Introduction

Madiba Hudson-Quansah

Contents

Chapter 1 **The Internet** — **Page 3**

1.1 Nuts and Bolts Description of the Internet 3
1.2 Services Based Description of the Internet 4
1.3 Protocols 5

Chapter 2 **Network Edge** — **Page 6**

2.1 Access Networks 6
Home Access — 6 • Hosts — 6
2.2 Physical Media 7
Guided Media — 8
2.2.1.1 Twisted Pair Copper Wire . . . 8
2.2.1.1.1 Advantages . . . 8
2.2.1.1.2 Disadvantages . . . 8
2.2.1.2 Coaxial Cable . . . 8
2.2.1.2.1 Advantages . . . 8
2.2.1.2.2 Disadvantages . . . 8
2.2.1.3 Fiber Optics . . . 8
2.2.1.3.1 Advantages . . . 9
2.2.1.3.2 Disadvantages . . . 9
Unguided Media — 9
2.2.2.1 Terrestrial Radio Spectrum . . . 9
2.2.2.2 Satellite Radio Channels . . . 9

Chapter 3 **Network Core** — **Page 10**

3.1 Packet Switching 10
Store-And=Forward Transmission — 11 • Queueing Delay and Packet Loss — 11 • Forwarding Tables and Routing Protocols — 11
3.2 Circuit Switching 11
Multiplexing in Circuit-Switched Networks — 12
3.2.1.1 Frequency Division Multiplexing (FDM) . . . 12
3.2.1.2 Time Division Multiplexing (TDM) . . . 12
3.3 Circuit Switching vs Packet Switching 12

Chapter 4 **Delay, Loss, and Throughput in Packet-Switched Networks** — **Page 13**

4.1 Delay 13
Types of Delay — 13
4.1.1.1 Processing Delay . . . 13

4.1.1.2 Queueing Delay 13
4.1.1.3 Transmission Delay 14
4.1.1.4 Propagation Delay 14
4.1.1.5 Transmission vs Propagation Delay 14
4.1.1.6 Total Nodal Delay 14
Queueing Delay and Packet Loss — 14 • End-to-End Delay — 14
4.2 Throughput 15

Chapter 1

The Internet

Definition 1.0.1: Network Edge

Hosts, access networks physical media

Definition 1.0.2: Network core

Packet/Circuit switching, internet structure

Definition 1.0.3: Performance

Loss, Delay, Throughput

Definition 1.0.4: Host

End Systems, Clients/Servers

Definition 1.0.5: Internet

Network of networks, interconnected ISPs

1.1 Nuts and Bolts Description of the Internet

The internet is a computer network that interconnects billions of computing devices throughout the world.

End systems are connected together by a network of communication links and packet switches. Different links can transmit data at different rates, with the **transmission rate** of a link measured in bits/second. The resulting units of information, known as packets, are then sent throughout the network to the destination end system, where they are reassembled into the original data.

Definition 1.1.1: Transmission Rate

The rate, in bits/second, at which a link can transport data from one end to the other

Definition 1.1.2: Packet Switch

Forwards packets arriving on incoming communication links to outgoing communication links.

Packet switches interconnect communication links, with the two main types of packets switches being:

Routers - Typically used in wide-area networks/network core e.g. the Internet

Link Layer Switches - Typically used in access networks e.g. LANs

Both types forward packets to their ultimate destinations, and interconnected packet switches comprise a route/path through the network

Definition 1.1.3: Route / Path

The sequence of communication links and packets switches traversed by a packet from the sending end system to the receiving end system.

Definition 1.1.4: Internet Service Provider (ISP)

A network of packet switches and communication links that provides access to the Internet.

End systems access the internet through **Internet Service Providers (ISP)**. ISPs can be classified into tiers:

Lower-Tier ISPs (Tier 1) - Small regional ISPs that provide access to end systems and small businesses.

Upper-Tier ISPs (Tier 2) - Consists of high-speed routers interconnected with high-speed fiber-optic links. They provide access to lower-tier ISPs and large businesses, and are connected directly to one another.

Each ISP on either tier is managed independently, runs the IP protocol, and conforms to certain naming addresses and conventions.

End systems, packet switches and other network devices run **protocols** 1.3 that manage the sending and receiving of information within the internet. The two most important protocols on the internet are:

Transmission Control Protocol (TCP) - Manages the sending and receiving of messages, ensuring reliable delivery of data.

Internet Protocol (IP) - Specifies the format of packets that are sent and received among routers and end systems.

Together, these two protocols are referred to as TCP/IP.

The **Internet standards** are developed by the Internet Engineering Task Force **IETF**, and are documented in **Request For Comments (RFCs)**.

1.2 Services Based Description of the Internet

The internet can be described in terms of the services it provides, i.e. as an infrastructure that provides services to applications. These applications include email, web surfing, internet messaging, mapping, video/music streaming, etc. These applications are said to be **Distributed Applications**

Definition 1.2.1: Distributed Applications

Applications that involve multiple end systems that exchange data with each other.

Internet applications run on end systems, not on packet switches in the network core, i.e. the network core is not concerned with the applications that are the sources and sinks of data.

End systems connected to the internet provide a **socket interface** to applications for the purpose of sending and receiving data.

Definition 1.2.2: Socket Interface

Specifies how a program running on one end system asks the internet infrastructure to deliver data to a specific destination program running on another end system. It is a set of rules that the sending program must follow so that the internet can deliver the data to the destination program.

1.3 Protocols

Definition 1.3.1: Protocol

Defines the format, order of messages send and received among network entities, and actions taken on message transmission/receipt.

All activity in the internet that involves two or more communicating remote entities is managed by a protocol.

Chapter 2

Network Edge

Definition 2.0.1: Network Edge

End systems and access networks that connect end systems to the network core.

2.1 Access Networks

Definition 2.1.1: Access Network

The network that physically connects an end system to the first/edge router on a path from the end system to any other distant end system.

2.1.1 Home Access

The two most common types of broadband residential access are:

Digial Subscriber Line (DSL) - Operates over existing telephone lines provided by a telephone company (Telco), with maximum rates ranging from hundreds of Kbps to several Mbps. The speed is distance dependent, with shorter distances yielding higher speeds. When DSL is used a customer's telco is also their ISP.

Cable - Makes use of the cable television (CATV) infrastructure to provide broadband access to homes. Cable modems provide data rates ranging from several Mbps to 100's of Mbps. Cable is a shared medium, with the bandwidth being shared among all users in a local area.

Definition 2.1.2: Modem

Converts digital signals from a computer to analogue signals for transmission over telephone lines, and vice versa.

Each DSL modem uses existing telephone lines to exchange data with a digital subscriber line access multiplexer (DSLAM) located in the telco's local central office (CO). The home's DSL modem takes digital data and translates it to high-frequency tones for transmission over the telephone wires to the CO, with the analogue signals from many houses being translated back into digital format at the DSLAM. The residential telephone line carries both data telephone signals encoded at different frequencies:

High-Speed Downstream - 50kHz to 1MHz

Medium-Speed upstream - 4kHz to 50kHz

Two-Way telephone - 0 to 4kHz

2.1.2 Hosts

In transmitting data a host:

1. Takes application data
2. Breaks it down into smaller chunks known as packets of length L bits
3. Transmits these packets into the access network at a transmission rate of R bps, this is also known as the link's transmission rate or bandwidth.

The packet transmission delay is given by:

$$\text{Packet Transmission Delay} = \frac{L \text{ bits}}{R\text{bps}}$$

Example 2.1.1

Consider a packet of length $L = 1000$ bits, and a link with a bandwidth of $R = 1Mbps$, what is the packet transmission delay?

$$\begin{aligned}\text{Packet Transmission Delay} &= \frac{L \text{ bits}}{R\text{bps}} \\ &= \frac{1000 \text{ bits}}{1,000,000 \text{ bps}} \\ &= 0.001 \text{ seconds} = 1 \text{ ms}\end{aligned}$$

Now assume there are two packets, each of length $L = 1000$ bits, that are sent back-to-back on the same link with a bandwidth of $R = 1Mbps$. What is the time taken to send both packets?

$$\begin{aligned}\text{Time to send 2 packets} &= 2 \times \frac{L \text{ bits}}{R\text{bps}} \\ &= 2 \times \frac{1000 \text{ bits}}{1,000,000 \text{ bps}} \\ &= 0.002 \text{ seconds} = 2 \text{ ms}\end{aligned}$$

2.2 Physical Media

Definition 2.2.1: Bit

The smallest unit of data in a computer network, represented as a 0 or 1. Propagates between transmitter and receiver pairs over a physical medium.

Definition 2.2.2: Physical Media

The physical materials that carry bits/signals between end systems and packet switches (transmitters and receivers). This is not excluding wireless media, which passes through the air which is also physical.

Examples of physical media include:

- Twisted Pair Copper Wire
- Coaxial Cable
- Multimode Fiber Optics
- Terrestrial Radio Spectrum
- Satellite Radio Spectrum

Physical media falls into two broad categories:

Guided Media - Signals are guided along a solid medium e.g. twisted pair copper wire, coaxial cable, fiber optics.

Unguided Media - Signals are transmitted through the atmosphere and in outer space e.g. terrestrial radio, satellite radio.

2.2.1 Guided Media

2.2.1.1 Twisted Pair Copper Wire

The least expensive and most commonly used guided transmission medium. It consists of two insulated copper wires twisted together to reduce electrical interference from similar pairs close by. Typically a number of pairs are bundled together in a cable by wrapping the pairs in a protective shield. A wire pair constitutes a single communication link.

Unshielded Twisted Pair (UTP) is commonly used for computer networks within a building, i.e. LANs. Data rates for LANs using twisted pair range from 10 Mbps to 10 Gbps, depending on the thickness of the wire and the distance between transmitter and receiver.

2.2.1.1.1 Advantages

- Inexpensive
- Easy to install

2.2.1.1.2 Disadvantages

- Susceptible to electromagnetic interference
- Limited bandwidth and distance

2.2.1.2 Coaxial Cable

Consists of two copper conductors, but the two conductors are concentric, rather than parallel. The inner conductor is surrounded by an insulating material, which is then surrounded by a cylindrical conducting shield. The shield serves to reduce electromagnetic interference and allows for higher bandwidth and longer distances than twisted pair. Coaxial cable is commonly used for cable television and broadband internet access. Can be used as a guided **shared medium**, where multiple transmitters and receivers are connected to the same coaxial

2.2.1.2.1 Advantages

- Higher bandwidth and longer distance than twisted pair
- Less susceptible to electromagnetic interference than twisted pair

2.2.1.2.2 Disadvantages

- More expensive than twisted pair
- More difficult to install than twisted pair

2.2.1.3 Fiber Optics

A thin, flexible medium that conducts light pulses with each pulse representing a bit. A single optical fiber can support large bit rates over long distances, with data rates of 10 Gbps or more over distances of tens of kilometres. Fiber optics is commonly used for high-speed data transmission in the core of the internet and for long-distance telecommunications.

2.2.1.3.1 Advantages

- Very high bandwidth and long distance capabilities
- Immune to electromagnetic interference

2.2.1.3.2 Disadvantages

- More expensive than twisted pair and coaxial cable
- More difficult to install than twisted pair and coaxial cable

2.2.2 Unguided Media

2.2.2.1 Terrestrial Radio Spectrum

Radio channels carry signals in the electromagnetic spectrum through the atmosphere. They are used for wireless communication between end systems and packet switches, and between packet switches in the network core. Examples of terrestrial radio channels include Wi-Fi, cellular networks, and Bluetooth. These can be classified into three groups:

Short Distance - Operate within on or to meters, e.g. Bluetooth

Local Area - Operate within ten to a few hundred meters, e.g. Wi-Fi

Wide Area - Operate over large areas, spanning tens of kilometres, e.g. cellular networks

2.2.2.2 Satellite Radio Channels

A communication satellite links two or more Earth-based microwave transmitter/receivers known as ground stations. The satellite receives signals on one frequency band, regenerates the signal using a repeater and transmits the signal on another frequency. There are two types of satellite radio channels:

Geostationary Earth Orbit (GEO) - Satellites that orbit approximately 36,000 km above the Earth and remain fixed in the same position relative to the Earth. GEO satellites are commonly used for television broadcasting and long-distance communication.

Low Earth Orbit (LEO) - Satellites that orbit much closer to the Earth, typically at altitudes of 500 to 2,000 km. LEO satellites are used for applications such as satellite internet and remote sensing.

Chapter 3

Network Core

Definition 3.0.1: Network Core

The mesh of interconnected packet switches and communication links that interconnects end systems.

3.1 Packet Switching

Definition 3.1.1: Packet Switching / Forwarding

A network in which messages are broken into smaller chunks known as packets that are sent from source to destination through packet switches.

Definition 3.1.2: Message

A chunk of data sent from a source to a destination end system.

In a network application end systems exchange messages with each other. To send a message from a source end system to a destination end system, the message is broken down into smaller chunks known as packets. Between source and destination, each packet travels through communication links and packet switches, (link-layer switches in the case of the network edge). Packets are transmitted over each communication link at a rate equal to the full transmission rate of the link, i.e. if a source system or a packet is sending a packet of L bits over a link with a transmission rate of R bps, the time taken to transmit the packet onto the link is given by:

$$\text{Packet Transmission Delay} = \frac{L\text{ bits}}{R\text{bps}}$$

Definition 3.1.3: Forwarding

The local action of transmitting a packet from the outbound link of one packet switch to the inbound link of the next packet switch.

Definition 3.1.4: Routing

The global action of sending a packet from a source node to a destination node through multiple packet switches.

The local action of transmitting a packet from the outbound link of one packet switch to the inbound link of the next packet switch is known as **forwarding**. The global action of sending a packet from a source node to a destination node through multiple packet switches is known as **routing**.

3.1.1 Store-And=Forward Transmission

Store-and-forward transmission, requires a packet switch receive the entire packet before it can begin to transmit the first bit of the packet onto the outbound link. Thus, the time taken to transmit a packet from the source end system to the destination end system is given by:

$$\text{Total Transmission Time} = N \times \frac{L \text{ bits}}{R\text{bps}}$$

where N is the number of links (or packet switches) between source and destination. A packet switch will always have an inbound and outbound link so with k packet switches there will be $k + 1$ links. Given P packets the total time taken to transmit all packets from source to destination is given by:

$$\text{Total Transmission Time} = (N + P - 1) \times \frac{L \text{ bits}}{R\text{bps}}$$

Due to store-and-forward transmission, the first packet must be transmitted completely before the second packet can be transmitted, hence the $P - 1$ term in the equation above.

3.1.2 Queueing Delay and Packet Loss

Definition 3.1.5: Packet Loss

The dropping of packets that arrive at a packet switch when the output buffer/queue is full.

Each packet switch has multiple links attached to it, and for each attached link there is an **output buffer/ queue**, which stores packets that the router is about to send into that link. The output buffer allows busy links to temporarily store arriving packets until the link is ready to transmit them. Thus in addition to the store-and-forward delay, packets also experience **queueing delay** at each packet switch. This delay depends on the level of congestion in the network. Since the amount of buffer space is finite, an arriving packet may meet a full buffer and be dropped, this is known as **packet loss**.

3.1.3 Forwarding Tables and Routing Protocols

Packet forwarding is done in different ways depending on the type of packet switch. In general, each packet switch maintains a **forwarding table** that maps destination addresses to that router's outbound links. When a packet arrives a router the router examines the address and searches its forwarding table using the destination address to find the appropriate outbound link for that packet.

The forwarding tables are populated using **routing protocols**, which are distributed algorithms that determine the route that packets take from source to destination.

3.2 Circuit Switching

Definition 3.2.1: Circuit Switching

A network in which a dedicated path (circuit) is established between a sender and receiver for the duration of the communication session.

In circuit switched networks the resources needed along a path to provide for communication between end systems are reserved for the duration of the communication session between the end systems. Traditional telephone networks are examples of circuit switched networks.

In a circuit switched network, when a user wishes to communicate with another user, a dedicated end-to-end path (circuit) is established between the two users. This path consists of a sequence of links and switches, with each link along the path having a fixed amount of bandwidth reserved for the duration of the communication session. The bandwidth used by a user during a communication session is a fraction of the total bandwidth of each link along the path, for example if a link where has a bandwidth of R bps and there are N circuits established on that link, then each circuit is allocated a bandwidth of $\frac{R}{N}$ bps.

Due to the reservation of resources along the path the maximum number of active users at any time is the total number of circuits that can be established in the network.

Considering the case of sending a file of 640000 bits from Host A to Host B in a circuit switched network, with 24 slots and a bit rat of 1.536 Mbps. Also assuming it takes 500 ms to establish an end-to-end circuit between Host A and Host B, the time taken to send the file is given by:

$$\begin{aligned}\text{Single Transmission Rate} &= \frac{1.536}{24} \\ &= 64 \text{ kbps}\end{aligned}$$

$$\begin{aligned}t_t &= \frac{640}{64} \\ &= 10 \text{ seconds}\end{aligned}$$

$$\begin{aligned}t_{\text{total}} &= t_{\text{establish}} + t_t \\ &= 0.5 + 10 \\ &= 10.5 \text{ seconds}\end{aligned}$$

3.2.1 Multiplexing in Circuit-Switched Networks

Circuit switching can be implemented with either **Frequency Division Multiplexing (FDM)** or **Time Division Multiplexing (TDM)**.

3.2.1.1 Frequency Division Multiplexing (FDM)

The frequency spectrum of a link is divided up among the connections established across the link, i.e. dedicating a frequency band to each connection for the duration of the connection, the width of this known as the **bandwidth** of the connection. For example, if a link has a bandwidth of R bps and there are N connections established across the link, then each connection is allocated a bandwidth of $\frac{R}{N}$ bps.

3.2.1.2 Time Division Multiplexing (TDM)

Time is divided into frames of fixed duration, and each frame is divided into a fixed number of time slots. When the network establishes a connection across a link, the network dedicates one time slot in every frame to this connection. These slots are dedicated for the sole use of that connection, with one time slot available for use to transmit the connection's data, in every frame. For example if a link has 5 connections established across it, then each connection is allocated one time slot in every frame, and thus can transmit data in one time slot in every frame, i.e. a frame would consist of 5 time slots, with each time slot allocated to one of the 5.

3.3 Circuit Switching vs Packet Switching

An argument against circuit switching is that it is wasteful because the dedicated circuits are idle during silent periods, and thus the resources reserved for a connection cannot be used by other ongoing connections.

Packet switching is more efficient than circuit switching because it allows the resources of the network to be shared among all users, with the resources being used only when there is data to send. However, packet switching can lead to congestion and packet loss when the demand for resources exceeds the available resources, whereas circuit switching can lead to blocking when there are too many connection requests and not enough resources to establish all the requested connections.

Chapter 4

Delay, Loss, and Throughput in Packet-Switched Networks

4.1 Delay

Definition 4.1.1: Delay

The time it takes for a packet to travel from source to destination. The most import of these delays are

1. Nodal Processing Delay
2. Queueing Delay
3. Transmission Delay
4. Propagation Delay

Together these give the **Total Nodal Delay**

4.1.1 Types of Delay

Most packet switched networks experience four types of delay at each packet switch / node:

Nodal Processing Delay - When the packet arrives at the inbound link

Queueing Delay - When the packet is waiting in the output buffer

Transmission Delay - When the packet is being transmitted onto the outbound link

Propagation Delay - When the packet is moving through the physical medium

4.1.1.1 Processing Delay

The time required to examine the packet's header and determine where to direct the packet. Processing delay can also include other factors, like error checking. Processing delays in high-speed routers are usually on the order or microseconds or les.

4.1.1.2 Queueing Delay

The time a packet spends waiting to be transmitted onto the outbound link. The queueing delay of a specific packet depends on the number of earlier arriving packets that are queued and waiting for transmission onto the link. If the queue is empty and no other packet is currency being transmitted the queueing delay is zero.

4.1.1.3 Transmission Delay

The time required to push all the packet's bits onto the link. It is given by:

$$\text{Transmission Delay} = \frac{L \text{ bits}}{R\text{bps}}$$

4.1.1.4 Propagation Delay

The time required for a bit to propagate from the beginning of the link to the destination router at the other end. The propagation speed depends on the physical medium of the link and is in the range of

$$2 \cdot 10^8 \text{m/s to } 3 \cdot 10^8 \text{m/s}$$

This is less than or equal to the speed of light in a vacuum. The propagation delay is the distance between two routers divided by he propagation speed of the link i.e.:

$$\text{Propagation Delay} = \frac{d \text{ meters}}{s \text{ meters/second}}$$

4.1.1.5 Transmission vs Propagation Delay

Transmission delay is the amount of time needed for a router to push out a packet's bits onto the link, whereas propagation delay is the amount of time it takes a bit to traverse the link from one end to the other.

4.1.1.6 Total Nodal Delay

Given all these delays, the total nodal delay at a router is given by:

$$\begin{aligned} d_{\text{nodal}} &= d_{\text{proc}} + d_{\text{queue}} + d_{\text{trans}} + d_{\text{prop}} \\ &= \text{Processing Delay} + \text{Queueing Delay} + \frac{L}{R} + \frac{d}{s} \end{aligned}$$

The contribution of these delays can vary significantly from one packet switch to another, and from one packet to another at the same packet switch.

4.1.2 Queueing Delay and Packet Loss

Unlike the three other delays the queueing delay can vary from packet to packet, therefore when characterizing queueing delay it is common to use statistical measurements, like average queueing delay, variance, and the probability that the queueing delay exceeds some specified value.

4.1.3 End-to-End Delay

The total delay from source to destination is the sum of the nodal delays incurred at each node along the path from source to destination. If there are $N - 1$ routers between source and destination then there are N links and the total end-to-end delay is given by:

$$d_{\text{end-end}} = N\left(d_{\text{proc}} + d_{\text{trans}} + d_{\text{prop}}\right)$$

Taking into account heterogeneous delays at the nodes and the present of an average queueing delay at each node gives:

$$d_{\text{end-end}} = \sum_{i=1}^{N} \left(d_{\text{proc}_i} + d_{\text{queue}_i} + d_{\text{trans}_i} + d_{\text{prop}_i}\right)$$

4.2 Throughput

Definition 4.2.1: Throughput

The rate (in bits/second) at which data is transferred between source and destination. There are two types of throughput:
Instantaneous Throughput - The rate at which data is transferred at a given point in time.
Average Throughput - The rate at which data is transferred over a longer period of time.

Definition 4.2.2: Bottleneck Link

The link on the end-to-end path that constrains the end-to-end throughput to be less than the end-to-end throughput of all other links on the path.

Throughput can be measured at any point along the path from source to destination, but the end-to-end throughput is constrained by the link with the lowest throughput, this link is known as the **bottleneck link**. Given a path with 2 links with throughputs of R_s and R_c, i.e. server to link throughput and client to link throughput, the end-to-end throughput is given by:

$$R_{\text{end-end}} = \min(R_s, R_c)$$

Having determined the throughput we can approximate the time it takes to transfer a large file of F bits from server to client as:

$$t = \frac{F}{\min(R_s, R_c)}$$

Example 4.2.1

Question 1

Given a file of $F = 32$ MB and the server has a transmission rate of $R_s = 2$ Mbps, and you have an access link of $R_c = 1 Mbps$. How long will it take to transfer the file from the server to your computer?

Solution:

$$\begin{aligned} t &= \frac{32}{\min(2, 1)} \\ &= \frac{32}{1} \\ &= 32 \text{ seconds} \end{aligned}$$

This generalizes to a path with N links with throughputs of $R_1, R_2, \ldots, R_N$ as:

$$R_{\text{end-end}} = \min(R_1, R_2, \ldots, R_N)$$

And the time to transfer a file of F bits from server to client is given by:

$$t = \frac{F}{\min(R_1, R_2, \ldots, R_N)}$$

In the case where multiple clients (N clients each of transmission rate R_c) and servers (N servers each of transmission rate R_s) are connected to the core of the computer network, and there are N simultaneous downloads in progress involving all N client-server pairs. There is a link at the core traversed by all N downloads with a transmission rate of R. If the rate of R is much much larger than R_s and R_c then the throughput is still given by:

$$R_{\text{end-end}} = \min(R_s, R_c)$$

But in the case where R is comparable to R_s and R_c then the common link divides its transmission rate equally among the N downloads, giving a throughput of:

$$R_{\text{end-end}} = \min\left(R_s, R_c, \frac{R}{N}\right)$$

Possible making the throughput much smaller than the case where R is much larger than R_s and R_c.