

Unix

Introduction

- What is UNIX?
 - A multi-user, multi-tasking, machine-independent computer operating system
 - An operating system controls other parts of a computer system by allocating system resources such as memory and scheduling CPU usage.
 - Designed as a software development environment
 - UNIX trademark: AT&T → Novell → X/Open Co. Ltd.
- History
 - 1969 – developed at AT&T Bell Labs in New Jersey
 - 1975 – AT&T makes UNIX widely available, especially to educational institutions at minimal cost.
 - 1984 – UC in Berkeley adds many major enhancements.
 - Now – OS of choice for engineering and scientific workstations. Two main versions – AT&T System V and Berkeley Standard Distribution
 - Future – Portable version.
- Philosophy
 - An individual program accomplishes one task well and can pass on output to be the input of another program without knowing what that program might be. Solving several small problems to solve a large one. Uses terse commands and messages – reduces typing and screen output.
- Why UNIX
 - Hardware independent
 - Written in C, not a platform specific assembly language
 - Good for software development
 - Rich set of tools
 - versatile command language

Organisation

- Kernel:
 - Core of system; Memory-resident control program; Manages resources (memory allocation, CPU scheduling, data transfer, executes commands from the shell, enforcing permissions)
- Shell:
 - Placed in the shell on login; enter commands at the shell prompt; shell interprets commands and sends them to the kernel then displays results on the screen;
 - Several available shells – different users may use different shells. E.g., Bourne (sh), C (csh), Korn (ksh), TC (tcsh), Bourne Again (bash). Each shell includes its own programming language. Shell scripts can be used to accomplish a series of tasks.
- Utilities:
 - UNIX provides several hundred utility programs (commands) that can accomplish many functions, e.g., editing, file maintenance, printing, sorting, etc. Single functions can be grouped together to perform more complex tasks (modular)

Usage

- UNIX commands are executable files and built-in programs
- Everything in UNIX is case-sensitive
- Command syntax: `command option argument`
 - Options: usually a hyphen followed by a single letter
 - some commands accept multiple options
 - Arguments: most commands are used together with one or more arguments
 - if no argument is supplied, a default one is often assumed.
- E.g., `wc -w file1` `wc` is the command (count); `-w` is the option (*word* count); `file1` is the argument
- Multiple commands can be put on the same line using a semicolon as a separator.
- Commands can continue beyond one line using the backslash
- Some common commands:

Revision Notes

ls, mv, mkdir, rm, cp, cd, man (followed by another command or with -k + keyword), whatis (also followed by another command), chmod, passwd, finger, who, logout

- Some common special characters:

; - executes multiple commands on one line
| - pipe (e.g. who | sort)
> - redirects output
* - wildcard for any number of characters (e.g. ls *.txt)
? - wildcard for a single character
[] - denotes a character within a certain range (e.g., ls file[1-4].txt)

- Some common control keys:

CTRL-u erase current command line
CTRL-c stop/kill a command
CTRL-h backspace
CTRL-z suspend a command
CTRL-s stop scrolling
CTRL-q continue scrolling
CTRL-d exit from an interactive program (signals end of data)

Network Related Utilities

talk, finger, ping, ftp, telnet, rlogin, rsh, rcp

File Permissions

r – read(4), w – write(2), x – execute(1)

u – user, g – group, o – other

E.g, User rw, Group rw, Other --. = 4+2 4+2 0 = 660

To show permissions, use the -l option with the ls command

Hierarchical File Structure

All files organized into a multi-leveled hierarchy called a directory tree.

At the top of the tree is the root directory, represented by /

File Types

- Ordinary files: store information. Always located within a directory. Do not contain other files.
- Directories: Branching points in the file tree. Organizes groups of files, may contain files or other directories. All descend from /
- Special files: Represents real physical devices, e.g. printer. UNIX considers any device to be a file, including the terminal. Usually found only in directories called /dev
- Pipes: Piping commands to another one creates a temporary file with the output data from the first command to be passed to the second command.

File Names

- File names can include most characters, but avoid spaces, tabs and special characters.
- Case sensitive
- Up to 256 characters
- Extensions: .a – archive, .c – C source code, .f – Fortran source code, .o – Object/executable code; .z – compressed file.
- Hidden files start with .
- No two files in the same directory can have the same name
- Reserved names: /, ., .., ~

Pathnames

- Specify location in the file tree
- Absolute pathname – relative from root, always begins with /, e.g., /usr/local/doc/sample.f
- Relative pathname – relative from current directory. Never begins with /, e.g. ../bin or ~/directory

Revision Notes

Common file commands

- ls (options: -a (displays hidden files too) , -l (details), -F (marks directories with / and executable files with *))
- more, pg, less
- head -n filename (displays the first n lines of a file)
- tail -n filename (displays the last n lines of a file)
- cat – dumps entire file to the screen without paging (-b – shows line numbers)
- cp – copy; cp -R dir1 dir2 (copies contents of dir1 to dir2)
- mv
- rm (-r – removes all files and the directory of a directory)
- -i can be used with cp, mv and rm to give a prompt
- file filename (displays file type)
- find path -name filename -print (finds a file named filename starting the search from path and then prints the results to the screen)
- diff file1 file2 – compares two files (-iw ignores case and spaces) or directories
- ln source linkname – creates a link (shortcut) to the file source named linkname
- sort options filename – sorts the lines in filename. (Options include -o (output to another file named afterwards), -u (remove all duplicate lines from the file))
- pwd – returns the current directory
- mkdir
- cd
- rmdir (remove directory as long as it's empty), rm -r (removes all files and directory)

Emacs

- is a text editor for UNIX.
- has tight integration with the command shell, so it can run programs.

Window divided into three areas: Edit, Status line, Mini-buffer (holds commands you give).

Common commands:

CTRL-v – scroll down one screen

ESC, v – scroll up one screen

CTRL-x, CTRL-s – save current file

CTRL-x, f – switches between buffers

CTRL-x, u – undo

Incremental Search - Esc, x, isearch-forward (or isearch-backwards), TAB (command completion)
Then enter search string

Global Replace - Esc, x, query-replace, TAB
Search string, then target string

Spell Checker – Esc, x, spell-buffer

Window Split – CTRL-x, 2 – split horizontal; CTRL-x, 3 – split vertical; CTRL-x, 1 – maximize current window; CTRL-x, o – move cursor to next window

Modes – Esc, x, fundamental mode; Esc, x, text-mode; Esc, x, C++ mode

The Internet

- Connection of networks (LANs and WANs) throughout the world using the TCP/IP protocol
- Local, global, flexible, fast, effective, ownerless, poorly organised.

History

- 1950s - US DoD works with Rand Corp. to create a network with no single 'command' machine
- Messages would be divided into packets, each labeled with its destination address, that would, probably by different routes, end up at the destination machine. The idea is that if one machine was destroyed, the others could still function by using alternative routes. The software that controls the packeting and reassembly of the packets is TCP/IP (Transmission Control Protocol/Internet Protocol)
- By 1969, the new network was established, under the name of ARPANet. Before long, research universities and defense contractors joined the network.

Revision Notes

- 1990 – Tim Berners-Lee used the technology to work with his far-flung colleagues. He saw the set of links from computer to computer as a spider's web. CERN, Geneva is generally seen as the birthplace of the web. Berners-Lee also established the idea of hyperlinks.
- 1993 – Marc Andreessen led the team that invented the GUI *browser*. First browser was *Mosaic*.
- As TCP/IP is publicly available, anyone can just hookup to the network.
- Summary: Internet development is due to four factors: i) the universal TCP/IP standard, ii) web-like ability to link between sites, iii) ease of a GUI browser, iv) growth of PCs and LANs.

Email

- Transmission of electronic messages
- Most popular service on the net

Newsgroups

- Collection of discussion groups accessible with specialized software.
- Groups are divided into the following groups: alt (alternative subjects/lifestyles), biz, comp, rec (recreation and hobbies), soc (society and culture), talk (daily news), misc
- In total, there are about 30,000 newsgroups and growing.

Telnet

- Remote login to another computer.
- Can use specialized software or use telnet://robin.cs.nott.ac.uk in the browser.

FTP

- File Transfer Protocol enables copying of files to/from another computer.

Web

- Access through ISP or an online service in conjunction with a browser.
- Websites are addressed by a URL (Uniform Resource Locator). Uses the http (HyperText transfer protocol). URL ends with a top-level domain name (.com, .net, .edu, etc.)
- Shockwave, Flash, Java, etc.
- Search Engines: AltaVista (1995 DEC), Excite (1993 Stanford), Hotbot (UC Berkeley), Lycos (1994 Carnegie Mellon), Yahoo (directory). Query Languages (boolean operators, e.g. and, or, not, "", -, +, -)

Security

- PGP file encryption
- SSL secure channel
- Cookies: text files on client computer used to track users. Can invade privacy.

Static Electricity

All components can be damaged by ESD – Electrostatic discharge. Damage by i) discharge from person to device, ii) charged device being discharged, iii) proximity to charged object.

Latent failure.

- Precautions:
- ESD sheilding
 - Mats, footwear, clothing
 - Wristbands, work surfaces
 - Tools
 - Air ionisers

Software development and Application development software

Development Process:

Requirements => Specs => Implementation => Validation => Maintenance =>

Revision Notes

Operating Systems

- Fundamental set of programs that lie between applications software. Controls the interactions among all applications and devices and establishes a user interface.

Personal Computers

- First OS were text-based command-line interfaces.
- 1981 saw the first GUI
- Apple Mac OS released in 1984
- Microsoft's Win 3.x were not true operating systems, but 95 and ME, NT and 2000 are.
- The newer OS, Linux, is based on UNIX

Applications Software

- Application Suites, e.g. MS Office.
- Business software
- Communications software
- Multimedia Applications
- Ethical Issues (software piracy, CD-R, counterfeits)

Storage

Magnetic Media

- Tape (long, narrow plastic strips coated with iron oxide)
- Floppy Disk (circular tracks around a disk)
- Hard Disk (rigid platters (thus data can be placed at higher density), head does not touch plate, multiple platters)
- Zip disk (metal not plastic)

Optical Media

- CD (compact disc (optical media: disc, magnetic media, disk)
- CD-Rs
- CD-RWs - data-bearing layer can take a reflective or non-reflective state which can be reversed.
- DVD (digital versatile disc) – greater density of data, more efficient encoding, can be double-sided

Input Devices

- Keyboard, Mouse, Touch, Pen-based, Scanners, Voice, Digital cameras
- Wireless technology
- Scanners – MICR (magnetic), OCR (optical character recognition), UPC (universal product code)

Output Devices

- Hardcopy: Impact: Dot-matrix printers (75-100 dpi)
Non-impact: Ink-jet printer (300-600 dpi), laser printer (600-1200 dpi), plotter
- Softcopy: CRT/LCD screen
Voice Synthesis, MIDI
- Monitor properties: Size (diagonally), Resolution (width x height), Dot pitch (distance between pixels), Color depth (number of colors).

CPU

Two main parts are ALU and CU. ALU – Arithmetic Logic Unit – performs basic calculations (+/-) and comparisons (<, >, <=, =>). CU – Control Unit.

Memory Types

RAM – write/read. Volatile – loses contents on power down.

ROM – read only. PROM – ready programmed. EPROM – can be erased by UV light. EEPROM – can be erased by an electrical charge.

Revision Notes

Buses: transports data and instructions within the computer

Ports: transports data and instructions between computer and peripherals. Serial: slow devices (keyboard, mouse); Parallel: faster (printers, scanners); USB: allows devices to be 'daisy-chained'; FireWire (IEEE 1394): very fast (video).

Speed and Power: Processor speed is not just a matter of clock speed. Also consider how many bits it can process at one time. E.g., Pentium 8B but Apple G4 16B. Also bus speed and width and memory speed.

Programming Languages

- Universal (can accomplish any programmable task)
- Natural, Expressive
- Implementable
- Efficient and Reliable
- Maintainable

1958-62: FORTRAN: Scientific programming

1959: LISP: AI applications

1960: COBOL: Business applications

1964: BASIC: Very limited, but very easy

1971: PASCAL: Teaching language

1972: C: Little new development, but a much more efficient implementation.

1972: Prolog: Non-procedural, goal-orientated.

1985: C++: Based on C, but with an Object Orientated emphasis.....

1988: Perl: Partially compiled scripting. UNIX debugging, CGI programming

1995: Java: Simplified descendant of C++.

Scheduling and memory management

- Allows multiple tasks to run simultaneously.

Running a process: Load executable file from secondary storage into memory.

Swapping: CPU swaps between processes. It must therefore preserve the state of the outgoing process and restore the state of the incoming process. The state is captured in its *core image*. May swap i) when process completes, ii) until I/O request, iii) until OS interrupts. The *scheduler* (part of the Kernel) chooses which process to run next (round robin or dynamic priority) from the process table.

Relocation: Allocation of virtual memory address can be done when program is loaded (load-time relocation) or every time an address is referenced (runtime relocation).

Single process system.

Fixed partition system: Different sized, but fixed sized partitions. Largest partition must be able to hold the largest program. Results in fragmentation of memory due to 'holes' of unused space.

Segmented Systems: memory split into multiple, variable size segments.

Paging and virtual memory. Page table. Processes can be larger than physical memory. Page faults.

Networking

- Allows multiple machines to be connected together, but each machine can accomplish their own tasks.

Classification: By Size (LAN, MAN, WAN); By Type (PointToPoint, Broadcast); Mobility (Mobile, Fixed)

Issues: Scale, Speed, Management, Security, Billing

Transmission Media

Copper Wire – Coaxial cable (BNC connection): Cheap, easy to install; Can interfere, can be tapped

Unshielded Twisted Pair (RJ-45 connection): Twisting shields against EM interference.

Revision Notes

Glass / Optical Fibre: No interference, light can be transmitted a long way, light can encode more info than electrical wire therefore greater bandwidth, only need one fibre, unlike electrical wires where a pair is always needed; Difficult to install, locating problem sections of wire is hard; repair of broken sections is difficult.

Microwave: Directional - Work well if no obstructions, e.g.. roof to roof; Won' t pass through metal

Infrared: Can equip a room, such that all computer are attached to network. But not permanently (physically); Short range; Line of sight required.

Laser: Very powerful, directional light; Works very well with optical fibre as the light source; Must have a line of sight, with out an optical fibre, e.g. vegetation, fog etc will break a link.

Radio Wave: No physical connection needed; In order to transmit large distances, need large aerial; Wont bend round earth.

Satellite: Radio waves, can be bent round the earth; One satellite can process many radio frequencies; Expensive (gotta put it in space... Best (only) idea is to lease bandwidth on one)

Switching

1. *Circuit switching*: Establishes dedicated link between points. Unused capacity cannot be used by other connections. Data is not delayed at the switches. Involves Circuit Establishment – Data Transfer – Circuit termination. E.g., Telephone network, ISDN.

2. *Packet switching*: Splits data into packets (with a header and trailer specifying length and error checking mechanisms, etc.) which are passed from node to node until it reaching the destination – *Routing*. At each node the packet is fully received then forwarded – *Store-and-Forward*.

3. *Virtual Packet Switching*: Sends packets along a single specified path. Guarantees in-sequence delivery.

LANs

Invented 1960s. Low cost.

Star; Ring; Bus

The Internet is a packet-switching network. Stations called Hosts; Nodes called Routers.

ASDL

- Uses existing telephone lines. Does not change into audio signal. Does not interfere with telephone signals.
- High data rates, no sharing of bandwidth, continuous and instantaneous connectivity
- Limited range (degradation of signals) – less suitable for rural locations.

HTTP

Http protocol is built on top of TCP/IP.

Client sends http request to server → Requests specified URL → Server sends response.

Although, GET (retrieve URL) is by far the most widely used command, http can also HEAD (retrieve response header), POST (post data to server), PUT (put page on server), DELETE (delete page from server).

In http 1.0, client opens a separate TCP connection for each requested object.

In http 1.1 supports persistent connections – all requests use the same connection.

HTML file sent first, then any embedded objects.

Dynamic content: CGI and ASP (active server pages) use programs hosted on the server.

Caching

Problems: If page changes or location changes, page in cache will be wrong. 'Expiration date' – time-to-live. 'if-modified-since' query.

Databases

Structured collection of data.

Managed by the DBMS (Database Management System)

DBMS: Data entry, Querying, Access control, Authentication, Concurrency control, Report generation, Views, crash control and recovery, logging and accounting.

Types of Database

Hierarchical databases – records arranged in a tree structure.

Easy to implement, fast access for pre-defined queries; redundancy ad-hoc queries may be nearly impossible, changing structure is expensive and difficult.

Relational databases – each relation stores information about one type of object. Rows = tuples, Columns = attributes.

Key to a RDBMS is a powerful query language that can retrieve, e.g., “all pupils whose department is computer science”

Excellent for ad-hoc queries, flexible and easy to restructure, sound theoretical basis; complexity of implementation, may require considerable computing power.

Networked Databases – not so hierarchically strict - each child can have several parents.

Less redundancy; Complex to implement and to model with, changing structure is still expensive and difficult.

Object Orientated databases – data consists of objects linked via relationships. Objects define both data and operations on the data.

Flexible data model, integration with programming languages, immature - no really good systems yet.

The History of Computing

Charles Babbage (b. 1791): Difference Engine, Analytical Engine, with Ada, Countess of Lovelace (first computer programmer)

1880: Herman Hollerith - Tabulating machine enabled the census to be announced very much quicker.

1896: Hollerith founded the successful Tabulating Machine Company, which in 1924, merged with two other companies to form the International Business Machines Corporation - IBM.

From 1924 to 1956 Thomas J Watson rules IBM – calculators and computers.

1940s: ENIAC, for Electronic Numerical Integrator and Calculator

1944: Harvard Mark I was born.

The First Generation, 1951-1958

The Vacuum Tube

Generated a great deal of heat, causing many problems in temperature regulation and climate control. In addition, although all the tubes had to be working simultaneously, they were subject to frequent burnout, and the people operating the computer often did not know whether the problem was in the programming or in the machine. Used magnetic cores to provide memory. These consisted of small, doughnut-shaped rings about the size of pinheads, which were strung like beads on intersecting thin wires. To supplement primary storage, first-generation computers stored data on punched cards. In 1957 magnetic tape was introduced as a faster, more compact method of storing data.

The Second Generation, 1959-1964

The Transistor

Bell labs: Develop the transistor, a small device that transfers electronic signals across a resistor.

Needed no warm-up time, consumed less energy, and were faster and more reliable.

Assembly languages.

After the development of symbolic languages came high-level languages, such as FORTRAN (1954) and COBOL (1959). Also, in 1962 the first removable disk pack was marketed. Disk storage supplemented magnetic tape systems and enabled users to have fast access to desired data.

Revision Notes

The Third Generation, 1965-1970

The Integrated Circuit

An integrated circuit (IC) is a complete electronic circuit on a small chip of silicon. In 1965 integrated circuits began to replace transistors in computers. The beginning of the third generation was trumpeted by the IBM 360 series in 1964.

The Fourth Generation, 1971-present

The Microprocessor

The Fifth Generation, Onward

Artificial intelligence, expert systems, and natural language.
Connectivity

The Story of Personal Computers

1975: MITS Altair - loaded with switches and dials but with no keyboard or screen.

1977: Apple Computer: Apple II

1981: The IBM PC which was launched – industry standard.

Microsoft supplied the operating system for the original IBM personal computer.

There are two critical points to be understood regarding the history of the Internet. The first is that the Internet was started as ARPANet, a network of equal computers that could survive a nuclear attack. Second, the Internet was made attractive to the average user by Dr. Tim Berners-Lee, who came up with the notion of links, and Marc Andreessen, who produced the first browser.

Security and Privacy

Authentication, Access control, Data integrity, Confidentiality.

Threat analysis

Attack goals: Control the host, Disable the host

Link encryption: maintaining confidentiality between machines; outsiders are blocked.

IP security – protects traffic between machines – IPSO (IP security option – adds a label to each packet); PPTP (PointToPoint Tunneling protocol); SSL (secure socket layer).

Firewall: Controlled access to the net – packet filtering (blocks certain packets); circuit filtering (blocks specific circuits)

Worm: moves from machine to machine; plants itself as a separate file.

Viruses: illicit instructions that pass themselves on to other programs.

Defence: fingerprints, checksums, backups.

UNIX security: passwords, file access, super-users, process control, audit controls.

Cryptography: conceal content of message. Encipher: Plaintext (through cipher) to Ciphertext.

Decipher: Ciphertext (through key) to Plaintext.

CERT – Computer Emergency Response Team.

Social Impacts of Computer Technology

Exclusion of the poor and elderly

Domination of UK/US – 87% of websites are in English

Loss of cultural identity – globalisation

Increases productivity, but also increases unemployment. Not just the number of jobs that change – also the function of people's jobs.

Privacy – workplace surveillance.

Computer Graphics

Pixel: smallest component of an image. More pixels per given area, crisper the image – resolution.

Colour depth: 2 colours (1 bit), 256 colours (8 bit), etc.

Spatial resolution: E.g., scanners 1200 dpi; monitors 80 dpi; printers 600 dpi; press 3000 dpi.

Spectral resolution: Perceptable frequency range.

Radiometric resolution: amount of values a data file can take.

Temporal resolution: how often (I.e. take a picture every two days)

Raster graphic: represented by a series of values in a grid (what colour is each pixel).

Vector graphic: represented by series of lines and shapes. Require less memory.

Colour perception: Hue (frequency), Saturation (how far away from grey at same intensity), Brightness (perceived intensity - how many photons per unit time)

Colour synthesis: Additive (Red + Green + Blue make a colour)
Subtractive (reflects all colours except the desired colour (cyan, magenta, yellow))
Optical – blending of moving colours – not used in computer graphics.
Diffuse – small dots of colour in varying densities.

Imaging hardware: Drum scanners (reflected light measured by photocell) 2000dpi
Charge-coupled device scanners (image read by long thin strips of CCD).

Monitors: CRT, LCD, Electroluminescent (emits light when exposed to high magnetic charge), Plasma (tiny neon bulbs).

Printers: Dot-matrix, Electrostatic (puts charge onto paper), Laser (laser puts charge on drum), Ink-jet, Thermal transfer (wax to paper).

Colour reproduction.

Picture formats: bmp (uncompressed raster – windows), tiff (uncompressed raster), gif (8-bit compressed), jpg/jpeg (8- or 24-bit compressed), ps (postscript – page-description language),

Virtual Reality

Distinction between VR and CG (interaction)

'Cyberspace' – global communications and computing infrastructure

VE (Virtual Environments): Any surrounding other than the one in which the participant actually exists, but in which the participant can believe they are present or can behave and operate as if they are present.

Augmented reality – takes what is real and adds to it in such a way that the user obtains more information. E.g., fighter pilot, brain surgeon.

Distributed VR – over a network.

Immersive VR – helmet mounted, etc.

Desktop VR

Simulation systems

Immersion Systems – HMD

Projection systems – 'Cave'

VR generator

Manipulation and Control Devices: trackball, 'SpaceOrb', instrumented glove, full body suit, mechanical armature.

Revision Notes

Input Processor → Simulation Processor (handles interactions, object actions, simulation of physical laws, etc.) → Rendering Process (visual, auditory, haptic (touch/force)) with World Database.

AVRIL – A Virtual Reality Interface Library (supports 3D rendering and provides drivers)

Multiverse – multi-user, non-immersive, X-windows based VR system.

VRML – internet VR language.

Alice – Interactive Graphics Programming Environment for Windows.

Multimedia

- Discrete (text and images)
- Video and audio (continuous – time-dependent)

Video – constant time between frames. 16 fps is just sufficient for smooth motion.

RSVP – resource reservation protocol

Quality of Service: Video - frames/sec, resolution, aspect ratio, colour, grayscale, width , height

Audio – telephone quality – 64Kbps; CD quality – 1.4Mbps

Compression: Size, Time-sensitive media, unpredictable network performance

Lossless: removes redundancies

Lossy: removes redundancies *and* non-essential information
