



Review for Exam I

CIS 1.0 review for exam I, by Yuqing
Tang

Hardware and Software

- Hardware – Physical components of computer
 - E.g. CPU, RAM, keyboard, monitors, printers, speakers, etc.
- Software – Programs (series of computer instructions) that execute on computer, and the associated data
 - E.g. Microsoft windows, Internet Explorer, Netscape, Microsoft Word, etc.

CPU

- CPU is the brain of the computer, responsible for controlling the **internal workings** of the machine.
 - CPU is made of circuitry
 - E.g. A 1.8GHz CPU can execute approximately 1.8 billion simple instructions in a second
- Two tasks:
 - Fetching program instructions from memory
 - Executing the fetched instructions
- CPU is also called processor.

Memory

- **Memory** is the part of a computer that stores programs and data.
- Digital computer stores and processes information as **binary digits**, or **bits**.
- **Cache** is the memory which is built into the CPU chip; it utilizes high-speed-circuitry to provide extremely fast access to data.
- **RAM** (short for Random Access Memory) is the memory which is packaged on separate chips, communicates with CPU using lower-speed circuitry.
- **Main memory** is composed of Cache and RAM.
- Main memory is **volatile**, meaning that it requires a constant flow of electricity to maintain its stored values. When the computer is turned off the values stored in the main memory will be lost.
- **Secondary memory**
 - A **hard disk** is a metal platter that stores bits as magnetized and interprets them as bits. A hard disk is capable of permanently storing vast amounts of information (usually measured in gigabytes), which can be transferred into main memory when needed.
 - Other memory
 - Floppy disk
 - CD, etc.

Input/Output devices

- **Input devices** allow the computer to receive data and instructions from an external source, such as a person entering commands at a keyboard.
 - E.g. Keyboards, mice, track pads, microphones, scanners, etc.
- **Output devices** allow the computer to display or broadcast its result.
 - E.g. monitors, printers, speakers.

Software

- **A software program** is a collection of instructions for computer to carry out in order to complete some task.
- **Application software** is the software program which is designed to carry out tasks within a particular application area, such as word processing, graphical design, or Web access, etc.
- **Systems software** is the software program which is designed to manage the resources and behavior of the computer itself.
- **Operating system** is a systems software, a collection of programs, which controls how the CPU, memory, and I/O devices work together to execute programs.

Memory organization

- **Bit** – Units of data that correspond to one of the two potential values: **0** and **1**.
 - Byte – a collection of 8 bits.
 - KB = 1024 bytes
 - MB = 1024 KBs
 - GB = 1024 MBs
- A **file** is a document that stores information, such as text (e.g., a term paper), and image, sound, or a program (Internet Explorer). The operating system keep track of where individual files are stored so that they can accessed when needed.
- A **directory**, or a **folder**, is a collection of files that are organized together and labeled with a common name.

Data representation

- A **binary** numeral system is a numeral system in which all values are represented using only **two binary digits**, **0** and **1**; these digits are called **bits**.
- **ASCII** is the standard code for representing characters; it maps each character to a specific 8-bit pattern.
- A document that contains only plain text (such as notepad file, html file) is called **ASCII** file or a **text** file.
- Size of ASCII file = number of characters stored in the file
- **Unicode** is a 16-bit encoding system capable of supporting most foreign-language character sets.

Binary numbers to decimal numbers

$$13_{10} = 1 \times 10^1 + 3 \times 10^0$$

$$1101_2 = 13_{10}$$

$2^0 = 1\text{s place}$
 $2^1 = 2\text{s place}$
 $2^2 = 4\text{s place}$
 $2^3 = 8\text{s place}$

$$\begin{aligned} 1101_2 &= 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \\ &= (8 + 4 + 0 + 1)_{10} \\ &= 13_{10} \end{aligned}$$

- Multiply each bit (either 0 or 1) with the corresponding power of 2 and then sum the results.
- The corresponding power of the right most bit is 0, then 1, 2, 3,...

An overview

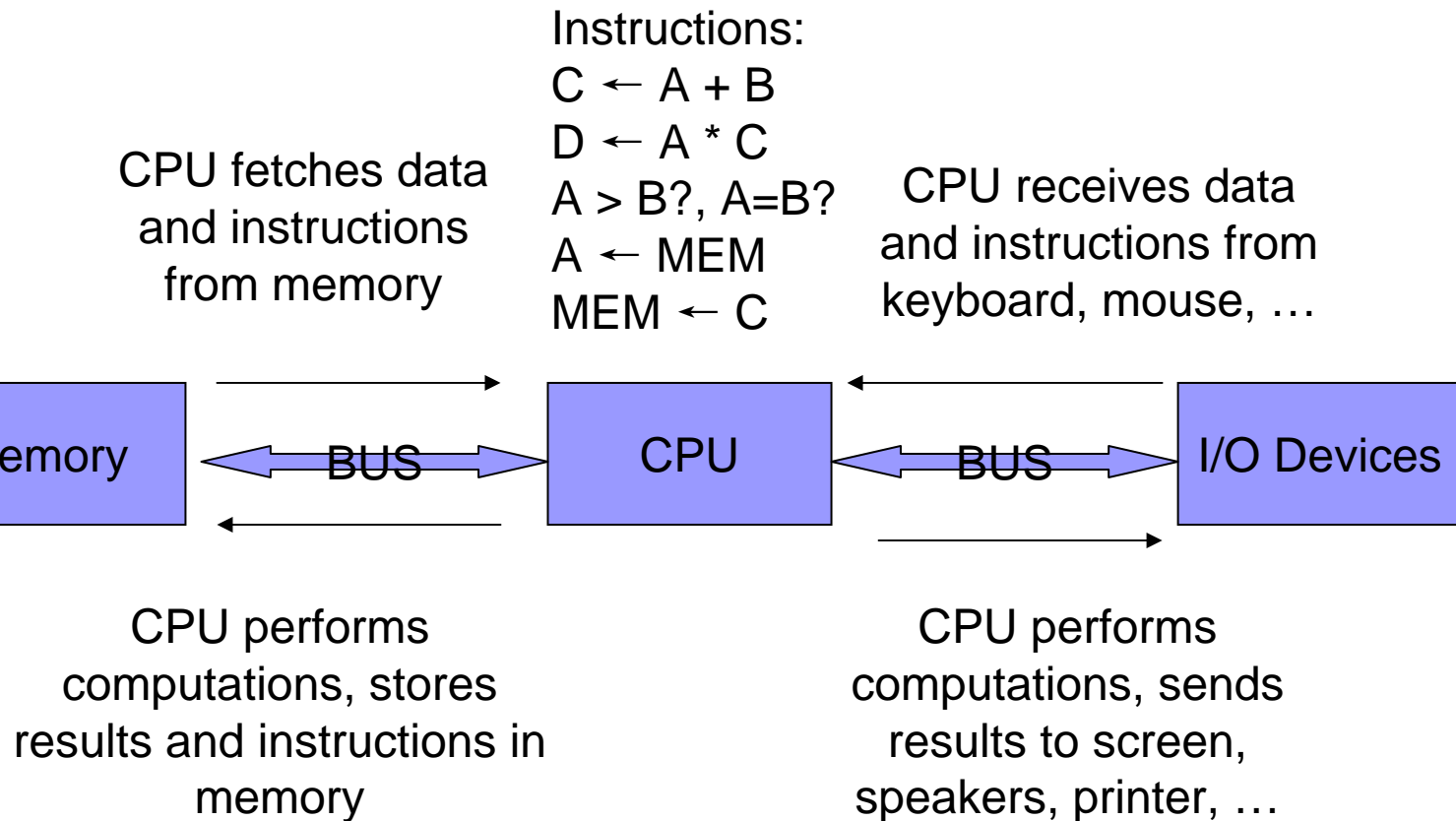
■ Hardware

- Central Processing Unit (CPU, or simply called processor)
- Memory
 - Internal memory (volatile)
 - CPU registers
 - CACHE
 - RAM (Random Access Memory): computer's main memory (or primary memory)
 - External memory (permanent), also called secondary memory
 - Hard drive
 - Floppy
 - Diskette
- Input / Output devices: keyboard, mouse, monitor, printer, etc.

■ Software

- Program
- Operating system
 - Manage the CPU and RAM allocation
 - File system: manage the secondary memory, directory, file
 - I/O devices management
 - Human-machine interface: GUI (Graphical User Interface)
- Application program: A computer program run on a computer directly targeting a task that the user wishes to perform

The von Neumann Architecture



LAN, WAN and Internetworking

- **LANs** (short for Local Area Network) are used to link computers over short distances, such as within the same room or building.
 - Ethernet is the most popular technology to build LANs.
- **WANs** (short for Wide Area Network) are used to connect computers over long distances, so it must include built-in controls for routing messages and adapting to the failures that will inevitably occur.
 - Internet as a whole is an example of WAN.
- **Internetworking** involves connecting two or more distinct computer networks together into an **internetwork** (often shortened to **internet**), using devices called routers to connect them together, to allow traffic to flow back and forth between them.
 - Historically, *Internet* and *internet* have had different meanings, with *internet* being a contraction of *internetwork* or *internetworking* and *Internet* referring to the worldwide network.

History of the Internet

■ ARPNET

- **Pre-birth:** J.C.R Licklider (MIT), 1960s, the “Galactic Network” idea: share computers (expensive), share and access information
- **Name obtained:** Larry Roberts’ team, 1967, finalized ARPANET plan (ARP – Advanced Research Project Agency, a U.S. Department of defense agency)
- **Became reality:** 1969, linking 4 computers at UCLA, UCSB, SRI (Stanford Research Institute), University of Utah
- Growth
 - 23 computers in 1971
 - 100 computers by 1980
 - More than 1,000 computers by 1984

■ Internet

- NSF (National Science Foundation) became involved in ARPNET in 1984 and NSFNET was created, and later becomes the backbone of Internet. **Term Internet** was coined.
- In mid 1980s, the NSFNET became open to commercial interests
- In 1991, CA*net (Canadian) and CERN (European) were connected to the Internet backbone.
- Internet Society (ISOC) – Some nonprofit organizations
 - IETF (Internet Engineering Task Force)
 - IAB (Internet Architecture Board)
 - IESG (Internet Engineering Steering Group)
 - IRTF (Internet Research Task Force)

Packet switching

- It is the central idea of ARPANET architecture.
- In **packet switching**, messages to be sent over the network are first broken into small pieces known as **packets**, and these packets are sent independently to their final destination.

Internet Protocols: TCP/IP

- **Communication protocols** are sets of rules that describe how communication takes place.
- **IP addresses** are unique identifiers assigned to the computers (and devices) on the Internet. An IP address is a number, usually written as a dotted sequence such as “146.245.201.20”.
- The manner in which messages are sent and received over the Internet is defined by a pair of protocols called the **Transmission Control Protocol (TCP)** and **Internet Protocol (IP)**.
- **TCP** controls the method by which messages are broken down into packets and then reassembled when they reach their final destination.
- **IP** is concerned with labeling the packets (with IP addresses) for delivery and controlling the packets' paths (routing) from sender to recipient.
- In internetworking and computer network engineering, **Request for Comments (RFC)** documents are a series of memoranda encompassing new research, innovations, and methodologies applicable to Internet technologies.

TCP/IP software and Routing

- When a message is sent over the Internet, **TCP/IP software** uses the rules of TCP to break the message into packets and label the packets according to their sequence (e.g. packet 2 of 5).
- Then **TCP/IP software** follows the rules of IP to label these packets with routing information, including IP addresses of the source and destination computers. The labeled packets are called **IP datagram**.
- Once labeled, packets are sent independently.
- Special-purpose machines, called **routers**, receive the packets, access the routing information, and pass the packets on (possibly via the other routers) toward their destination. Routers use various types of information, including statistics on the network traffic pattern, to determine the best path for each packet to follow.
- When the packets arrive at the destination, **TCP software** running on the recipient's computer then reassembles the packets in the correct sequence to recreate the original message.

Domain Names and Domain-Name System (DNS)

- **Domain name** is the name assigned to each individual machine; it can be used in place of the machine's IP address.
- Domain names are hierarchical in nature
 - The leftmost part specifies the name of the machine
 - Subsequent parts indicate the organization to which the computer belongs to.
 - The right-most part is known as the **top-level domain** and identifies the type of organization with which the computer is associated.
 - E.g www.brooklyn.cuny.edu
- **Internet Corporation for Assigned Names and Numbers (ICANN)**, a nonprofit coalition of businesses, academic institutions, and individuals, accredits companies, known as **domain-name registrars**, which sell domain-name rights directly to consumers.
- **Domain-name System (DNS)** is a system where machine is assigned a name.
- **Domain-name servers** are used to store mappings between domain name and their corresponding IP addresses.

Email, Mailing lists, Email Viruses

- **Electronic mail**, abbreviated **e-mail** or **e-Mail** or **email**, is a method of composing, sending, storing, and receiving messages over electronic communication systems.
 - The Internet e-mail system is based on the Simple Mail Transfer Protocol (SMTP).
 - A modern Internet e-mail address (using SMTP or Usenet) is a string of the form *jsmith@example.com*. The part before the @ sign is the **local-part** of the address, often the username of the recipient, and the part after the @ sign is a domain name which can be looked up in the Domain Name System to find the Mail transfer agent accepting e-mail for that address.
- A **mailing list** is a collection of names and addresses used by an individual or an organization to send material to multiple recipients.

Web, URL, HTML

- **WEB:** An authorized individual on one computer can have access to files stored on another computer.
- Berners-Lee designed for the Web relied on two different types of software running on Internet-enabled computers.
 - **Web server** – a computer software that stores documents and serves them to other computers.
 - **Web browser** – a computer software that allows users to request and view the documents stored on the servers.
- **HTML (HyperText Markup Language):** Authors define the content of Web pages using HTML tags. The Web browser read HTML tags and render pages accordingly.
- **URL (Uniform Resource Locator):** A **Uniform Resource Locator** (URL) is a string of [characters](#) conforming to a standardized format, which refers to a [resource](#) on the [Internet](#) (such as a document or an image) by its location.
 - *Generic syntax: scheme://authority/path?query#fragment*
 - E.g. <http://www.cs.gc.cuny.edu/~tang/teaching/cis10/cis10.html>
- **HTTP (HyperText Transfer Protocol)** determines how messages are exchanged between browsers and servers using TCP/IP.
 - When the user clicks on a link in the browser, the browser identifies the Web server and sends a request for that page according to URL.
 - The server locates the specified page in its directories, and sends the page back to the browser for display using HTTP.
- A **hyperlink** is an element on a web page that connects to the page to another a webpage.
- Text that contains embedded hyperlinks is referred to as **hypertext**.

HTML tags

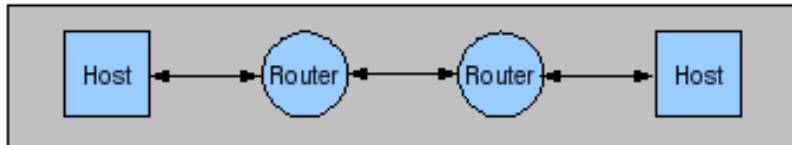
- `<html> ...</html>` defines a webpage.
- `<head>...</head>` defines a head of a webpage. The head element can contain information about the document. The browser does not display the "head information" to the user except `<title>...</title>`. The following tags can be in the head section: `<base>`, `<link>`, `<meta>`, `<script>`, `<style>`, and `<title>`.
- The title of a webpage defined by `<title> ...</title>` must be defined inside `<head>...</head>`, which is not part of the display of the webpage; it is displayed on the browser's title bar.
- `<body>...</body>` contains all the displayable elements of a webpage; it is inside `<html>...</html>`.
- `<H1>...</H1>`, ... define the headers inside a webpage.

<http://web.cs.gc.cuny.edu/~tang/teachings/cis10/html-lab-1.html>

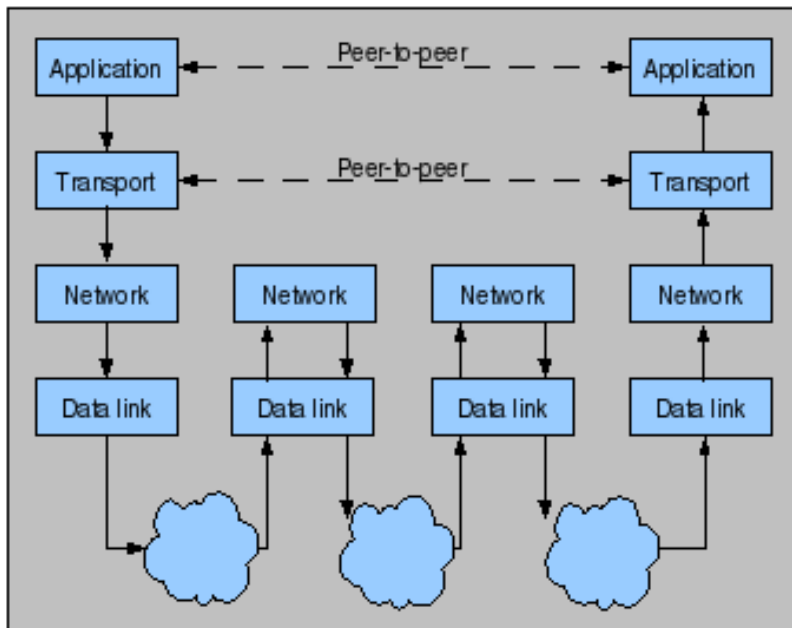
<http://web.cs.gc.cuny.edu/~tang/teachings/cis10/html-lab-2.html>

Layers in the Internet protocols

Network connections



Stack connections



- **Application:** [DNS](#), [TLS/SSL](#), [FTP](#), [HTTP](#), [SMTP](#), ...
- **Transportation:** [TCP](#), [UDP](#), ...
- **Network:** [IP](#) ([IPv4](#), [IPv6](#))
- [ARP](#) and [RARP](#) operate underneath IP but above the link layer so they belong somewhere in between.
- **Link:** [Ethernet](#), [Wi-Fi](#), [PPP](#), [FDDI](#), [ATM](#), [Frame Relay](#), ...

Machine language and high level language

- **Machine language** is a collection of different patterns of binary bits that correspond to different computer machine instructions.
- **High-level programming languages** are the languages that provide high-level abstractions and constructs (which correspond to machine level operations) for solving problems using computers.

Algorithm and Computer Program

- An **algorithm** is a series of steps that can be followed to solve the problem.
- A **program** is a series of instructions that specify exactly what the computer is supposed to do. Instructions can be
 - in high level programming languages
 - Or in machine language (binary coded instructions)
- **Stored program scheme:** The programs in machine language are stored in computer memory, the CPU fetches and executes the instructions in the program one by one.

Compiler and Interpreter

- An **interpreter** reads the statements in high-level language one at a time, immediately translating and executing each state before processing the next statement.
- A **compiler** translates the entire high-level language program into its equivalent machine-language instructions.

Software life-cycle

- Analyze the problem
- Specify the problem strictly
- Devise algorithm (and software architecture: how to organize the data, and organize the program.)
- Coding: implement the algorithm in some computer language(s)
- Testing
- (Documentation, and software training and support)
- Maintenance

Search engines, Subject Directories, meta-search engines

- A **search engine** is a searchable database of Internet files collected by a computer program (called a wanderer, crawler, robot, worm, spider). Indexing is created from the collected files, e.g., title, full text, size, URL, etc. There is no selection criteria for the collection of files, though evaluation can be applied to ranking schemes that return the results of a query.
- A **subject directory** is a service that offers a collection of links to Internet resources submitted by site creators or evaluators and organized into subject categories. Directory services use selection criteria for choosing links to include, though the selectivity varies among services. Most directories are searchable.
- **Meta search engines** simultaneously search multiple search engines. They are also sometimes referred to as **parallel search engines, multithreaded search engines, or mega search engines**.

Other concepts in Internet Searching

- Deep web
- Advanced searches
- **Spider**: Program that traverses the Web from link to link, identifying and reading pages.
- **Index**: Database containing a copy of each Web page gathered by the spider
- Query logic: AND and OR
- Field search
 - TITLE:
 - URL:

History of computer science

■ People:

- ☐ Pascal
- ☐ Jacquard
- ☐ Babbage
- ☐ Ada Lovelace
- ☐ John von Neumann
- ☐ Alan Turing
- ☐ Grace Hopper
- ☐ Hollerith
- ☐ Ritchie and Thompson: UNIX and the C programming language

■ Machines and things

- ☐ Pascaline
- ☐ ENIAC
- ☐ Moore's Law
- ☐ RFC
- ☐ Internet 2: **Internet2** or **UCAID** (University Corporation for Advanced Internet Development) is a [non-profit consortium](#) which develops and deploys advanced [network](#) applications and technologies, mostly for high-speed data transfer.
- ☐ Unix
- ☐ Difference Engine, Analytical Engine