yz709-network-sup1

Question 2

Consider a communication network consisting of a room full of people, where one or more people are exchanging thoughts with one or more others by talking.

- (a) For each of the abstract terms *node, channel, layer, transmission* (the act thereof), *coding, addressing and multiplexing*, identify one or more corresponding concrete components or activities within the system. If the correspondence is not exact, give the closest approximation you can.
    - Each person is a node within the system; within each node, there are several layers: brain layer (produce thoughts), vocal cord layer (make voices corresponding to the thoughts), mouth-ear layer (node-to-node layer, mouth for transmitting out voices, ear for receiving voices). Peer mouth-ear layers are "physically" connected via air.
    - The channel is the abstract voice transmission pipe between node entities; the voices of human speech are unique, so assume each voice has a unique frequency. The bandwidth b of the channel is the maximum data rate - 20 kHz; although we can have a throughput between 20 Hz and 20 kHz, the propagation delay d of the channel is the distance between the sender and the receiver; therefore, the capacity of the channel is $b \times d$.
    - Words are coded as sound waves before transmitting and decoded into words at the receiving end; the sender and receiver have to agree on the same protocol(i.e., language) to understand each other.
    - Multiplexing is when many people speak simultaneously; the air can carry all the different sound waves as a common transmission media. Although not precise, we can think of a frequency-division multiplexing model. Each person's voice has a

different frequency so that air can transmit voices with different frequencies together.

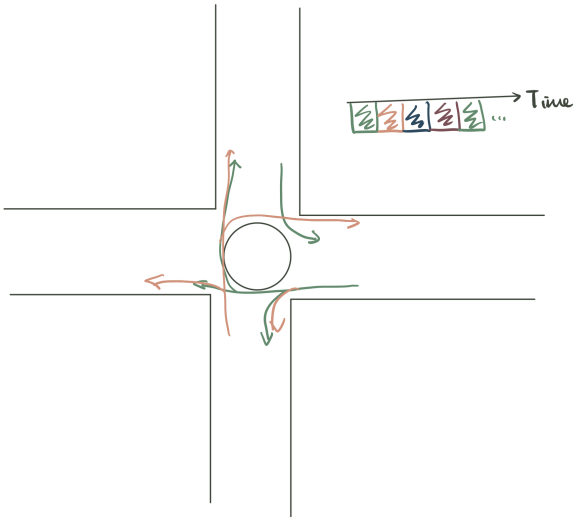
- (b) For one of the layers you identified above, describe
  1. the abstract interface it demands of the lower layer;
     - The vocal cord layer needs to turn words into sound waves, then output the sound waves to the lower layer (mouth-ear layer), so it requires the mouth-ear layer to act (change shapes or focus) to output the voices.
  2. the abstract interface it provides to a higher layer;
     a. The vocal cord layer accepts words from the higher layer (brain layer), so the brain layer needs to construct sentences precisely describing the thoughts before passing them down to the vocal cord layer.
  3. The symbols which the layer transmits to its peer layer across the corresponding channel.
     - For the vocal cord layer, the sound waves are transmitted across the channel, the sender would encode the words into sound waves, and the receiver would decode the sound waves to words.

Comments:

Question 3

Consider a roundabout.

- (a) What is being multiplexed?
  - For a typical roundabout, there are flows of cars (i.e., queues of cars) on each side, so for four sides, there are four flows of cars, each car wants to go in a particular direction.
- (b) What is the multiplexing strategy (e.g., time division)?
  - It is synchronous time-division multiplexing (STDM), although the time allocated for each flow may be different. For instance, in the following diagram (which only showed two-time slots), two flows of cars can go ahead for each time slot, and there are four possible destination directions. We allocate time slots in a round-robin fashion.



- (c) How is the access control agreed?
  - The access control determines the time allocation for each flow of cars, and we use traffic lights as signals for drivers.
  - The time allocation depends on the traffic load in each direction; usually, by analysing past road traffic statistics, a reasonable time slot will be assigned for each flow of cars. There might be dynamic adjustments at different times of the day, for instance, a lot of traffic from the countryside to the city centre in the morning because people are all driving to work, and in the evenings, the traffic load reversed.

> 💡 Comment:

Question 4

Consider a 10-Gbps link on a copper wire with a speed of propagation of 2.3 x 10^8 m/s:

- How “wide” is a bit (i.e., in time)?
  - Consider transmitting 10^9 bits per second, the width within the unit 1 second of a bit is $\frac{1}{10^9} = 10\ ns$
- How long is a bit?
  - Bit length is the distance in meters the bit occupies on a transmission medium.
  - Since the transmission speed on a copper wire is $2.3 \times 10^8\ m/s$, then bit length in the copper wire will be $10\ ns \times 2.3 \times 10^8\ m/s = 0.23\ m$.

💡 Comments:

Question 5

For each of the following operations on a remote file server, discuss whether they are more likely to be delay sensitive or bandwidth sensitive:

- (a) Open a file.
  - Delay-sensitive since only the file name or several other operation flags need to be exchanged.
- (b) Read the contents of a file.
  - It depends on the size of the file contents; if the file size is big, it will be bandwidth-sensitive; if the file size is relatively small, then it is delay-sensitive.
- (c) List the contents of a directory.
  - It depends on the size of the directory; if there are a lot of nested folders and many files, then it will be bandwidth-sensitive; otherwise, it will be delay-sensitive.
- (d) Display the attributes of a file.
  - File attributes are meta-data that describes how files and directories in a file system behave. These include flags indicating whether a file is visible, modifiable, compressed or encrypted. The bandwidth requirement is small, so it is delay-sensitive.

💡 Comments:

Question 6

Discuss the relative performance needs of the following applications in terms of average bandwidth, peak bandwidth, latency, and loss tolerance:

- (a) File server.
  - Average bandwidth: a file server normally handles a request at a time, a Unix file is usually small, so the average bandwidth requirement is small.

    - Peak bandwidth: at peak time, a file server might encounter multiple requests; a file server usually has a high peak bandwidth to allow it to serve multiple file requests simultaneously.
    - Latency: a file server is usually shared within a small firm; if users access files that are locally set up, then the latency is relatively small; however, if users access files remotely, then latency might increase.
    - Loss tolerance: The file server needs to guarantee the messages sent will be received and resent if packets are lost.
- (b) Print server.
    - Average bandwidth: a print server usually has one request at a time, the print jobs have varied sizes but are generally small, so the average bandwidth requirement is relatively small.
    - Peak bandwidth: a print server in an organisation may frequently encounter multiple print jobs; it will put them into a print queue and handle one job at a time. So an extensive buffer is more important than a high peak bandwidth.
    - Latency: A print server is usually locally located, so that the latency will be relatively small, but we can accept a much higher latency than a file server.
    - Loss tolerance: Low loss tolerance, especially for printing texts. Because it will be not understandable and readable if some of the words are lost.
- (c) Routine monitoring of remote weather instruments.
    - Average bandwidth: It depends on the amount of information that needs to be sent over. The amount of data is extensive for a remote weather measurement station, so high bandwidth is required. However, only a limited amount of information will be sent for a family weather measurement tool, so a much lower bandwidth.
    - Peak bandwidth: The peak bandwidth is similar to the average bandwidth because the amount of information sent at any time will not change much.
    - Latency: high latency because remote monitoring might require sending a message on a WAN network.
    - Loss tolerance: Low loss tolerance because the message contents sent will usually be texts; it will become not readable if some of the words are lost.
- (d) Voice.
    - Average bandwidth: low bandwidth is required because voice messages can be compressed a lot, and only unicast is usually needed.

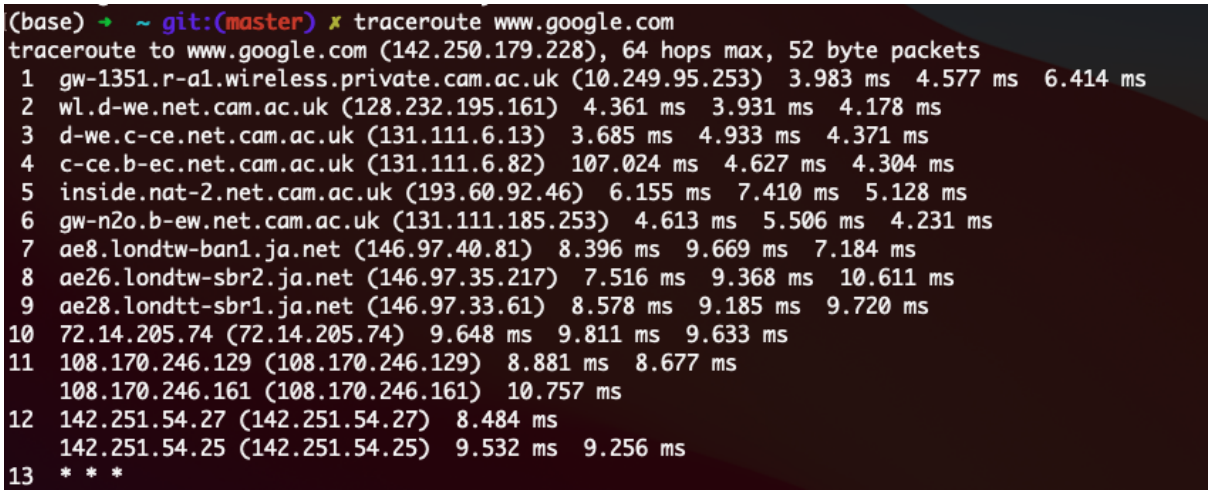
    - Peak bandwidth: When many people are speaking together, the bandwidth requirement dramatically increases, so the peak bandwidth is high.
    - Latency: Low latency is required because interactions between users always require timely responses; for instance, if someone wants to interrupt a conversation, we need to make sure the receiver receives the message and stops the conversation as soon as possible.
    - Loss tolerance: high loss tolerance because the receivers could still understand the message even if some packets were lost during transmission.
- (e) Video monitoring of a waiting room.
    - Average bandwidth: bandwidth depends on the frame size. It will be bigger than sending texts or pictures.
    - Peak bandwidth: The peak bandwidth is similar to the average bandwidth because the scene does not change too much from time to time for a video monitoring of a waiting room.
    - Latency: low latency is required because the purpose of monitoring is to reduce the response time when anything suspicious comes up.
    - Loss tolerance: high loss tolerance because we do not need an extremely high-quality video, as long as we can distinguish the objects in the recordings.
- (f) Television broadcasting.
    - Average bandwidth: usually takes up 6 MHz ~ 8 MHz of bandwidth for over-the-air transmission.
    - Peak bandwidth: The peak bandwidth might be massive.
    - Latency: high latency because television broadcast uses have a wide range.
    - Loss tolerance: some loss tolerance because we do not need an extremely high-quality video, as long as we can understand the contents.

> 💡 Comments: I am not too sure about this question; I used a lot of intuition.

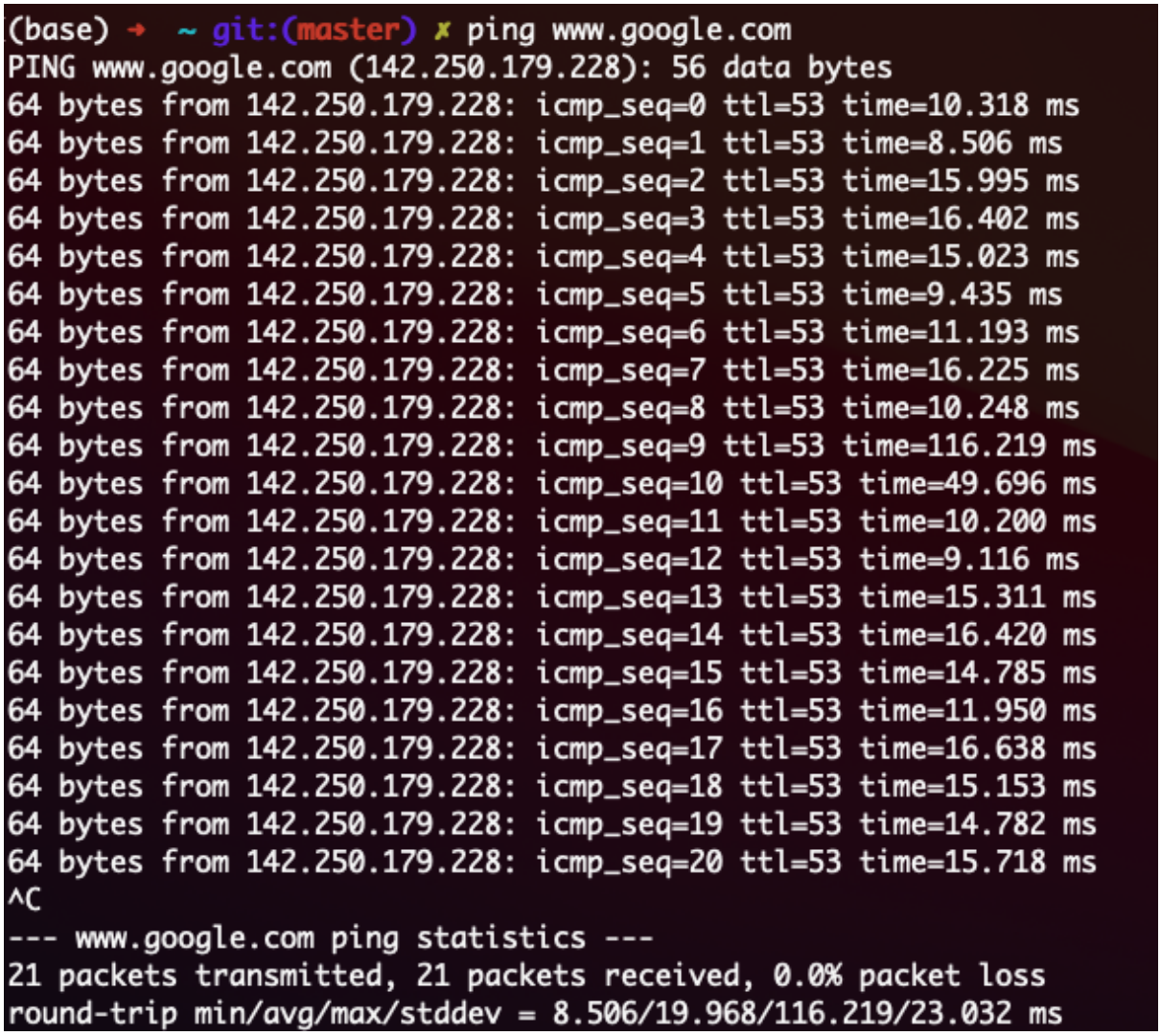
Question 7

Use ping and traceroute to a host such as www.google.com, and explain the relationship between time to live (TTL) to hops. Take a screenshot, if that will be helpful in explaining your answer.

- The ping command operates by sending Internet Control Message Protocol (ICMP) echo request messages to the destination computer and waiting for a response.
- Time to live (TTL) indicates the amount of time or "hops" that a packet is set to exist inside a network before being discarded by a router. So to send a message from the source to the destination, TTL must be strictly smaller or equal to the hops the packets need to travel.
- In the following diagrams, the packets have to travel 12 hops, and the TTL is 53, so we can successfully communicate with the Google server.



traceroute



ping

💡 Comments:

Question 8

About how long did it take you to get through these questions?

- 8.10 am - 9.40 am (Finish question 2 - 5)
- 10 am - 11.30 am (Finish question 6 - 7)

💡 Comments:

Appendix - Chapter 1 Notes

- A network is a system of links that interconnect nodes to move information between nodes. (e.g., Telephone network, Optical networks, and Cellular networks)
    - It is specialised to handle one particular kind of data (e.g., keystrokes, voice or video) and typically connect to special-purpose devices (e.g., terminal, hand receiver, and television sets).
- A computer network(Internet) handles many different types of data and supports a wide range of applications. Different applications have a wide range of different requirements
    - World wide web: (1) Uniform Resource Locator (URL) to identify possible objects that can be viewed from your web browser, (2) Messages exchanges to (2.1) translate server name into IP address, (2.2) set up the TCP connection between the browser and the server, (2.3) sending requests and responses, (2.4) tearing down the TCP connection
        - URL: a protocol used to download the page(HTTP) + machine name that serves the page + folder & file
    - Audio and video streaming: (1) do not download the entire file at one time, (2) more timely transfer of messages from a sender to a receiver, (3) discontinuity is not acceptable, different from www, (4) uni-direction
    - Real-time audio and video (e.g., voice-over-IP Skype): (1) much tighter timing constraints, (2) bi-direction
- A network can be defined recursively from networks(either directly or indirectly connected), but at the bottom level, the network is directly connected via a physical medium.
    - Directly connected: Devices are nodes and the physical medium (e.g., coaxial cable or optical fiber) directly connect them are linked. Link types include point-to-point links and multiple-access links.
    - Indirect connectivity (connectivity among a set of cooperating nodes) is designed to overcome the non-scalability of physical links, including

▼ (1) switched network: (1.1) circuit switch establishes a dedicated link between source and destination, employed by the telephone system, (1.2) packet switch(**more efficient**) for internet communication, uses store-and-forward strategy to forward packets, split a message into packets and reassemble at receiving end.

Circuit switching	Packet switching
Guaranteed performance	No guaranteed performance
Fast transfers once circuit is established	queues and queuing delays
	Header overhead per packe
Wastes bandwidth if traffic is bursty	efficient use of bandwidth
Connection setup adds delay	No overhead due to connection setup
Recovery from failure is slow	Resilient - can route around trouble

- and (2) internet where a set of independent networks are interconnected, uses a router to forward messages
  - Currently, operational TCP/IP Internet uses capital I, and a generic internetwork of networks as internet (i)

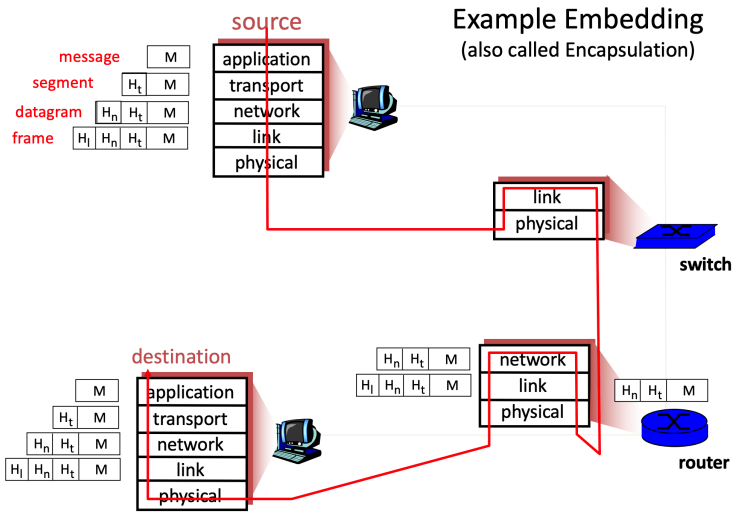
- Networks can be classified by their sizes

Local area network (LAN)	≤ 1 km
Wide area network (WAN)	worldwide
Metropolitan area network (MAN)	~ 10 km
Strong/System area networks (SAN)	in a single room that connects various components of a large computing system

- An address is a byte string that uniquely identifies a node.
- Routing is the process of determining systematically how to forward messages from the source toward the destinations (i.e., unicast, multicast or broadcast)
- Multiplexing is when a system resource is shared among multiple users.
  - Synchronous time-division multiplexing (STDM): divide time into equal-sized quanta, giving each data flow a chance in a round-robin fashion.
  - Frequency-division multiplexing (FDM): transmit each flow over the physical link at different frequencies (e.g., different TV station signals are transmitted over airwave at different frequencies)

    - Statistical multiplexing (very cost-effective for multi-users): (1) physical link shared over time like STDM, (2) data is transmitted on-demand with an upper bound (i.e., packet size) rather than a predetermined time slot, (3) routing path different packet-by-packet
        - Overcome limitations of STDM and FDM: (1) idle host occupy time or frequency, (2) the maximum number of flows known in advance
        - But may have (1) transient overloading when all flows of data arrive at the same time → solved using a buffer but creates queuing delay overhead, (2) persistent overload when buffer overflows → congestion lead to packet drops
- Application programs running on the hosts connected to the network will communicate over logical channels, with varied channel service requirements:
    - Request/reply channel (used by file access such as FTP or NFS, has a client-server pair): (1) guarantee message sent would be received, (2) deduplication on receiving end, (3) protect privacy and data integrity
    - Message stream channel (used by video conferencing and video on demand): (1) no need to guarantee all messages delivered, (2) but need to guarantee same ordering, (3) protect privacy and data integrity, (4) support multicast and broadcast
- Network failure: (1) single or multiple bit errors (e.g., due to data interference with the outside world), (2) packet loss (e.g., due to congestion on the switch buffer or due to switch-forwarding errors), (3) node or link failure
- The protocol defines format, order of messages sent and received, and actions taken on message transmission and receipt.
    - Specifies the header format, which is a small data structure that instructs how receiving peers can handle the message.
- The protocol provides a communication service which (1) it exports locally on the same machine (the service interface), (2) along with a set of rules for peer communication (the peer interface). e.g., request/reply protocol abstracts the request/reply channel.
    - A service interface vertically to allow one higher-level object uses its service on the same machine: define operations that can be performed on the protocol (e.g., HTTP protocol supports hypertext page fetching from a remote server)
    - A peer interface horizontally to allow communication between peers on different machines: define the form and meaning of messages exchanged (e.g., HTTP protocol defines the formats and arguments of a GET command and corresponding responses)

▼ Encapsulation/embedding: Messages will be passed from an application → go through a stack of protocols vertically down, attaching one header per layer → arrive at the hardware-level, transfer the message through direct link → pass on the message vertically upwards, stripping one header per layer



- For each protocol specification, there can be various implementations, the same protocols with different implementations need to communicate.
- Differences between routers and switches:

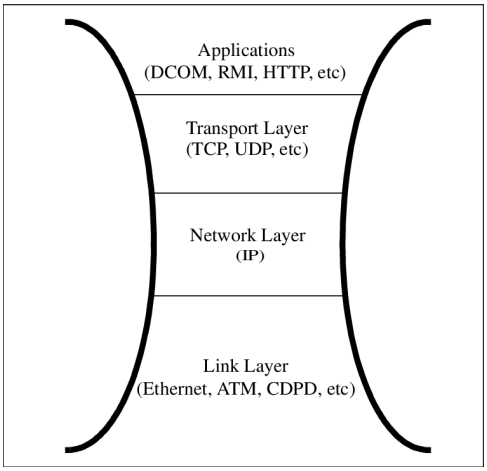
	Router	Switch
Layer	Network layer	Data link layer
Usage	LAN, MAN	LAN
Type	- Adaptive routing - Non-adaptive routing	- Circuit switching - Packet switching
Handle header	Does strip the header, adds a new link layer header, then sends to the destination	Does not strip the header, forward frames based on MAC address

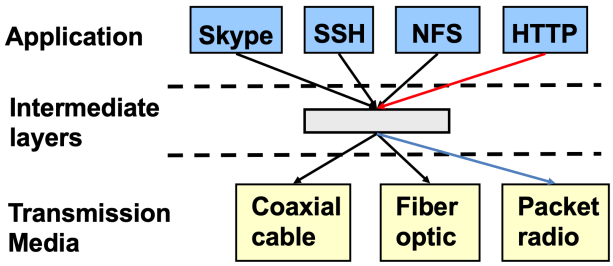
- The TCP/IP model partitioned network functionality into five layers, each having one or more protocols that implement the functionality of that layer.
  - (1) **physical layer** handles the transmission of raw bits over a communications link
  - (2) **data link layer** collects a stream of bits into a frame, the layer is implemented by a combination of hardware (e.g., network adaptors) and software (e.g., device drivers).
    - Protocol examples include Ethernet and wireless protocols (e.g., 802.11 WiFi standards)

    - (3) **network layer** handles routing among nodes within a packet-switched network (i.e., frames and packets are the same)
        - The Internet Protocol (IP) is a set of rules for routing and addressing packets so they can travel across networks and arrive at the destination
    - (4) **transport layer** implements a process-to-process channel, transferring messages
        - The Transmission Control Protocol (TCP) or the User Datagram Protocol (UDP), both provide logical channels to application programs. TCP - reliable byte-stream channel, UDP - unreliable datagram/message delivery channel
    - (5) **application layer** enables an application to access the network
        - HTTP, FTP, Telnet for remote login, Simple Mail Transfer Protocol (SMTP)
- OSI reference model has two additional layers between the application and transport layers: (1) presentation layer allows applications to interpret the meaning of data, do encryption & compression etc, (2) session layer does synchronisation, checkpointing and recovery of data exchange.

▼ Hourglass design structure reflects: (1) IP interoperability: defines a common method for exchanging packets among networks, but (2) various transport protocol, each offering a different channel abstraction to applications, (3) various data link protocol, ranging from Ethernet to wireless to the single point-to-point network, (4) allows the internet to adapt rapidly to new user demands and changing technologies because new application only implement once

```
application    (M) -> message
transport      (H_t, M) -> segment
network        (H_n, H_t, M) -> datagram
link           (H_l, H_n, H_t, M) -> frame
physical
```





- Network APIs (e.g., socket interface) can be used by programs to invoke protocol services for communicating over computer networks. Operating systems provides network API to connect to their networking subsystem.
- Placing network functionality with end-to-end principle (some application requirements can only be correctly implemented end-to-end, such as reliability and security):
  - (1) Only-if-sufficient (not in low-level unless can completely implemented)
  - (2) only-if-necessary (not in the network but in host unless necessary)
  - (3) only-if-useful (low-level gives performance enhancement but ensure not a burden for applications that do not need this functionality)
- The bandwidth(data rate) of a network (e.g., 10 million bits/second = 10 Mbps) is the number of bits that can be transmitted over the network in a time period.
  - Bandwidth is the maximum data rate, throughput is the actual measurement of the data rate at a time
- The latency(delay) of a network is the time taken for a message to travel from source to destination. $\text{Latency} = \text{Propagation delay} + \text{Transmit} + \text{Queuing Delay}$

    - Round-trip time (RTT) of a network is the time taken for a loop trip (from one to another and then back).
- $\text{Bandwidth} \times \text{Latency}$ = maximum number of bits in transit. If we use RTT, then $\text{Bandwidth} \times \text{RTT}$ = the amount of data a sender can send before the receiver starts receiving data.