**Definitions**

1. **Compliers**

Compliers takes a higher-level programming language (like Java or C++) and translates them to a lower level language like machine code so that the CPU can execute.

1. **Interpreters**

Interpreters execute or interprets instructions in their written programming language without translating them. This is usually done on a complied programming language so that the execution can be done in the CPU.

1. **Machine code**

Machine code is a sequence of individual instructions that are executed by CPU. CPU can only execute machine code.

1. **Memory**

Memory is a sequence of locations that store one value of data. Each value has fixed number of bits (say 16 bits for all). A CPU can change and read the value from the memory from time to time, which is done through the location’s address.

**MARIE**

MARIE consists of five registers. Accumulator (AC) holds the general-purpose registers, all arithmetic operations will be performed here. Memory Address Register (MAR) holds the address of the memory that is to read from or write to the memory. Memory Buffer Register (MBR) holds the data read from or write to the memory. Instruction Register (IR) holds the current instruction. Program Counter (PC) holds the address of the next instruction.

Subroutine is a piece of function that can be called often.

**Circuits**

1. **Combinational circuits**

Combinational circuits are circuits whose outputs directly depend on the inputs. Such example can be seen in the implementation of Boolean functions like AND, NOT and OR gates.

Half-adders or adders adds two one-bit numbers. It has two inputs and two outputs - carry which is represented by the AND gate and the result of the adders which is represented by XOR gate. On the other hand, full adders adds three one-bit numbers, which is constructed by combining the half-adders together. It has three inputs and two outputs. One of the inputs is known as carry-in combines together with the ‘result’ from the first half-adder and then creates another result. The carry-out in these full adders is represented by AND gate while the result is represented by the XOR gate. Ripper-carry adder merges multiple full adders together to add up multiple individual bits. For example, adding two 3-bit numbers would require a chain of three full-adders.

Decoder converts a binary number to unary representation. It activates the output that corresponds to the value of the input. It computes the value of the inputs and then output its corresponding value. For example, if the first input is 0 and second input is 1, combining the inputs would result in 10, which is a value of 2 in decimal. Hence, it would output the position 2.

Multiplexers chooses which input data to activate, it selects one inputs out of several inputs. To do this, it has an additional input called selector, that selects which input values to output. If the selector is 1, then it chooses the data 1’s values to output.

1. **Sequential circuits**

Sequential circuits are circuits whose outputs depend on the sequence of inputs. Flip-flop is an example of a sequential circuit. It consists of a set and reset button and stores one bit of data at a time. This data can be reset by toggling the reset button. The NOT gate will then negate the bit stored back to zero.

D flip flop is an improved version of flip-flop. It consists of two inputs and output. The input of data to store, and the input of choosing the data to read or write.

**CPU**

CPU consists of three components - ALU, registers and control unit.

ALU mainly performs the MARIE operations like arithmetic operations. It does operations like addition, subtraction, incrementing by 1 and comparison and uses multiplexers to select the operation to output (through opcodes).

Register file combines all the registers into a single circuit and temporarily store the data and instructions. This storing of data is typically done through flip flops and uses a multiplexer to select the register to read from. In addition to that, it also has the read and write feature that chooses which register to read and write from.

Control unit performs the fetch-decode-execute cycle through control signals. The control signals do function like choosing which register, memory to read and write and also performs operations in ALU. *Control signals is a set of “wires” that switch on and off for every component to implement each RTL step.*

**Fetch-decode-execute**

In the fetch phase, the PC points to the next instruction that is being executed. It copies the address in the PC to MAR. Then, the MAR reads the memory and stores it at MBR. The content of the MBR is then being fetched to the IR and lastly increase the PC by 1. In the decode phase, the control unit looks for the next instruction to be executed (IR) and loads the instruction into MAR. The memory stored at MAR is then written to the MBR. In the execute phase, it depends on the actual instruction. If the instruction is Load X, then the value stored at MBR will be loaded into AC.

**Memory**

There are two types of memory architecture - byte-addressable and word-addressable. In byte-addressable, each memory location represents 8 bits (one byte) while in word-addressable, each memory location represents 16 bits (one word). The number of addressable locations in each type of architecture can be defined as 2^n where n is the number of bits.

**I/O Devices**

Each I/O device have their own set of registers. There are two types of I/O architectures - memory mapped I/O architecture and instructions based architecture. In memory-mapped I/O, the CPU’s memory and the I/O share the memory address space, meaning the I/O registers are mapped into the memory space of the CPU. An example would be MARIE architecture, for each instruction, it is stored in the memory space shared with the CPU. An advantage of such method is that it is relatively simple, as no new instructions are needed to perform in order to access the devices. However, in such scenario, the RAM available for the CPU will be decreased, as the I/O might take up a portion of the memory. On the other hand, in instruction based I/O, the CPU has special instructions to access I/O such that the CPU and the I/O have individual memory address space, this makes the memory performance much efficient, because they do not need to share the spaces. An advantage this is that less logic is needed to decode the I/O addresses because special instructions are used for that.

There are two ways of accessing an I/O device - programmed I/O and interrupt based I/O. In programmed I/O, the program checks the I/O registers at regular intervals, this makes programmers easier to control the frequency of the intervals. Such method is said to increase the CPU’s power usage because it constantly checks the arrival of I/O. On the other hand, in interrupt based I/O, it notifies the CPU for I/O requests and execute interrupt handlers. The I/O device will first activate the interrupt signals by setting a bit in a special register and causes the CPU to execute the interrupt handler code. It calls a special subroutine and returns back to the state as it left before through context switch. This is done by shadow registers, where it switches to a separate register and returns back to the main register after executing the interrupt handler code. Another way would be via programming, where the CPU saves the previous location to memory and returns back to this location.

One of the ways of distinguishing the I/O devices is by assigning an identification number to each device. When an I/O is being performed, it saves the number in a special register. The handler then uses this specific number to jump to interrupt vectors that consists of various subroutines for various device types.

DMA allows CPU to delegate the memory transfer task to a special controller, which could help to reduce the CPU’s work. For example, the hard disk controller could copy data directly into RAM and graphics cards could load data into RAM. This increases the CPU’s productivity.

**BIOS and UEFI**

At start, the power supply gives energy to the motherboard and other components. Then, the clock starts to tick and CPU gets powered to initialize the initial software - BIOS/UEFI.

BIOS provides the initial system configuration and starts up controls. It starts with power on self test where it checks the parts of the CPU. After that, it initializes the video card to show some initial messages on the screen. During this time, heavy work like graphics cannot be run because it requires drivers to execute and BIOS storage space is limited during this time. After initialing the video card, it checks other hardware components and starts them like networking, sound, RAM timing, DVD drive etc. It then searches for the operating system in a bootable drive. If the OS is loaded from a bootable drive, the boot loader will be executed. On the other hand, if the OS is loaded from a hard drive, the Master Boot Record will be executed. This boot loader will first load the OS and the kernel drivers and then lastly the graphical user interface.

The difference between BIOS and UEFI is that BIOS only works up to 2.2terabytes of hard drive while UEFI works up to 9.4zettabytes. BIOS does not allow network access before the OS is loaded while UEFI allows network access before the OS is loaded. In addition to that, BIOS has lower security and authentication, while UEFI provides security and authentication before the OS is loaded.

**Operating system**

OS introduces a level of abstraction between the hardware and the software. This level of abstraction is achieved mainly through virtualizing the CPU and memory. To allow multiple process running continuously without interruptions, the process can be in these three states: ready, running or blocked. In ready state, the process is ready for execution but not executed yet. When the OS schedules all the processes in order, it puts the process into running state. In running state, the process is executed and de-scheduled to return back to ready state. If the process requests for some I/O, the process will be put into the blocked state to prevent time consuming. In the blocked state, the I/O is being executed and returns back to ready state and being scheduled again.

There are two modes in the OS, user and kernel mode. In user mode, only a subset of instructions is allowed to be executed. I/O instruction is not allowed to be performed and certain memory instructions are restricted. While in kernel mode, codes are run without restrictions. The OS runs in kernel mode and when an interrupt happens (system calls), CPU switches from user mode to kernel mode and jump to special subroutine. When the subroutine has completed, the CPU will switch back to user mode and interrupt handler restores back to the position as it left.

Process can be switched from one to another through two methods: cooperative timesharing or preemptive timesharing. In cooperative timesharing, all process must cooperate with the OS and make system calls in regular intervals. However, in such timesharing method, some process might not cooperate with each other and malicious process might cause bug. In preemptive timesharing, timer interrupts are used to generate interrupts in regular intervals and allow the OS to switch to a different process when the interrupt completes. In the meantime, preemptive also allows the OS to kill malicious process as it regains control of the system after a few time intervals.

There are three ways to decide when a process switch to another process. In first-come first served, the processes are ordered according to sequence, creates a poor turnaround time because there is no fairness. In shortest job first, the processes are ordered according to the shortest time first, this creates an average turnaround time. Lastly, in round robin scheduling, the processes are split into time slices. These time slices create a good fairness because each process gets a fair amount of time to run. Due to massive process switching, the OS needs to compromise between fairness and efficiency.

Memory virtualization has a couple of goals, to protect the memory from malicious process, to ease programming and to allow process to use more memory that the physical available RAM. Multiprogramming allocates each process with a fixed address space and allows the process to think that the address starts at 0.

Virtual memory translates the virtual address to physical address by adding a base register to it. On the other hand, the OS uses memory protection through a bound register to prevent a process to use more memory that the allocated space. This creates a fixed memory address space for each process. The OS also implements realistic virtual memory systems where it allocates the memory into smaller chunks called pages. Each process has its own page to access and could use those extra pages from the hard disk if needed. Those unused pages (if process uses little memory) will be stored in hard drive and allocated for other processes. If a process tries to access its page, the OS will activate an interrupt and loads back the page from the disk.

**Application architectures**

There are four types of architectures. In client-based architecture, only the server runs the data storage while client runs the rest. In client-server, the client runs the presentation and application logic while the server runs both the data storage and data access. In thin-client, the client runs only the presentation logic. In server-based, the server runs all four logics. In multi-tier, more than one server is used to handle more users. Server 1 is used to handle logic, while server 2 handles the data access and data storage. Presentation is managed by client.

**Internet layers**

1. **Application layer**

Application layer deals with the application that the user interact with. The WWW is run by the HTTP protocol. HTTP protocol defines how web browsers talk to web servers. HTTP protocol is stateless. There are two ways to solve this problem - session identifier and cookies. Session identifier is random numbers added at the end of the request and the user will send that session ID as part of every single future request. On the other hand, cookies is a small piece of data that is attached to the response header given by the sender. When the client requests again, it will send the cookie along.

Two tier email architecture is a client-server architecture where the client implements the application logic using SMTP and POP or IMAP. At first, the client sends the mail to the mail server via SMTP. The mail is then forwarded to the recipient’s mail server via SMTP. Then, the recipient accesses the mail in the mail server through POP or IMAP. POP downloads the messages while IMAP allow users to access the same mailbox on multiple devices.

Three tier email architecture is a thin-client that runs on web application, where the server implements the application logic and the client handles the server. At first, the client sends the mail through the website through HTTP. The mail is then sent to the mail server through SMTP, POP or IMAP. The mail is then forwarded to the recipient’s mail server via SMTP. Then, the mail will be transferred to the web server through SMTP, POP or IMAP. The recipient accesses the messages via the website through HTTP.

1. **Physical layer**

Physical layer specifies how the signals are transmitted over the cables or radio waves.

Digital data refers to discrete values while analog data refers to continuous values. On the other hand, digital signal is wave with discrete states, usually represented in unipolar or bipolar encoding. Analog signal is wave with sinusoidal shape, usually represented in the form of wave modulation.

Unipolar encoding allows data to be encoded with one polarity (positive or negative) while bipolar encoding allows data to be encode with two polarities (both positive and negative). An example of bipolar would be NRZ. It switches position when there is a change in polarity; 0V to +V is for 0 and 0V to -V is for 1. Another example would be NRZI, switches position when a 1 is detected. Lastly, Manchester encoding is used to synchronize the bits from the sender to the receiver via a self-clocking system. It switches position in the middle of the bit, 0 is from high to low while 1 is from low to high.

There are three types of wave modulation - frequency, amplitude and phase. In FM, 1 is represented as high and 0 is represented as low frequency. In AM, 1 is represented as high and 0 is represented as low amplitude. In PM, 1 is represented when the wave starts by going upwards and 0 is represented when the wave starts by going downwards. The device that does the conversion of digital signal to analog signal is modem, undergo the process called modulation.

1. **Data link layer**

MAC controls when a device is allowed to transmit. In controlled access MAC, only one device is allowed to transmit at a time. In contention-based MAC, any device can transmit at any time, which operates on first come first serve basis. On top of that, it also avoids collision by carrier sensing and detect collisions then retransmit the frames when the frames are destroyed.

Data link layer deals with the interface between the hardware and the software by sending frames over a physical medium and controls the physical medium through media access control. There are two types of frame transfer - via ethernet or wireless. In ethernet medium, the MAC is CSMA/CD. CD is collision detection. Whenever it detects a collision, it sends a jam signal and waits for random time then retransmit the frame. A switch is used to connect multiple LANs together. It has a forwarding table that stores the devices’ MAC address and the port that is connected to. When it receives a frame, it checks the forwarding table and sends the frame to the correct port. If the MAC address is not in the forwarding table, the switch will broadcast to all devices and the device that replies with its port will be stored in the forwarding table. There are two properties of switch ethernet, it is full duplex (multiple devices are allowed to sent at the same time) and it has buffer memory, meaning that it stores the second frame until the first frame is done transmitting.

On the other hand, in wireless medium, the MAC is CSMA/CA. CA is collision avoidance. It constantly avoids collision by detecting it. It does this through two methods - ARQ and controlled access. ARQ ensures that the AP has received the frame by sending ACK, if the ACK is not received, it will wait for random time and retransmit the frame. Controlled access lets the device to send Request to Send and only sends if the AP replies Clear to Send.

There are two ways of setting up a wireless network. In independent BSS, devices talk to other directly. While in infrastructure BSS, devices communicate via an AP and the AP transmits the frame. Roaming works by combining multiple BSS together, ensuring the APs transmit the traffic between the BSSs. Roaming is done through the extended service set. One of the advantages of ESS is that it allows user to stay connected to the same network without losing signal strength, it creates a wider coverage of signals.

1. **Network layer**

Network layer deals with giving directions for the packets to route between different LANs. This is achieved through IP addresses. IP addresses consists of three parts: the network, subnet and hosts. Subnet mask is the combination of network and subnet. It shows whether if the device is within the same local area network. Devices have different subnet would have to transmit packets via a router.

A router is a device that connects different LANs. It has a routing table that stores the destination IP address. When a packet is received, it looks up the destination IP address in the routing table. It sends the packet directly if it’s connected directly to the router or sends the packet to another router until it reaches to its final destination IP address. Static routing configures a fixed routing table that are manually set. On the other hand, dynamic routing allows routers to configure their routing table by exchanging information. This is done either through the distance vector, which exchanges the transmission distance and choosing the shortest route (done by BGP, RIP or EIGRP) or link state that exchanges that quality of the links and choosing the fastest route (done by OSPF - Open Shortest Path First).

There are two types of address resolution, DNS and mapping IP to MAC. Both functions to map higher layer address to lower level address. DNS is mapping human readable addresses to IP addresses. This is done either through iterative DNS or recursive DNS. In iterative DNS, it iteratively asks the root server which directs to the top layer domain until it gets the IP address. On the other hand, in recursive DNS, it searches the IP address in the cached DNS server, which consists of the common addresses used.

On the other hand, mapping IP address to MAC address, is for packets sending within the same LAN (to the router). This is done through the ARP. The device will first broadcast the destination IP address to everyone and the router will reply with its MAC address. Then, the device sends the packets to the router through the MAC address.

1. **Transport layer**

Transport layer establishes a logical connection between the sender’s application and the recipient’s application, this is done by breaking up the packets and assemble them at the recipient’s side. Port numbers identifies the application that handles the message. It distinguishes the destination where the messages are sent. Each application has a unique port number, a website can have different port numbers to address mail servers, web application, etc.

TCP creates a virtual circuit between applications and splits the messages into segments and reassemble them in the correct order. If the segments have errors, it will request the sender to resend the segments. An example of such usage of TCP is HTTP, SMTP and IMAP. It uses ARQ to ensures that all packets are arrived. There are two types of numbers in TCP ARQ, sequence numbers, which is the number of bytes sent by A and acknowledgement numbers, which is the number of bytes received by B.

TCP establishes this virtual circuit through two sessions, three-way handshake and four way handshake. In three-way handshake, it is used to set up the connection. At first, the client will send SYN (with seq A) and the server replies with SYN (seq B) and ACK (ack A+1). The client will then send ACK (seq A+1 and ack B+1). This ends the open connection process. After the full duplex data transmission ends, four-way handshake is used to close the connection. The client sends FIN and the server replies with ACK and FIN (without changing the number of bytes). Then, the client sends FIN to complete the connection.

There are a few factors that affect the TCP transmission - the segment size and the transmission speed. For segment size, path MTU discovery is used to determine the maximum segment size. The length of the packets will be increased until error occurs, the length is then decreased to use the last error free size. For transmission speed, the packets are sent slowly and receives the acknowledgement from the server. If server doesn’t acknowledge, then the sending speed will be decreased again.

**The internet**

The internet is made up of network of networks built on AS and run by ISP. Internal routing occurs when packets are sent between devices within the same AS. While external routing occurs when packets are sent between different AS through the Border Gateway Protocol (BGP).

There are three ways that could reduce the traffic flow:

1. **Load balancing**

In load balancing, DNS-based is used to map users with different IPs and different locations to servers that are geographically closer to them, this is to reduce latency. In special hardware, load balancing can be implemented by allowing all devices to enter a single server and route them to distribute the requests to other servers, which is mostly done for small networks.

1. **Content caching**

Content caching is used to route the users to local cache that stores the frequently used webpages, that is implemented transparently.

1. **Content delivery networks**

CDN operates servers in multiple locations that run on their own high speed network. Users will be routed to the nearest server to increase the transmission speed.

**Transport layer security (TLS)**

Transport layer security stays between the transport layer and the application layer. It establishes a shared key to protect the confidentially, integrity, and authenticity of the message. It has three main sub protocols - handshake, record and alert. In TLS handshake, it uses the Diffie Hellman Key Exchange to authenticate the server and the client to establish a shared key and session information. In TLS record, it encrypts the data that is exchanged using the ChangeCipherSpec messages. Lastly, in TLS alert, it immediately closes the session.

**Certificates**

Certificates provide additional information for authenticity that are usually signed by certificate authorities (like chrome, safari). However, there are a few problems for certificates. Certificates can be purchased; thus they are not trustworthy. In the meantime, some certificates have failed, but users still need to accept those certificates to access the website. Some certificates are granted to websites that have similar domain names and principal names (like bank’s name similar to a domain name might cause a certificate to be valid).

**Virtual private network**

Virtual private network creates a logical connection between a client to a network via an encrypted channel, which is implemented by TLS or IPSec. VPN only provides encryption service between the client and the gateway to the internal network (tunnel endpoints), this means that the traffic outside of the tunnel (like internal network) is not encrypted.

**Firewall**

There are two types of firewall systems. A normal firewall and packet filter firewall. A normal firewall creates a barrier between a more secure and a less secure network and filters traffic according to security rules on what can enter and what can’t enter. A router is an example of a firewall placement. On the other hand, packet filter firewall operates on the network layer and above, to provide greater security. It uses static filtering set that filters packets according to source and destination IP address, protocols, ports and current stage of connection. For example, it searches the port number of the specific application and denies entry for that.

Demilitarized zone is one of the ways of placing firewalls in a company. It is a zone between two firewalls that is considered to be less secure from the internal network but still protected from direct access to it. This zone usually consists of servers that need to be accessed by the public (like google services). Such firewalls could be placed with one facing towards the internal network and the other facing towards the internet and the DMZ is in between the two firewalls. DMZ also filters outgoing traffic, it prevents malicious software from being sent out and block outbound traffic from critical network areas. On top of that, it also blocks IP spoofing that could potentially cause infections.

Firewall operates on NAT and proxies. NAT hides the internal network of the IP address from outsiders while proxies hide the IP address of individual devices. However, usually firewalls don’t help when the intrusion happens within the internal network. Once an attack is launched internally, firewalls cannot control them. It only acts as a barrier between the “outside world” and the “internal world”.

There is one security feature that helps to secure the IP packets in the network layer - IPSec. In IPSec transport layer, the payload in the packets are encrypted, but not for the header. This is mainly used for device-to-device communication. On the other hand, in tunneling mode, the entire IP packet (which includes the header and the payload) is encrypted and a new header is added. This is mainly used for network-to-network communication.

There are two systems that control the security beyond firewalls, IDS and IPS. In IDS, it monitors the network, alerts potential malicious activity and logs information about activities. On the other hand, IPS has extra feature than IDS. When malicious activity is found, it blocks the activity.