**Computing Theory**

Algorithm- Sequence of st eps that can be carried out to perform a task

For algorithm to be carried out it have to be unambiguous and effectively computable

User interface [Human Computer Interface]

HCI might keep in mind who is going to use the system, what tasks the computer is performing, what environment in which the computer is use, what is 2 feasible

Command Line Interface (CLI)

* allow anyone with a knowledge of the commands access to the entire operating system so they have access to working of the entire computer
* CLI can only be used by people who are computer literate so user need to understand the commands and their uses and they also need to understand something out how the computer operates and how information is stored
* Advantages: Provides more concise and powerful means to control the operation of a program or the computer system
* Disadvantages: Difficult to use, not user friendly especially for beginners and users has to remember many commands

Graphical User Interface(GUI)

* Uses icons, menu and pointer and user can select command by icon
* Advantages: Simple, intuitive, attractive, arouse interest when using the software
* Disadvantages: Visually impaired people are unable to benefit from using this interface, expensive in producing the program in term of effort, time and memory space, many advance commands are not provided and less flexible

Menu Interface

* A list of choices is made available
* Advantages: User does not need to remember the format of commands and save time for keying in, manufacturer can save cost and space since only a keypad and selection keys are required
* Disadvantages: Cumbersome when the list of a menu for selection is too big or multiple menus are used, provides less advanced options for those who require more advanced to be done

Web user interfaces (WUI)

* Type of GUI that accepts input and provide output by generating web pages
* Advantages: Fast and responsive, support on-line application
* Disadvantages: More prone to viruses attack and other computer crimes, older version of the operating system may cause the website to lag and hang

Form based interface

* Type of user interface used to organize the questions or option for user to fill in by pointing to the fields and typing text, or by choosing from a list
* Advantages: Provides something that people are familiar with, data can be captures on one page
* Disadvantages: Less flexible, unable to provide additional information other than what are states in the form

Touch User Interface

* Type of graphical user interface that use touchpad or touchscreen display as combined input and output device
* Advantages: easy to use the program, minimise input typo errors
* Disadvantages: Not all platforms offer touch interface, an expensive technology currently and electronic devices need to be trained for accepting different kind of strokes and writing

Attentive user interface

* Manages the user attention deciding when to interrupt the use, kind of warnings, and level of detail of messages presented to user
* Advantages: Easily captures user’s attention and alert user of some action
* Disadvantages: The interface might prompt the user at the wrong time

Voice user interfaces

* Accepts input and provides output by generating voice prompt and the user input is made by pressing keys or buttons, or responding verbally to the interface
* Advantages: Very intuitive and easy to use interface, good for disable people
* Disadvantages: need time to train the system

Multi-screen interfaces

* Employs multiple displays to provide more flexible interaction
* Advantages: allow people to see more at 1 go
* Disadvantages: Not all platform offer multi-screen

3 characteristics of a user interface: Types of interface/Hardware devices/Design of interface

Validation – automatic checking of data entered into computer system and to check that the data is sensible

* Presence Check 🡪 check if the field is not blank
* Existence check 🡪 check is certain value is present in specified value
* Type check 🡪check if data is of a particular type
* Length Check 🡪Check if data has corrected no of characters (min and max)
* Range check 🡪 check If data in pre-determined range
* Format Check🡪 check if individual characters are valid (NRIC)
* Check Digit 🡪 allow number to be self-checking
* Integrity check🡪confirm the value of a piece of data by comparing it with other data
* Lookup check 🡪 check if data matches one of a limited no of valid entries
* Batch header check 🡪concerned with batch processing

Check Digit 🡪 digit used to check the validity of code numbers

Modulo 11

Digit = 3 4 5 6 ?

Weight = 5 4 3 2 1

Check digit = 11 - rem of ((3x5 + 4x4 + 5x3 + 6x2)/11)

Verification🡪used to check that data is entered correctly and that there are no transcription errors

Transcription error

* Specific type of data entry error made by human operators(Mistype)

Transposition error

* Mistaken for transcription errors, occurs when the characters have switched position due to typing fast (eg. sda--sad)

How to solve transcription and transposition error?????

* Watch the screen when they type and to proof read
* If entry is in data capture forms, coder should use validation rules and input mask(string expression)
* These errors also occur in syntax when coding or programming and should be checked by proof reading

Other verification method

* Double entry 🡪 enter data twice and data is compared 🡪 computer compare 2 set of data to see if they match
* Screen verification 🡪 (proof reading) 🡪 user read the date entered to see if correct
* Check digit 🡪 check if data is correct by comparing it with the check digit
* Batch totals 🡪 total value of one or more files in a batch data they are calculated and compare with the total
* Control totals 🡪 batch totals that have a meaningful value
* Hash totals 🡪 batch totals that have no other meaning.

Testing🡪Process of detecting errors in a piece of software

Array 🡪 data structure make up of a number of variables or data items that all have the same data type and are accessed by the same name

Operations of array 🡪 load initial information into an array which is later required for processing, process the elements of the array and to store results of processing into the elements of an array and to store information which is then written to a report

1. Load a set up of initial values into elements of an array, process the elements of an array, search an array and wrote out the contents of an array to a report

Desk Checking (Dry run)🡪 involves tracing through the logic of algorithm with some chosen test data

When selecting test data to desk check an algorithm, you must look at program specification and choose simple test case only, based on the requirements of the specification, not the algorithm

6 steps in desk checking an algorithm:

1. Choose simple input test cases that are valid
2. Establish the expected value for each test case
3. Make a table of relevant variable name within the algorithm
4. Walk the first test case through the algorithm, keeping step-by-step record
5. Repeat the walk-through
6. Compare the expected result and the result calculated

Desk checking will detect early errors and will eliminate most errors but it still cannot prove it is 100% correct

Program testing

1. Syntax error

🡪occur when a command does not follow the expected syntax of the language

🡪Situations: Missing punctuation key or reserved word spelt incorrectly

🡪how it is detected: Complier or interpreter will often detect and highlight

🡪Strategies: Consistent naming convention, use meaningful variable names and consistent use of letter case and thorough design before coding

1. Semantic Errors

🡪It will run but the computer will not generate any error messages but the program will not do the right thing

🡪Producing wrong answer due to incorrect solution implemented

🡪How is it detected: Programmer requires to work backward by looking at the output of program and trying to figure out what it is doing

🡪Strategies: programmer must fully understand problem so that he can tell if his program properly solves it

1. Runtime

🡪Error that only occur when program is running and it is difficult to foresee before a program is compiled and run

🡪How it is detected: Program will crash

🡪Strategies: Dry run algorithm with realistic test data before coding, thorough design of algorithm before coding

1. Logical

🡪error that causes a program to output an incorrect answer

🡪Occur when mistake is made in their logic for some part of the program

🡪How detected: Test with test data that has predicted output and compare with actual output

* Strategies: Dry run algorithm with realistic data before coding, Through testing of code with a test plan

1. Linking

🡪Error that occur when programmer call function within a program and correct library has not been linked to the program

🡪How detected: Complier or interpreter will highlight, program will crash when function call in attempted to execute

🡪Strategies: Test individual chunks of code that contain any reference to external libraries, Thorough testing of code with a test plan

1. Rounding/Truncation

🡪When a number is approximated to nearest whole no/ nearest zero

🡪How detected: Test with test that that has accurate predicted output

🡪Strategies: Use integer data type wherever possible, use data types that will store larger and accurate no than will ever be used in the program

Debugging features: Breakpoint and stepping

Breakpoint = Signal that inform the debugger to temporarily suspend execution of the program at certain point, point where the program can be halted to see if the program works at this point

Stepping = When program in break mode- program paused at a breakpoint, you can control how execution continues.

|  |  |
| --- | --- |
| White-box testing | Black-box testing |
| 🡪the internal structure, design and implementation of the data item being tested is known to the tester  🡪Software developers are responsible for doing white-box testing  🡪Programming implementation and knowledge is not required  🡪It means structural test or inferior test | 🡪the internal structure, design and implementation of the data item being tested is not known to the tester  🡪independent software testers are responsible for doing black-box testing  🡪Programming implementation and knowledge is not required  🡪It means functional test or external test |

Modular Design

Module = section of an algorithm that is dedicated to a single function

Module must be large enough to perform its task, and must include only the operation that contribute to the performance of that task

Difference between functions and procedure:

* Function returns value while procedure may or may not return value or may return more than one value using output parameters
* Functions are normally used for computations whereas procedures are normally used for executing business logic
* Procedure is bundle of code; it does not have return type while function have return type
* Procedure accepts input or output parameters but function only accepts input parameters

Difference between parameter and argument:

Parameters are temporary variable names within function, argument is a value that is assigned to that temporary variable

Within the function, parameters act as placeholders for the argument it is passed

Parameters can be passed by value or reference:

Value: the actual value is passed into the procedure

Reference: Address of the variable is passed into the procedure

Difference will only matter is the value of parameter is changed

Global variable: Longstanding variable that is accessible anywhere throughout the main program and all subroutines, lifetime and storage of a global variable exists until the program terminates

Local variables:

🡪 If variable defined within subroutine/function/procedure, scope of this variable is only accessible within the things it is declared on

🡪It is not possible for variables declared locally to be modified or accessed by external module lifetime and storage of a local variable exists only when needed

🡪Memory allocation occurs within the scope of the procedure/function and the memory is deallocated once the procedure/function returns

Recursive subprogram is one that includes, among the statements making up the subprogram, a call to the same subprogram and it will continue calling itself recursively, so it must have a means of finishing and continuing the calling program and this is done by stopping condition, which causes the subprogram to exit rather than call itself again.

Advantage: When solution to problem is recursive, it enables the programmer to write a program which mirrors the solution, recursive solutions are often much shorter than non-recursive ones

Disadvantages: If recursion continues too long, stack of return address may become full and the program will crash, recursive routines can be difficult to follow and to debug and recursive routines are sometimes very slow in execution owing to the overheads involved in repeatedly calling the subroutine, and storing and retrieving return address and parameters

Recursion vs iteration:

Iteration allows multiple blocks of data and instructions to be processed repeatedly in sequence using loops, by using iteration construct, these loops will execute repeatedly until a condition is fulfilled

Recursion on other hand is way of programming that uses the selection construct where a function calls itself one or more times in its body, and then terminates when it reaches a base case

Dynamic programming:

Build up solutions simpler instances from small to large, save results of solutions of simpler instances, useful when problem can be solved using solution of two more instances that are only slightly simpler than original instances

Standard modules- one which carry out common/standard task/can be used for s standard situation in a program

Built-in functions – made available by programming language, are already made and tested, benefits is no need to write again, less likely have to errors and likely to be high quality

Module library-module can be kept here and reused

Library program – program contained in a program library and there may be programs in a program library but more often they are subroutines that programmers can use in their programs

Modularisation:

Simple solution algorithm-define the problem, write down the control structures required to reach a solution and devise a solution algorithm, which uses a combination of sequence, selection and repetition control structures

Top-down design/functional decomposition- process of identifying first the major task, then further subtasks within them

Modularisation-process of dividing a problem into separate tasks, each with a single purpose

Modularisation allows:

* Provides abstraction-process of paying attention to important properties while ignoring nonessential details and modules or subroutines are way to achieve abstraction
* Allows multiple programmers to work on problem
* Allows you to reuse your work
* Allows you to identify structure more easily when you combine several programming tasks into modules

Structured programming-Methodical approach to design program that emphasises breaking large and complex tasks into smaller subtasks

In structured program, only 3 basic constructs are used, sequence, selection and iteration and each program module are generally not very long and is easily comprehensible

Top-down design methodology allow programmer to concentrate overall design of algorithm without getting too involved with details of lower level modules and it also can separate modules, once unidentified and written it is easily understood and can be reused and can be independently modified

Top-down programming is type of structured programming where overall problem is defined in simple terms and then split into number of smaller subtasks

Modular design is method of organising a large computer program unto self-contained parts, modules, which can be developed simultaneously by different programmers or teams or programmers

Modular programming is software design technique that increases extent to which software is composed of separate called modules by breaking down program function into modules

Advantages of modular programming- individual modules can be separately tested, modules can be kept in module library and reused in other programs and long programs that are split into modules are easier to read, debug and maintain and modular approach means that several people in team of programmers can each work on separate modules, thus shortening the total development time for a large project

Advantages of sorting- if data is stored and sorted in pre-defined order, data searching can be optimized and by sorting data, information can be made more readable, processing of data can be performed in a defined order

Internal sort is performed when no of elements is small enough to fit into the main memory, sorting is done by modifying order of elements within the sequence

External sort is when all elements that needs to be sorted cannot be places in memory at a time so addition memory is required in order to perform the sorting and it is used for massive amount of data (Merge sort)

Divide and conquer is important algorithm design paradigm based on multi-branched recursion, it works by recursively breaking down a problem into 2 or more subproblems of same type, until these become simple enough to be solved directly

|  |  |  |  |
| --- | --- | --- | --- |
| Sorting method | Best | Average | Worst |
| Insertion | O(N) | O(N^2) | O(N^2)-N^2 |
| Bubble | O(N) | O(N^2)-N^2/4 | O(N^2)-N^2/2 |
| Quick | O(Nlog2N) | 0(Nlog2N) | O(N^2)-N^2/2 |

Linear search

worst case = n

average case = n/2

Binary Search

Worst case = log2n

All algorithm falls under iterative and recursive

Iterative- use loops and conditional statements

Recursive- use divide and conquer strategy so it breaks down large problem into small pieces and make it more small, straightforward and simple to understand

Analysis of an algorithm provides information that gives us a general idea of how long an algorithm will take for solving a problem

Analysis does not five a formula that helps us determine how many seconds or computer cycles a particular algorithm will take to solve a problem

It is important to decide what to count while analysing an algorithm and 2 classes of operations th at are typically chosen for the significant operation -

Comparison or arithmetic

Database structure

Database consists of a large collection of data items and links between them.

Entity – information contained in a table in database

Each column in table contains information relating to a particular attribute

Flat file database that can be held as a table and stored in single test file

Difference between flat file database and relation database:

Flat file has no way to link the information geld in separate file, and where there are changes to any field, you need to rewrite the program. While relational database can be done quite easily

Primary key- Unique identifier that does not allow duplicate and null values

Composite key- primary key consisting of more than one field

Foreign key- filed in relational data table that match the primary key columns of another table and it cannot contain values that do not appear in the table that it refers to

ER diagram

One to many 0 - 0

Many to one 0 - <0

Many to many 0>-<0

Data redundancy – data organization issue that allows the unnecessary duplication of data within the database

Update anomaly- When a data of an entity is being updated, that entity will have 2 values

Insertion anomaly- Inability to add data to the database due to absence of other data

Deletion anomaly- When a record is deleted that may continue contain attributes that should not be deleted

Normalization- Technique organizing the tables in a database to reduce data redundancy and prevent inconsistent data. There at least three normal forms associated with normalization: 1nf 2nf 3nf

Tables should be organized so that:

-no data is unnecessarily duplicated

-data is consistent throughout the database

-Structure of each table is flexible enough to allow you to enter many ir as few items as you want to

-Structure should enable a user to make all kinds of complex queries relating data from different tables

1NF – contains no repeating attributes or groups of attributes (Usually have 1 more pk)

2NF – It is 1NF and no column that is not part of a primary key is dependent on only a portion of the primary key (Make one more table with only PK)

3NF – Contain no ‘non-key dependencies’ and there are no transitive dependencies in the table (Make one more table)

Flat-file is program-oriented while database is data-oriented

Advantages of relational database approach:

* Data independence-change in structure of database do not affect the programs that access the data
* Consistency of data- each item is held only once so there is no danger of an item bring updated on one system and not on another
* Control over redundancy- redundancy is minimized compared to flat-file which contain several same information
* Greater security and integrity of data- only authorized users are allowed to access data
* Increased productivity and more information available to users-user have access to information and DBMS provides easy to use query language

Disadvantages of relation database approach:

* Larger Size – require more disk space and power computer
* Greater Complexity- it must be carefully designed and this require considerable expertise
* Possible inefficiency or poor performance-spmetimes much less efficient for some processing tasks
* Greater impact of system failure – if database system fails, everyone using it will be affected
* More complex recovery procedures – Recovering from a system failure require a complex procedure to ensure no data loss

Database management system

DBMS – application program that provides an interface between the data and the users, for access to data as simple as possible

Functions of DBMS: Data storage, retrieval and update, Creation and maintenance of the data dictionary, Managing the facilities for sharing the database, Backup and recovery, Security.

Database Administration- responsible for supervising both the database and use of DBMS

DBA tasks are: Design of database, monitoring the performance of the database, keeping users informed of changes in the database structure that will affect them, maintenance, implementing access privileges for all users of database, allocating passwords to each user, providing training to users in access and use of database

Data Dictionary- file containing descriptions of and other information about the structure of the date geld in a database. It contains, table and columns, name and description of each data item, characteristic of each item of data, such as length and data type.

Indexes – Speed up access to data but will slow down record updating when indexed filed is changed index entry will have to be changed as well and when a new record is added, all indexes have to up updated. Foreign keys are index usually

Query – allow users to combine into one table the information from two or more related tables, select fields that are to be shown in Queried table, specify criteria for searching, save the query so that it can be executed whenever necessary, save the Queried table so that it can be displayed or used as the basis for a report

Reports- can be generated as the result of a query, usually possible to format the output to the user’s requirement

Security issues

Databases must be protected from deliberate hackers, disk crashes fire, floods

Physical security:

Regular backup- making back up regularly and back-up data stored in fire proof safe and other at different building

Only authorized personnel are allowed in main computer room where disks or tapes are stored

Prevention of unauthorized access

Transaction logging-information about every updating transaction si recorded on different transaction file, before and after image of any record being updated is saved so that if part of database is destroyed by disk failure, up-to-date copy can be created

Checkpoints – marker placed in transaction log at a point in time when all outstanding updates in temporary buffers are written permanently in database so if data failure occur, use after-image to recover

Passwords to restrict entry

Encrypting the data to make data unreadable and unintelligible

Use DBMS to assign privileges to each user so that user can access data that are necessarily to their jobs

Back up and Archive

Archive is storage of information for long periods of time and data are compressed to take less space on a cheaper storage medium. It can be accessed but not easily available

Backup file is duplicate of existing file in use and it is kept as a security measure in case the original is corrupted in any way

Archive files are kept in secure location away from the main computer system

Network：

Internet is a network of networks set up to allow computers to communicate with each other globally

Use backbone cables connected by trans-continental leased lines fed across the sea bedsok

URL – full address of an internet resource, which is usually a file and has reference that in full

Internet registrars hold records of all existing website names and the details of those domains that are currently available to purchase, they are company that act as resellers for domain names and allow people and companies to purchase them. All registrars must be accredited by their governing registry

Internet registries are 5 global organisation governed by ICANN with worldwide databases that hold records of all domain names currently issued to individuals and companies, and their details

Domain Name System – Registries also allocate IP addresses and keep track of which domain name is associated

IP address – Unique address that is assigned to a network device and it performs similar function to a home mailing address by indicating where a packet of data is to be sent or has been sent from

Sub-netting- Allow further structure in the addressing

It will allocate just one Class C netID instead of 6

On the internet, all of allocated IP addresses have netID pointing to router and the router has to interpret the hostID to direct the transmission to appropriate host on one of LANs

Domain name – location or site of resource on Internet, each location has a unique domain name.

Once registered by internet registry, domain names stored in number of large publicly accessible databases know as DNS

Domain names are structured into hierarchy of smaller domains and written as string separated by full stops

Each domain name has one or more equivalent IP address

DNS catalogues all domain names and IP addresses in series of global directions that domain name servers can access in order to find the correct IP address location for a resource

When URL is being typed, browser requests corresponding IP address from local DNS and if that DNS does not have correct IP address, the search is extended up the hierarchy to another larger DNS database. The IP address is located and a data request is sent by user’s computer to that location to find web page data and webpage can be accessed within a browser by entering IP address if it is known

Computer which has the same IP address are potentially connected to internet at all times. Computer which accesses Internet through an ISP are assigned a dynamic IP address when logging in

Fully qualified domain name – includes the host server name, for eg. www, mail, docs

LAN and WAN

Network classified into local area networks and wide area networks

In LAN, computer system situated relatively close to each other for example in same building or cluster of buildings like schools. Network connections are normally wire cables, wireless links or fibre optic cable. LANs can transmit data very fast but only over a short distance and LAN usually serves a single organisation. Many LANs are divided into logical groups called subnet and ethernet is the most popular physical layer LAN technology in use today

In WAN, computers are geographically remote. WAN make use of range of connection methods like public telephone links, undersea cables and communication satellites. These connections are dependent upon telecoms providers, cost is high and speed are limited, LAN links LAN in geographically spread through organisation

Network topologies

LAN can use different layouts or topologies

Physical bus topology: All computers are connected to a single cable where its end is plugged into a terminator. When there is a signal from a source, it will travel in both direction to all machine connected on cable until it finds the intended recipients. It is inexpensive to install as it requires less cable and additional hardware.

But if main cable fails, network can no longer be transmitted to any of the nodes. Performance degrades with heavy traffic and low security as all computer on network can see data transmission

Physical star topology:

It has a central node which act as a router to transmit messages. Switch keeps a record of unique MAC address of each device on network and can identify which particular computer on the network it should send the data to.

If one cable fails only one station is affected and performance is consistent despite network heavily being used. Higher transmission speeds will give better performance than bus topology. It is more secure as messages are sent directly to central computer and cannot be intercepted by other station. Easy to add new station without disrupting network

Ring topology:

Computers are connected to one another in circular fashion. It allows data to travel only in one direction.

Network fails is cable connection is broken or if a computer shuts down

Mesh network topologies:

Each node in mesh network has connection to every other node by transmitting data across any intermediate nodes. Only one node required connection to internet and all others can share this connection .

No cabling cost and more the nodes installed, the faster and more reliable the network. New nodes are automatically incorporated into the network and communication is faster since data packet do not need to travel via a central switch

Physical vs logical topology:

Physical is its actual design layout which is important when selecting a wiring scheme and design the wiring for a new network

While the logical topology is shape of path the data travels ins and describes how components communicate across the physical topology.

Both physical and logical topology are independent of each other, so that network physically wired in star topology can behave as bus network by using protocol and appropriate physical switching

Wi-Fi

Local area wireless technology that uses radio frequency transmission and enables you to connect device to network resource or to the internet via a wireless network access point. An access point had a range of about 20 metres indoor and more outdoor

Computer need a wireless network adaptor to connect to wireless network. Station is the combination of computer and interface controller. All stations share single radio frequency communication channel and each station is constantly tuned in this frequency to pick up transmissions and transmissions are received by all station within range of wireless access point

Intranet

It is a communication system solely within particular company or organisation, it provides similar services to those offered by the Internet, such as web pages, but only to close group, normally controlled by a password.

For security it is usually only available to users physically connected to company’s network or possibly accessed using a virtual private network. Intranets are increasingly used to provide access to collaborative working tools such as customer relations management and project management and unsecured access is not permitted across the internet

VPN- secure, point to point connection from device to network over Internet. It establishes encrypted channel that ensure sensitive data is safely transmitted and it prevents unauthorised people from eavesdropping on traffic and allows user to conduct work remotely. VPN technology is widely used in corporate environments to protect information being shared

Communication and transmission

Data flow modes – data flow along individual link is simplex, half duplex, full duplex. In simplex mode, the flow is one-way. In a duplex mode flow is both ways but only occurs simultaneously in full-duplex

Message types – message can be sent as a broadcast (one-to-all communication) or multicast (one source to many destinations) or unicasting (one-to-one communication)

Transmission nodes – There is circuit switching or packet switching approached in communication over an internet.

In circuit switching, one endpoint creates single path connection to another and the circuits refer to connection between endpoints

Circuit switching is method used in traditional telephone system and this method can be provided for data transfer rather than voice communication.

If 2 computers use circuit switching principle, a bandwidth is wasted during periods when no data is being sent. The 2 devices must transmit and receive data at same rate so circuit switched networks can only connect computers or devices that operate at same transfer rate. Since this is exclusive connection between 2 device, data segments arrive in same order that they are sent so simplifying process of reconstructing the message at recipient end.

Packet switching, data transmitted is broken into small units (packets) that move in sequence through network. Each packet takes best route available at any given time rather than following an established circuit path. Packet switching assumes that network is constantly changing and adjustments need to be made to compensate for network congestion or broken links

When packet switching is used, there are 2 ways that network can provide service, connectionless or connection-oriented service. If connectionless service is provided, a packet is dispatched with no knowledge of whether or not the receiver is ready to accept it. In a connection-oriented service the first packet sent includes a request for an acknowledgement and if this is received, sender transmits further packets and if no acknowledgement is received the sender tries again with the first packet

Data packets

Size of each packet in transmission can be fixed or variable, each packet contains a header and a payload containing the body of data being sent. Some packets may also use trailer section with a checksum or Cyclical Redundancy Check to detect transmission errors by creating and attaching a hash total calculated from the data contained in packets. Hash total commonly involved adding up total number is 1s in transmission. CRC checksum is recalculated for each packet upon receipt and matched to help verify that payload data has not changed during transmission. If CRC totals differ, the packet is refused with suspected data corruption and new copy is requested from the sender

Parity check – simpler error-detection and an extra bit is sent along with each byte

When even parity is used, number of 1-bits sent are even

Odd parity check will check for odd number of 1-bits.

Parity check will not prevent or fix errors but it enables receiver to detect that an error has occurred

Header includes sender’s and recipient’s IP addresses; protocol being used with this type of packet and number of packets in sequence being sent. Include Time to Live or hop limit, after which point the data packet expired and is discarded. Upon receipt of all packets, they are reassembled in correct order and data is extracted

Routers

They are used to connect at least 2 networks, commonly 2 LANS or WANS, or to connect LAN and ISP’s network. The act of traversing between one router and another across a network is called hop. The router reads the recipient’s IP address in each packet and forward it on to the recipient via the fastest and least congested route to next router, which will do same until packet reaches its destination. Routers using routing tables to store and update locations of other network devices and most efficient routes to them. A routing algorithm is used to find the optimum route. When router is connected to Internet, the IP address of port connecting it must be registered with the Internet register because this IP address must be unique over the whole Internet

Gateways

Routing packets from one network to another requires router if networks share same protocols. Where these protocols differ between networks, a gateway is used rather than router to translate between them. All header data is stripped from the packet leaving only raw date and new date is added in format of new network before gateway send packet on its way again. Gateways otherwise perform a similar job to routers

Media Access Control(MAC) address:

Every computer device which is capable of being part of network must have wired or wireless Network interface card(NIC). Each NIC has unique MAC address, which is assigned and hardcoded into card by manufacturer and which uniquely identifies device. Address is 48 bits long and is written as 12 hex digits.

TCP/IP Model

It is important to have protocols and standards to ensure the proper transfer of data between devices on networks. Protocols exist that specify the format of data, and signals to start, control and end the transfer. HTTP has become standard protocol for browsers to render web pages.

TCP/IP is used worldwide and enables communication with any other computer connected to Internet regardless of its location.

TCP/IP specifies how individual signals are sent over network and provides transport and routing for data, Originally TCP/IP was designed for internet but has proved very flexible and is used by majority of networking systems, reducing the need for special gateways to convert the signals into different protocols and allowing easy connection on internet. This protocol splits the data into small packets which may be sent, or routed, using different physical network lines

TCP/IP protocol is et of networking protocols that work as four connected layers, passing incoming and outgoing data packets up and down the layers during network communication and they are Application layer, Transport layer, Network layer, Link layer

Various protocol operates at each layer of the stack, each with different roles. In each layer, the data to be sent is wrapped, or encapsulated in an envelope containing new packet data as it descends the layers and is unwrapped again at receiving end in a networking equivalent of a game of pass the parcel

Application layer:

This layer sits at top of stack and uses protocols relating to application being used to transmit data over network, usually the internet. If this application is browser, it would select an appropriate higher-level protocol for the communication

Transport layer

It used Transmission Control Protocol to establish an end-to-end reliable connection with the recipient computer. The data is then split into packets and labelled with the packet number, the total number of packets and the port number through which the packet should route. This ensures that the recipient computer handles it with the correct application.

If any packet got astray during the connection, the transport layer requests retransmission of lost packets. Receipt of packets is also acknowledged \Beside TCP, User Datagram Protocol is used at transport layer. The smallest packet using UDP ais 60 percent less than that of using TCP. Hence, UDP is simplest communication protocol in this layer. But, UDP is considered unreliable because the receiver does not generate an acknowledgement of packet received and the sender does not wait for any acknowledgment of packet sent. Video streaming generally uses UDP as the bandwith needed is significantly huge and packets loss may not be obvious in video quality

Network layer

Referred to IP layer or internet layer. It adds source and destination IP addresses. Router operate on this layer and use these IP addresses to forward the packets on the destination. The addition of an IP address to the port number forms a socket.

Link layer

Physical connection between network nodes and adds the unique Media Access Control addresses identifying the Network Interface Cards of the source and destination computers. Once the packet finds the correct network using the IP address, it can then locate the correct piece of hardware. The destination MAC address is that of the device that the packet is being sent next. This means that is the 2 computers are on different networks, the destination MAC address will initially be that of the first router that packet will be sent to

Once, that packet reach the receiving end, the MAC address is stripped off by the link layer, which passes the packets to the network layer, The IP addresses are then removed by the network layer which passes them on to the transport layer to remove the port numbers and reassemble the packets in the correct order. The resulting data is then passed to the application which resents the data for the users.

Since routers operate on the network layer, source and destination MAC addresses are changed at each router code. Packets, therefore comove up and down the lower layers in the stack as they pass through each router or switch between the client and server

Client-Server Architecture

One or more computers known as clients are connected to powerful central computer known as server. In large network, there may be several servers, each performing a different task such as: File server holds and manages data for all clients, print server manages print requests, Web server manages requests to access the Web, Mail server manages the email system, Database server managers database applications

In client-server network, client makes a request to server which then processes the request.

Advantages:

* Security is better as all files are stored in central location and access rights are managed by server
* Backups are done centrally so no need for individual users to back up their data. If there is breakdown and some data is lost, recover procedures will enable it to be restored
* Data and other resources can be shared
* It is easy to replace, upgrade or relocate the clients because they are all independent and request data only from the server

Disadvantages:

* It is expensive to install and manage. Professional IT staff needed to maintain the servers and run the network
* If all client simultaneously request data from server, it may overload, leading to congestion in the network
* If server fails for any reason, none of requests of clients can be fulfilled leading too failure of client server network

Cloud computing

It is a growing service-based industry which provide access to software or files via the Internet using the client-server model File storage companies such as Drobox offer file storage facilities where users ‘files are kept on remote servers. Other companies offer software via the cloud, a provision know as software as a service. This software is available through website logins where all company data and application are stored offsite

Data processing

In client-server model, data may be processed on either side

Client will send request message to server which should respond with the data requested or suitable message otherwise

Client-side processing

Data is processed on client computer rather than on server and this may happen because client computer has specific software that can process the information, or to lighten the load on the server’s processor. Validation of data entry can be done at client’s side. Processing data on client-side processing also improve security as it can avoid unnecessary data transfer. Client-side processing can also adjust styles for different platforms or screen sizes. Java script is a client-side language used frequently to provide interactivity on a web page

Server-side processing

Servers often process enormous volume of data on behalf of multiple clients. They can process data much faster than client computer. There are specific languages used for server-side processing such as SQL. Search requests may be sent to server where they may be applied to database using SQL. Database search results are then sent back to client browser. Validation may also be carried out on server where an invalid entry must be compared with data already on serve database. Example are checking user credentials, or looking up valid airport location. JavaScript may also be circumvented maliciously so server-side validation is important for the integrity of server data

|  |  |
| --- | --- |
| Client-side processing | Server-side Processing |
| Initial data validation | Provides further validation |
| Provides web page interactivity | Used to query a Database |
| Manipulates user interface elements | Update server databases |
| Applies styles (CSS) | Performs complex calculations |
| Reduces the load on the server | Encodes data to readable HTML |
| Reduces the amount of web traffic | Keeps data secure |

Thin vs thick client computing:

Thickness of client computer refers to level of processing and storage that it does compared with the server it is connected to. The more processing and storage that a server does, the ‘thinner’ the client become. If all the processing and storage is done by the server, then all that is required for the thinnest-client computer is a very basic machine with very little processor power and no storage. Known as dumb terminal

Application Programming Interface:

API is set of protocols that governs how 2 applications should interact with one another. An API set out format of requests and responses between a client and a server and enables one application to make use of the services of another. An organisation may use Twitter APL to enable relevant tweets to be regularly fed through to display window within their own website. Price comparison websites may also use an API to gather data from individual company websites in order to display a list of each of them for the consumers

Peer-to-peer networks

There is not central server. Individual computers are connected to each other, either locally or over a wide area network so that they can share files. In a small local area networks, peer-to-peer network is good choice as it is cheap to set up and enables users to share resources such as printer or router, and is not difficult to maintain

Peer-to-peer networks are used by companies providing video on demand. A problem arises when thousands of people simultaneously want to download the latest episode of a particular TV show. Using peer-to-peer network, hundreds of computers can be used to hold parts of the video and so share the load. This is the main principle behind dozens of torrent website that enable sharing of files, often containing the copy right material.

Peer-to-peer networks have widely been used for online privacy, since it is impossible to trace files which are being legally downloaded

Hashtable

Hash table – Unordered collection of key-value pairs

Key is sent to hash function that performs arithmetic operations on it. The output of hash function aka hash value is the index of key-value pair in the hash table

Storing records in array allows direct access, and this in turn will facilitate towards faster searching, insertion and deleting operation

Main requirements of hash functions:

Always produce same hash value for same key, provide uniform distribution of hash values. This means that every value has an equal probability of being generated. Minimise clustering; this will arise when many different keys produce the same hash value. Where 2 or more different keys produce the same hash value, we say a collision has occurred. Be efficient to compute

Array – data is stored in array where each data value has its own unique index value. Size of array should be set according to amount of data expected. As array allows direct access, access of data become very fast if we know the index of desired data

Collision occur when hash function produces same hash value for 2 or more keys. Collision are real problem when using hash table because no 2 sets of data can be stored at the same location.

Linear Probing

Creates hash value that references a position that is already occupied, then the data be referenced to next free position in hash table. This can be performed by probing sequentially or using an interval until an empty slot is found.

Clustering occur when hashing algorithm produces indices that are not randomly distributed and this increases the possibility of collisions as well as number of comparisons needed when searching for a record

There need to be enough slots to store column of data. Hash table have load factor which is the number of keys divided by the size of database. A high load factor means it will become increasingly difficult for the hashing algorithm to produce unique result and hence, increasingly the possibility of collision

Searching for item in Open addressed hash table:

Hask key generate index value, examine indexed position to see if key matched key of data you are looking for, if there is no data at the hash value position, it can be concluded that data is not found, and there is no need to search through the whole array to confirm, if there is data at the hash value position but the search key of the data does not match with they key stored at the indexed position, check each location that follows until the matching record is found or until an empty slot is found

Deleting item

Removing a record from an open hash table is problematic. Replacing the removed record with NULL values may interfere with subsequent search operations. A possible solution solves this is to not delete the item, but rather mark the item to be deleted with a tombstone value to indicate that the data in that position has already been deleted, but it is kept so that reference to the next item can be referenced

Advantage of Linear probing:

Fast rehashing, fast access, better memory locality and cache performance. All elements laid out linearly in memory

Performs better when number of keys is know in advance

Close Addressing:

Separate changing is close addressing system as address given to a key is fixe. This method deals with collision using linked list. Instead of storing the actual data in hash table, a pointer to allocation where the data is stored is maintained. each data item is stored as a node with three attributes: the key value, the data, a pointer to the next node

The first value to be stored will have a null pointer as it is the only item in the list. When there is a collision, the pointer for the first node at that location will be updated to point to the new node. This will create a chain of nodes whose keys resolve to the same hash value

Searching an item in close addressed hash table:

When collisions have been dealt with using chaining, the data will always be accessed through a single location in the hash table. The process for retrieving it is as follow:

Generate the has key to find the index value, examine the indexed position to get the pointer to the node at the head of the list

if head is empty, return item not found,

else, examine the node to see if it matched the search item.

If it is not, follow the linked list until the node is found or not found

Deleting an item in a closed hash table will first involve searching for it, the subsequent steps will be similar to how a node is deleted from a linked list. In the worst-case scenario, the impact will be just performing a linear search which has a time complexity of O(n). Therefore, it is always important to ensure that the hash function is designed and tested to minimise the number of collisions.

Advantages of Chaining:

A chained hash table can grow infinitely and will not need to be reseized as long as there is enough space, easier removal, typically performs better with high load factor, no issue with clustering

A tree is abstract data type that stores elements hierarchically as a non-linear data structure with exception of top element in tree has a parent element and 0 or more children’s elements. It organises its data hierarchically as collection of elements know as nodes that are connected by branches

Root node – topmost node in the tree

Parent node – every node in a tree is connected by a directed bracmh from exactly 2 other node

Child node – node directly connected to another node when moving away from the root

Leaf node = node with no child

Subtree- subtree rooted at x can be defined as tree of all descendants of x that has node x as root

Depth – number of edges from n to the tress’s root node

Height – number of edges on longest path form root to leaf

Every node has at most 2 subtrees

BST is binary tree where each node contains value from a well ordered

Socket – power mechanism provided by most operating systems

Internet socket deliver data using same transmission control protocol and internet protocol suite that is used to transmit data over the internet. This means that internet sockets can deliver data between any 2 programs, even programs that are running on different computers, as long as the two computers can access each other over the network

Data is transmitted through an internet socket may pass through multiple devices before reaching its destination.

As networks can become congested, we cannot assume that data sent over internet socket will be transmitted instantaneously. Program may receive only the first gald of a message before the second half arrives some time later. To avoid working with incomplete data, we will need to define a protocol so that the start and end of messages can be detected unambiguously

IP address and ports

Each end if a socket is associated with a running program and is uniquely identified by a combined IP address and port number. The IP address identifies which device that end of the socket is attached to and the port number identifies which program on that device is using the socket

There are 2 kinds od IP address IPv2 and IPv6 where IPv4 have 32 bits and are usually presented as 4 denary numbers separates by dots. Each denary number can range from 0 to 255 and corresponds to one byte (8 bits) of the IP address

!Pv6 have 128 bits and presented as 8 groups of 4 hexadecimal digits departed by colons/

On each device, port numbers are used to distinguish between attached sockets. The device also keeps track of which program is associated with each port and which port numbers are still available for use by new sockets

Creating a socket programming

Server creates passive socket, binds it to pre-chosen port number and listens for an incoming connection

Next the client initiates a connection request using the server’s IP address and port number. If no server is listening on the chosen port, the connection will be refused

On the other hand, if the connection request reaches an IP address and port number that a server is listening on, the server accepts and creates a new socket for the requesting client using dynamically assigned port number.

The passive socket goes back to listening for new connection while the client and sever can now exchange data using the newly-created socket

We must be able to encode the raw bytes into a sequence of 8-bit characters using Python’s bytes type

This encoding and decoding is important as internally, python str is actually treated as sequence of number called Unicode code points, there are over a million possible code points, so it is not always possible to represent each code point using just 8 bits. Instead, the Unicode standard defined an encoding called UTF-8 so code points can be represented using bytes in a space—efficient t and consistent manner

To enter a sequence of bytes directly in code, we can use bytes literal that starts with the letter b, followed by a sequence of bytes enclosed in matching single or double quotes.

Character sets

Byte – number of bits used to store one character

Character Set – complete set of characters that computer uses

ASCII – character set that is a 7-bit set of codes that allows 128 different characters. It is stored as 8 bits, but the eight bits could be used for error checking in data transmission

128 different characters represented in 7 bits are used for upper case, lower case, digits and symbols

ASCII cannot be used for arithmetic and would use unnecessary space to store numbers and numbers for arithmetic are stored as pure binary numbers

Unicode- uses between 8 and 32 bits per character so it can represent characters from all kind of languages. Support more than 1 million characters. Since it is larger than ASCII, it takes up more storage space when saving documents.

It allows localisation of software, where standard software can be adapted for use by different cultures by modifying layout and features of interface. Multiple languages

1 byte = 8 bits

n bit = 2^n of characters

Data privacy – Requirement for data to be accessed by or disclosed to authorised persons only

It is easier than ever to gather a person’s data and use it to obtain valuable insights, track movements or to commit fraud. Eg. Impersonate by registering for new phone line on telco website, perform unauthorised or fraudulent baking transactions via phone baking

We want personal data to be private and protected so that unauthorised people cannot use personal data for unauthorised use

In Singapore, personal data is protected under Personal Data Protection Act (PDPA)

Personal data – data, whether true or not, about an individual who can be identified from that data

Personal data in Singapore is protected under PDPA 2012

PDPA – Data protection law comprising various rules that govern the collection, use, disclosure and care of personal data. It recognises both rights of individuals to protect their personal data, including rights of access and correction, as well as the needs of organisations to collect, use or disclosure personal data for legitimate and reasonable purposes

PDPA takes into account:

Consent to CUD

Notification Purpose of CUD

Appropriateness Only CUD for appropriate reason

Accountability Make information about their personal data protection policies available on request, they should also make available the business contact information of the representatives responsible for answering question relating to organisations’ CUD of personal data.

Aim of PDPA is to regulate flow of personal data among organisations and to strengthen and entrench Singapore’s competitiveness and position as a trusted, world class hub for businesses

Do Not Call (DNC) Registry:

To prevent unnecessary marketing call, individuals can register in DNS registry to opt out of marketing messages and calls. PDPA prohibits organisations from sending marketing messages to Singapore telephone numbers or other numbers that are registered with the DNC registry

Three type of DNC registers:

No voice Call register

No text messages register

No fax message register

Registering the phone number in each register is to opt out of receiving marketing messages through voice calls, text messages and fax messages.

Organisations which have an ongoing relationship with a subscriber or user of a Singapore telephone number may send marketing messages on similar or related products, services and memberships to that Singapore telephone number via text or fax without checking against the DNC Registry.

But each message will contain an opt-out facility that recipient may use to opt out from receiving such messages and if recipient opts out, organisation must stop sending messages to that number after 30 days

DNC registry (For organisations):

Even if you register your number in DNC, you may still get marketing message but that message will include a link for you to opt out from receiving their marketing messages, once you opt out the mobile phone company can no longer send you marketing messages on that number

Use of NRIC:

NRIC/FIN unique number identifier assigned to citizens/foreigner and it contains personal information about the person, such as date of birth and address. As unique identifiers like NRIC and FIN are permanent, irreplaceable and used in variety of government transactions, we need to be careful with such data

Individuals should not readily provide their NRIC/FIN and personal particulars to companies. Consent is required before organisations can obtain a person’s data. Under PDPA, organisations are generally not allowed to CUD NRIC numbers except CUD of NRIC numbers is required under the law or CUD or NRIC numbers is necessary to accurately establish or verify the identities of the individuals to a high degree of fidelity

If police ask for personal data, you have to comply to them. PDPA does not apply to public agencies and organisations acting on behalf of these public agencies.

Data obligations and Data protection:

Organisations are required to abide by the following 9 main personal data obligations:

* Consent Obligation:

🡪Only CUD personal data for purposes for which an individual has given their consent

* Purpose Limitation Obligation:

🡪Organisations can only CUD personal data about individual for purposes that a reasonable person would consider appropriate in the circumstances and for which the individual has given consent

* Notification Obligation:

🡪Notify individuals of the purposes for the organisation to CUD personal data

* Access and Correction Obligation:

🡪Upon request, personal data of individual and information about the ways in which his or her personal data has been CUD within a year before the request should be provided. Organisations prohibited from providing individual access if provision of personal data could reasonably be expected to cause harmful effects. Organisations required to correct any error or omission in individual’s personal data that is raised by the individual

* Accuracy Obligation

🡪Make reasonable effort to ensure personal data collected is accurate and complete, if it likely to be used to make a decision that affects the individual or disclosed to another organisation

* Protection Obligation

🡪Make reasonable security arrangements to protect personal data that your organisation possess or controls to prevent unauthorised access, CUD

* Retention Limitation Obligations:

🡪Cease retention of personal data or remove means by which personal data can be associated with particular individuals when it is no longer necessary for any business or legal purpose

* Transfer Limitation Obligation:

🡪Transfer personal data to another country only according to requirements prescribed under the regulations, to ensure that standard of protections provided to the personal data so transferred will be comparable to the protection under the PDPA

* Accountability Obligation:

🡪Make information about data protection policies, practices and complaints process available on request. Designate a Data Protection Officer to ensure that organisation complies with the PDPA

Protection of Personal Data:

Company can call you once you agreed that they can contact you even if your number listed in DNC. If there are any queries on personal data, or to withdraw consent, you can contact data protection officer (DPO). Under PDPA, company need to appoint one or more persons to be DPO to verse data protection responsibilities within organisation and ensure compliance PDPA.

Tear or shred paper with personal data so that people cannot obtain personal data about yourself

Decision Table

Precise way of modelling logic

|  |  |
| --- | --- |
| Conditions | Condition Entries |
| Action | Action Entries |

Decision table rules

* To every decision, there are usually 2 possible outcomes
* For n decisions, there are 2^n possible outcomes
* Usually, several possibilities may result from these related decision

Advantages of using decision table

* Possible to check all combinations
* Show the cause-and-effect relationship
* Easy to trace from actions to conditions and vice versa
* Use of standardised format
* Compact and easier to follow one column
* Alternatives can be grouped to facilitate analysis – Provide concise descriptions of logical complex situations

Limitations of using decision table

* Large decision tables can become incomprehensible and difficult to modify
* Flowcharts are better to express the total sequence of events needed to solve a problem
* Flowcharts are more familiar to many programmers and beginners

Decision Trees – Show possible choices to take in reaching an outcome. It is identical to probability tree and easy to use

Remove Redundancies:

f.