）
- Solutions【s1 p35】
    - Synchronous Time-Division Multiplexing (STDM)
        - 分时间段使用，这段时间归谁，那段时间归谁，不能乱用。
    - Frequency-Division Multiplexing (FDM)

        - 分出多个lane，每个lane分配一台电脑，同样是专属。这样效率还是不高，可能一条道没用，另一条道却堵。
    - Packet switching (Statistical Multiplexing, On-demand Time-division)
        - 将各个电脑的数据包分成小块，小块混杂着传输。
        - 各个小块过程中可能线路不同，但最终到达的目的地相同。

- Circuit vs Packet switching【s1 p37】

Architecture

- Support for common services【s1 p42】
    - request/reply channel
    - message stream channel: videos, sending message continuously
- Layering【s1 p43】
    - Application programs, Process-to-process channels, Host-to-host connectivity, Hardware
- **Layers of network applications**【s1 p44】
    - Users
    - Applications
    - Network services
    - Network hardware
- Simplified file transfer architecture

---

2. Network architecture

Layering and protocols

- **Protocols**【s2 p7】
    - Standards at a layer
    - a library function, or a running process
    - protocol stack: set of consecutive layers
- **Interfaces**【s2 p8】
    - service interface
    - peer-to-peer interface
- Protocol graph
    - most peer-to-peer communication is indirect, direct only at hardware level.
    - RRP: request/reply protocols
    - MSP: message stream
    - HHP: host-to-host
- Encapsulation （封装）

    - Data $\rightarrow$ (RRP, Data) $\rightarrow$ (HHP, RRP, Data) $\rightarrow$ (RRP, Data) $\rightarrow$ Data

Architecture

ISO/OSI Model

- 7 layers
    - 7: Application
    - Presentation
    - Session
    - Transport
    - Network
    - Data link
    - 1: Physical
- 太麻烦了，仅供参考，实际并未使用。
- Nodes 至少要有3个layer: physical, data link, network
- Brief introduction of the 7 layers【s2 p14~p16】

TCP/IP architecture

- 漏斗形 hourglass shape【s2 p17】
- Does not imply strict layering
- 4 layers【s2 p19】
    - 4: Application
    - Transport
    - Network
    - 1: Physical

Application programming interface (API)

- How can we build network applications?【s2 p24】
    - applications: inter-process communication
    - common services
    - A **network architecture** is a layered arrangement of common services, which support inter-process communication.
- The most widely used network API: **socket interface**
    - A socket is defined as an endpoint for communication
    - Host num: Port num, 161.25.19.8: 1625
- The client/server paradigm【s2 p28】

    - steps

- Further notes about socket API【s2 p37】

---

3. Direct connection

Network performance

- Performance metrics【s3 p4】
    - Bandwidth
        - data transmitted per time unit
        - Mbps = 10^6 bits per second
    - Latency
        - time to send message from point A to point B
        - RTT: round-trip time. RTT == Propagation delay $\times$ 2
        - **Latency = Propagation + Transmit + Queue**
    - Throughput = $\frac{\text{TransferSize}}{\text{TransferTime}}$
- Latency 和 Bandwidth 可以看作 fix cost 和 unit cost。
    - For high speed networks, latency dominates.
- Delay $\times$ Bandwidth product
    - amount of data "in the pipe"
    - **the amount of data the sender can send before the receiver receives**.
- 当算大小时，用2的次方换算，如 $1KB = 1024 Bytes$；当算速度时，用10的次方换算，如 $1Kbps = 1000 bits/s$

Direct link

- One can send message to the other without the third computer.
- Direct link networks【s3 p14】
    - connecting
    - encoding
    - framing
    - error detection
    - reliable delivery
    - media access control

Connecting hardware

- Node architecture, network adapter, links【s3 p18~p20】
- Typical cables【s3 p22】
    - twisted pairs

    - coaxial cables
    - fiber optic cables

- Wireless links【s3 p26】

    - Mobile phone system
    - Wireless LAN/WAN
    - Short range device connection (infrared, bluetooth)

Encoding, framing, and error detection (NOT required)

Reliable transmission

- Transmitted frames may be delayed or lost.
- Solution: Automatic Repeat reQuest (ARQ)

    - Correctness: each packet is released to the network layer once and only once, without error
    - Efficiency

- Stop and wait【s3 p51】

    - Acknowledgement and timeouts
    - It does not keep the pipe full

- Improvements【s3 p54】

    - Concurrent logical channels

- **Sliding window**【s3 p55】

    - allow multiple outstanding un-ACKed frames
    - upper bound: window.【s3 p57】
        - Sender window size (SWS): Delay $\times$ Bandwidth product.
        - Receiver window size (RWS): = 1 or = SWS

---

4. Direct link networks

- Ethernet, Token rings, Wireless【s4 p4】
- IEEE 802.x specifications【s4 p6】
- Media Access Control (MAC)

    - Ethernet: CSMA/CD
    - Token ring
    - Wireless: CSMA/CA

Ethernet

- Physical connection【s4 p9】

- Physical topologies【s4 p10】
    - Linear bus topology
        - not so reliable, many single failure points
        - difficult to identify the problem if the entire network shuts down
    - Star topology
        - require more cable length, but more reliable
- Ethernet transmission【s4 p14】
    - Broadcast
        - all stations receive all transmissions
    - CSMA/CD medium access control (MAC) scheme【s4 p15】
        - all hosts compete for the same link. possibility of collisions
    - Transmission algorithms
        - 1-persistent protocol
        - p-persistent protocol
- Ethernet addresses【s4 p18】
    - addresses are assigned to network adapters
    - 48-bits
    - address recognition【s4 p19】
    - Possible destinations【s4 p20】
        - unicast, broadcast, multicast
    - MAC address is used to distinguish between the destinations
    - Promiscuous mode

Token rings

- Ring topology【s4 p26】
    - prevent collisions before occurring
- Token ring transmission【s4 p27】
- Token passing paradigm
    - travel in a unidirectional fashion
- When the node fails, the relay closes bypassing the node.
    - 在物理角度与 Ethernet 并没有本质区别，同样可以用 Hub
- Token ring maintenance【s4 p30】
    - Ring can elect a new monitor. Highest MAC address wins.
- FDDI (Fiber Distributed Data Interface) dual ring operation
    - no single failure point
    - 但 token ring 太复杂了，现在更多用 Ethernet 的结构。

Wireless

- Wireless links【s4 p33】
    - similar to Ethernet structure, sometimes called "wireless ethernet".
- Access point (AP)
- Ad hoc (mesh structure)【s4 p35】
- Most radios in wireless networking can't transmit and receive at the same time, so we can't detect collisions. Instead, we'll do CSMA/CA (collision avoidance).
- Problems
    - The hidden node problem【s4 p37】
    - The exposed node problem
    - Solution: Multiple Access with Collision Avoidance (MACA)【s4 p39】

---

5. Packet switching

- Problems of direct link networks【s5 p5】
    - limitation: scalability
    - key problem: inter-connecting networks
    - solution: packet switching

Inter-networking

- General working mechanism: switching and forwarding
- Inter-connecting links of the same type: Bridges/LAN switches
- Inter-connecting links of different types: Gateways/routers, the Internet Protocol

Switching and forwarding

- Ethernet【s5 p9】
    - shared medium: Bus
    - Hub
- Switched LAN
    - 用的是 switch 而不是 Hub
    - NO shared medium
    - no direct link between hosts. 因为switch算是一台电脑
    - 现在实际上只用 switch 而不用 hub 了，因为成本也不高。
- Switch
    - Each one runs a **data-link protocol**.
    - Each input or output is a port (interface). 这里的port指的是物理意义上的接口，而不是数字。

    - Layer 3 switch (**router**): links may run different data link protocols

- Hub, Switch/Bridge, and Router/Gateway【s5 p13~14】

How does a switch decide what output port to use?

- Datagrams
    - connectionless
    - forwarding table
    - For connectionless networks【s5 p17】
        - routing is important
        - it is not known whether the network can deliver the packet
- Virtual circuit【s5 p18】
    - connection-oriented
    - Two part process:
        - connection setup
        - data transfer
    - virtual circuit table
    - connection establishment【s5 p20】
        - Permanent Virtual Circuit (PVC)
        - Switched Virtual Circuit (SVC)
    - Typical virtual circuit networks
        - Frame relay (support VPN)
        - Asynchronous transfer mode (ATM)
- Source routing【s5 p22】

Bridges/LAN Switches【s5 p24】

- LAN Bridge operation
- Learn table entries based on source address.
    - 一开始是空表，例如A从p1来，就知道它在p1，此时broadcast，因为不知道目标在哪，如此反复直到填出全表。
- Extended LANs can have loops
    - solution: minimum spanning tree
- Limitations【s5 p30】
    - do not scale
    - do not accommodate heterogeneity

---

6. Basic internetworking

The network layer

- Connecting heterogeneous networks【s6 p8】
    - requirements【s6 p6】
    - Router/Gateway
- Key functions of the network layer【s6 p12】
    - global addressing
    - data conversion
    - fragmentation
    - routing

The Internet Protocol (IP)【s6 p14】

- IP datagram delivery【s6 p18】
    - connectionless
    - Best-effort（尽力而为）
        - unreliable service
- Packet format
- Datagram size
    - Maximum Transmission Unit (MTU)
    - How to choose the max datagram size for IP?
        - fragment IP datagram into multiple MTUs
- Fragmentation and assembly【s6 p21】

Address and datagram forwarding

- IP addressing
    - independent of MAC addressing
    - used by: higher layer protocols, applications
    - virtual
    - 32-bit (IPv4)
    - unique value for each host/interface
        - prefix identifies the network (assigned by global authority), suffix identifies the host/interface (assigned by local administrator).【s6 p28】
- Dotted decimal notation
    - 4 decimal values per 32-bit address
    - each number represents 8 bits, the value is between 0 and 255.
- Classes of IP addresses【s6 p30】
    - class A: large networks, host 24
    - class B: medium, host 16
    - class C: small, host 8

- Special address
    - 0, booting
    - 255, broadcast
- Router addresses
    - need one address per router connection
- Datagram forwarding/routing
    - next hop
- Subnetting【s6 p38】
    - subnet masks 分割，后面为0的部分为host可以使用的位置。

---

7. IP related protocols

- Address translation (ARP)
- Host configuration (DHCP)
- Error reporting (ICMP)
- Virtual networks and tunnels

ARP

- Address translation【s7 p7】
    - map IP address into physical (MAC) addresses
    - ARP (Address Resolution Protocol)
- ARP【s7 p8】
    - ARP Request: who has this IP address?
    - ARP Reply: I have xxx. My MAC address is xxx.
    - Reverse ARP Request (RARP): who has this MAC address?
    - RARP Reply: I have the MAC. My IP address is xxx.
- ARP details【s7 p11】
- ARP spoofing attacks【s7 p16】
    - Denial of Service (DoS)
    - Man in the middle
    - MAC flooding
    - mitigating factor
        - only local attackers can exploit ARP's insecurities
    - How to prevent?【s7 p20】
        - cannot be completely fixed

Host configuration (DHCP)

- Dynamic Host Configuration Protocol (DHCP)【s7 p22】
  - IP addresses bound to workstations dynamically
  - Benefits and limitations【s7 p28】

Error reporting (ICMP)

- Internet Control Message Protocol (ICMP)【s7 p31】
  - Route test
    - Echo (ping), trace route
  - Error reporting
  - Redirect

Virtual networks and channels

- Private networks【s7 p35】
- Virtual Private Networks (VPN)
  - sharing some of the same physical links, but make it private. 因为单独买physical的线太贵了……
  - 在R1信息加密（包括地址也加密，别人拿到也看不懂），传到R2解密，然后再按地址传。（于是就可以躲过某些地址筛查）

---

8. Routing and the global Internet

Distance vector routing

- Routing table
- Network as a graph
  - all-pairs shortest path problem
- Static solution【s8 p9】
  - problem
- Distance vector routing【s8 p11】
  - algorithm【s8 p13】
  - deal with link failures【s8 p15】
    - Periodic updates
    - Triggered updates
  - count-to-infinity problem (2-node loops)
    - solution【s8 p18】
  - Routing Information Protocol (RIP)【s8 p19】

Link state routing【s8 p21】

- spread the local knowledge of each node to all other nodes in the network, so that every node can construct a network weighted graph
- Strategy【s8 p22】
    - send to all nodes (not just neighbors) information about directly connected links (not entire routing table)
    - Link State Packet (LSP)
- Link state routing mechanism
    - reliable flooding【s8 p24】
    - Dijkstra
- Properties【s8 p26】
- Open Shortest Path First Protocol (OSPF)

Global Internet

- Multi-backbone Internet【s8 p30】
    - challenges【s8 p32】
- Related topics【s8 p33】
    - Subnetting
    - Network address translation (NAT)【s8 p35】
    - IPv6【s8 p37】
        - 128-bit addresses
        - notation【s8 p38】
        - address space allocation
        - packet format

---

9. End to End Protocols

Transport protocols【s9 p6】

- Provide application-to-application (process-to-process) communication. Called "end-to-end"
- Optionally provide:
    - Reliability, Flow control, Congestion control
- End to end protocols
    - Simple demultiplexer (UDP)
    - Reliable byte stream (TCP)

UDP

- User Datagram Protocol【s9 p9】

    - **Unreliable** message delivery
    - Connectionless
    - No flow control
    - No error recovery (**no ACKs**)
    - Application multiplexing

- Addresses for applications: protocol ports【s9 p11】

    - server: lower port numbers
    - client: higher port numbers

- Simple asynchronous demultiplexing【s9 p13】

- Basic firewalls【s9 p16】

TCP

- Transmission Control Protocol【s9 p18】

    - Most popular layer 4 protocol
    - Connection-oriented protocol

- TCP feature summary

    - provides a stream transport service
    - allows two application programs to...

- Reliable byte stream【s9 p20】

    - reliable, in-order delivery

- Segment format【s9 p21】

- An apprent contradiction?【s9 p22】

    - IP and TCP
    - How is this possible? ACK mechanism.

- Achieving reliability

    - reliable connection setup
        - must be reliable
    - reliable data transmission
    - reliable connection shutdown
        - must be graceful

- Why startup/shutdown is difficult?【s9 p25】

- TCP Startup solution: **3-way handshake**【s9 p26】

    - use three-message exchange
    - SYN, SYN+ACK, ACK
    - Segments are labeled with a **sequence number**.【s9 p29】

        - Protect from out-of-order delivery.
        - Initial Sequence Numbers (ISN)

- TCP Shutdown: **四次挥手**【s9 p30】

- Reliable data transmission【s9 p31】

    - Guarantees a reliable delivery of data: **ARQ**
    - Ensures that data is delivered in order: **SeqNum**
    - Enforces flow-control between sender and receiver: **Sliding window**

- Reliable delivery of data

    - Positive Acknowledgement
    - **Retransmission**【s9 p33】

- Retransmission

    - Problem: how long should wait?
        - distance to destination, current traffic condition. 在TCP层面上，不知道中间的路是怎样的（不像data link layer）
    - Solution: **adaptive retransmission**
        - User current estimate to set retransmission timer
        - $\text{EstimatedRTT} = \alpha \text{EstimatedRTT} + (1-\alpha) \text{SampleRTT}$
        - $\text{TimeOut = 2 \times \text{EstimatedRTT}}$
        - 这里的SampleRTT指的是最近一次检测到的数据

- Pipeline operation

    - 由 **receiver** 决定 window size

- TCP flow control【s9 p38】

    - sliding window protocol
    - receiver: **advertise availble buffer space**
    - sender: can send up to entire window before ACK arrives

- Window advertisement【s9 p40】

    - Each acknowledgement carries new window information

- Sliding window【s9 p41】

    - Sending side: buffer bytes between LastByteAcked and LastByteWritten
    - Receiving side: buffer bytes between LastByteRead and LastByteRcvd

- Flow control【s9 p42】

    - Receiving side
    - Sending side

---

10. End-to-End Data

- End-to-End data
    - Presentation formatting
    - Markup language (XML)
    - Multimedia data

Presentation formatting

- Problem【s10 p7】
    - at the transport layer (TCP/UDP), a message is sent as an uninterpreted string of bytes.
    - How to encode the different kinds of data into byte strings?
- Encoding/Decoding
- Byte order: big-/little-endian
- Data types
- Formatting approaches【s10 p11】
    - Argument mashalling（参数编组）
        - Remote Procedure Call (**RPC**)【s10 p12】
    - Markup documents

Markup languages

- XML【s10 p19】
- Difference between XML and HTML【s10 p22】
    - XML is designed to carry data, not to display data
    - HTML is designed to display data
- Flexibility of XML【s10 p23】
- XML element, prolog, syntax, tree
- Displaying XML【s10 p28】
    - CSS
    - **XSL** (eXtensible Stylesheet Language) preferred!
- XML validation【s10 p32】
    - DTD (Document Type Definition)
- JSON
    - a lightweight data-interchange format
- YAML (Yet Another Markup Language)
- Application of XML

    - From HTML to XHTML
    - General document formats
        - docx
    - MathML
    - SVG
    - Synchronized Multimedia Integration Language (SMIL)
    - RSS (Really Simple Syndication)

Multimedia data

- The amount of data is very large
- Compression methods【s10 p53】
    - Image compression: JPEG
        - 记一次颜色，然后标出哪些坐标对应哪些颜色
    - Video compression: MPEG
        - I frames (intrapicture): self-contained, 保存完整的图片
        - P frames (predicted picture): depending on earlier I frames, 只存储不同之处
        - B frames (bidirectional predicted picture): depending on earlier and later frames (可以是I也可以是P)

---

11. Applications: World Wide Web

- Separation of duties【s11 p4】
    - network
    - application
- Typical application protocols【s11 p7】
    - HTTP (WWW)
    - FTP (file transfer)
    - SMTP, POP, IMAP (emails)
    - SSH, Telnet (remote shell)
    - RDP, VNC (remote desktop)

World Wide Web (HTTP)

- HTML (HyperText Markup Language)【s11 p11】
- URL (Uniform Resource Locater)【s11 p12】
- URL and MIME types【s11 p15】
- HTTP (HyperText Transfer Protocol)【s11 p16】
    - default server port: 80
    - HTTP request message【s11 p18】
    - HTTP request methods【s11 p21】

    - HTTP response codes【s11 p23】
- HTTPS【s11 p25】
- Cookies【s11 p26】
    - to identify the client
    - private mode
- Documents【s11 p28】
    - Static: fast, but not customized
    - Dynamic: slower, but more flexible
- Client side elements
    - HTML5
- AJAX (Asynchronous Javascript and XML)
- Web services: XML + HTTP【s11 p31】
    - XML-RPC
    - example: Google search API
- HTTP server operation
- Multi-tier servers
- Browser operation【s11 p37】
    - browser kernels (layout engines)

---

12. Applications: DNS and Email

Domain name service (DNS)

- DNS functionality【s12 p5】
    - given name of a computer, return computer's IP address
    - Example【s12 p6】
- Configure of the network connection
    - IP address
    - netmasks
    - gateway
    - name server
- Domain name syntax【s12 p7】
    - DNS: domain name $\rightarrow$ IP address
    - ARP: IP address $\rightarrow$ MAC address

- Top-level domains
- DNS client/server interaction【s12 p14】
    - client known as resolver
- Name resolution【s12 p15】
    - local name server不需要知道全世界所有的name所对应的IP地址，按照图中的顺序问就行，问过的存为cache，更快。
- Root servers【s12 p18】
    - 13 root name servers
    - multiple server hosts
        - load balancing【s12 p20】
        - content distribution
- Global server load balancing (GSLB)【s12 p22】

Internet Email (SMTP, POP, IMAP)

- Internet Email【s12 p28】
    - Email address
    - Mail message format
- MIME (Multi-purpose Internet Mail Exchange)【s12 p29】
    - sender inserts additional header lines to indicate the encoding method
- MIME encoding
    - quoted-printable
    - base64
- SMTP (Simple Mail Transfer Protocol)【s12 p34】

Questions

s1

- What is packet switching? What are the advantages of packet switching?
- Why is layering useful?

s3

- If you want to connect your home to the Internet, what are the common access technologies available for you to choose from?
- Why is the sliding window transmission “reliable”?

- Why should the sender window size (SWS) be set to bandwidth*delay product estimate?

s4

- Why is Ethernet more popular than Token Ring in LANs?
- Why might a mesh (ad-hoc) topology be superior to a base station topology for communication in a natural disaster?
- How can an ethernet communicate with an FDDI network?

s5

- What are the advantages and disadvantages of datagram switching?
- What are the advantages and disadvantages of virtual circuit switching?
- How can broadcast and multicast of messages be implemented in the extended LAN (connected with bridges)?
- What is ATM? What do you think about the future of ATM?
- How can an ethernet communicate with an FDDI network?

s6

- Why is IP able to connect various link technologies to form a larger internetwork?
- What aspect of IP addresses makes it necessary to have one address per network interface, rather than just one per host?

s8

- What are the differences between distance vector and link state routing approaches?
- What are the limitations of the Network Layer?

s11

- Why the client's port numbers are different but the server's are the same?【s11 p6】
- What made WWW so successful?
- Why does HTTP run over TCP (rather than UDP)?

s12

- Is is possible to map multiple domain names to one IP address? If yes, what kind of problems should be addressed?