

COMP 3

Problem Solving

Abstraction and Information Hiding

Abstraction - Simplifying a problem that is complex by removing the unnecessary details

Information - Hiding complex parts of a system behind a simple to use interface

Interface - The boundary between the front end of a system and its implementation

Comparing Algorithms

Algorithms can be compared by expressing their complexity as a function relative to the size of the problem

Big O

Time complexity - How long the algorithm takes to complete a task with a given input.

Space complexity - How much memory (RAM) the algorithm needs with a given input

Order of Growth - How quickly the complexity of a function grows as n increases.

Asymptotic behaviour - What happens to the complexity for very large n

Big O notation - for time taken for algorithm

Fastest $O(n^0)$ constant time $O(a^n)$ exponential

$O(\log n)$ logarithmic $O(n!)$ factorial slowest

$O(n)$ linear

$O(n \log n)$ Linearithmic

$O(n^2)$ Quadratic

$O(n^a)$ polynomial

Types of Problem

Non-computable - Cannot be solved using an algorithm

Intractable - Cannot be solved in a reasonable amount of time (more than polynomial complexity)

Decision - One with a Yes/No result

Problem

Undecidable - Non-computable decision problem

Heuristic - Trial and error, informed guess, approximate approach solution to an intractable problem.

Halting Problem - It's impossible to always know whether any program will eventually stop given particular input

Finite State Machines

Mealy - FSM where output is based on state and input (output on arrows)

Moore - FSM where output is based on present state only (output on states)

Deterministic - Each state has a unique trigger for each possible outgoing transition

FSA - Finite State Automaton. No output, only an accepting/halt state which means 'Yes' for the decision problem

Turing Machine - A machine which describes an algorithm. It has an I/O alphabet, an infinitely long tape in one direction, a read/write head and an FSM which acts as the set of instructions (program). All machines can be reduced to an equivalent Turing Machine. No device can be more powerful. Is the

TM can't solve a decision problem, no machine can.

Universal - A machine capable of simulating any other Machine.

Universal Turing - A ~~not~~ machine which simulates another Machine TM, with its instructions and data on the tape. This machine can solve any algorithmic problem.

Language and notation

Natural - Real Spoken and written language (English) Language

Formal - designed by an alphabet and rules of syntax Language

Regular - Can be processed by a computer (i.e. an FSM accepts it)

Regular expressions (Reg ex)

a^* 0 or more a's $a^?$ 0 or 1 a's $[ab]$ a or b

a^+ 1 or more a's $a|b$ a or b $[a-z]$ lowercase

$\backslash d$ single digit $\backslash w$ single alphanumeric character
• Single character

Used for validation and searching

Backus-Naur Form

Another way of expressing rules for constructing strings.

e.g. $\langle \text{number} \rangle ::= \langle \text{digit} \rangle | \langle \text{digit} \rangle \langle \text{number} \rangle$ names of things are in $\langle \text{tags} \rangle$
 $\langle \text{digit} \rangle ::= 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9$

$::=$ means "consists of" | means 'or' $\langle \text{digit} \rangle$ is terminal because it can't be broken down further

Reverse Polish Notation

This is also called postfix notation. Normal maths sums use infix. The 'in' or 'post' describes where the operation goes relative to the two operands. It's easier for a computer to calculate using postfix, since a stack can be used and the input sequence needs only to be read once. RPN doesn't need brackets

Infix	Postfix (RPN)
$3 + 4$	$34+$
$12 * 4$	$124*$
$(y-2) * (x+3)$	$y2-x3+*$

Programming Paradigms

Pronounced "Paradimes"

Functional Programming

Treating computation as the evaluation of mathematical functions. e.g. Haskell, F[#]

Logic Programming

Defining facts and rules. e.g. Prolog could be used to recognise people or animals from known characteristics. This is called an expert system.

Event-Driven programming

The flow of the program is determined by events such as mouse clicks, which trigger sub-routines. e.g. VB.NET

Procedural / Imperative programming

Code is executed one line after another. Code can be split into procedures / functions / sub-routines. e.g. VB.NET

Object-oriented programming (OOP)

Using instances of classes in order to create objects

Object - an instance of a class

Instantiation - The act of declaring an object as an instance of a class

Class definition - The template from which objects are created

Encapsulation - Combining the data with methods manipulating that data to form a new data-type : the class

Inheritance - A sub-class will inherit all the public properties and methods of the base/parent class

Polymorphism - ~~Over~~ An inherited method can be defined differently in different sub-classes (overriding)

Class definition

How to define a class in an exam

ClassName = Class(ParentClassName)

Public

obviously
optional, but don't
forget this!

Procedure ProcedureName (override)

Function GetValue ← There are public functions
for retrieving all private variables

Private

Value1: Value1 DataType ← e.g. String

End

Actual variables are normally private

There is no need to repeat data/ methods from parent/base classes.

Recursive Techniques

Recursive - A routine designed in terms of itself

Routine

General case - The solution in terms of itself for a value n

Base case - A value that has a solution which does not involve any reference to the general case

For example:

Function Factorial(n)

IF $n = 0$ RETURN 1

ELSE RETURN $n * \text{Factorial}(n-1)$

Base case

General case

End Function

print Factorial(4)

Output: 24

Stack Frame - Locations in the stack used to store the values regarding to each invocation of a routine.

Abstract data types / Data Structures

ADT - A data type whose properties are designed independently of any programming language

List - A collection of elements with an inherent order

Pointer - A variable that contains an address. The pointer points to the memory location with that address.

Null Pointer - A pointer that does not point to anything (\emptyset or -1)

Dynamic Data - The memory taken up by the data structure structure is variable at run time

Static Data - The memory required to store the data structure structure is declared before run time

Heap - Area of memory (pool of unused memory locations) used for allocation to dynamic data structures.

Linked Lists

Use records. Each item has the actual data, as well as a pointer to the next item in the ordered list. The list as a whole needs a pointer for the first item in the list, as well as one for the next free memory location. The final item has a null pointer.

Stacks

A LIFO (Last-In-First-Out) abstract data type. These can be only accessed using the top with two main functions
Pop - Taking an item off the ~~the~~ top of the stack
Push - Putting a new item onto the top of the stack

Errors

Overflow - Trying to push onto a full stack

Underflow - Trying to pop from an empty stack

Stacks are used to store return addresses and parameters when subroutines are called

Queues

A FIFO (First-in-First-Out) abstract data type.

Linear Queue - Elements join the queue at one end and leave at the other

Circular - The Final location links back up with the first. Pointers are used for the front and rear, and memory locations can be re-used (overwritten)

Priority Queue - Each job has an associated priority, and it is ordered by this.

Graphs

Graph - A diagram consisting of vertices joined by edges/ arcs

Degree - (of a vertex) is the number of vertices connected to that one

Labelled/ - edges are given values/ weights
weighted

Automation - Turning an abstraction into a form that can be processed by a computer

Directed - Edges have arrows, indicating direction

Simple - Undirected + each edge connects two different vertices uniquely

A graph can be represented as an:

Adjacency Matrix - Grid with "1" for connection. Good when there are a lot of edges.

Adjacency List - List of vertices connected to each vertex.
Good when there are few edges.

Trees

Tree - Connected, undirected graph with no cycles.

Rooted Tree - Tree in which one vertex is the root/start/top.
Each edge is directed away

Tree Traversal

Breadth-First - Look at all of the vertices on a level, then go down to the level below. Go left to right

Depth-First - Look at all left children until you stop at a leaf, then go back and look at the nearest right vertex.

Standard Algorithms

Insertion Sort $O(n)$

Start at the second item of the list

For each item, insert it into the correct position of the new ~~list~~ list. Each iteration, the new list gets one bigger. The new list starts as just the first item in the original list

e.g. sort the list 34, 7, 16, 23

34, 7, 16, 23	new list
7, 34, 16, 23	original list
7, 16, 34, 23	
7, 16, 23, 34	

Binary Search

Start with an ordered list

Guess the middle value

If you are too low, you now only need to search the upper half of the original list
Opposite for lower half

Guess middle of new half-sized list

Repeat

Better than linear search

Hashing

For a hash function $H(k) = v$,

$H()$ is the hash function which converts a large amount of data irreversibly into a smaller amount (fixed)

k is the key which is input into the Hash function

v is the Hash value, which is the result of the Hash Function. This is normally used as an index to an array

Collision - When two or more different keys hash to the same hash value

Open Hashing - Store a record in the "next available" location when its hash value collides. This can be solved by rehashing

Rehashing - Hashing the hash value again in an attempt to find a free memory location.

Closed Hashing - The common hash value location contains a pointer to a linked list containing all of the records Hashing to that value.

Binary Trees (searching)

These can be represented using three arrays: One for data, one for left child pointers and one for right child pointers. There are three recursive binary tree search methods

In-order search

Output result as Left child Root Right child
/Branch /Branch

Each branch can be more than one node

Draw around the tree ~~clockwise~~ from left of root.

Output node data as you pass the bottom of it

Pre-order search

Output result as Root Left child Right child
/Branch /Branch

Draw around the tree from left of root. Output node data as you pass left of it

Post-order search

Output result as Left child Right child Root
/Branch /Branch

Draw around the tree from left of root. Output node data as you pass right of it

Simulations

Model - An abstraction of a real or imaginary situation that enables an automated solution. The abstraction is a representation of the problem that leaves out unnecessary detail.

Time-based - Using a timer with set intervals

Activity-based - Queuing up different activities, one after another, ignoring time

Simulation - A computer program or network that attempts to simulate a model of a system. They allow users to study or try things that would be difficult or impossible to do in real life

Entities - The components that make up a system

State History - consists of state ~~the~~ descriptions at each of a chronological succession of instants

Simulations often use **Pseudo Random Numbers**, **Timers** and **Queues**.

Real Numbers

Floating Point Numbers

These are used in normalised form. This provides the greatest precision and range for the number of bits given. The size of the mantissa decides the precision, and the size of the exponent decides the range. The mantissa has a point between the first two digits, which must be different. Two's compliment is used.

Mantissa - Holds the significant digits of the number

Exponent - The number of places the binary point has been moved to the left to normalise the number

Significant Digits - All digits except from leading and trailing zeroes of a number

Precision - The maximum number of significant digits that can be represented.

Range - ~~flow~~ The difference in scale of possible represented numbers

Absolute Error - The difference between the actual number and the nearest representable number

Relative Error - The absolute error divided by the actual number

Cancellation - when adding a very large number and a very small ~~large~~ number.

Overflow - The number is too large to be represented by the available number of bits

Underflow - The number is too small to be represented by the available number of bits.

Decimal \rightarrow Binary

- convert number to fixed point two's compliment binary;
- count the number of positions the point moves left so that it's inbetween the first two digits (this can be negative).
This is the exponent;
- After moving the point, the mantissa is the actual number

Binary \rightarrow Decimal

- Convert exponent to decimal;
- Move binary point that many places to the right (left for <0);
- If negative, convert back to positive;
- Convert to decimal and add '.' if needed;

Operating Systems

System Program - A program that manages the operation of a computer!

OS - The software that supports a computer's basic functions:

1. Processor Scheduling
2. Memory management
3. I/O management
4. File management.

Virtual Machine - The apparent machine that the OS ~~provides~~ provides to the user, achieved by hiding the complexities of the hardware behind layers of OS software

API - (Application Programming Interface) A layer of software that allows application programs to call on the services of the OS

OS Classification

Interactive - User and computer are in direct two-way communication

Real-Time OS - Inputs processed in a timely manner (quickly) so that the output can affect the source of inputs

Network OS - OS extension for networked computers. It intercepts commands that reference resources elsewhere on the network (e.g. file server), then redirects in a manner completely transparent to the user

Sandbox - Tightly controlled set of resources for guest programs to run ~~in~~

Embedded - Dedicated computer system with very limited computer OS (or no) user interface. Designed to operate largely or completely autonomously

Desktop OS - Allows user to carry out a broad range of general purpose tasks

Server OS - Optimised to provide one or more specialised services to networked clients.

Databases

Conceptual data modelling

Database - A structured collection of data

DBMS - (Database Management system) Software which enables definition, creation, maintenance of a DB as well as providing controlled access to it

Data Model - Method of describing the data, including structure, relationships, constraints.

Conceptual Model - Representation of the data requirements independent of any DBMS

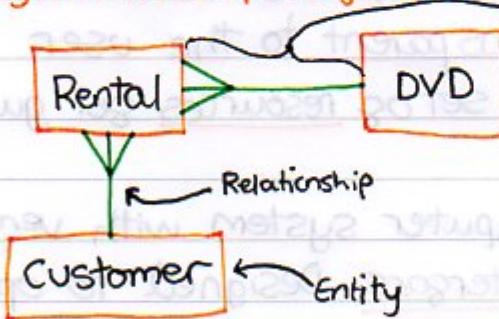
Entity - A thing about which data is recorded

Relationship - Association between two entities

Degree of Relationship - Many to one, one to one or one to many

Relationship

Entity-Relationship Diagrams



The degree of relationship between DVD and Rental is "one-to-many". Each DVD can belong to many rentals, but each rental may only be associated with one DVD.

Design

Relation - set of attributes and tuples, modelling an entity

Attribute - Entity's property or characteristic

Tuple - Row in a table (group of attributes)

Primary Key - Attribute which uniquely defines a tuple

Composite Key - Primary key, but with more than one attribute

Foreign Key - Attribute in one table which is Primary Key in another

Referential Integrity - If a value appears in a foreign key in one table, it must also appear in the primary key in another table

Relational DB - Collection of tables

Normalised entities - Set of entities that contain no redundant data

Normalisation - Act of producing a set of normalised entities. This involves adding entities to split many to many relationships into two one to many. After normalising, the DB is in Third Normal Form

SQL (Structured Query Language)

This is the declarative 4th generation language used to design and manipulate databases

DDL - (Data Definition Language) used to create a database structure. Creating tables

```
CREATE TABLE tableName (
    field1 data_type(size) PRIMARY KEY,
    field2 data_type(size)
);
```

Some data types include:

varchar - array of characters (string)

int - whole number

text

date

decimal(x, y) - x = size, y = no. decimal places

DROP TABLE tableName;

DROP DATA BASE databaseName;

DML - (Data Manipulation Language) Used to query and manipulate tables created with DDL

```
SELECT tbl1.field1, tbl2.field3 FROM tbl1, tbl2
WHERE tbl1.ID = tbl2.ID AND tbl1.field1 > '1'
ORDER BY tbl1.Date DESC;
```

`INSERT INTO tableName (field1, field2, field3)`

`VALUES ('value1', 'value2', 'value3');`

`UPDATE tableName`

`SET field1 = 'newValue'`

`WHERE field1 = 'oldValue';`

`DELETE FROM tableName`

`WHERE field1 = 'value1';`

Communication and Networking

Data Transmission - Movement of Data from one place to another

Serial Data - Bits are sent one after another along a single wire.

Transmission Reliable for long distances. e.g. USB

Parallel Data - Bits are sent down several wires simultaneously.

Transmission This is faster, as more data is sent at once, but data is prone to skew (not all data arrives at the same time) is used for long distances

Band Rate - Rate at which signals change on a wire/line

$1 \text{ Baud} = 1 \text{ Signal change per second}$

Bit Rate - Number of bits transmitted per second

Bandwidth - Range of possible signal frequencies that a transmission medium may transmit. A higher bandwidth could increase bit rate if more bits transmitted per signal

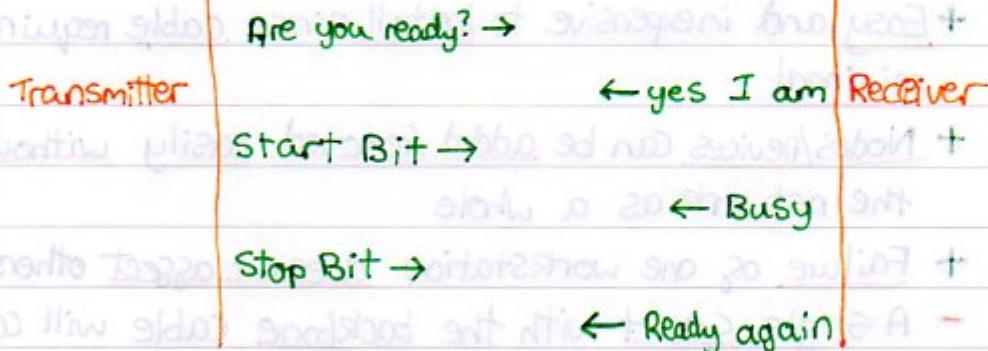
Latency - Time delay between the moment some data communication is initiated and the moment its first effect begins

Synchronous - Parallel transmission occurs usually under the direction of a clock to keep ~~every~~ everything synchronised.

Asynchronous Data Transmission - The arrival of data cannot be predicted by the receiver. Start bit is used to signal the arrival of data and to temporarily synchronise the transmitter and receiver. Stop bit used to signal the end of data transmission. Parity bits can also be used for checking data integrity.

Communication Protocol - set of pre-agreed signals, codes and rules to be used for data exchange between computers (or between a computer and a peripheral (e.g. printer)) that ensure the communication is successful.

Handshaking - transmitter and receiver exchange signals to establish connection and to ~~to~~ send data.



Baseband - A single data channel is used in which the whole bandwidth is dedicated to one channel at a time.

Broadband - The bandwidth of the line carries several data streams at the same time.

Networks

LAN - (Local area Network) Collection of computers and peripherals confined to one building or site with common electrical connection. Can be connected to other LANs, or to a **WAN**

WAN - (Wide Area Network) Set of links that connect computers and LANs that are geographically remote

Stand-alone - Computer that's not networked. Requires own computer software and peripherals

Network - Shape, layout, configuration or structure of the **Topology** connections that connect devices to a network

BUS



All workstations linked by a single 'backbone' cable. No central computer; PC's used as file server.

- + Easy and inexpensive to install since cable required is minimal
- + Nodes/devices can be added/removed easily without disrupting the network as a whole
- + Failure of one workstation doesn't affect others
- A single fault with the backbone cable will cause the whole network to fail. This could be difficult to isolate and repair
- Performance degrades as traffic and cable length increase

CSMA/CD - (Carrier Sense Multiple Access with Collision Detection) A commonly used Bus protocol for handling data collisions

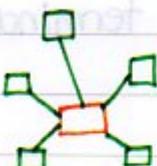
Step 1. If the bus is quiet, transmit a frame.

Step 2. If bus is busy, continue to listen until it's idle then transmit immediately.

Step 3. While transmitting, monitor bus for collision. If detected, transmit brief jamming signal (stops other computers transmitting) then stop transmitting.

Step 4. Wait a random amount of time, then attempt to transmit again, starting from the first step.

STAR



Each node is connected to a central computer, which is a hub or switch. Hub transmits received data to all nodes. Switch ~~or~~ intelligently sends packets to the right node at the MAC (Media Access control) address level.

- + A cable failure on one branch will only affect that branch
- + Collisions won't occur if two computers transmit at once
- + More secure since workstations cannot intercept messages due to there being a central switch
- + Data transmission rate can vary on each branch
- If the central switch fails, whole network is affected
- More cable is needed. The cost of cable can escalate with expense also in disguising it

¶

Segmentation - Solution to congestion on an Ethernet bus network. It involves splitting a larger non-switched network into network segments linked by bridges or routers. Smaller segments means fewer collisions happen

Network - a run of Ethernet cable to which a number of segments workstations are attached

Peer-to-peer - No dedicated servers. All computers are equal and function simultaneously as clients and servers. A computer's resources can either be shared with all others on the network, or none at all. Appropriate with few users located in the same area and where security is not an issue

Server-based - A network in which resources, security, administration and other functions are provided by dedicated servers

Most/

Thin-client - All processing and storage takes place on a central network ~~soft~~ server. Clients are merely dumb terminals with little or no processing power or local hard disk storage.
opposed to rich/thick-client network (server)
E.g. military - users don't need to carry heavy equipment, and not much loss if terminals get destroyed

Web 2.0 - ~~soft~~ Software that becomes a service accessed over the Internet. Users have the ability to publish and customise content

Web - Self-contained, modular applications that can be described, published, located and invoked over a network (normally the Web) (basically an app)

SaaS - (software as a service) Model of software development. Application is hosted as a service. No need to install/run the application or download updates. Often need to pay each time the service is used

AJAX - Technology that allows part of a web page to be updated instead of reloading the entire page. Only getting the necessary chunks results in faster loading and a smoother experience.

Wireless - A LAN with which devices and computers use radio waves/microwaves to transmit data between them.

WAP - (Wireless Access point) Allows devices operating wirelessly to connect to a wired network

Wi-fi - Trademarked technologies that support wireless networking. WLANs use WEP (wired equivalent Privacy) security, or newer and more secure WPA (wi-fi protected Access)

Bluetooth - wireless protocol for exchanging data over short distances from fixed and mobile devices. A PAN (Personal Area Network) is created

Inter-networking - Connecting multiple LANs through the use of gateways that provide a common method of routing information packets ~~between~~ between the networks

Router - A device that receives packets from one host (computer) or router and uses the destination IP address on the packets to pass them on, correctly formatted. Each router maintains a table of routes (e.g. knowing what country to send packets to). A router will know about its sub-networks and their IP address ranges, but not completely about its super-networks

Gateway - Device used to connect networks using different protocols so that information can be passed from one system to another

Subnet Mask - For a network/sub-network, defines how many possibilities for each IP address number there are. It's used to check if a computer is on the same segment by sending the network identifier (fixed part with 0.0 etc.) 255.255.255.0 means that the first three numbers are fixed

Routable (Public) - Assigned by RIPE NCC in Europe. These are usual IP Address IP addresses used to ~~be~~ be accessed over the Internet.

Non-routable (private) - can't be used on the Internet. Organisations use IP Address them for LANs, because they can't be hacked from the outside. e.g. 10.0.0.0 to 10.255.255.255 or 192.168.0.0 to 192.168.255.255

Server-side Scripting

Web server - Program, an executable or a script that is interpreted by an interpreter running on the web server, which extends the functionality of the server and allows it to generate content at the time of the HTTP request

CGI - (common Gateway Interface) A gateway between a web server and the Web Server extension that tells the server how to send information to it, and what to do after receiving information from it

Dynamic Web - Content that is generated when the web browser page content request is received

Internet Security

Firewall - Hardware device or program that controls traffic between the Internet and a private network or computer. It controls which data packets should be allowed through. Traffic can be blocked from specific IP addresses, domain names or port numbers

Packet Filtering - Firewall analyses packets sent against filters (rules) that determine whether they are allowed through or blocked

Proxy Server - Requesting computer doesn't immediately come into contact with the server. This allows info to be filtered first

Encryption - Using an algorithm ^{and a key} to convert a message data into form that is not understandable without the key to decrypt the text.

Cryptanalysis - ~~Trying to~~ Trying to find the plain text without the decryption key

Break the Code - Guessing or deducing the key

Key Encryption (RSA Method)

Each computer has a public key and a private key. These undo each other. e.g. encryption with the public key can be decrypted with the private key

For two computers, A and B, A sends a message to B

1. A encrypts the message with B's public key
2. A sends encrypted message to B
3. B decrypts using its private key

For more security, A could first encrypt with its private key to

1. A encrypts using A's private key and B's public key
2. A sends message to B
3. B decrypts using B's private key and A's public key

This method makes the transmission more unique to the two computers involved.

Digital Signature

This is a "Stamp" added to documents and other data which is unique to the sender and authenticates the message by making sure that changes cannot go undetected.

1. Message is hashed to produce a message digest
2. Digest encrypted using A's private key. This is the signature.
3. Signature appended to the message
4. Encrypted using B's public key
5. Sent to B

1. Message decrypted using B's private key
2. Signature decrypted with A's public key. This is the original digest
3. Message is hashed
4. Is decrypted message digest = hashed message, message has not been tampered with

Digital ~~cert~~ - Something issued by a certification ~~organisation~~ authority used to verify the ~~organisation~~ person is who they claim to be. It contains a copy of the certificate holder's public key, name, serial number and expiration dates. It's signed using the certification authority's private key to verify that the certificate is real. A digital certificate is used to verify that a public-private key pair belongs to the sender of the encrypted message.

Virus Detection

Worm - Small program that exploits a network security weakness to replicate itself through networks. It may attack computers and allow someone else to remote login to the computer.

Remote - When someone connects to a computer via the Internet. They could erase files and execute programs.

Spam - unwanted junk e-mails sent to many e-mail addresses.

Phishing - Pretending to be someone else in order to steal passwords, bank account info and other personal information.

Pharming - When a phisher changes DNS server information so that customers are directed to a different, incorrect site.

Trojan - A program that hides in useful or desirable software, but attacks the computer it infects.

Virus - checks files against a dictionary of known viruses. If a virus is found, deletion is attempted. If this doesn't work, it is quarantined (moved to a separated area of the hard drive where it can't infect other files).

Computer Security Procedures

Authentication - Legitimate users can be authenticated using passwords, biometric data, or digital certificates. An organisation may only accept e-mails which have been digitally certified and signed.

Authorisation - An authorised user has a user ID and password. Passwords and encryption are used to keep data secret from unauthorised people.

Accounting - Creating an account of every activity made on a network or computer, making it easier to locate security breaches. e.g. keeping a network Internet activity log

Simulations (continued.)

Methodology of Simulation (7 steps)

① Problem Formulation

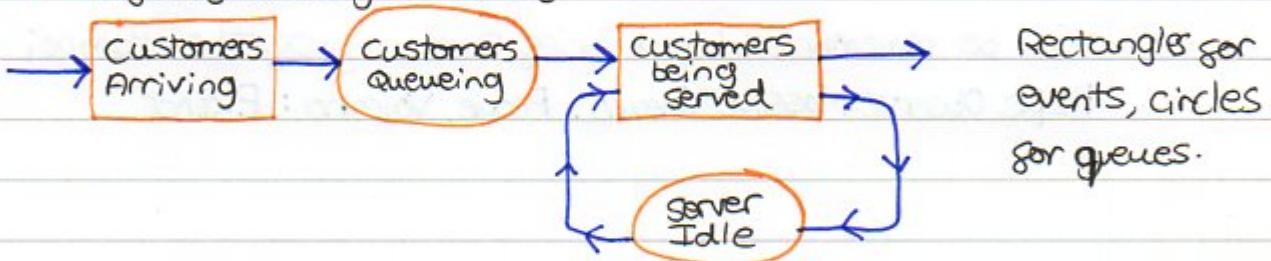
- State objectives
- State criteria by which results will be measured

② Observation of the real System

- Decide what data is needed (e.g. average arrival rate)
- List parts of system and their relationships

③ Formulation of the Model

- Decide important parts and define their properties and states
- Use cycle diagrams, e.g.:



Time-Driven - Master clock incremented by regular time unit and program checks if activity should have started or finished

Activity-driven - The next event to occur occurs and master clock is incremented by an appropriate amount of time.

④ Model Validation

- Ensure model is correct before making program
- Do a hand simulation (dry run) of the life cycle diagram

⑤ Program Formulation

- Use object-oriented design for components
- Include lots of diagnostics and comments
- Visualise results using tables, graphs or animation

⑥ Program Validation

- Verify program using known historical data, and compare with actual outcomes
- Or forecast a future scenario and see how good the prediction is.

⑦ Design of Experiment and Analysis of Results

- Run the simulation a lot with different values for variables to try to find the true effect of the variables
- Use this to improve the real system

Remember, POMVPE,

Problem Observe Model Validation Program Validation Experiment

Can be remembered by a brief summary of the Pompeii incident:

People Observe Massive Volcano. People, Volcano : Extinct