HOW Computers Work

CM1030

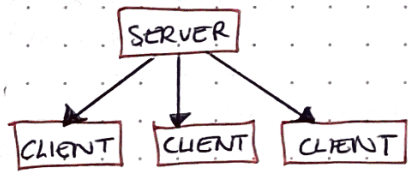
ABSTRACTION

An abstraction is where you highlight key information without going into all the detail. Or taking something to a higher level.

A virtual machine is an abstraction of a computer.

How does a webpage work?

- A server is where the information for the webpage is stored.
- A client accesses that information either via us or via software eg browsers.



http://blahblah.com/computing/search=computer

- http → the language
- blahblah.com/computing → the webpage
- search=computer → the query

These days it is very rare for a webpage to be static, it is much more likely that the pages are created as we go. Usually off the back of a database.

Websites are typically modular with lots of different components working together eg video, pictures, flash plugins etc.

IP - internet protocol

HTTP - hyper text transfer protocol

COOKIES - small amounts of data that keep track of the state we are in eg are we logged in? What is our browser history?

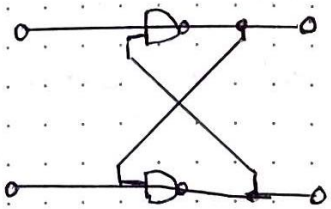
API - application programming interface. Kind of like a data access point that you can plug into. eg embedded google maps

MODULARITY - having lots of different modules that you can plug into each other to build up the application eg printer drivers that can be used by the OS.

Data Representation

At its core the language a computer uses is binary. 1/0, on/off, true/false with a single digit being known as a BIT

GATE can be constructed out of gears, relays or optical devices. In a modern computer they are usually small electronic circuits where 1 and 0 is represented as voltage levels. A very important gate is a flip flop it accepts a 1 and 0 and will store that information until another stimuli is introduced.



= NOT

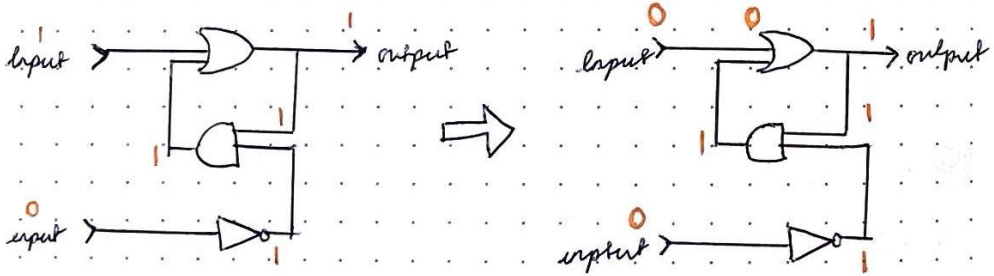
= AND

= OR

= XOR

looking at the flipflop gate, you pass a current through the upper gate and it will produce an output of 1.

The bottom gate at 0 will change to a 1 through the not gate. Which in turn means the and gate is producing a 1. This means that even if the upper gate is turned off the output remains a 1.



Temporarily changing the bottom input to 1 will force the output to be 0 until changed.

Circuit gates are important in CS because

- It helps visualise digital circuit design and how they're constructed
- Provides an abstraction for data storage
- Provides an abstraction of the underlying circuitry

MAIN MEMORY

A computer contains a large number of flip flops capable of storing single bits within main memory.

We group these bits into 8, otherwise known as a byte and contains them within a cell. The left most bit of a byte is the 'high order' end (or the most significant) and the right most end is the 'low order' end (or least significant)

Each cell has a unique name known as its address

Because the cells can be accessed independently the main memory is usually called RAM (random access memory)

Kilobyte $= 2^{10} = 1,024$ bytes $= 1KB$

Megabyte $= 2^{20} = 1,048,576$ bytes $= 1MB$

RGB colour is made up of 3 byte strings or 24 bit per pixel (24 pixels x 3 bytes = 72 bytes)

MASS STORAGE

Due to the limited size and volatility of main memory most computers will have a mass storage, eg flash drives, cds, SSD and more. The advantages are low cost to storage size, less volatility and the ability to remove and archive them.

MAGNETIC - eg harddrive, where a thin disc with a magnetic coating is used to store information. The disc circumference has tracks that are split into sectors. And the read-write head positions itself over the sections

OPTICAL - eg a CD, information is recorded on them by creating variations in the reflection which is read by a laser as the disc spins. CDs perform best when reading long continuous strings of data such as music due to the spiral design of the disc.

FLASH DRIVES - works by capturing electrons in small chambers of silicon dioxide, these are stored for many years without power and so perfect for stable non-volatile data storage. Though with repeated use the silicon loses its ability to capture electrons as effectively, in devices such as SSD that are meant to stay replace harddrives the issue of wear is resolved by rotating the data.

REPRESENTING INFO

TEXT - this is coded into bits typically the most common words have shorter bit patterns when they go through compression, Standard letters use the ASCII (American Standard Code for Information Interchange). It uses 8 bit strings to encode the alphabet, punctuation and numbers.

There are extensions to ASCII that cover Ü, ä etc but it became clear that there weren't enough variations to cover languages such as chinese, japanese etc. UTF-8 uses 24 or 32 bit patterns to represent these symbols.

NUMBERS - It requires more data to store numbers using ASCII so binary digits are used.

IMAGES - many images are made up of a collection of dots, each RGB of that dot is then encoded in a bit map. This is the most common method, Red, Blue Green and Luminescence (essentially the sum of the RGB components).

This pixel bit map isn't great for scaling images (eg pixelated) and so images are sometimes described as a series of geometric shapes that are scaled in relation to each other, This is common in CAD

SOUND - is captured by recording the amplitude up to a rate of 44,100 times per second, each sample is 16 bits long.

MIDI is an alternative which looks more at what the sound should be and stores the average rather than constantly sampling.

COMPRESSION

Compression is either lossy (lose information) or loseless (information isn't lost)

IMAGE - There are a couple of different ways to compress images, GIF images there is a palette of 256 colours, this means each pixel can be encoded with 1 byte.

JPEG has several different methods of compressing pictures, the human eye is more sensitive to brightness than changes in colour so it first averages chrominance over 2x2 pixel squares. Then it takes 8x8 blocks and using discrete cosine transformation relates each pixel to it's neighbour rather than directly storing the information. It is then compressed using run length encoding

AUDIO AND VIDEO - the most common method of compression of video and audio is MPEG (motion picture experts group).

Video compression must cover a wide range of uses, video conferencing where speed is important vs HDTV where quality is important. The encoding works by encoding key frames fully and then referencing between frames to those fully encoded. Discrete cosine transformation is then used to further compress the images.

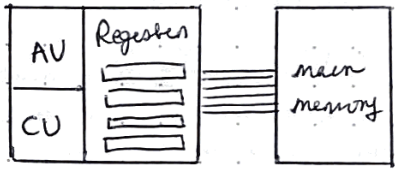
Mp3 is a common method of compressing audio and works by removing information that the human ear can't hear.

Computer Architecture

The part of the computer that controls the manipulation of data is the CPU or central processing unit.

A CPU consists of 3 parts

- Arithmetic/logic unit
- Control Unit
- Registers



THE BUS - a series of wires between the CPU and main memory that communicates bit patterns

THE REGISTERS - can either be general or special purpose. General registers hold information for the input and outputs that the arithmetic unit is currently working on.

CONTROL UNIT - transfers the data from memory into the registers and tells the logical unit where the information is and where the result should be stored

STEP1 - Collect value1 from RAM store in register
STEP2 - Collect 2nd value and store in 2nd register
STEP3 - Activate addition logic on registers
STEP4 - Store the result in 3rd register
STEP5 - Stop
~~STE~~

Previously the arithmetic component the steps of the program were inflexible, until it was discovered that programs could be stored in main memory and the control unit could go and collect the information for the instructions from there, STORED-PROGRAM CONCEPT.

MACHINE LANGUAGE

Instructions are encoded as bit patterns and known as machine language. The types of instructions machines use are limited, adding additional functionality beyond the basics adds convenience but not additional capabilities

RISC - Reduced instruction set, small number of instructions. Fast efficient and cheaper

CISC - Complex instruction set, complex set of instructions, deals with complex software

DATA TRANSFER instructions, perform LOADS of data into registers, it is worth noting information is copied rather than transferred. As well as STORE instructions for results

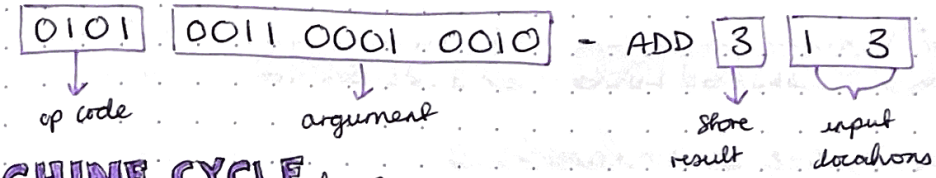
I/O (input/output) are also considered under the data transfer header for communicating with peripherals

ARITHETIC/LOGIC instructions perform the operations eg OR, AND or arithmetic. It can also perform shift or rotate operators for moving information in the registers

CONTROL instructions direct the next step is to be executed rather than controlling data. Can implement jumps to skips the next instruction either conditional (if this then that) or unconditional (go to next step)

example 1

156C register 5 = register 6C LOAD
166D register 6 = register 6D LOAD
3056 Add register 5 and 6 store in 0



MACHINE CYCLE has 3 steps

1. FETCH
2. DECODE
3. EXECUTE

FETCH - CPU requests the data from main memory for the instruction stored at the location indicated by the program counter. CPU then places the instruction in the instruction register and then increments the program counter by 1.

DECODE - The CPU then decodes it into its component parts

EXECUTE - The programs then performs the instruction by LOADING data or engaging the logical unit to perform the calculation.

THEN REPEAT

PERIPHERALS

Communication between peripherals such as keyboard, mice etc is typically handled by a controller. Sometimes these controllers have their own basic memory and CPU. USB and firewire are examples of controllers that can handle multiple devices.

The controller connects to the bus and so has access to the CPU and main memory, and so the CPU can send instructions to the controller in much the same way that it would to main memory.

The controller can also act out instructions on behalf of the CPU that doubles the processing power on these moments the bus is available.

Operating Systems

Graphical User Interface is how the system communicates with its users. Objects can be manipulated, program files and folders can all be displayed as pictures and inputs such as keyboards and mice can move them around.
The command prompt is another shell alternative to the main windows OS

WINDOW MANAGER is an important part of the GUI. It allocates blocks of the screen called windows, and the program associated with each window. It identifies where the mouse is when it is clicked and the program it is hovering over

Kernel is the central part of the operating system it contains the very basic functions of the operating system

FILE MANAGER - is the part of the kernel that helps manage mass storage
- → where the file is stored
- → what storage is available for use
- → which users are allowed to access what

The file manager allows users to group files into folders (directories) and create subdirectories called file paths.

DEVICE DRIVERS - is another component of the kernel this handles communication with the device drivers and sometimes the peripherals themselves

MEMORY MANAGER - coordinates the use of main memory and important functions in multitasking machines. It must assign each running program its space and ensure it can't access memory it's not supposed to
Paging is where the system needs more memory than you can actually have and so it creates the illusion of space by shuffling data between main storage and memory. This additional space is called virtual memory

Bootstrapping happens whenever the computer is turned on. The CPU program counter starts with a predetermined location wherever it is restarted, this is in main memory. But as main memory sucks at holding onto information this is pointed to a special part of main memory which tells it to load the operating system from mass storage.
This main memory is a form of flash storage.

The OS is stored in mass storage because it wouldn't be efficient to store it in RAM due to modern OS ~~stay~~ size. It is also needed to be updated regularly and this isn't straightforward in RAM.

Priviledged Instructions help interrupt malware from getting control over other programs data. When an attempt is made to run a privileged instruction when not in privileged mode an interrupt command is given. But the process is complex and open to exploitation

Admin User one way of having a secure system is to have users and accounts where one user cannot access the files of another on the same machine. The admin is the only one with overall access, the user identifies themself on logon.
The administrator can use auditing software to identify attempted access and any unusual behaviours

Co-ordinating Processes

A process is the act of running the program. You can think of the program as a piece of sheet music on a shelf and processing being the act of playing it.
While a musician plays one song after another, a program will play several at the same time.
The program being run is kept track of by the scheduler.

SCHEDULER - (part of the systems kernel) adds new processes and removes them when they are done. It keeps track of them in a block of information in main memory called the PROCESS TABLE.
The process table contains such information such as the processes priority, the memory location where the process is located. As well as if the process is READY to be run or WAITING for something to happen.

MULTIPROGRAMING - is where each program is divided into time slices the CPU runs the program for the allocated time slice and then is stopped by an interrupt signal and the next process time slice is run. The act of switching is known as process/context switch.
When the interrupt handler is run it:
- → Records where you are with the current task
- → Interrupt handler hands control to the dispatcher

→ dispatcher then starts the next priority process

It is worth noting that multiprocessing improves the computers efficiency.

HANDLING COMPETITION among processes is a key part of the operating system, how it allocates the resources.

→ FILE MANAGER, manages access and storage to files
→ MEMORY MANAGER, manages access to memory space
→ SCHEDULER, allocates space on the process table
→ DISPATCHER, allocates time slices

Semaphores if a resource eg a printer is available the OS needs to decide if it can give access to a process request to use it or identify if it is being used by something and deny the request.

If the printer is in use it might deny the request and mark it as waiting for the printer to be available. It could do this by assigning a 1 for in use and a 0 for available.

However the interrupt of processes mean it's possible for the system to check the status and go to print then be interrupt and assign the printer to another task before restarting the first process.

This can be counteracted by not allowing the process to be interrupted when changing the flags by disabling the interrupt signal and then enabling it. Or by testing and setting the flag in one machine language instruction. The process of setting flags is known as semaphore.

Deadlock is where processes are blocked by each other from proceeding.

STEP1 – locks the audio file then wants the audio card
STEP2 – locks the audio card then wants the audio file

Neither process can continue and are deadlocked. A deadlock can only occur if all of these conditions are satisfied.

1. Competition for resources that can't be shared
2. Resources are requested but by bit
3. Once a resource is allocated it can't be given up

Most detection is at the third point and is known as DEADLOCK DETECTION and it can be resolved by forcibly retrieving the resources.

Those that look at the 2nd event are known as DEADLOCK AVOIDANCE, this can be done by making a process use a resource in one go or make unshareable resources shareable eg SPOOLING by storing the instruction to the resource in memory then executing it when is available

Network Classes

→ PAN (PERSONAL AREA NETWORK) eg wireless mouse and a PC
→ LAN (LOCAL AREA NETWORK) eg office computer network
→ MAN (METROPOLITAN AREA NETWORK) eg local community
→ WAN (WIDE AREA NETWORK) eg opposite sides of the world
  → OPEN NETWORK aka the INTERNET
  → CLOSED NETWORK / PROPRIETARY aka restricted access
    ↳ A bus network, star network, peer to peer

PROTOCOLS

Are standards that allow devices to communicate and understand each other, without standard protocols vendors would have to construct their own and then make sure the devices can communicate with other brands.

CSMA/CD (Career sense multiple access, collision detection) is for use on a bus network. All machines receive all the messages but only keep those addressed to itself. If a machine tries to submit a message the same time as another they both stop transmitting and wait a random amount of time and then try again.
This doesn't happen on a wireless network as

1. Computers only have to be in range of the router not each other
2. As a result they won't know when each machine is transmitting

Wireless therefore uses CSMA/CA (Career sense, multiple access, collision avoidance) which works on avoidance rather than detection. When collisions do occur, messages can be resubmitted. The main method of avoidance is to give priority to messages that have been waiting longest. When the channel is silent it transmits, when it is busy it waits a random period before trying again.

REPEATER the simplest method of connecting 2 bus networks. It simply passes messages between the two networks.

BRIDGE similar to a repeater but only passes messages between networks if they are addressed to a computer on that network (more efficient)

SWITCH effectively same as a bridge but for more than 2 networks

ROUTER allows connection between different network types eg WIFI and ethernet. Builds a network of networks (an internet)

Distributed Systems is one where different devices perform different processes for a piece of software. There are several different types of such systems.

CLUSTER COMPUTING - many different computers work together to provide the computing power of that of a much larger machine
GRID COMPUTING - more loosely defined than a cluster, and allows people to opt in their PCs computing power when not in use. For example assisting with large calculations on Research
CLOUD COMPUTING - where huge numbers of computers are connected and their computing power is utilised. Amazon rent out such machines, but does raise issues around security

INTERNET

The Internet (capital I) is a collection of networks maintained by ISPs. ISPs can be constructed into tiers

TIER 1 → very few high speed high capacity WAN
The backbone of the internet

TIER 2 → more regional less capacity

TIER 1 and TIER 2 ISPs networks of routers that provide infrastructure

TIER 3 → How the individual connects to the internet ISPS

THE LAST MILE PROBLEM — references the fact that the connection to the internet in most homes is typically via analogue phone lines. It is easier to update larger infrastructure, but that last little bit is proving costly.

IP ADDRESS is effectively the address of each computer on the internet. The ICANN (international corporation for assigned names and numbers) gives batches of IPs to ISPs, the ISPs then assign them to machines.

application
HTTP SMTP
Specific to applications eg HTTP for webpages, SMTP - Email

transport
TCP UDP
Handles transportation of the message

network
IP
Handles navigation between networks

link
ETHERNET WIFI
Handles transmission over individual networks

Common Protocols

SMTP → for sending email (simple mail transfer protocol). The sending server contacts the receiver using DNS to convert the email to the IP. The protocol dictates that it must identify itself to the sending server. SMTP was only intended for coding messages in ASCII so other protocols are used for data in different formats.

POP3 → (post office protocol 3) and IMAP (internet mail access protocol) are used for accessing mail that has already arrived on a computer. POP3 facilitates the download of the mail onto the users computer to be manipulated. IMAP covers the manipulation on the server

VOIP (voice over internet) There exists several forms of VOIP, mainly it uses a P2P network to connect the machines using existing speaker and microphone hardware aka SKYPE.

The second uses analogue phone adapters to connect to the existing telephone infrastructure.

ARTIFICIAL INTELLIGENCE

Artificial intelligence is the field of computer science seeking to build machines that do not require human intervention for complex tasks.

INTELLIGENT AGENTS - are devices that respond to stimuli from their environment eg. an autonomous airplane or a character in a video game

They will have sensors: microphones, cameras and actuators for affecting their environment: wheels, legs, grippers.

Most AI Research focuses on building AI that behaves intelligently by responding with its actuators to information from its sensors and can be categorised into several classes of intelligence.

SIMPLE - the simplest response is a reflex action, a predetermined response to inputted information

GOAL DIRECTED the result of forming a plan of action based on the current inputs eg walking to the end of a hall, playing chess. In some instances the machine can learn from trial and error and positive/negative reinforcement

TURING TEST used to serve as the benchmark for measuring progress in AI.

His proposal was the test should be if a human communicated with the device would they be able to determine if they were talking to a machine or not.

This is no longer considered relevant as it's fairly easy to mimic human response with basic rules

PERCEPTION - IMAGES identifying a grid of 8 digits would rely on the software matching bit patterns in the image to a prerecorded template and trying to compare the images. This is easily falliable especially when it comes to handwriting.

The second method is to match geometric features eg 1 would be a single vertical line, but this is still error prone depending on the quality of the images.

Trying to identify general pictures will be even more complex, it will involve edge processing (trying to view it as a line image), region matching (identifying bits of the image that are related), smoothing (to remove flaws)

LANGUAGE PROCESSING provides another tricky problem, while translating programming language to machine language happened fairly quickly. Spoken language is different as it relys on additional knowledge.

eg CINDERELLA HAD A BALL
→ Cinderella had a ball like a football
→ Cinderella had fun
→ Cinderella threw a party.

SYNTACTIC ANALYSIS → identifies the grammatical role
SEMANTIC → identifies the actions
CONTEXTUAL → understanding what is meant by the sentence

REASONING rather than programming the machine to find all possible outcomes of a a problem it is more efficient to program the machine to construct its own solutions. Each reasoning problems consist of PRODUCTION SYSTEM.

1. A COLLECTION OF STATES, the start state, the desired/goal state each state is a solution that might occur when solving the problem
2. A COLLECTION OF RULES/MOVES how we move from one state to another.
3. A CONTROL SYSTEM solves the problem of how to get from the start state to the goal state.

A state graph is a map of all the different states of the problem and has arrows determining which way you can move through the states. The control system determines the path to take to get to the goal state.

SEARCH TREES is the route the computer builds to get to the goal, it tests each possible path from current state and identifies how to get to the goal.

However sometimes constructing search trees can be time consuming, so the program will select the direction that looks the most promising. To do

thus we use a heuristic, which in some cases is a quantitive value measuring the distance between the current state and the goal. The heuristic shouldn't be onerous to calculate, it's meant to be less intensive than constructing a complete search tree