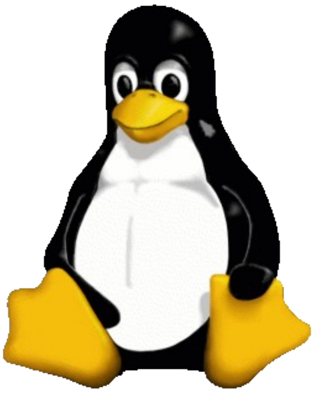
Embracing the Penguin

Embracing the Penguin



A short guide to the new LINUX PC, how it works (or should), and some things that might be helpful. The penguin reference is for Tux, the the official mascot of the Linux kernel. This, and some other images here, have been shamelessly stolen from Wikipedia.

Table of Contents

The Basics

The new LINUX PC can do most things you would want except for run Windows based games to any useful degree. However, you should consider the following before criticising this:

1. A dedicated games console like the PlayStation / Xbox / Wii provides a much better gaming experience than any comparable-cost PC.
2. What you loose in software compatibility you gain by being highly immune to viruses, etc.
3. Life is short, get out and live more rather than playing on your PC!

First we cover the basics, which is enough to get by with most things. Finally there is a how & why explanation, this is not essential but helps you avoid problems and ultimately become more productive with your PC.

Starting up & Logging In

To start the computer switch on the mains power (AC supply), press the 'power' button on the front of the PC and also on the monitor and the loudspeakers (if desired). If you need to use the printer or speakers, etc, also press their power buttons. Also note:

- The laser printer takes around 1m30s to start up and should only be switched on ***after the main power block is on*** if you expect to use it.
- Switch the laser printer off ***before you switch off the mains power block***. If you leave things on for long periods then just leave the printer on as it will go in to a power saving mode 5 minutes after it was last used (configurable).

After around 30 seconds or so you will be presented with the main log-in screen. Enter your user name (type it in, then press the 'return key'), then your password.

Both are case sensitive and it seemed unhappy with capitalised user names, hence all the user names are all in lower case. Good passwords[1] with a safe written down copy[2] at home are generally a better bet that easy to remember ones that can be guessed but you did not write down, but do not use the same password for your email accounts as your main log-in[3].

Even if you know the password of the other accounts, please make use of your **own account** for doing stuff. This keeps your own preferences and last web site visited, etc, all to your own liking.

The default permissions allow you to read others' files (and vice-versa) in any case, so you can print others photos, etc, if you need to. If you want more privacy, then use the permissions to make certain directories inaccessible to other users (see later for the details of how that works).

[1] A good password is not a dictionary word or common name, is at least 6 characters long, and maybe has unusual capitalisation, so it is difficult for a computer to find by brute-force searching.

[2] The yellow post-it note with passwords stuck to the monitor is a classic example of how not to keep a secure system!

[3] Most email systems send the user name and password unencrypted to the servers, so they are visible to others on any network used, hence the advice to keep them separate from the ones protecting your computer. This prevents a network snoop logging in to your PC with those details.

Logging in a 2nd Person

If you are using the PC, and someone else needs to use it temporarily, you have two options:

- Close everything you are doing, log out, then allow them to log in.
- Go to the top-right and in the drop-down user menu select the other account for them to log in with.

If you have finished what you were doing, then the first option is best. If not, then use the second option and have two (or more) users active at one time. When the 2nd person is done, they should log out.

If the person is not an existing user but, for example, a visitor wanting to check some web site, then use the 'Guest session' account as this limits them to very basic stuff and they can't access your files[4]. This is also a very safe option for general browsing.

Finally, you might well want to consider have a couple of user accounts for yourself, to keep work and play separate (see Users and Groups on page 52).

Shutting Down

Properly shutting down is very important if you do not want to loose data! This applies to all types of computer (not just LINUX based ones) so please read the boring section later in this document to understand *why* this is so important.

As the screen-saver powers off the monitor after 40 minutes or so, it is easy to assume the PC has stopped already and go and switch off the AC supply, so if you are not *absolutely sure* then wiggle the mouse to check it is really off first!

Generally you just go to the user name drop-down menu (top right) and select the bottom “Shut Down” option, or use the top-left menu System → Shutdown and select “Shut Down” to stop the PC in an orderly manner.

NOTE! There could be more than one person logged in so please be considerate before forcing a shut down. Ideally log out everyone else first.

I configured the PC's power switch to also shut down in an orderly manner if pressed briefly (see System → Preferences → Power Management 'general' tab). If you hold it down for 5 seconds it will force an 'unexpected' shut down. Only use that action as a ***last resort!***

If no-one has logged in (the main log-in screen is waiting for you) then go to the bottom left and there is an option to shut down there, please use that as the civilised method.

If the printer is doing nothing it is safe enough to switch the printer off, however, ***if the printer is busy then wait until it stops!*** Once the laser printer is off you can switch off the mains power.

[4] This creates a temporary home folder for them in `/tmp` which you need not worry about. It is cleared automatically when the PC starts up.

Accessing the Internet

There are two web browsers installed, FireFox and Opera. Generally FireFox works on more web sites, but Opera is just *nicer to use* than most other web browsers, especially once you get used to the mouse gesture stuff.

Choose the Applications → Internet menu and select FireFox or Opera (as preferred), or just click once on the icons along the top left-ish bar.

Both browsers support 'tabs' that allow you to have several web pages open at once (e.g. choose the File → New Tab or use the Ctrl+T key combination, then enter the web address or search topic).

Installed to FireFox is the “Adblock Plus” plug-in, this blocks most of the annoying advertisement banners that can take longer to load than the page you want to read.

The default I set for Opera is not to animate images (e.g. jumping ads), to block plug-ins (the Adobe flash player is commonly used for ads and is time consuming to load), and not to play sound in web pages. You can edit these preferences per-site, for example, to tun on plug-ins for YouTube if you want to use that site, etc.

While LINUX/FireFox/Opera are ***much less*** open to Internet attacks than Windows/IE has been[5], you should still act sensibly as no computer system is immune to vulnerabilities. So far Opera has been a bit more secure than any of the other modern web browsers, so I use it for general surfing where you could end up in the seedier side of things or, just as likely these day, to visit a reputable web site that unfortunately has been compromised for such “drive-by attacks”[6].

As a rule, ***nothing*** that appears in a web browser should be trusted, and ***never*** follow links in emails that claim to be from a bank, etc. Most are frauds and designed to take you to a copy web site that is there for “phishing” - the process of getting log-in details for something profitable like your bank or sites you shop at.

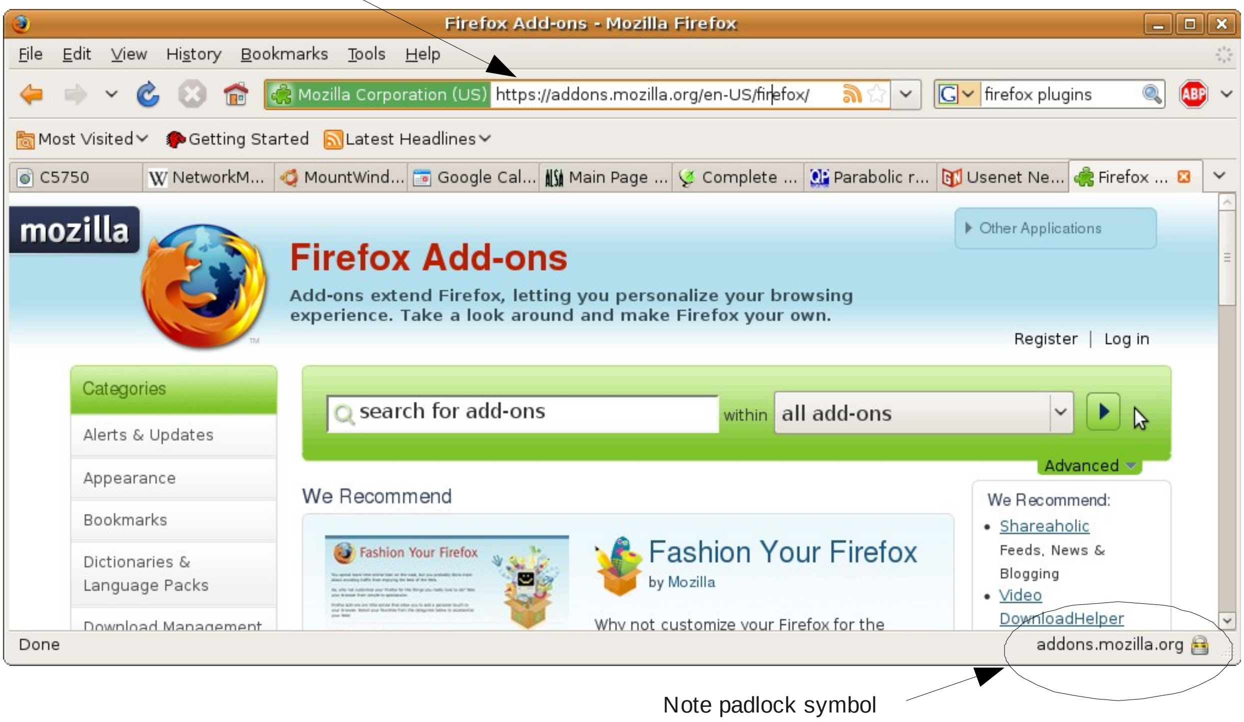
Ignore any offers of updates, virus checking, or any sort of system scans from websites (see the later section for how this should be done). If anything web or email related asks you to urgently visit a site *appearing* to be one you use, or tells you install something, run it, or enter your password then DON'T. As you have already guessed, they never are the police, system administrator, etc, and sadly they do not have $50M in an African bank that needs some help to liberate it...

[5] Details from http://www.webdevout.net/browser-security with easy to grasp graphs at the bottom.

[6] Flaws in the company's web site allow the attacker to place stuff on the 'legitimate' web site that attempts to infect any visitors, or maybe to send them to other sites, etc.

When it comes to web site forms (e.g. for credit card details) then check that the page address starts with 'https' and not 'http' as that extra 's' implies it is a secure (encrypted) link between you and the site[7] to make it hard for any eavesdropping.

Use of https in address



Add-ons extend Firefox, letting you personalize your browsing experience. Take a look around and make Firefox your own.

Why not customize your Firefox for the

Note padlock symbol

Email & Newsgroups

Electronic mail (email) and newsgroups work in a similar manner.

In the case of email, you traditionally would type a message in to an *email client*, such as Thunderbird (or Outlook on Windows, etc). When you click on the 'send' button the email client sends the message to your out-going mail server which is normally provided by your Internet Service Provider (ISP) (e.g. BT, Tiscalli, Virgin, etc). That server then attempts to have the message routed to the recipient's incoming mail server. Finally, the recipient can then use an email client to log-in and download the message to their PC in order to read it.

These days, most email servers may support a 'web mail' system where you can go directly to them via a web page log-in to read and write messages. The basic idea is the same, but your messages are held only on the servers and not on your PC. The advantage of web mail is you can access them from anywhere, the disadvantage is you trust the mail servers not to loose them!

Newsgroups work along similar lines, but the servers that holds the posted message are open for anyone to download and/or reply to.

[7] For https to work and be trustworthy, they also have to have a valid certificate. This is a system to make sure that the web address you are talking to ties in with the security keys they are using, usually the browser shows this in the address bar.

To read or write an email, select the Thunderbird email client using Applications → Internet → Mozilla Thunderbird Email/News. There may be a case for migrating to Evolution (Ubuntu's preferred email system), but Thunderbird was in use before so I kept it. Also change System → Preferences → Preferred Application to have Thunderbird as the 'Mail Reader' choice.

For other email accounts, and generally you should have at least two, then use either Yahoo! or Google for their free web-based email service. For example, go to the www.yahoo.co.uk web site for creating an account.

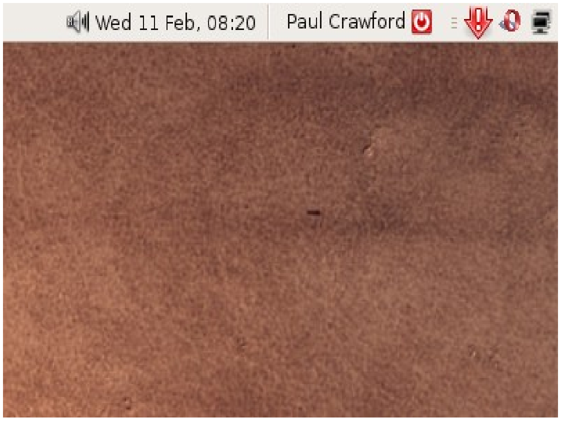
Keep one account only for moderately trusted use, friends, and web sites you plan on dealing with for business. Use another for any web sites you expect will spam you, don't care about, and for use in a cyber café where you don't know if the PC is secure or not.

Thunderbird also supports newsgroups[8], but quite a number of them are of dubious content. Select the news group account in Thunderbird (left hand panel) then 'Manage Newsgroup Subscriptions', at which point it takes a minute or so to download the list of groups. You can then search for topics of interest, such as 'golf', and tick the box(s) and press 'subscribe' then 'OK'. Finally click on the group to read (left panel again, like 'inbox' etc, maybe you have to click on the [+] box to open the list of them) and download the messages, usually just the headers for the most recent 500 or so is sensible.

Generally 'binaries' type of group names often have photos and downloadable stuff, but beware they are all of ***very low*** trust, and a common route to passing on malware (although that has been almost non-existent for LINUX uses *so far*). Even for something of legitimate interest, half or more of the “discussions” are spam and, it seems, general mud-slinging between people with nothing better to do with their life.

System Updates

Nothing is perfect, in fact, software is often far from perfect and there is always a list of mistakes to be corrected. As most of the software on your PC came from the Ubuntu package manager, it will handle their updates automatically.



Thus almost all updates for your system will appear as a prompt in the top-right of your screen (red arrow thing with “!” shown opposite, or a yellow star), but only the administrator-enabled accounts allow that to run the update process (and they ask for your own password again just to be sure). Just click the mouse on the arrow and it will then tell you what updates are available, generally just select 'Install' with all selected (default).

Sometimes it will tell you a re-start is needed, but usually not. Even then, don't worry as if you turn off at night (as I suspect you normally will) then the next day's start-up will suffice.

[8] You can also try the Pan newsgroup reader (can be installed from System → Administration → Synaptic Package Manager) as it handles attachments better than Thunderbird, but currently has a security hole relating to execution permissions of malicious attachments that is still to be fixed (have my own fixed version if you need it).

The only exceptions to this of note are the FireFox plug-ins, sometimes they will ask at the time you start FireFox but they 'just happen' when you click yes (but always check what you are agreeing to).

Opera also offers updates from time to time, but that need you to confirm the download type (it should detect Ubuntu 8.10 is in use, this uses the Debian package format). Either save and run later, or select the 'package manager' rather than saving. Again, you will need to use an administrator-enabled account for this update and provide the password when prompted.

Exploring the File system

There is the Nautilus file explorer in Ubuntu[9] that does the same job as one would expect from Microsoft's Explorer (but lacking certain MS stupidities). You can start quickly in any of your most likely areas from the top start bar using the Places drop-down menu.

The home folder is the area where you *should* keep all of your data and is located in `/home/paul` for user 'paul', etc, and to make some things easier you can also start in other sub-folders such as `/home/paul/Documents` using the 'Documents' option in the drop-down menu and so on.

If you go to Places → Computer you will see any removable media devices that are present and the file system. In fact, all of the available storage is in the one file system, so if you have a CD in the drive, it will normally be automatically mounted and available for access under `/media/cdrom` and so on. This is a different approach to Windows where different physical storage system usually have different drive letters (such as C: for the 1st hard disk, D: for the CD-ROM, etc).

If you plug in a writeable device such as a USB drive, it should automatically be detected and the Nautilus explorer started there. Usually it appears on your desktop as well. You can then use the usual methods of copying/moving files on and off that storage device (copy/paste from the right-click menu, or drag-and-drop). Before you pull the device out, you ***must*** unmount it first!

In the explorer or desktop, just right-click the corresponding device or folder and select 'Unmount Volume' and wait for it to be removed from the view (this could take a few seconds if you have just tried to copy a large file to it). By doing so you ensure the data stored on that device is safely committed to it, and that is a Jolly Good Thing (see later for *why*).

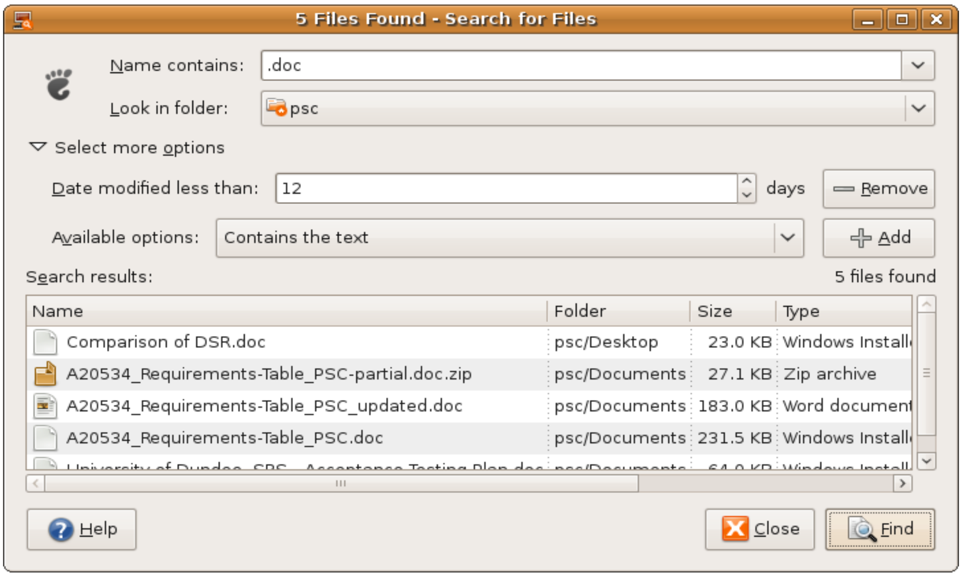
Finding Files

The Nautilus file explorer has a 'find' feature but it is slow and really only works OK for files in the current directory (folder) that you are looking at.

A much better way of finding stuff is to use the drop-down menu Places → Search for Files... menu, this allows you to enter parts of the name and is case insensitive, i.e. it will find ".DOC" and ".doc" to be the same (which is not the normal behaviour for LINUX where case is important).

You enter as much of the name as you know/remember, and then below that you tell it which starting directory to search (typically your home area). You can also add other things by clicking on the "select more options" line, for example, by choosing "Date modified less than" then clicking on the '+ Add' to add this option, then enter the required value (for example, 12 days ago). Then click on the 'Find' button to search for recently modified files. The result looks like this:

[9] Ubuntu is an African word meaning 'Humanity to others', or 'I am what I am because of who we all are' and it is one of the more popular and easy-to-use of the numerous LINUX distributions.



5 Files Found - Search for Files

Name contains: .doc

Look in folder: psc

Select more options

Date modified less than: 12 days

Available options: Contains the text

Search results: 5 files found

Name	Folder	Size	Type
Comparison of DSR.doc	psc/Desktop	23.0 KB	Windows Install
A20534_Requirements-Table_PSC-partial.doc.zip	psc/Documents	27.1 KB	Zip archive
A20534_Requirements-Table_PSC_updated.doc	psc/Documents	183.0 KB	Word document
A20534_Requirements-Table_PSC.doc	psc/Documents	231.5 KB	Windows Install
University of Dundee - SPS - Acceptance Testing Plan.doc	psc/Documents	64.0 KB	Windows Install

Notice this search finds the phrase “.doc” in various places within the names, so you won't just find documents that end in .doc this way.

It is also possible to add other combinations of searches, for example, such as the modification date being 'more than' 2 days ago to exclude very recent files, or to include hidden and backup files, etc. You can end up with quite a slow search if you apply some of them though.

A final point to remember with this tool is it makes use of several lower-level tools to do the job:

- The program **locate** is used for most searches, this is fast but relies on a database of what is on your disks. Typically this is updated once per day (a common reason for odd disk activity) so you might not find a recently created or moved file with a simple search.
- The program **find** can be used to locate files with numerous properties, this typically takes longer as it searches on request, but it will find any new entries.
- The program **grep** is used to find text inside files. This is the slowest of all searches as it has to read through everything looking for your request. If used, it is a *very good idea* to use other tests (e.g. part of the name, date range, etc) to limit the number of documents you try to search though.

Remember that too general a search will produce a *huge* number of matches!

File Sharing

One reason people get a PC/broadband set-up is the promise of “lots of free stuff on t'Internet lad”. In reality, most of that is copyrighted and thus should not really be passed around, also a lot of stuff circulated by file-sharing systems is rubbish at best, and laden with malware at worst[10].

A Brief History

The earliest 'big name' in this area was Napster, they appeared once it became practical to compress music files to a size that could be practically passed around the internet by home PC users (by the MP3 method[11]). Napster offered people a system around 1999 where they could upload a list of their own CD collection (suitably MP3 converted) to a central server, and download other's CD track listings to allow them to copy the files on the list.

Controlling copying and performing is, of course, the whole point of copyright law and when faced with an alternative to buying physical CDs, the major record labels reacted with the help of the Recording Industry Ass. of America (RIAA) in the way any true American company would - they sued and lobbied for law changes to protect their interests.

This led to a court-ordered closure of Napster and, partly piqued by the court's actions, it then lead to the development of peer-to-peer (P2P) networking applications that lacked the central server aspect of Napster (there are/were others such as eDonkey, Limewire, Guntella, Kazaa, and so on). Without a centrally owned 'head' for court attention, the RIAA turned to suing individual users in a legal whack-a-mole game that was more about an intimidation-based deterrent than practically attempting to stop the system.

In a number of cases this failed to produce the result the RIAA wanted; the courts would not outlaw P2P as it had other legitimate uses (ironically, given Sony's involvement here, the argument was similar to the case in the 1980's brought by the movie industry against Sony for their VCR), and often the accusations turned out to be wrong due to errors in translating the IP address of the user to a person, or due to the widespread adoption of WiFi networking that was normally supplied without any access control, hence was often used by other people without the owner's knowledge or permission. They often still are, as ISPs frequently don't want the service support calls needed to get stuff working if they supplied them as secure 'by default'.

Napster returned as a brand after its fall from legal grace, but now it offers music with Digital Rights management (DRM), also referred to as Technology Users' Rights Denial Systems (TURDS). These are systems to allow the copyright owner to control what the user can and cannot do. Often this goes far beyond copy-protection and includes aspects such as on what machine can play it, if you can skip adverts, and so on. In the modern Napster case, you can download all you want, but if you stop paying the subscription, it all stops working. In some cases when businesses that

[10] The 2006 paper “Malware Prevalence in the KaZaA File-Sharing Network” by Seungwon Shin, Jaeyeon Jung, and Hari Balakrishnan (http://nms.lcs.mit.edu/papers/imc145s-shin.pdf) found about 15% of executable programs on the KaZaA network had viral code in them (for Windows of course...)

[11] For most users, the ease of downloading smaller files and lack of hi-fi quality sound on their PC made the MP3 compressed format successful, so much so that “mp3” has become synonymous with computer stored music, just as Hoover became to vacuum cleaners

apparently sold (rather than 'renting') tracks closed or changed their plans[12], the users found that they two had been shafted by the TURDS once the DRM servers stopped working.

The movie industry had a better deal for a while, the latter arrival of the DVD compared to the audio CD slowed the ease of copying video to PC, and the much bigger file size helped slow transfers. They had hoped that the odious Digital Millennium Copyright Act[13] passed in the USA would help along with DVD's (and later BluRay's) encryption, but the technological measures were quickly defeated[14] and files shared nevertheless.

The UK's equivalent of the RIAA, the British Phonographic Industry (BPI), has taken a more sensible route to combating piracy[15] but still has failed to get truly workable agreements between the music labels, ISPs and consumers. Unlike the RIAA's somewhat suicidal route of suing it customers, the BPI has pushed for a '3 strikes and your out' agreement where they contact the ISP to have warning letters sent prior to any action against the alleged infringer.

Sadly there are a lot of people that would pay for music and video under reasonable terms & prices, but the industry as a whole has failed to agree on a workable model to compensate the artists and satisfy the consumer. The only real success is Apple's iTunes store, driven by the success of the iPod, but even there it is moving from being iPod/Mac/Windows limited to having TURDS-free music as an option that can be played anywhere (just like the old CD in fact!).

There are still some new attempts at viable, legal, internet-based music services appearing, one is 'spotify' which is an advert-sponsored streaming system, basically like an on-demand radio. How well it works or how long it lasts remains to be seen.

Bit Torrent - What is it?

A common system used for P2P sharing is the BitTorrent protocol. This is a de-centralised system where there is normally a small metadata[16] file (the SomethingToShare.torrent file) that identifies those computers (known as 'peers') which have copies of the file you are looking for.

The BitTorrent client then uses the information in the metadata file to locate the various parts from

[12] See http://www.theregister.co.uk/2008/04/30/eff_msn_music_open_letter/

[13] The DMCA criminalises production and dissemination of technology, devices, or services intended to circumvent Digital Rights Management (DRM) whether or not there is actual infringement of copyright itself. It has been used by companies for anti-competitive ends, e.g. attempts by the printer maker Lexmark to prevent others from producing replacement ink cartridges. It has also led to the arrest of cryptography and security researchers who visited the USA for conferences for work they conducted legally in their country of residence.

[14] The movie companies did not support LINUX which was the original reason for Norway's 'DVD Jon' (Jon Lech Johansen) cracking the DVD Content Scrambling System in the first place to produce the 'de-css' software.

[15] It has become common to refer to copyright infringement as “piracy” and “theft” but it is rather different. If someone steals your CD, you are deprived of its use, but if it is copied then you are not. However, the artists and so on behind the CD are deprived of the income they would have received had it been sold, so there is a strong moral case for copyright law. When one hears about the terms & conditions imposed on artists by the industry, and their resulting pitiful share of the sales, the argument looses some of its gloss. The industry estimated figures for lost revenue due to piracy, both in music and software, are often grossly over-inflated as they normally assume every copy made is a lost full-price sale.

[16] Metadata is “data about other data” such as information on a book (e.g. title, author, number of pages, etc) rather than the book's contents.

the other computers and to copy them to your PC and verify the data, while adding you to the set of 'peers' that share the file to others.

Unlike a conventional file copy/transfer, where you start at the beginning and copy until the end of the file from one source, the BitTorrent system divides the shared file in to a large number of segments (the “swarm”) and they can be independently gathered from multiple sources, ultimately to be put together in to a copy of the wanted file.

The name “torrent” comes from this ability to join up a large number of smaller transfers into one big file, so the incoming data is not a single uniform stream but a torrent of multiple streams.

It is ultimately a cooperative system, as it relies on someone 'seeding' the torrent with a whole file, and then others who copy it to also share it, thus providing a redundant set of sources for anyone else wanting to make a copy. If a few of the peers are missing for any reason (PC switched off, user deleted the file, etc), the rest can provide enough data to produce a whole copy and so on. It is therefore considered 'good practice' to keep a copied file sharing for a while to allow others to download it from you, those who don't are often refereed to as 'leeches'.

Bit Torrent - Example Use

If you were looking for a file you need to find the `some_file_to_download.torrent` file of whatever you are looking for, and then use a suitable client to collect the data referred to in it.

As for all P2P systems, a lot of stuff is copyright-controlled and so should not be out there, so some of the torrent tracker sites get legal threats from time to time. A few that seemed to work include http://isohunt.com/ and http://btjunkie.org/ and http://thepiratebay.org/ (who clearly nail their colours to the mast, but are closing/changing following the recent court case) and http://www.mininova.org

Take particular care to avoid torrent search sites that offer pay-services or prompt you for user information, they are almost certainly scams! Choose to download the torrent file, do not try “faster downloads” or whatever is being pushed.

The world of P2P sites and legal position is constantly shifting, so should you be interested it is a good idea to actively follow what is going on. The site http://torrentfreak.com/ offers articles with a pro-P2P bias that are intelligent and well written articles, which is more than can be said for the majority of commentators on the site! The technical news site http://www.theregister.co.uk/ covers most IT things, including good articles on music, P2P and copyright law by the journalist Andrew Orlowski, but with generally a more pro-copyright bias.

Ubuntu LINUX is supplied with a bit torrent client (Applications → Internet → Transmission BitTorrent Client) which can make use of an existing '.torrent' file you already downloaded, and is called by FireFox automatically if you download a new .torrent file. However, the version with Ubuntu 8.10 is not that good and you should have the latest one and configure basic options (see Appendix B – Transmission BT client configuration on page 61).

Some systems are moving to use “magnet” links, these keep just one item of information on the file(s), known as a *hash* value, and depend on querying the P2P peers to find the matching data. This has the advantage of no single tracker to depend upon, but currently (as of 17 Nov 09) the TransmissionBT client has no support for it, You can start with a simple Google search, for example[17] “paris hilton torrent”, or try a torrent search site directly. When you find one, the site might indicate if it is still active (i.e. non-zero seeds), then just select the option to download the torrent file and then Transmission will ask if you want to add the files. Click 'yes' to start then be prepared for a ***long*** wait.

Due to a number of factors, one being the ISP's data throttling[18] activities, P2P is very slow and you may have to leave the computer on for a day or more to transfer a big file. Typical transfer speeds are in the dial-up modem range of 2-20kbit/sec! Finally, some repeated warnings:

- A lot of P2P files are copyright controlled, and you could get in to trouble for copying/sharing them. Admittedly only likely if you are a doing it a lot and choose to ignore any cease and desist letters, but a consideration nevertheless.
- A lot of stuff found on P2P is utter rubbish or malware, so treat it as suspicious and try it under LINUX and not Windows. If you are told to download/install something to view the content[19] then just ignore it. Really, DON'T do it!
- If leaving your computer on for long times, make sure you save/close every application other than the BitTorrent one so if power fails or you accidentally switch off, you don't risk loosing too much data[20].

Safe surfing folks...

Camera & Photos

Choose the account to use with the camera so you keep all photos in the one palace ***and back that up!*** The account used here as Julie's. Try to make sure the camera is well charged then plug in the special USB cable to the Olympus camera and plug the other end it in to the PC's USB socket (easiest option is the USB extension cable).

The camera should notice this and its rear menu should show the option for PC connection, press the right-arrow part of the selector/menu button (one with the flash symbol) on the back of the camera, then the down arrow to choose 'MTP' rather than mass storage.

The PC should then recognise it as a camera device (rather than memory stick, though that should work as well) and offer to start the photo package Applications → Graphics → F-Spot

After a short while it should show you all of the camera's images, you can then select 'Copy' to copy them over to the PC. You should delete them from the camera's memory card ***after you have backed up the PC copies***, to save space and transfer times.

[17] There are some cases where the moral argument for copyright providing remuneration just don't work.

[18] Euphemistically referred to as “traffic management” or shaping. ISPs don't like P2P because it allows users to actually use that “unlimited” package they were over-sold by the ISP and thus cost them more money than anticipated, but often these days it is the BBC iPlayer that uses the ISP's bandwidth up.

[19] There are occasions where the audio or movie player will need to get a newer codec to play something, these are handled by the normal program/system updater and you should recognise this. If anything looks out of place, e.g. a Windows-style menu on your LINUX PC, it is definitely a scam!

[20] To be extra safe, after saving/closing all non-essential stuff, open a terminal (Applications → Accessories → Terminal), run the command **sync**, then close the terminal before leaving the computer to get on with the slow process of downloading stuff. See the later chapter on file systems for an explanation.

Once the copying appears to be over, and the camera's LED has stopped flashing, it should be safe to unplug it, but using unmount first is always a good idea!

NOTE: The F-Spot program defaults to the `~/Photo` directory (here the '~' character implies your home folder, such as `/home/julie`) but I changed it to use the Pictures directory already made easy by Ubuntu (such as top menu bar left-ish Places → Pictures to browse this area with ease). Inside F-Spot I used the option: Edit → Preferences → Import Settings to change this.

When you select the photos in the camera and tell it to copy them, it appears to organise them in folders by year/month/day and and then are named as the camera had automatically done. For example:

```
/home/julie/Pictures/2009/02/04/p2040056.jpg
```

and so on. It also provides a time-line sort of slider to search through the image 'thumb-nails' (they are small images roughly the size of your thumb for easy browsing).

If you double-click on an image it calls up the 'Eye of GNOME' viewer, this also offers a slide-show style of going through all images in a folder. However, I found its printing a touch unreliable.

I found a quick and easy option was to drag images in to an OpenOffice word processing document and then size it, add text (as required) then print. Applications → Office → OpenOffice.org Word Processor

To edit an image, for example, to crop it or make it light/darker, then you can use Applications → Graphics → GIMP Image Editor

This is a complicated package to use due to the sheer range of stuff it can do (just like the Adobe Photoshop used commercially) so you need some patience and effort to get to grips with doing stuff. **Make sure you save any modified file as a different name!** That way you can go back and try again and **for safety you should make a copy of your key images on to a CD/DVD as soon as practical.**

Emailing a Photo

Most cameras take very detailed photos these days, in the several mega-pixel range (with the resulting file size typically in the couple of MB range), and so are often rather large for sensible email use (say, 100kB). Thus is makes sense to make a smaller version of any images prior to emailing them, an action that also allows you to do other things like crop the photo to the region of interest.

If you have configured System → Preferences → Prefered Application for email to suit, you can easily create a reduced size image ready for email from inside the F-Stop Photo Manage (after selecting an image with the mouse, use Photo → Send by Email).

The other route is more tedious, but ultimately more powerful. Open the image with GIMP (e.g. right-click on the file in the Nautilis explorer) and choose Open With → Open with “GIMP Image Editor”. To crop an image, choose the top-right square option in the “Toolbox” page:

Go to the main image, move the mouse to your chosen top-left point, then press down the left mouse button and drag the box to enclose the area of the image you wish to keep. If you get it wrong, press the 'Esc' key to cancel and try again.

Then go to the main image menus and select Image → Crop to Selection to trim the region of choice. Then use the menu option Image → Scale Image to allow you to re-size the number of pixels in use. A sensible limit for an emailed image is 1024 pixels in the biggest direction, so go to the largest (if it is bigger than 1024 in Width or Height) and change the number to 1024 (then press return). For reasonable quality I normally recommend using the Cubic interpolation mode, but feel free to play with this if bored. Finally press the 'Scale' button so it is done.

Finally go to File → Save as... to bring up the menu for saving the resulting image as a **different** file. You could type in the path/name as something like this:

```
~/Documents/test.jpg
```

To save in your home directory's Documents folder, as test.jpg name, or use the options to explore the file system and then choose an extension type. When you select 'Save', in this case it will prompt you for the desired image quality (as JPG format is lossy - it looses information to make the file smaller) and something like 80-90% is generally a good choice.

As a general rule, do all your image editing first, and save last, as repeated saving in a lossy format can ultimately result in a very poor quality image. If you want to do a lot of editing in steps, save intermediate versions in a lossless format such as TIFF.

Finally, compose your email (the 'Write' button in Thunderbird, top left-ish) and attach the saved (now much smaller) image using the 'Attach' button of the email (middle-right of normal buttons at top of email composer).

Music

Music on Disk

If you already have some music on your hard disk, then open Applications → Sound & Video → Rhythmbox Music Player and then from the File menu select Import Folder, then navigate you where the music files are stored. It then creates a list of the music with (guesses?) of the artist, track title, and album name.

Copying from an Audio CD

The new PC can 'rip' tracks off an audio CD, play them, and you can make a 'best of' CD if you want. To rip tracks off a CD, first make sure it is clean (always wipe from centre out (radially) and **never** in the direction of rotation) and place it in the CD/DVD drive. Typically it will recognise this and offer to open the Applications → Sound & Video → Rhythmbox Music Player for you, otherwise start that from the top left menu bar.

Typically it recognises the CD (I guess from a database of track times, etc) and you should see a CD icon in the left column under 'devices'. If you right-click the mouse on this you can select 'Extract to Library' and it will transfer all tracks to your Music folder.

It seems to do this by creating a folder for the band, then the CD title, and finally names the tracks according to this (I guess if it is an unrecognised disk you might be prompted for names or it will use something dull and simple). For example:

```
/home/paul/Music/Queen/Greatest Hits I & II/01 - Bohemian Rhapsody.oga
```

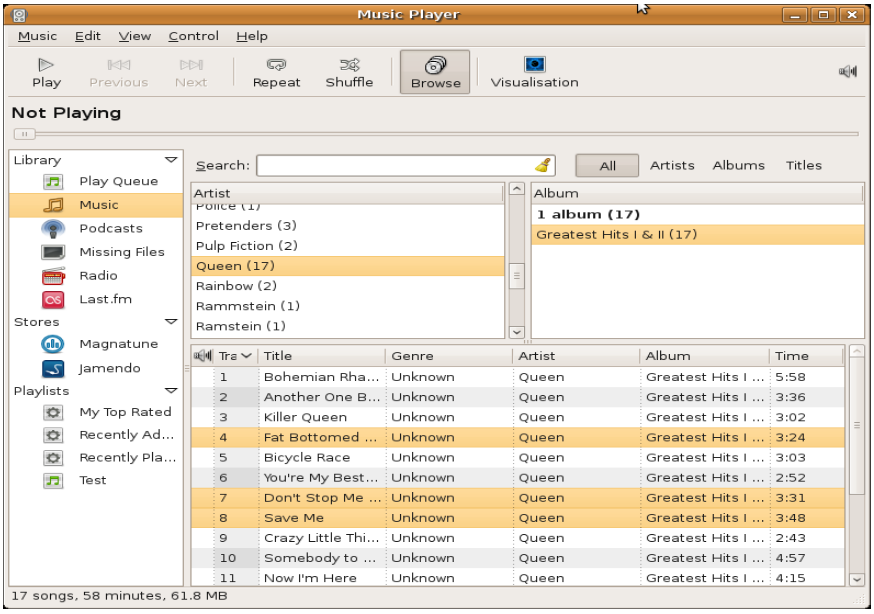
The default setting is to save tracks in .oga format, this is an open/free format with relatively little loss of quality (near CD quality, thought there is lossless .flac format for CD quality). Most people think of computer-based music as beginning and ending with the mp3 format, which is lossy (i.e. quality degraded compared to the original) but this attribute allows a trade-off in terms of file size and sound quality.

You can configure the format using Rhythmbox's menu Edit → Preferences → Music (tab) → Prefered Format 'Edit' box, this allows you to choose mp3 format (however, you have to install the package “ubuntu-restricted-extras” first using System → Administration → Synaptic Package Manager).

There is an ipod import tool that works OK, ***but do not try to change the iPod's data***! It seems Apple have deliberately made the system rather difficult, and this will upset it. Possibly the gtkpod-aac program will work, and remember to **unmount** (or shut the PC down) before disconnecting the iPod.

Creating an Audio CD

To create your own 'greatest hits' CD, select the track(s) you want from the main (lower) track listing that appears if you click on the Library/Music (left vertical area). You also have the areas to navigate by artist (but allow for multiple listings such as “The Rolling Stones” and “Rolling Stones”) and by album title. By holding down the 'Ctrl' key you can select multiple tracks with the mouse if required, and then right-click the mouse and use 'Add to playlist' to create a selection for later CD generation. This can be repeated to add all of the tracks you want in the list.



Artist
Police (1)
Pretenders (3)
Pulp Fiction (2)
Queen (17)
Rainbow (2)
Rammstein (1)
Ramstein (1)

Album
1 album (17)
Greatest Hits I & II (17)

Tra	Title	Genre	Artist	Album	Time
1	Bohemian Rha...	Unknown	Queen	Greatest Hits I ...	5:58
2	Another One B...	Unknown	Queen	Greatest Hits I ...	3:36
3	Killer Queen	Unknown	Queen	Greatest Hits I ...	3:02
4	Fat Bottomed ...	Unknown	Queen	Greatest Hits I ...	3:24
5	Bicycle Race	Unknown	Queen	Greatest Hits I ...	3:03
6	You're My Best...	Unknown	Queen	Greatest Hits I ...	2:52
7	Don't Stop Me ...	Unknown	Queen	Greatest Hits I ...	3:31
8	Save Me	Unknown	Queen	Greatest Hits I ...	3:48
9	Crazy Little Thi...	Unknown	Queen	Greatest Hits I ...	2:43
10	Somebody to ...	Unknown	Queen	Greatest Hits I ...	4:57
11	Now I'm Here	Unknown	Queen	Greatest Hits I ...	4:15

17 songs, 58 minutes, 61.8 MB

You can select a new list or use an existing one. If you create a new list (Add to playlist → New playlist) , then it will be waiting for you to enter a name, so in this case I typed 'Test' then pressed the enter/carriage return key.

If you click on the Playlist/Test entry it will tell you the playing time (above shows the Queen listing as 58 minutes). Generally stick to 60 minutes maximum for an audio CD you wish to create. Finally right-click on the play list and select 'create audio CD' or use the program's Music → Playlist → Create Audio CD option. After a short time (minute or two) preparing the data, it will attempt to write it to a blank CD (which, of course, you have ready to insert in the CD/DVD drive...)

Just Playing Music

The Rhythmbox software will play what you have in any of its lists, this could even be the whole collection you have! Simply select the first track of the list, and press the big 'play' button (top left of program). To create a list, do it just like the above section on creating an Audio CD, and if it is a real favourite then save it so it is available every time you want to listen.

Note: There are (at least) 3 volume controls on your PC / sound system:

- Rhythmbox's has one (loudspeaker icon top right of program, just below the 'X' close icon on its top bar), this is a slider but only takes effect once you release it.
- The overall system has another slider (top bar, right-ish side) which alters all sound sources. Right-click on this and choose open and make sure the various input control sliders are high up (as well as the master volume control).
- The external speakers have a control, in this case it is the '-' and '+' buttons either side of the central power button. On power-up this starts at minimum (zero?) volume.

So if you don't hear any sound, check all are non-zero, and also power is on to the speakers, and (just in case) that the speaker's sound cable is in the light green jack socket on the back of the PC.

Word Processor / Spreadsheet

Ubuntu comes with the OpenOffice suite, which has a drawing package (under Applications → Graphics) as well as the more commonly used spreadsheet (Calc) and word processor (Writer) both in Applications → Office.

Before using them, I suggest you open and turn off some irritating 'help' features:

Spreadsheet: Tools → Cell content → AutoInput (make it un-ticked)

Word processor: Tools → AutoCorrect → Replace (try 'i' and then delete the i->I change as it breaks i.e. and similar)

Tools → AutoCorrect → Word Completion (I un-ticked this to stop the yellow suggestion as you type thing)

When saving a document, it defaults to the OpenOffice format which is good, but not very usable to folk with MS Word and similar. You can choose the format for saving, but do not expect formatting and style to be accurately preserved if you change to Word format!

If you need to send a document to someone else for reading (not editing) the best option is to use the word processor's “File → Export as PDF...” option to create a portable document that most users can read & print with good predictable results.

Writing a DVD and Making a Backup

Your PC cost a couple of hundred quid, a new hard disk maybe £50. That is no big deal to replace in the grand scheme of things.

Your ***data*** on the other hand, such as hours of work, or photos of past people/events that you can never re-visit, is priceless.

A lot of folk have data lost through a combination of:

- Ignorance - not knowing how to make a backup, maybe not trying to find out.
- Effort - too much effort and there is always tomorrow, isn't there?
- Cost - A few DVDs needed, or maybe an external hard disk, just too much...

A 'backup' in computer terms means a 2nd copy (or greater number) of your data kept on physically separate media from the 1st (primary) copy you use, thus protecting it from deletion mistakes or media failure. This should really be in a ***physically different location*** to protect against fire, floods, theft, etc.

Making a copy of a file or files to a CD or DVD is an obvious way to protect important data, but that is not quite the same as a proper backup. Why? Well if you copy a file generally only the contents are guaranteed to remain the same, other important information such as the path (i.e. where it normally lives), owner and group setting, permissions, etc., may not be kept thus making it hard to restore a blank or damaged computer file system to the original state.

A backup program on the other hand is designed to make a copy of everything in key locations (e.g. you normally ignore /tmp for backing up as it should never have important stuff in it) to a single file or tape device. If required, you can then restore an individual file, or all files, back to their original state.

There are some scripts for automating back-up, see Appendix E – Automatic Back-up.

But as an alternative, here is a quick guide to using the CD/DVD drive to save some important files:

By default, Ubuntu 8.10 comes with the Brasero CD burning program, but it did not work OK for me. So I installed K3b instead (to install, try System → Administration → 'Synaptic Package Manager' then search for K3b).

Start Applications → Sound & Video → K3b and then choose the “New Data CD” or New Data DVD” option, depending on the size of data set and what sort of blank disks you have.

If you have a lot of files with common data (i.e. could compress well) you might want to use the option in Nautilus (select file(s)/directory(s), then right-click) to 'Create archive...' to generate a single file with everything packed in to it. If you use the .tar.gz option then it will preserve all of the file permissions and can deal with very long or unusual file names that CD/DVD file systems cannot.

BUG NOTE: The .zip option seems to fail on files >2GB but the .tar.gz works OK and can be more efficient in space saving.

However you create/choose the files, you can drag & drop them in to the K3b 'current project' window. The lower scale shows you how much of the disk it uses, and once done you just select

'Burn'.

BUG NOTE: While it is good to verify any disk after writing it, K3b seems to eject and then just wait pointlessly if this option is selected for data projects (a known bug, but I think it worked OK for an iso though?) but if you *previously* select the main program's options Settings → Configure K3b → Advanced then tick the option “Do not eject media after write process” then Apply, it works with a verify, though you need to manually eject the disk on final completion. Still to confirm that...

After selecting 'burn' and inserting a blank CD/DVD (and ignoring any pop-up from the system asking you what to do with a blank disk, which you can turn off if it bothers you) and then 'Burn' again, it should just work. Write on the completed disk what it is, and then check you can read it on another PC.

Remember, a backup is no good if you can't use it, and ***never*** wait until it is an emergency to find that out! I have seen CDs written on one drive that were OK to read on the original, but not on any other CD/DVD drive, so try it out on ***another*** PC.

Running a Windows Program

Sometimes you need to use Windows software, maybe you have a unique application that is only available for Windows, maybe you need strict format compatibility with a MS program, or maybe something in LINUX just sucks (e.g. printer support).

To support this, you can use a **virtual machine** running Windows XP. This is software that pretends it is a whole PC, so you can start another OS and have a computer running inside a computer. Cool or what?

This has the advantage that you have LINUX file system stability and safety for web browsing / email, but with the ability to run windows stuff almost for real. Of course, it is no good for fast game play and so on.

To use this select Applications → System Tools → VMware Player

This should allow you to start (play) the existing Windows XP virtual machine and then you have it available. Of course, should you be using something else (e.g. CD drive or printer) then it won't be connected to the VM when it starts, but there is an option to re-connect once it is free.

But please, please, don't use the Windows VM for web browsing, etc, as then you are back facing the Diet of Worms again! Should damage occur to the windows system you can replace the virtual machine with the backup copy and start fresh.

The Windows VM was configured such that the “My Documents” area is shared from the LINUX host, so that you can access those files from /home/vmuser/share under LINUX as an easy way to put files in to and out of the VM.

Recovering from Problems

Typically with a Windows PC if it goes wrong then you reboot, maybe using the “three fingered salute” of Ctrl+Alt+Del or perhaps using the power button. While often effective, this is not really a good solution in all cases, and LINUX computers are expected to keep running even if there is a problem. Here are a couple of scenarios where there are less drastic remedies than a forced re-start.

A 'hung' System

While very rare, you may find at some point something goes wrong and your PC appears to have frozen. This could be something like a web browser using **all** of the CPU time in some manner that should not, but unfortunately has, happened. None of the windows responds to the mouse and you fear that an open document, etc, might be lost. What should you do?

The first thing is wait for 20-60 seconds in case it just responds really slowly.

The second thing to try is to log-in without using the normal graphical interface. To do this use the 3 simultaneous key combination Ctrl+Alt+F1 (that is function key 'F1' on the top row of your keyboard) and, maybe after trying twice after waiting for a few seconds, you should be presented with a command prompt asking you to log in.

Try using the account you had just been working with (for example user name 'paul') and then your password. Then use the command 'top' to show what is running. If the top line shows something at 100% CPU time you have the culprit! Press the 'q' key to quit 'top' and then run the command 'sync' and wait 5-10 seconds to tidy up the disk system before you do anything serious.

If it is ***your*** process running (most likely) then try to stop it by using:

```
kill ProcessID
```

Where the numerical value for process ID is the left most on the highest entry in top's listing (under the heading PID). If you run top again and it still shows the process running at around 100% CPU time 20-30 seconds after this (quite likely, given there is some fault), then try the more brutal version:

```
kill -9 ProcessID
```

If that fails, or it if is not your process, you may have to use the power of the root user with:

```
sudo kill -9 ProcessID
```

In order to force a process stop. If 'top' is then showing all is OK, you can use the command **logout** to exit. If you wish to check the user interface without logging out, then use Ctry+Alt+F7 to go to back to the graphical terminal window[21] which should now be responding as usual.

If even this has not recovered the system, you could try the **reboot** command. Of course, you loose any stuff that was open but should recover with clean file systems (i.e. avoid simply powering off if you can!)

Network Down

Occasionally the Ethernet link to your router/modem may stop working as expected, usually when it should have automatically renewed its agreed address, etc. If you suddenly find you have no internet connection, here are some suggestions:

- Check that the modem/router/etc has not been switched off. Occasionally they get knocked about and the power lead can come out or the 'on' button accidentally pressed. If off, then

[21] There are 7 'terminals' in total, using Ctrl+Alt+F1 through to F6 you open text-mode terminals tty1 to tty6, while F7 is the normal graphical user interface you are familiar with.

obviously turn it on and allow 30 sec or so for it to negotiate its settings with your ISP, then try renewing the network connection (below). It is a good idea to take note of any lights that are normally active so you can tell if it is behaving oddly when you have a problem.

- In the top-right there is a small black icon of two monitors behind each other (I think that is what it represents) for your wired network connection. Click on that and it should show you at least one connection, probably “Auto eth0”. Click on that and the black icon should do a little 'dance' as it renews the link settings. Fixed now?
- Open a terminal and try the **ping** command to see if you have a link to your router, typically with `ping -c 4 192.168.1.1` to probe the router (assumed here to be using the normal address 192.168.1.1) 4 times. It should have reported response times of around 1ms or less. If not, check the network cable(s) are all OK (assuming of course you have already checked the power was on), and maybe try powering off your modem/router for 30 seconds before trying the renew the network once more.
- Try accessing the Dundee web site by its numerical IP address, enter in to your web browser's address bar (**not** the Google search box!) the value `http://134.36.22.54/` and see if the page comes up. If so, the DNS servers are down so probably best to try again tomorrow.

If it all looks bad, maybe your ISP is having problems (more common than one would like) so before doing anything drastic, consider just giving up and trying again tomorrow! Phoning an ISP help line is typically a waste of time & money unless you absolutely must get help.

Nautilus Sluggish, or Change of User Settings

If you find the Nautilus file explorer is slow (or not responding), or if you have just changed the user settings, you may find that logging out then back in is enough for resolve the problems without needing a reboot.

Problems with Video / Audio Playback

Sometime you will find a video file that just won't play well, either the audio is missing or garbled, or the video is juddering/jumpy, etc. Usually this is down to certain combinations of file formats and codec used in LINUX, and the two common things to try are:

- Try a different movie player. Often VLC works more reliably than the default Ubuntu movie player.
- Is the file high-definition video? If so you may struggle to play it on a PC with a slower CPU and/or standard “on board” graphics system. Short of paying for a hardware upgrade, see if the same file is available in standard definition format (say 720 pixels x 576 lines, or less).

Security & Privacy

The Basic Rules

There is much said about 'cyber crime' and all of the problems that afflict PCs, and usually LINUX or MacOS users just point and laugh at the average Windows user's pox-ridden internet experience. But that is foolish as it gives non-Windows users an over-inflated sense of their own invulnerability. What the fail to realise is this:

No useful general purpose computer can <u>ever</u> be made completely secure!

This sounds like an absurd statement, after all we don't have problems with toasters or CD players (yet?) so why can't a PC be made similarly safe? The answer is that a *useful* computer is so by virtue of all of the different things it can do, it is not a single-function appliance like a toaster. Think of it as a tool like an electric drill. You can make a good electric drill that passes all of the safety requirements needed by law, but would that stop someone drilling their own leg due to stupidity or lack of care & attention?

So here are the basic things for safe computer use:

1. Educate the users. This is the ***single most important*** aspect! They need to recognise scams and be very sceptical of the on-line world. Further more, they need to follow good practice which includes the following steps, and to ***think*** before they act!
2. Keep the system up to date with all patches. Why leave a known crack open?
3. Have a securely configured computer. This starts with good use of user accounts and passwords (see later for details), and continues to adding and configuring software. By default Ubuntu has little enabled so it is relatively safe, so think before adding stuff like networking and if you do, consider using a firewall to keep it private.
4. Avoid wireless access point that you do not know who they are and why they are open. They could be a "honey pot" trap for monitoring your attempts at logging in for stuff.
5. Regularly back up your data. Sooner or later you ***will*** suffer from data loss due to deleting by mistake, suffer from a hard disk failure, or have a software error or malware attempt to trash your documents. This way you have a method of recovering.

Sounds simple, and it is not that hard in reality. Remember it is a moving target as the bad guys adapt, and that security is ultimately about a manageable risk through layers of protection, and not the unachievable goal of "total security" claimed by some commercial players.

Note that it is not only your PC they want, as recently it has been found they are targeting your modem/router[22] for the same ends. Also by fiddling with certain types of home router's DNS settings[23] they can silently redirect your PC's web browser to phishing sites.

[22] For the article see: http://www.theregister.co.uk/2009/03/24/psyb0t_home_networking_worm/

[23] For example http://www.theregister.co.uk/2008/01/23/pharming_attack_in_the_wild/

Know Your Enemy

It is a common misconception that somehow security & safety in the 'virtual world' of electronic data and communications represents a whole new area of crime, but in fact the various scams & threats that exist have all been done before. True, the details of how some are done are technically novel and unintelligible to the average user, but that is distracting from the important points. As in most TV detective dramas, you need to look for the same points: motive, method and opportunity.

Motive

Originally most malware was created out of nerdish curiosity, an intellectual show-off competition if you like, but now that is the minority case. The real motives now are usually the classic ones: Money, Sex, Power (including region / politics), and Revenge (usually from feeling they did not get enough of the first three).

So why do they attack your home computer system? There are two primary reasons and they relate mostly to money, but sometimes to power or spite:

- To get personal information from you, such as bank log-in information (for example, using key-logging software to record what you type), or enough personal data for "identity theft", where they pretend to be you for buying goods, or taking out loans in your name. Putting it simply, theft or fraud.
- To use your computer to perform illegal activities, basically your computer becomes part of a huge 'botnet' (the term is shortened from 'robot network').

Typically botnets are used for sending spam emails as crap advertising (it only takes one idiot in a million to buy the goods to make it worth the while, as the infected users are paying for the computer and internet connection and not them).

The other trick is to use your computer along with others to attack a web site simply by overloading it with requests (known DDoS from 'distributed denial of service'), usually as part of a protection racket (Mafia style "pay us or else") or for political ends (in effect, to silence sites they don't agree with). Really just the crimes of old with a novelty wrapper.

Method

There are a couple of categories that most malware can be filed under, and it pays to take a quick look at them and how they work.

The "Trojan"

If you study Greek history, there is the tale about the war in which the Greek army had laid siege to the city of Troy for a long time, and failed to defeat their defences. So they made a large wooden horse and hid some soldiers in it, appeared to give up and go home while leaving the horse outside the gates to the city. The Trojans thought it was some sort of gift, took it inside and then during the night, the Greek soldiers got out and opened the gates, allowing their army in to defeat the Trojans.

While we may now smile at the apparently dumb act of the Trojans taking the horse in at face value, *exactly the same ploy works today* in a different guise. The malware type named after the Trojan

Horse has the same method, it is a malicious payload disguised as something the user will want, and by running it their PC is duly compromised.

So how do they dupe the user? Again, the motives used are fairly predictable (this process has the name 'social engineering' these days):

- Sex/Romance: For example, it appears to be a video of someone famous naked (or is offered as something you need to view the video, such as a flash player or codec update), or it claims to be a 'gift' from an admirer.
- Money/Dishonest Gain: Maybe taking advantage of folk desperate for a job who are more susceptible to a scam, or maybe it is offered as a way of getting something 'for free', such as an unlocked version of some commercial software.
- Fear/Authority: It claims to be from an authority (police, your system administrator, a bank, etc) and demands you check something to avoid the consequences. Or maybe some frightening (but invented) news such as terrorist attack nearby[24].
- Curiosity/Novelty: It is a game, or a flying toaster screen saver, a torrent search toolbar, or similar "Oh shiny!" stuff, or news about some breaking event, celebrity, etc. As they say, the Devil makes for for idle hands...

It is clear that this can be avoided by never downloading and running stuff[25] that you are not absolutely sure of.

Remember that no authority relies on email for official communications, and if you do have a worry just phone the bank, etc, and ask them directly. Do **not** rely on any contact information offer on the web site or email, no matter how convincing, as that is likely to be faked as well.

The 'Virus'

The difference between the computer virus and the trojan is the virus runs without the user having to knowingly execute any program. It relies on flaws in any part of the system that allows it to start running where it would not be expected to happen, for example on opening a document, or visiting a web site. As often as not, it is now the applications (PDF reader, flash player, word processor), rather than the operating system components, that are targets.

While the Trojan relied on duping the user, the virus needs only 'contact' with a vulnerable system. The protection against this is obviously to keep the system updated so the window of opportunity for an attack is kept down, also to be wary of visiting web sites that are pushed by spam or scam-like offers since they could be engineered to try to infect, and don't open unsolicited documents, etc.

By far the most easy (and in retrospect stupid) method of propagating was the 'autorun' feature of Windows where if you insert a CD or USB stick, certain programs would be executed automatically. Here the 'virus' is really a Trojan Horse where front gate opened on demand! Sadly Ubuntu has a similar feature but at least it asks you first. I suggest turning this off: in the Nautulus explorer go to Edit → Preferences → Media tab → Software (change to "do nothing").

[24] A good example of such a scam is covered here http://www.theregister.co.uk/2009/03/16/geo_located_malware/

[25] Here the LINUX default is safer than Windows downloaded stuff wont run unless you set the 'execute' permission, but that is not fool proof protection as fools can be so damn inventive.

The 'Worm'

The worm is a self-propagating virus, one that moves (burrows) from computer to computer over a network (either an internal network or the internet) usually relying on flaws in some network access point, or weak easy-to-guess administrative passwords.

Protecting against worms is really like the virus; keep up to date and also don't enable network stuff unless you really need it, in which case have a firewall to restrict access (not 100% fix, but helps reduce the opportunities to attack the system).

The Trojan, virus and worm are all basically doing the same thing, the only difference is how they get in to your system to start executing. Of course, some of the more sophisticated malware uses all of the above methods to spread.

Notability in my comments is the lack of a mention of anti-virus (AV) software, the usual recommendation for Windows computers. Why? Well you have to consider the following:

- Most[26] executable malware is Windows-only *for now*, and those for LINUX have rarely been seen 'in the wild'. So AV stuff really will do little good for a LINUX PC (maybe just help your Windows-using friends though should you attempt to forward some infected document, etc).
- Anti-virus software works by monitoring all activity on your PC and scanning any new files, this slows the PC down by a large amount, and faults in the AV can seriously disrupt[27] your PC.
- Most detection is based on recognising ***existing*** patters of malware, the flaw is obvious when you consider that new 'worm' malware spreads to hundreds of thousands of PCs in a few hours, while the AV search patterns take a day or two to be updated to deal with them.
- It is not that effective: the very best[28] in a 2009 test missed 1 in 20 of known viruses!
- You normally have to pay a subscription to keep the AV updates.

So my own judgement is AV is not effective enough to overcome the downsides to its use.

Ad-ware and Spy-ware

These are programs rather like the Trojan/virus, but are less destructive. Their goal is to bombard you with adverts (locally generated spam if you like) and/or gather your browsing habits for sale to unscrupulous advertising agents (are there any scrupulous ones?)

[26] From http://www.viruslist.com/en/analysis?pubid=204792070 dated 4 Aug 2009: “The fact is that the majority of malicious programs identified to date (well over 2 million) target Windows. Linux, on the other hand, with a mere 1898 malicious programs targeting the operating system, appears to be relatively secure.”

[27] It is not uncommon for AV software to seriously upset computers, for example http://www.theregister.co.uk/2009/07/03/mcafee_false_positive_glitch/ and http://www.theregister.co.uk/2009/08/12/ca_auto_immune_update/

[28] See http://www.theregister.co.uk/2009/08/06/vista_anti_virus_tests/ and http://www.virusbtn.com/vb100/RAP/RAP-quadrant-Feb-Aug09.jpg for the details and results.

Generally they are a Windows problem, and often even more specifically targeted to just the Windows Internet Explorer web browser. There is currently little risk for LINUX users, but again the golden rule is the same as the Trojan:

- Do not install anything you don't fully understand. In particular, only ever get plug-ins for your web browser directly from the original web site. For example, any new features for Firefox can be found at https://addons.mozilla.org/en-US/firefox/
- Assume all web site lie to you, there is never any need for a new feature / update that is worth the risk of such a casual download / install (the only exception is the occasional update for Firefox's plug-ins at start-up, but that will not appear in a web page).

Think before you click...

Phishing for your details

Phishing (pronounced "fishing") is a reference to a trap where an imitation web site is set up with the intention of tricking unwary visitors in to entering useful personal details.

While your bank log-in is the most obvious thing to guard against, email is a problem few think about. However, often a commercial web site will email you if you forget your password, so by gaining access to your email account they can in turn start to access sites you shop with.

Add to that the tendency for users to have the same[29] user name and password on a number of sites, and you start to see why getting one can open up the doors to other mischief.

The rules to minimise this are relatively simple:

- Always carefully type in the wanted address (or use your book marks feature in the web browser) to access a web site. Remember some unscrupulous operators register sites that are similar sounding or based on common miss-spelling errors.
- Never follow email links as it is easy to generate web addresses that look like the real thing, but are in fact different and point to a fake site (a very simple example is 6 'v' characters look quite like the 'www' start of most addresses).
- Beware of using a search tool such as Google to find a site, you might find the wrong one!
- Have different hard-to-guess passwords for the key log-ins you need.

It is worth considering the password issue in more detail. You should have at least the following 4 ***different*** passwords, and maybe more:

1. Your computer log-in account(s).
2. Your main email account for generally trusted use (remember, it is common for email password to be sent in the clear over any networks).
3. Another email account (e.g. web-mail) that you use for untrusted sites and in situations where you should not trust the computer (e.g. a cyber café, friend with a dodgy Windows PC, etc).
4. A password (ideally a couple) you use for any on-line shopping, etc.

[29] Men are much worse at 47% compared to women at 26% according to http://www.theregister.co.uk/2009/09/02/password_security_survey/

The idea here is to restrict the damage done if one is compromised. Also note that a password, to be any good, has to be hard to guess. What is not obvious is that dictionary words and/or common names make up a few tens of thousands to try, trivial for a computer search (known as a 'dictionary attack' for obvious reasons). Adding in some numbers, punctuation symbols, or odd caPitaliSAatioN can greatly increase the problem of brute-force searches.

Spam emails

I do not know for sure why the trade name for a brand of chopped pork became associated with unsolicited bulk email, but now in computing terms 'spam' means junk mail of the electronic sort.

This is basically low-grade advertisement, often for products of dubious merit (fake watches, vigra substitutes, loans, etc) and sometimes a route to spreading infection between computers (usually by means of attached files). The irritation aspect is independent of your operating system (unlike infection, which is unlikely for LINUX) and sadly there is little prospect of it being controlled any time soon. Today the majority of emails sent through the internet are spam!

Most email clients (such as Thunderbird) allow you to set a 'junk filter' that will move dubious emails to a separate folder where you can occasionally look through the list then clear the lot. You train the filter by telling it when you find spam in your in-box and it gradually learns of the sort of things that are junk.

A lot of web-mail systems, for example Yahoo and Google's Gmail, provide spam filtering and AV scanning of any downloads for 'free' (well, advert sponsored really). While it is easy for humans to spot such crap, computers are less effective. Just delete such messages.

However, if you get spam emails that offer an 'un-subscribe link' it is generally best not to use it, as that just confirms to the spammers that your email is in use.

The 'Chain Letter'

An old trick, one the relies on superstition or a false interpretation of what the email 'letter' is about. This is a method of propagating messages (and possibly email attachments or web links) that relies on humans doing its dirty work.

The basic scam is to tell you that passing it on will do you (or some worthy cause) some good, and/or not passing it on will bring bad luck, plague of locusts, etc. ***Just ignore them and delete it!***

Opportunity

So how do they find your computer? Where is the chance to find you?

These days you find most attacks are internet-driven: file sharing, by spam email, and by infected web sites. Sometimes the web sites are of reputable companies but, unfortunately, not very technically secure companies!

The other worrying thing is the ease with which most folk will part with personal data. Clearly this is a big risk for children who don't realise that on-line “Anne, 13, schoolgirl in to Pokémon” could quite easily be off-line “Eric, 50, unemployed bus driver”.

However, it is important to put the risk in to perspective. Very few attacks/abductions have ever taken place that are the result of the internet meetings/stalking, compared, for example, to car accidents that are accepted as part of modern life's risks.

Again, the secret is ***user education***. Be aware of the scams, don't trust most web sites, be very sceptical of on-line persona, and never give out information without careful consideration of its possible misuse (e.g. true date of birth on social web pages such as Facebook – think of what would your bank might ask over the phone to help confirm your identity?)

Cross Site Scripting (XSS)

The original web pages consisted just of text along with some information that allowed hyper-links (the text, etc, you click on to take you to another web page), plus some image tags to allow you to add pictures to the page.

Then someone had the idea of adding *scripting*, that is text in the web page that runs like a program so you could have animations, checking of web forms before submitting them, etc. By limiting what the script was allowed to do, it would be perfectly safe. What could possibly go wrong with that?

History, of course, has shown us that whatever can be exploited for gain (legal and illegal) will be!

Soon the bad guys were looking at ways of using scripting to attack PCs. The two basic things they try to do are:

- Intercept your access to a high value web page, such as on-line banking, to allow them to access private, valuable, data as if they were you.
- Direct your web browser to access another site, without you knowingly doing anything, where a virus attack can be initiated (the "drive-by infection").

While the 2nd point (virus infection) is far less likely for a LINUX PC, there are several variations of the XSS attack and they represent a very significant threat[30] as a number of them work on most systems (LINUX, Windows, Apple MacOS, etc) by compromising the original web site to serve up malicious scripts.

Stopping XSS attacks is possible by limiting the freedom for scripts to be run on your PC, either disabling javascript totally, or using the NoScript plug-in for Firefox to select which sites can and cannot run scripts.

Unfortunately, a lot of web sites, including some financial and commerce sites, depend a lot on scripting and on those scripts being able to run from one site (say the seller) to another (say the credit card handler). In these cases, the security of script blocking makes the site unusable for most practical purposes.

There in lies the rub: you have to choose:

- Ease of use and things actually working as expected (scripts enabled), or
- Very high security (scripts disabled, or skilled judgement and selective enabling)!

While some improvements are always being made, it is such a fundamental legacy problem of web site technology and design that no universal solution is practical.

[30] As of 2007, around 80% of security vulnerabilities reported were relating to XSS attacks, see the Symantec report http://eval.symantec.com/mktginfo/enterprise/white_papers/b-whitepaper_exec_summary_internet_security_threat_report_xiii_04-2008.en-us.pdf

However, there are a few things to note before feeling this is a futile problem that cannot be stopped by any simple and practical steps. Firstly, most of the PC-side vulnerabilities can be greatly reduced by the usual combination of:

- **Not** using Windows as the operating system, as it is the most common target.
- Prompt patching of your system when any software fixes come out (see page 8). Web browsers and the Adobe flash player in particular!
- Keeping any valuable web browsing to a separate user account (see page 52), or by clearing the web browser history & cookies before ***and*** after each 'valuable' browsing session.
- Using the AdBlock plug-in to limit syndicated advertisements that could include malicious code[31] (flash animations, etc,) as well as making your web experience nicer!
- Use of multiple passwords (already mentioned).
- Regular checking of any credit card statements, etc.

Protection against compromised web sites is more difficult, as it is not always obvious when one has been breached. While this is more likely with small companies than big technically supported ones, that is no guarantee[32].

Banks know that fraud occurs but have to accept some fraud, as eliminating it all is too expensive, too impractical, or basically would stop the (rather profitable) internet business from taking place in the first place. Hence provided that consumers are being ***reasonably responsible***, they provide protection if you do suffer from any losses.

Finally, the use of Firefox's NoScript plug-in can be very effective, as long as you have the knowledge to know what sites to enable or not. That is not a trivial problem, and most normal users just end up saying “yes” to every request thereby nullifying the benefits, but still having the aggravation of using it!

[31] Advert banners seen carrying malware here http://www.theregister.co.uk/2009/07/20/digital_spy_malware/

[32] Big names to fall include PayPal http://www.theregister.co.uk/2009/02/10/paypal_xss_bug/ , American Express http://www.theregister.co.uk/2008/12/20/american_express_website_bug_redux/ , some UK banks and newspaper http://www.theregister.co.uk/2009/06/01/website_bug_plague/ and six software security & anti-virus companies http://www.theregister.co.uk/2009/05/12/av_xss_six/ to add insult to their injury!

Google, your Friend or Informant?

Background

Google is a USA company that has around 60% of the world's web search business, but what it sells is primarily advertisement services with revenues of over $20 billion in 2009.

They are not the only one in this business, but such is Google's success that the trademarked name has now become a common-use verb, as in "to google for something".

Companies pay Google for 'advertisement' to be delivered to those using its search engine, and in particular that Holy Grail for advertisers, the successful 'targeted' advert. The advertisement comes in two primary forms:

- First (and obvious) are the 'sponsored links' which are not necessarily relevant to your search, but the company has told Google of key works it thinks are linked to its business.
- Second, and more subtle, is the payment for 'page ranking' where a result that is not particularly relevant to your search is promoted to the top 10 or so where most users select from.

There are several others search engines that are also worth considering[33], such as Yahoo! and Microsoft's Bing, along with smaller and older search engines such as AltaVista. In some regions Google are not the top choice, for example, in Russia it is Yandex, while in China it is Baidu that are the people's choices. All do basically the same thing.

In order to deliver such targeted adverts, they gather and retains all of the information it can about their users: what they search for, what results they used, what sort of sites they have visited repeatedly, etc. Most people don't think twice about typing stuff in to the search box, even when those searches could extend in to personal questions of health, medication, relationships, financial difficulties, morally questionable subjects, idle curiosity... Even when deliberately anonymised, a user's search history can be far more revealing than expected, as AOL found[34] to their cost.

So that is the Faustian pact: you get free searches to locate what you want from unimaginable volumes of on-line human knowledge, and in return Google gets your on-line soul.

How They Follow You

Realistically, you cannot make effective use of the Internet without the assistance of a web search engine, so how can you retain enough of your privacy when doing so?

That, of course, depends on how much is "enough", and to some extent that depends on who is asking for your details. By and large Google (and others) run automated systems and don't care much about you as a person. You are a "demographic" to sell to advertisers, and in a lot of cases they don't have anything as explicit as a name to tag you with. But not always.

[33] I have personally found that Google is often best for random searches, while Bing turned up better results for some business searches (for example, with "restaurants in Dundee").

[34] See http://en.wikipedia.org/wiki/AOL_search_data_scandal for a summary of that event.

If you sign-in to one of Google's services[35], such as Gmail or Google documents/calendar/etc, then they have a name (not necessarily real) and a social connection web of your email contacts, etc. If you are using a paid service, then of course they have a real name & details of who paid.

So how do they (Google, advertisers, etc) know it is you visiting or searching each time? That comes down to something unique about your computer, but that is not quite so easy to get, and it is something you can take steps to avoid.

Return to Sender

Firstly, they have your IP address, which is a number assigned to your Internet connection by your ISP. Your IP address is not necessarily fixed, it could be changed to another user when your modem is turned off, or periodically renewed as a policy of the ISP, etc. Without asking your ISP to find out who it was assigned to (which ISPs would generally do only for a legally competent request, by the police, for example), it is not that meaningful.

But they can easily look up who owns the block of addresses from which your one was allocated, and that provides them with your ISP and hence a moderate degree of geographic location information.

Since your IP address is not necessarily fixed or even unique[36], they can get other information from your web browser, the most basic of which is your operating system and web browser type. Initially it was thought that your web browser was not very unique, and of course it changes with each system update, but recent research[37] shows it to have a far more 'unique' ID than most anticipated. So while it may not say who you are, it could be used to show the same person visited successive web sites.

Hand in the Cookie Jar

So what they rely on generally are the web browser 'cookies', small files used to store site-related information such as your login status, on-line shopping trolley contents, etc, together with Adobe Flash's Local Shared Objects (LSO[38]) as unique tags that allow them to track where ***you*** have been, and in particular, when you return to a given site. These tags have important and legitimate uses of course, but from a privacy point of view this is the main concern.

Also possible, but fairly uncommon, are javascript hacks to read your browser's history[39].

You are the One and Only

With the rise of cookie-removing to protect privacy, the advertisement and other motives for keeping tabs on you have not stayed still. An example of this is where you are given a unique web page that provides a way of identifying your last log-in. For example, if you go to the main page for Facebook to log-in, the web address bar reads:

[35] The same logic applies for Yahoo! web mail and Microsoft's Hotmail services, of course.

[36] You could be one person sharing accounts on a computer, or one of several computer users behind NAT for a company, university/school, or home network, that shares the one IP address

[37] See http://www.theregister.co.uk/2010/05/17/browser_fingerprint/ which includes a link to the test web site. My work PC was "1 in a million" when I checked it!

[38] See http://en.wikipedia.org/wiki/Local_Shared_Object for more details.

[39] See http://www.theregister.co.uk/2010/04/05/firefox_browsing_history_fix/ for progress in this area.

http://www.facebook.com/

But this address also works to log-in and appears exactly the same:

http://www.facebook.com/index.php

However, on logging out your browser is left with something like this

http://www.facebook.com/index.php?lh=5960802237810c0b592f6dcb742af17&eu=1eg...

Most web browsers come back after closing with the same web pages open, so they know its you again! Of course, with this Facebook example it is less important as you will be logging-in once more and obviously they then know it is you, but they could track who else shares your computer account that way.

Dirty Deeds

More worrying are the attempts by some ISPs to profile your web browsing habits to muscle in on the on-line revenue taken by Google, etc.

Much to the UK government's disgrace, they have failed so far to prosecute the ISP British Telecom (BT) for their secret trials[40] of user profiling with the company Phorm (formally an ad-ware / spy-ware company). The whole outcry over the secret spying on BT customers has put off others, as both Virgin Media and TalkTalk had been in discussion about this technique as well, but how long ISP neutrality will last remains open.

Such methods are far more intrusive than Google's, due to the ISP's knowledge of the account holder, and due to the much greater problems of avoiding it.

How To Avoid Tracking

There are a number of steps you can take to deal with Google-like profiling, but the failed BT/Phorm fiasco is a worry and more difficult (though not impossible) to avoid.

Fresh Every Day

The first approach to limiting the cross referencing of your browsing and searching habits is to ensure that cookies and LSO are deleted each time you exit from your web browser. It is not perfect anonymity, but seriously fragments any profiles that include your actions. Be aware though of the key side effect – you won't be able to return to a web site exactly where you left off after closing the browser, instead you will need to log-in again, etc, as for the first time.

With Firefox go to the Edit → Preferences → Privacy tab, and choose either “private browsing session” (everything forgotten when you exit), or custom history settings and make sure the cookies are set to be deleted when you close Firefox.

Deleting all LSO is not so easy, but the Firefox Add-on called “BetterPrivacy” with do the same thing as for cookies. After installing it, exit Firefox and tick the box to always delete all LSO on exit.

You also need to avoid the “unique web page” approach to tracking as well. Normally the best option here is to make sure you close any pages that have such odd web addresses once you have

[40] See http://www.theregister.co.uk/2010/02/25/bt_cps/ for recent news, and follow the links for earlier news such as the EU's formal warning to the UK over the lack of legal action in protecting privacy.

logged out. Use the bookmark feature to save the 'clean' address you want to return to.

If worried, then configure Firefox to close all tabs on exit, and to start with your simple choice of home page each time it opens. Got to Edit → Preferences → Main tab then choose the “When Firefox starts:” drop-down option to have either a blank page, or your home page.

Masked Ball

The second approach to the problem with Google in particular is to visit their search service via a proxy, so Google's profile is one for everyone using the service and of little personal use.

A good option along this line is Scroogle http://www.scroogle.org/cgi-bin/scraper.htm which essentially anonymises your Google search history. You can get a Firefox plug-in[41] for this to make it easier via the normal search box.

Tunnel to Freedom

A final option, and one that allows you to avoid a spying ISP or untrusted WiFi points, is to use a Virtual Private Network (VPN) connection. Here you set up an encrypted connection from your PC to some trusted server, and all of your internet traffic is routed through the encrypted “tunnel” so it appears to come from the server.

Anyone inspecting the encrypted traffic knows only the volume of data, and not where you are ultimately visiting or what you are accessing. Anyone trying to track you back from a web search or web site visit gets the IP address of the VPN server, and without the VPN provider's cooperation they have no knowledge of who you are or where you are located.

However, a VPN is still not a perfect solution for various reasons:

- You still need to remove all cookies and LSO, or they can still build up a profile of you (just without the geo-location).
- If you use personalised services (such as Gmail) while Google searching (i.e. cookies from one have not been cleared before accessing the other) you are known irrespective of VPN, clearing of tracking objects, etc.
- VPN are often slower, tricky to set up, and useful ones typically cost £5-10 a month.

If you want to use Google personalised services and to remain moderately anonymous, in addition to the previous precautions, your options are:

- Have two different[42] web browsers, such as Opera and Firefox, and use one only for Gmail, etc, and the other for Google searches and web site visits.
- Use Bing or Yahoo! as your search engine, this can usually be configured easily by the drop-down menu accessed by clicking on the left of Firefox's search box. Swap around if you use other services (e.g. if logged in to Yahoo! web mail, then use Google or Bing for searching, etc)

[41] See https://addons.mozilla.org/en-US/firefox/addon/12506 though it is best to check before installing.

[42] If you have two instances of the same browser, they will most likely share cookies. LSO are shared anyway, so disable flash in one (e.g. Firefox flash-block plug-in, or Opera's Tools → Quick Preferences → Enable Plug-ins (un-tick that option)

- Always go via Scroogle or similar for web searches.

There is virtually no way to get “total privacy” on the internet, and if a criminal act is involved then most ISPs and VPN providers will tell the authorities what they know. However, everyone should take some steps to preserve their privacy on a point of principle, if not from the risk of such identifiable data being released unintentionally (as AOL did) or otherwise sold or leaked.

Using Google Effectively

Finally, it is worth mentioning a few tricks from Google's search capabilities that are not initially obvious. Most people just type in something along the lines of what they are looking for, click on “Search” and hope for the best. Often, they find what they want, but sometimes it can be quite frustrating to weed out the irrelevant stuff.

Google has an “Advanced Search” set of boxes to allow you to refine your criteria, and it also supports a safe-search option to protect children, etc, from some of the excesses of the web. Unfortunately, it needs cookies to remember you opted in to that protection, another reason why you should have separate log-in for each user.

But there are some things you can use on what you enter to help. Firstly, think about what you are asking for, if you just say **restaurant** then the best a search tool can do is guess[43] your location from you IP address (i.e. ISP) which may not be that local at all. So think about the question and add relevant stuff such as searching for **Dundee restaurant** or even **Dundee Chinese restaurant** depending on how narrow a choice you are interested in.

Next in the list of Google tips is the use of special characters[44] such as + and – to specify you want stuff included or excluded, so you could have decided that you did not want Chinese food tonight, so you could search for **Dundee restaurant -Chinese** to drop pages mentioning 'Chinese'.

Then we have quote marks “” to tell Google that you want a specific phrase. In the previous examples you will be offered pages that have any of the wanted words anywhere on the page. You could be more specific and search for **Dundee “Chinese restaurant”** for a page that has Dundee somewhere and the two words 'Chinese restaurant' together in that specific order.

Google is good at recognising common mis-spellings and correcting them, but for some cases they have mapped word meanings, so you can not, for example, use **color** and **colour** to distinguish between USA and UK pages – they answer to the same meaning in the Google database.

A final tip is the use of expressions that google recognises to refine a search, and one common example is **penguin site:sat.dundee.ac.uk** which tells Google that you are looking for this guide (or related references to penguins) but know the general web site address to find it.

Another example is **penguin filetype:pdf** where you are looking for a document of the 'pdf' type that is penguin-related.

[43] Of course, on smart phones, particularly those with GPS, you can opt for very precise location information to be made available to the search application!

[44] See http://www.googleguide.com/crafting_queries.html for a guide to more of the options.

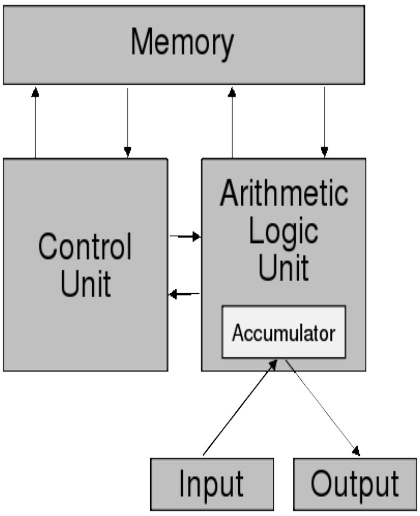
How it Works

What is a Computer?

Most electronic system have some sort of computer in them today, in washing machines as timers and cycle control, in cameras for focus and exposure control, etc.

However, when you say “computer” most people think of a general purpose computing machine: the personal computer (PC) in the office, or the mainframe or supercomputer in some far away room, something that can do a lot of different jobs depending on what you need done.

Such machines look, in a 'logical' sense of how the hardware works, like this diagram[45]:



The 'control unit' and the 'arithmetic and logic unit' are all part of one device in most cases, known as the Central Processing Unit (CPU). This does the “thinking” part, but it is in reality very dumb, it is only the sheer speed of doing stuff (in human terms) that gives the impression of complex actions. A typical PC's CPU executes up to a couple of billion instructions per second!

The 'memory' part of a computer holds the **data** you want to work on, for example the contents of a letter you are writing, and also the **program** (list of CPU instructions) that perform the desired actions on the data, for example, the word processor that allows you to edit the letter.

For practical reasons, the 'memory' shown in the simplified diagram is made up of two general classes of storage:

- There is the fast electronic memory often referred to as Random[46] Access Memory (RAM) which is *volatile*, in other words, if the power goes off, it 'forgets' the contents. The attraction here is the ability to read and write any data item in a time scale of nanoseconds (1ns = 1/1,000,000,000 second) thus allowing the CPU to operate with similar instructions times.
- And there is slower *non-volatile* memory, sometimes referred to as mass storage, that keeps stuff when the power is off. Examples of this are the floppy disk (now rare), Hard Disk Drive (HDD), CD/DVD disk drive, or 'flash memory' such as the USB drives. Those with mechanical operations such as HDD have access times measured in the several millisecond range (1ms = 1/1,000 second) hence are around a million times slower than RAM! However, once you have found the data reading/writing is much faster, but still orders of magnitude slower than RAM.

[45] Virtually all computers have the same basic building blocks and follow on from the early ideas of John von Neumann (1903-57) and Alan Mathison Turing (1912-54) on a universal computing machine, originally developed for code breaking and weapons simulation, but later to become one of the 20th century's most significant inventions.

[46] Today it seems odd to mention “random access” as being anything special, but early computers stored most data on tape which is a sequential access system.

Modern computers make clever use of RAM and mass storage to try and get the best of both worlds, but that is also why it is important to shut down a computer in an orderly manner (rather than just switching off the mains) so anything important in volatile storage can be moved to the non-volatile storage before it is lost.

Finally there is the 'Input / Output' system of the computer, this is the interesting part that most people are aware of. For human interaction there is the typically a keyboard and mouse for input, and a video display (monitor) and loudspeakers for output. Other common I/O devices include the camera, microphone, and printer, and there are less obvious devices such as the modem or network connection for transferring data in and out of other computers (e.g. for an internet search, etc).

At a low level, all a computer does is fetch an instruction from memory, decode what that requires to be done, and then to do that to the data and/or to go elsewhere for the next instruction. The secret then of doing useful stuff is the ***software*** part, the lists of instructions, while the ***hardware*** part is all about doing this fast enough without too much power/heat being involved.

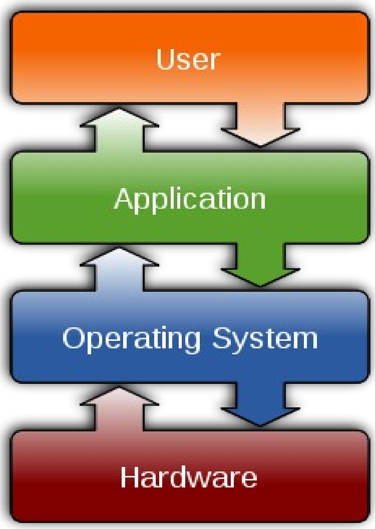
Operating Systems and Programs

When the computer starts up, the CPU has no idea what to do, so the very simple approach is to try the first location in memory for a valid CPU instruction.

In a typical PC, this starts running a very simple program often referred to as the Basic Input Output System (BIOS) which normally performs a basic self test of the PC's hardware (i.e. to find out how much RAM is installed and is it working OK, are there any CD/DVD drives? any HDD?, etc). This stage is normally over very quickly if there are no errors, and then it tries to boot from one of the mass storage devices such as the first HDD.

The expression 'boot' and booting' in computer usage comes from the expression “to pull one's self up by one's bootstraps”, a physically impossible gravity-defying act. But here one simple program (the BIOS) is able to load a much more complicated choice of program from another storage medium, increasing the complexity of what is executed by the CPU somewhat like lifting one's self up!

This is the point most people think of as something 'running', and that software is referred to as the Operating System (OS). The OS is not intended to actually *do* anything in terms an end user of the system would expect, but its job is to *manage* the system. Some of the key jobs it does include:



- Starting and stopping other programs (also referred to as 'applications') that do the business, such as a word processor, web browser, email client, etc.
- Managing the use of the computer's basic resources (memory, percentage time using the CPU, etc)
- Providing a uniform common 'view' of the hardware so that programs that need to do something like play music need not know what make/model of sound hardware is in use.
- Implement a **file system** that organises the storage of data in a logical manner so programs can request data by the file

name, without needing to know anything about the mass storage device(s) and the data's resulting location(s).

- Implementing the security model that controls who can access what.

The diagram shows the layers of a computer, the users does stuff with the 'application' programs, they in turn depend upon the OS to safely and reliably interact with the underlying hardware that is in use.

The early operating systems were developed in the days when human input was basically the keyboard, and output was a character printer or crude text monitor. What the user saw then was only the **command prompt**, that intimidating flashing cursor that waited for you to enter some cryptic command that would start something useful.

By the late 1980s that had changed as graphical interfaces became common, this added the mouse to allow a pointing action, and a larger display with high resolution graphics that allowed such nice features as menus listing actions that you could request.

The first approach to implementing this was to add a 'window manager' program that ran on top of the OS and provided the pretty and intuitive (well, *relatively* intuitive...) user interface. This approach[47] is still in use by most systems (by variety), for example LINUX, Apple's MacOS, Sun Microsystem's Solaris, etc.

However, from 1995 onwards Microsoft's Windows OS, with the most systems by total number of users, changed from Windows 3.1 (which essentially ran on top of MS-DOS) to Windows 95. This was more or less an integrated OS, window manager, and later on (regrettably) the web browser[48].

While the OS is the basic lower layer of software, most OS downloads or installation CDs come bundled with various applications that help use it and lead to a PC that can do something useful on its own (before you install any special applications).

What is LINUX?

LINUX is the name of an operating system, or more specifically it is the trademarked name of the OS **kernel**, which is the central core of the OS doing the most low-level and essential tasks.

This comes in a variety of distributions, of which Ubuntu is one. They are essentially a combination of the kernel and lots of other small programs that provide all the expected supporting tasks. As most people know of Microsoft's Windows system, here is a brief comparison:

[47] A minor advantage, should the user graphical interface crash, is the ability to log-in text-mode style and either restart it or issue a **reboot** command to restart everything in an orderly manner.

[48] The bundling of Internet Explorer within Windows was generally thought to be essentially an anti-competitive business move by Microsoft against Netscape, and the bad architecture of this from a security perspective has resulted in a significant proportion of the Windows/IE vulnerabilities in the last decade!

Aspect	LINUX	Windows
Cost & license terms	The basic software is free, and with the source code available (a strict requirement of the GPL license) so you can change anything if you are willing (or foolish enough) to try.	Costs around £50-£150 and you are most certainly not free to copy it. From XP onwards, Microsoft have a 'product activation' feature to prevent you changing PC without re-licensing your copy.
History	Started around 1991 by Linus Torvalds as a free and open implementation of the UNIX system favoured by large multi-user computer systems since the 1970s. Later versions became semi-commercially supported.	Developed by Microsoft around 1985 to provide an easy Apple-like option for PC users that were running their single-user MS-DOS system, but not significant until around 1990 (Windows 3.0) From Windows95 in 1995/6 it was developed into an integrated system.
Security	Very good (but not perfect). The underlying model of separating normal 'users' and the administrator account (that should be rarely used) limits the damage from a careless user and/or insecure application. In addition, the use of an 'execute' attribute for files, and less attempts to make life easy, greatly limits the opportunities for malware to infect the system.	Was utterly appalling, but has improved a lot. Origin in single-user and non-networked architecture lead to problems, later bad decision to integrate Internet Explorer into the 'protected mode' kernel operations contributed a lot of holes as well. Great popularity makes it more of a target as well.
Ease of Use	Was bad, but improving. Written by computer geeks for themselves (largely). As a result, the fundamental aspects are very well engineered (but often cryptic), but the end user side is not so good, more or less opposite to the Windows case. Learning basic ideas helps, both to use it and to avoid problems.	Generally good, as a lot of effort put in to the whole 'user experience' aspect. However, each new version (e.g. 2000 → XP → Vista) has re-arranged things and to some extent breaks the overall ease of use case.

Aspect	LINUX	Windows
Available software	Quite a lot, and mostly free. Some works very well, but quite a lot has a half-finished feeling. Limited support from some companies for their hardware due to small market share.	Very big range, but most is paid-for and a lot of 'free' offers on the internet have nasty side effects such as ad-ware and malware. Dominant market share ensures good support for hardware.

The Storage of Information

File Systems

Imagine you wanted to read Shakespeare's play “Macbeth” so you go to a reference library to find it. You might start with an index of what is in the library, and then look to any logical organisation of the contents. So you might look first for “English Literature”, followed by the sub-section of English literature “Plays” and of the plays, those by “Shakespeare”, before finding finally that they do indeed have a copy of Macbeth. Next the librarian sees that the index tells them the section, book case, shelf and index number to find the copy. They go and retrieve it by physical location, from the logically arranged name search, and return with it for you.

To the computer, the mass storage devices such as a hard disk look just like an array of millions of pages, each holding the same amount of data (typically 512 bytes). If you wanted to store some data, which pages are free? To recover previously stored data, where is it held?

The **file system** is the computer's equivalent of the librarian and their index book. It stores the name(s) used to locate the file (known as the **path**) and the file's name, this provides a uniform logical way of accessing stored information. It also manages the information about where that file is physically kept on disk, which is likely to be on multiple blocks scattered all over the underlying mass storage medium. From this it provides the data recovery process where by you open the file by its name and as you attempt to read (or write) that file, it transfers the data from (or to) the appropriate block(s) on the storage system.

Now remember that mass storage is very slow in electronic terms so if, as an example, you were writing a big file to disk you expect to wait as the disk's write head cycled backwards and forwards again and again over the disk as every different block's location is visited to write the appropriate contents.

That could take a lot of time (even in human terms, let alone computer terms) and early on that was simply how things were. However, as the size of affordable electronic memory increased dramatically over time, it became common to implement a **disk cache**. This allows you to write to 'disk' much more quickly, but in fact the data is temporary held in memory set aside for the cache and then the operating system arranges (in the background) the necessary write operations to the underlying disk in to an efficient order while you can get on with something more useful.

Thinking of a library, it is the difference between handing the librarian one book at a time and waiting until they return from storing it away (old way), or simply leaving a pile of books on their in-box and letting them store them away in an efficient manner after you leave (disk cache).

Even with flash memory storage that has no mechanical delay in accessing data (e.g. an external USB stick) the cache system is still useful for two reasons:

- The main memory is still orders of magnitude faster than the flash memory (even allowing for the lack of any mechanical access time).
- The cache operation minimises writes to the storage system during file copy/updates. This extends the life as flash memory has a finite number of write operations (no problem reading, though modern devices have some 'wear levelling' tricks to avoid premature failure of a very frequently accessed block such the file tables, for example).

The whole cache system works great, but has a very serious weakness: what if the power failed mid-operation? Or someone removes the mass storage device before it is done?

The result is Very Bad Stuff, typically with the data organisation tables being no longer consistent with what is stored on disk, and often being in an internally inconsistent state as well. Bad news all round, as not only have you lost the last data being written, you risk corrupting the tables needed to access the other (previously stored) files as well.

That is why you ***must*** unmount any removable device before pulling it out!

The unmount operation tells the operating system to complete the business of writing to that device, and to make *quite sure* the file allocation tables are finalised and committed to the disk as well, so it is then safe to remove the device. This is not a LINUX issue, as Windows does exactly the same using the bottom right 'safely remove hardware' icon (and pops up a warnings if you don't do it).

But what of a power failure or system crash? To help with those cases, the default choice of file system for LINUX internal storage is one called 'ext3' and it is a **journalling file system.** In this case, as it writes data to disk it keeps a journal (also on disk) of intended operations, and if they were successfully completed. If it all goes horribly wrong and the system is not cleanly stopped, the system performs a check on start-up[49] and goes through the journal to allow the system to be re-started in the last known consistent state.

Of course you will still have lost any data not committed to disk at the point of failure (which can be up to 30 seconds after you *think* it was written!) but you are at least back to something safe and usable in a minute or so, avoiding the tedium of a full file system check which can take literally ***hours*** to perform on large disks.

So please shut down properly!

There is a slight performance penalty in journalling, but most users consider data integrity more important. Of course, you can always get a faster disk or RAID system if you do find the file read/write speed is limiting your PC's performance.

Finally, if you are about to do something that might just result in an unexpected crash/reboot/power-

[49] Occasionally it performs such a file-system check at boot-up 'just in case', typically after a long time idle and/or a certain number of start-ups. This can be changed/disabled with the tune2fs utility, but I ***strongly*** advise against it. While tedious, the periodic check is all about making sure your data is safe in the file system!

off, then consider using the ***sync*** command[50] which tells the OS to flush to disk quickly, and then waiting a couple of seconds or so for it to complete before proceeding with your dodgy deeds...

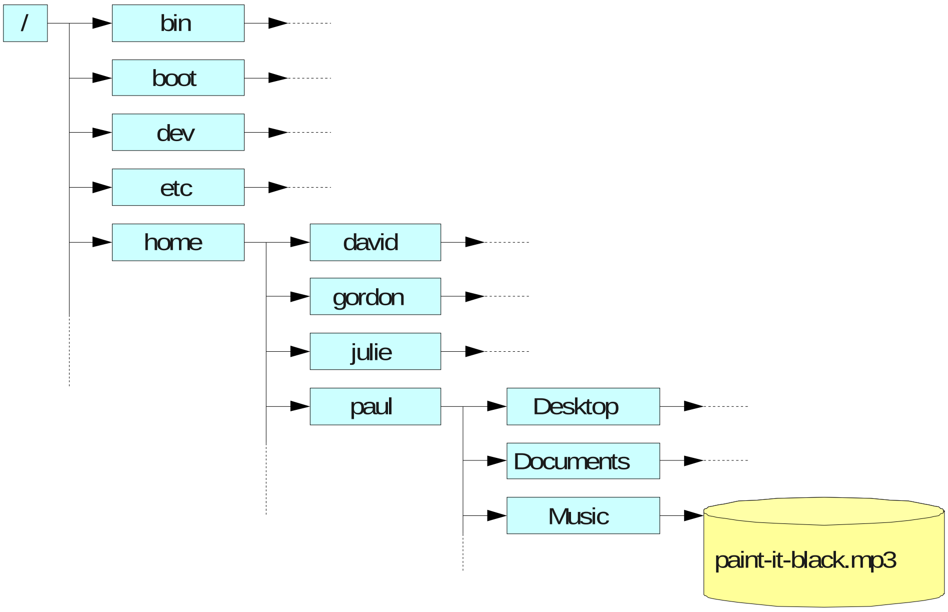
The Directory Tree and Storage Devices

Most file systems are arranged as a 'tree', starting at the root directory (folder) and branching out from there. An example might be:

/home/paul/Music/paint-it-black.mp3

The forward-slash character '/' is the directory separator, and this has a special function in linking steps the the hierarchy of the tree.

This example tells the system to start in the root directory, as shown by the '/' character at the very beginning. From there it looks for another directory located there called[51] 'home' (a directory is just a special file with information about other files stored there, which can include other sub-directories). Then in /home it should find another directory called paul and in there (/home/paul) is should find one called Music, before finally finding the file called paint-it-black.mp3 which we wanted to access. In diagram form:



The whole point of directories making up the path to the file is to allow a sensible hierarchy to help organise the data (i.e. the files), and thus to make it easier to locate and protect it, similar to having a logical index of books by author and/or by type, rather than having everything in a single alphabetical listing of title.

In LINUX there is one file system that rules them all. Everything appears as a file, with all ultimately branching from the first '/' root location. Any and all storage devices are 'mounted'

[50] The web site www.sysinternals.com provide a similar disk utility for Windows should you need it.

[51] Generally names for files & directories can use all of the alphabet (a-z and A-Z) and some forms of punctuation, such as the hyphen (minus sign) '-' and the underscore '_' characters, but other punctuation symbols are reserved for special functions. While the space ' ' is a valid part of a name, its use in a file or directory name can make some actions more difficult as spaces are also used to separate commands.

somewhere in this system. The main disk becomes the mount point for '/' and other storage devices, such as the CD-ROM drive, are mounted elsewhere, typically appearing as directories under the directory /media

Keep to your own /home/*username* directory area unless you really know what you are doing!

File Permissions: Protection & Privacy

Continuing the library analogy, you may find that books have different categories with respect to access: some might be available to borrow by anyone with a valid library card, others kept in the reference section where you can read them but not remove them, and some rare books might be available only with special prior permission.

As well as having a name, a file normally has other useful attributes as well, depending on the OS and file system in use, that basically control what you can do with them.

System using the old FAT16 or FAT32 systems (typical for USB drives[52] and cameras) only have those attributes originally supported by MS-DOS, however, these are useless for a multi-user system[53] as there is nothing to control how each <u>use</u>r can access the file!

So we will concentrate on how the LINUX file system permissions work. In this case there are 3 relationship categories:

User - the owner of the file (normally its creator).

Group - other in the file's group[54] (e.g. members of the same class, business section, etc)

Other - anyone else not in the user's group.

For each of these 3 categories, there are are 3 traditional attributes: read, write and execute. What they do depends upon the file, if it is a regular file (e.g. normal data or program file), or if it is a directory (i.e. also known as a folder, as they are just special files that list what is kept in that part of the file system path):

[52] While FAT32 is widely used due to its simplicity and wide compatibility, it is not a very good file system and also has the restriction that no ***single file*** can exceed 4GB, which can cause problems with large video files or backing up a lot of data to an external HDD.

[53] Microsoft addressed this with their NTFS file system which added Access Control List (ACL) permissions, however, they are far more complex in practice and rarely used to good effect.

[54] Normally a file's owner and group IDs are the same, but they can be changed to be different cases.

File type \ Attribute	r = Read	w = Write	x = Execute
Normal file	Permits reading of the file's contents.	Permits writing/editing the file's contents.	Permits the running (execution) of the file as a program.
Directory	Permits the reading of the files' names, thus allowing the identification of what is stored there.	Permits the creation, deleting or re-naming of files[55].	Permits enter to the directory.

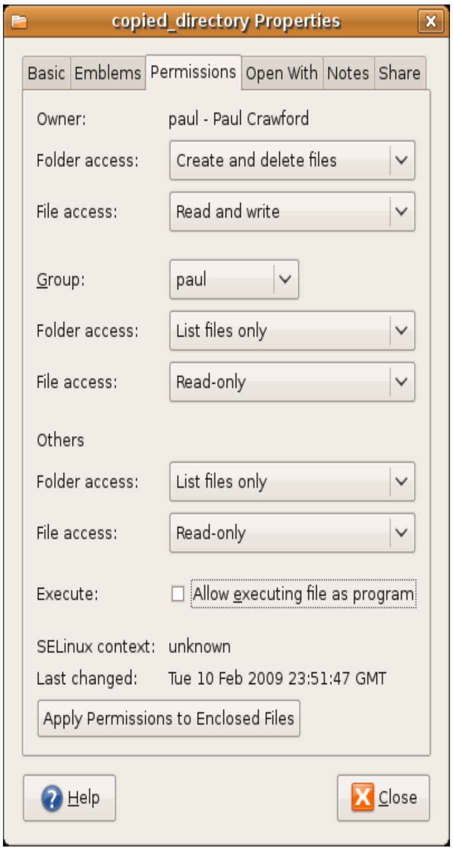
There are some odd aspects to this at first glance. For example, you can make a file read-only but it can still be deleted! Why? Because removing the file name is an operation on the ***directory that lists it***, and not on the file contents.

So if you wanted to allow others to read a file, but not change or delete it, you must make the file read-only (so they can't modify the contents) and also make the directory non-writeable for those users (so they cannot do anything to the name such a removing it).

Similarly, you can have a directory that is set to be executable-only to 'other users' so they can't see what is in there (no read permission) or make changes (no write permission), but they can enter it. So if you tell them the name of a specific file that already exists there, and the file is read-enabled for them, they can read it! Useful for web sites where you email a link for them to download it, but no one else is likely to guess its existence.

There are some special extra settings that replace the 'x' mode: there is the 't' mode that allows a directory to be writeable to others but they can not modify files they don't own (typically used for the /tmp temporary location that is globally accessible), there are also the 's' modes that allow others to execute a file with the owner's user or group privileges, a powerful and potentially dangerous thing if the file is owned by the root user. If you don't fully understand the implications of these modes, do not use them!

Note there is no 'hidden' attribute for LINUX file systems, but the convention is that any file name or folder starting with a '.' (full stop), such as the file `.bashrc` in the user's home directory, will not normally[56] be shown. You can of course change the preferences to show hidden files in Nautilus (View → Show Hidden Files).


copied_directory Properties
Basic
Emblems
Permissions
Open With
Notes
Share
Apply Permissions to Enclosed Files
Help
Close

Owner: paul - Paul Crawford
Folder access: Create and delete files
File access: Read and write

Group: paul
Folder access: List files only
File access: Read-only

Others
Folder access: List files only
File access: Read-only

Execute: □ Allow executing file as program

SELinux context: unknown
Last changed: Tue 10 Feb 2009 23:51:47 GMT

[55] The special execute = 't' case excepted, where you can only modify your own file.

Nautilus allows you to check and modify the file permissions in a nice graphical way, but there is the command `chmod` to provide a fast (and potentially dangerous!) way of doing this for lots of files.

It is possible to change the user (owner) and group settings of a file/directory as well (using `chown` and `chgrp` commands), but normally only the **root** user can do that. These commands, when executed by root, are also potentially dangerous.

While the internal LINUX file system (e.g. ext3) supports the added security of the 'execute' permission, external media such as CD-ROM and USB drives are typically without this option and so are *assumed* to be read/write/execute enabled. This is generally a bad idea so it is always safer after copying stuff to the HDD to set it to 'rw' only.

Using the Nautilus explorer, right-click on the copied file or directory and then choose the 'Permissions' tab. Change to something like this (image shown right) with owner 'read and write' file access, for group and others to 'read only' and **un-tick the "Allow executing file as program" box** before finally pressing the "Apply permissions to Enclosed Files" button.

If it is a single file, the idea is the same but there is no 'apply' box, it just happens as you change the settings (or so it appears).

Think before doing this, it should be only used in such a blanket manner with stuff just copied over from removable media (or any hard disk that lacks LINUX permissions such as an NTFS one taken from an older Windows computer).

Finally remember that file permissions are often there to protect you against both malicious and ***careless actions***, so do not just force them if you don't know why there were set!

Links: A File by Another Name

Going back to the library analogy, you might be looking under "Scottish History" and in that section you look for "Cultural References" and there you find another reference to Shakespeare's play Macbeth. How is this book found in the library?

There are 3 options that spring to mind:

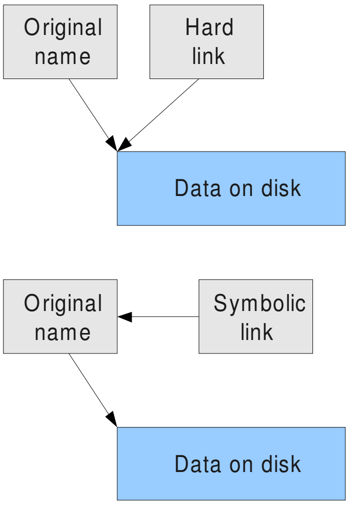
1. They have two copies of Macbeth: one in a location linked to by index from "English Literature → Plays → Shakespeare" and another copy in a location linked to by "Scottish History → Cultural References".
2. They have one copy, and the index book provides the same physical location from both "English Literature → Plays → Shakespeare" and from "Scottish History → Cultural References"
3. The have one copy, and when you look in "Scottish History → Cultural References" the entry for Macbeth has a cross-reference telling you to go and look up " English Literature → Plays → Shakespeare → Macbeth" for the actual location.

[56] Or use the option with the 'ls' command to list all files, for example with "ls -la" to list file, with 'all' (i.e. include those not normally viewed) and in 'long format' (which shows each file's type, permissions, hard links, owner, group, size, and name).

The same basic ideas can be used in the LINUX file system to provide multiple ways of locating data in a file:

- The first case is what happens if you make a copy of the file to a 2nd location. You then have two identical sets of data at that point in time, and from there you can do different things with them. This could be just what you want, but if you you do not want to independently modify the data (i.e. you expect the two copies to stay identical) then you are wasting disk space with a 2nd copy.
- The second case is what happens if you create a **hard link** to a file. Every file has at least one hard link, from its original file name, and you can create more entries in the file system for the same data. If you then delete the original file name, the data is still available until the last hard link is removed. However, this can be confusing as normally the multiple hard links are not obviously[57] referring to the same data.
- The third case is often the most useful and is referred to as a **symbolic link** under LINUX. This is a bit like Microsoft's "short cut" feature, but properly implemented[58]. If you delete the original file, the data on disk is gone and you end up with a link to nowhere. However, create a new file with the same original name, and once again the link allows you to access the new data by the alternative path/name. Delete the link, and only the link is gone (the original data is unchanged).

The basic operation of the two types of link can possibly be better illustrated by the diagram here:



In the Nautilus explorer you can right-click on a file and there is an option to create a link, but it just goes in the same directory so you generally have to cut/paste it elsewhere to be useful.

There is the command **ln** to create links, both hard and symbolic, to both files and directories. However, the use of this is not completely intuitive and it defaults to creating a hard link, which you should generally be wary of, so you need to use it this way:

```
ln -s original_path_name new_symbolic_link
```

Where the '-s' option tells it to create a symbolic link. As an example, let us assume that Paul wants to make his music collection available so sets his permissions to suit this. Julie can access these files without her having to explore the whole file system to find it with this sort of command:

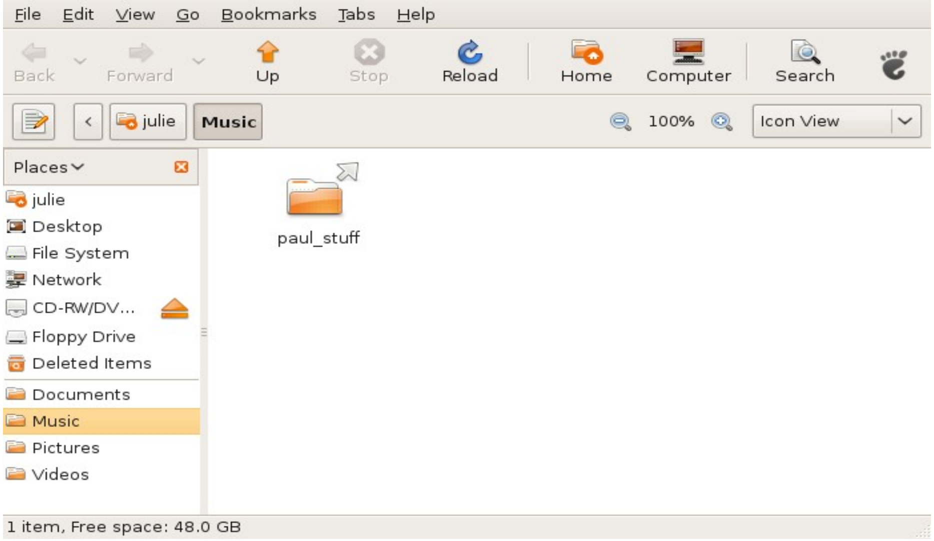
```
ln -s /home/paul/Music /home/julie/Music/paul_stuff
```

This will create the symbolic link paul_stuff in Julie's Music folder. Note the the original path name

[57] If you use the ls command with the '-l' (long format) option you get information about the number of hard links, but this is also non-obvious. Even the normal manual description of the ls command skips the meaning of the number showing hard links!

[58] The Windows short-cut is only understood by Windows Explorer and relies on that interface to use it as it is not part of the file system, but a crude hack added on top. If you try to use any basic file access methods, for example editing the `something.lnk` file, you basically get gibberish! The LINUX/UNIX symbolic link pre-dated Microsoft's "short cut" and is an integrated file-system feature.

should be unambiguous, so use the full path as shown in this example (i.e. starting from '/'). When viewing her folder using Nautilus it will look like something the following, with the arrow pointing top-right added to the new folder's icon to indicate it is a link (and so is not actually in Julie's Music folder tree):



Links work for individual data/program files and for directories (which this example shows), since both are just slight variations of a file in LINUX.

Deleting Data

In a library or office you may want to get rid of some paperwork that is no longer needed. The two questions most often asked (or, maybe, that should be asked) are:

- I have just had a “Doh!” moment and thrown stuff out I now need, can I get it back?
- I want to dispose of confidential data, can anyone else recover it?

With paper you can rake the bin and get things back very easily, as long as it has not been emptied yet. But if you can recover it that way, then others can recover it just as well. Do you have a shredder?

Generally, it is safest to use the Nautilus explorer to delete files. It places them in to a waste basket (the deleted items folder, bottom right of the desk top view) and they remain there until the system starts to run out of disk space, then older stuff gets deleted for real.

If you have just deleted something by mistake you can left-click once on the waste basket icon and it shows you what it still has, select the file/directory(s) you want (hold down the Ctrl key to click on multiple files to select them), and then right-click and choose “restore”.

You can empty the whole bin (using the “Empty Deleted Items” button, top-ish right), or individual files in the bin (select them but right-click and choose “Delete permanently”). This achieves the same as a normal delete[59] (which you can do directly to skip the waste basket by selecting a file normally, then using the Shift+Del combination).

[59] The command **rm** is used to remove files from the command prompt but take care with it!

However, the data itself has not been removed from your storage device(s)! What happens during the delete operation is the file tables relating to the data are marked as free so in the future another file can make use of the space.

So even after a 'permanent delete' you can still potentially recover the data! But it is not so easy in the LINUX file system case, and if you have used the storage much afterwards there is a danger that part of the original data will have been overwritten by new stuff. Generally you should never rely on an “un-delete” operation saving you from lack of care.

If you have a storage device that is not used much after the delete operation (or file table corruption) there are various free tool to attempt recovery, but professional help is very expensive (think £2,000+ to begin with) and it comes with no guarantee of success. **Keep a backup!**

What about data removal for proper privacy? Well you should properly wipe a disk, and by “wipe” I mean overwrite all of it with zeros (or random gibberish) as you might clean a blackboard with a damp cloth, before disposal. It is sometimes suggested to mechanically damage it as well, just to be sure, but that is potentially dangerous to your safety and often ***not even as effective as a minimum wipe***.

How good is wiping? Well, it depends upon who you are trying to prevent accessing your data. If you have the likes of GCHQ, NSA, or serious crime police forensics interested, it is very difficult to remove all traces, but in most cases you just need to avoid the “identity theft” type of criminal and they lack much capability beyond the simple data recovery tools. For that case simply overwriting with zeros will do, but to make life *really difficult* you can do overwrite several times with random data (gets a bit tedious though, as a typical hard disk takes ***hours*** to wipe per pass).

Before you ask me about wiping (hint: look up the **dd** program), you really should be pondering the question:

“Why is there sensitive data on an unencrypted device in the first place?”

After all, if the storage device was lost or stolen, all of the data would be compromised[60] and not just the file(s) you are thinking about properly deleting now.

Information on setting up an encrypted private directory on your PC:

https://help.ubuntu.com/community/EncryptedPrivateDirectory

For software that allows you to have encrypted systems on USB devices, etc, this may be of interest:

http://www.truecrypt.org/

However, to access the data on other computers you need to have TrueCrypt installed on them and to provide the password(s) used (you can also get TrueCrypt for Windows and MacOS). If that other PC has been compromised by a key-logging virus, etc, (most likely a Windows PC) both the data and your password will be exposed by that route!

Also consider hardware-encrypted storage such as those from https://www.ironkey.com/

[60] Think of HMRC loosing two CDs in the post with the details of 25 million UK citizens. They said it was “password protected”, but that usually implies the most feeble of protection that simple spreadsheet or database programs (like MS Excel or Access) offer for their files.

Disk Operations: Partition & Format

There will be occasions when you need to repair a corrupted 'disk', most often a USB stick that was not properly unmounted. Or you could have a new disk that must be prepared with a file system before you can use it. These operations are quite dangerous because if you do it to your ***main disk by mistake, your system is toast!***

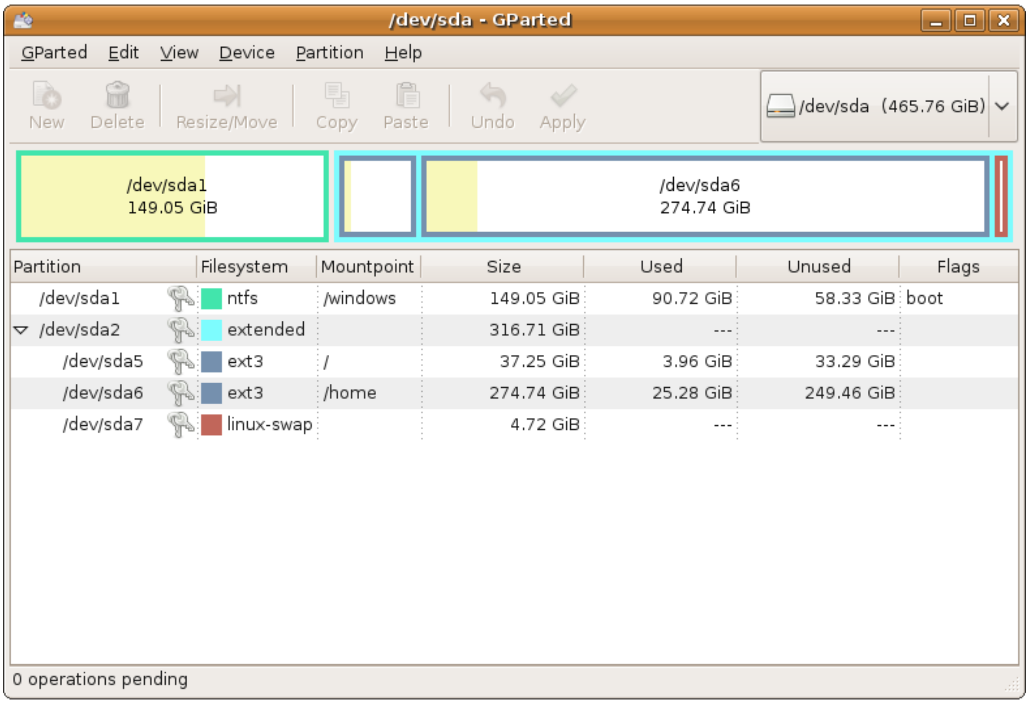
First the analogy: When you make a library, you have an area of land and you then build the room(s) to house the books, and finally you organise the book cases and index numbers, etc, to find things. The most efficient use of space is one single large room, as then you can put new books wherever there is a suitable gap.

However, you can see clear practical reasons why you might choose to have several rooms. For example, one for children's books to minimise any disruption from young readers, maybe another for rare books that need special care and environment, etc. In this case you have to decide in advance how to divide you your land for the rooms during the building process, and accept that some excess space on one room may have been more useful in another later on.

The storage device is the equivalent of the land, and they are given names like /dev/sda for the first physical disk in the system, then /dev/sdb for the second, etc. The creation of a **partition** on the disk is the equivalent of building a room. Once you have created the partition, the actual file system is created by the process known as **formatting** and that is the organisation of the bookcases.

In a lot of cases, you have a single partition using the whole of the disk, but there can be good reasons for having a couple of partitions. The system I installed has two partitions, one smaller one for the main system starting with '/' and a second one for all of the /home area where the user's files are stored. The first disk's primary partition has a name like /dev/sda1 and usually the other partitions are of the 'extended' type, they start with /dev/sda5, then /dev/sda6, etc.

The best way to do any disk operations is with the tool **gparted** which can be installed using the usual Synaptic Package Manager. It then appears as System → Administration → Partition Editor and you need to supply your password to use it (to help deter careless actions). When you start it, you should see something vaguely like this:



Partition	Filesystem	Mountpoint	Size	Used	Unused	Flags
/dev/sda1	🟩 ntfs	/windows	149.05 GiB	90.72 GiB	58.33 GiB	boot
▽ /dev/sda2	🟦 extended		316.71 GiB	---	---	
/dev/sda5	🟦 ext3	/	37.25 GiB	3.96 GiB	33.29 GiB	
/dev/sda6	🟦 ext3	/home	274.74 GiB	25.28 GiB	249.46 GiB	
/dev/sda7	🟥 linux-swap		4.72 GiB	---	---	

This is an example of a main disk which you **should not**[61] **edit!** You can tell it is the main disk as it has the mount point '/' (and '/home') listed, so the first thing to do is go to the top right drop-down menu and select a different disk to work on (the later example chooses /dev/sdb)

If you have an external device plugged in, and it is not totally unrecognised or broken, the system will have mounted it under '/media' and the first thing to do after selecting it (via top right drop-down menu) is to right-click on the disk picture and unmount the device. Once unmounted, the various options to edit the disk become available.

If you have a disk that has wanted data, but may be corrupted, then right-click on the partition of interest and chose the 'Check' option[62] then click on the 'Apply' button (green tick).

If the disk is unusable, or you want to change file system, then use Delete and New to create the new file system. Of course, you will loose all data on that disk so have it backed-up first if it has anything important on it!

[61] Gparted will generally not let you edit your main disk, but do not rely on that sensibility check. Think first!

[62] The **gparted** tool can fix FAT16, FAT32 and ext3 file systems, but not the Windows NTFS type unless you also install the **ntfsprog** tool kit. If you have a problem with a NTFS disk it may be best to use the ntfsclone tool to make a backup image (just in case), but then connect it to a Windows computer and use the chkdsk command to fix it.

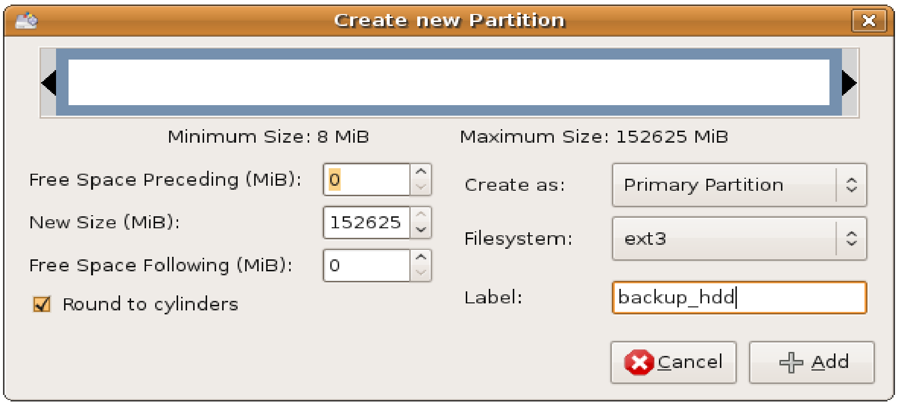
The typical choice of file system comes down to these choices:

File System	Advantages	Disadvantages
FAT16	Compatible with most OS (DOS, Windows, Apple & LINUX).	No security and little fault tolerance. Can't be used for drives above 2GB, inefficient for small files on large devices.
FAT32	Compatible with most recent computers (Windows 98 onwards, Apple & LINUX).	No security and little fault tolerance. Can be used for bigger drives, but no *single file* can exceed 4GB.
NTFS	Preferred choice for Windows 2000 onwards, supports large files.	Relatively slow access on LINUX machines, and security models are not compatible between OS. Not compatible with simple devices (i.e. most digital cameras, picture frames, etc).
ext3	Robust system for LINUX use, supports proper permissions.	Generally only understood by LINUX computer.

So the choice really boils down to these factors:

- If you want good performance and reliability for LINUX use (e.g. backup hard disk drive) then choose ext3.
- If you want compatibility for transferring data to other non-LINUX devices then choose FAT16 (disk of 512MB or below) or FAT32 (above 512MB). Remember the 4GB limit though, this is easy to reach with a DVD or backup file.

To create a new file system, having unmounted the device select the partition with a right-click on its picture and chose 'Delete' then 'New' and choose the file system type and a label name. A label is recommended, but not essential, and generally it is best not to have spaces in the name but to use the hyphen '-' or underscore '_' to join any words. The menu should then look like the following:



Create new Partition

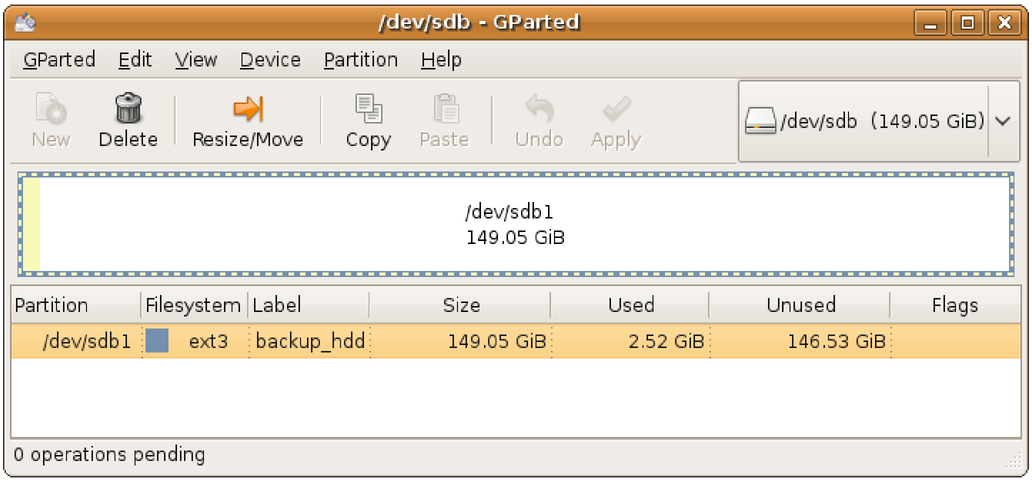
Minimum Size: 8 MiB
Maximum Size: 152625 MiB

Free Space Preceding (MiB): 0
New Size (MiB): 152625
Free Space Following (MiB): 0
☑ Round to cylinders

Create as: Primary Partition
Filesystem: ext3
Label: backup_hdd

Here we are creating a primary partition for the whole disk, choosing the 'ext3' file system, and the label will be 'backup_hdd' which then becomes the default name when mounted automatically as /media/backup_hdd

Click on 'Add' to make it happen, and of course confirm it when promoted that you should have backed up first! After a minute or so, it should be done and you will see something like this:



Partition	Filesystem	Label	Size	Used	Unused	Flags
/dev/sdb1	■ ext3	backup_hdd	149.05 GiB	2.52 GiB	146.53 GiB	

0 operations pending

Note that around 2% of the disk appears to be used before we have put any data on to the disk, this is normal for ext3 and is space reserved for directories, file tables, etc.

Users and Groups

A LINUX system starts with the **root** user, they are like God to the system and can do anything: create and destroy user accounts, format hard disk, install or remove software, etc. Due to such power, you should ***never be logged in as root*** unless you have a specific task that needs it.

Other users have restricted powers by virtue of the file permissions and ownership of system files created during set-up. They cannot make system-wide changes that impact on other users without an additional privilege being granted.

Users are often organised part of defined groups, for example others in a school class or business section of a company. Groups exist to provide ways of offering access and sharing in well defined ways, and users are normally a member of several groups that are not always other human users. For example, it makes sense to restrict children from having access to a printer or fax machine just to protect it and limit the operating cost, so by having a group for the printer you can have certain users as members, or not, of that group.

With Ubuntu you can use the graphical interface System → Administration → Users & Groups to add, delete and modify users properties. One important aspect of this is the option to perform administration tasks, such as installing software or updates. This should only be offered to responsible people, and generally one user should do it all so they keep track of what is happing.

Unlike Windows where often you have to log out and in again as Administrator[63] to install software, you can do it in LINUX with the package manager or from the command line using the **sudo** command. Both prompt you for your password as confirmation of what you are going to do.

Be warned though! Using sudo allows you to run any command with root privileges (think "super user do it") so it is quite possible to toast your whole system this way.

[63] Windows has a "run as" option a bit like sudo, but it rarely works properly for software installations.

Finally, it is worth remembering that each user account need not represent a different person. In fact, there are very good reasons why an individual might benefit from having two or more user accounts, the on-line equivalent of Dr Jekyll doing his good work by day, and Mr Hyde exploring the dark side of London by night... The reasons include:

- Keeping work and recreation time clearly separated by reducing the temptations to do the other.
- Improving security as work (or personal finance) information is not accessed via the same account / web browser as less trustworthy sites.
- Preserving settings for specific jobs where multiple work accounts can remember the last set of relevant web pages, documents used, etc.

In terms of the security aspect, you would set up only one or two accounts with administrative privileges. Maybe the work account, but possibly also an account used specifically for system administration. However, it is a good idea to have more than one admin account so you can fix the system should you manage to get locked out of the main one!

Finally, do not allow administration from accounts used for “dubious” activities (e.g. resolving late night drunken debates about something, where Google becomes the source of arbitration, and it goes downhill from there, etc), or for those who do not have the knowledge and/or responsibility to use it wisely.

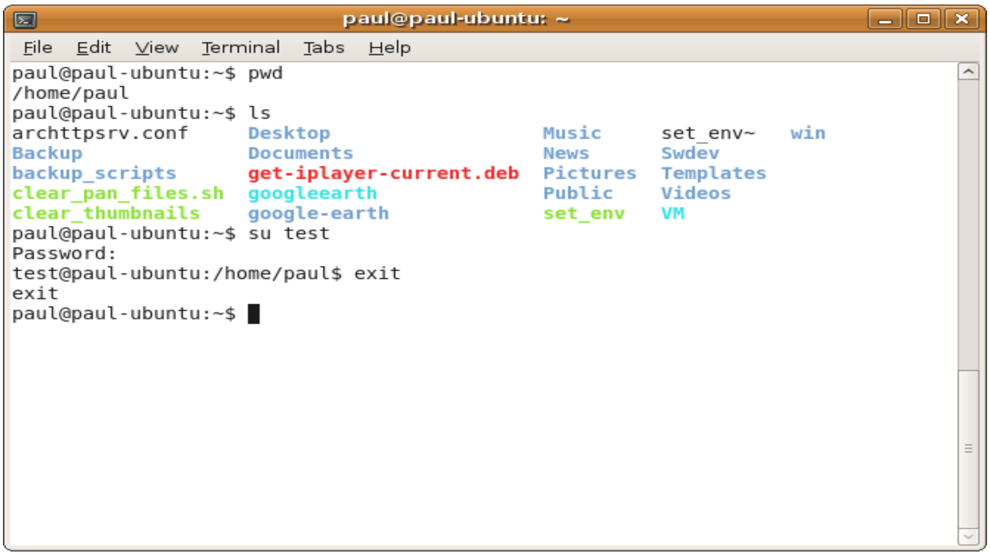
Of course, everyone will tell you they are responsible, and would never do drunken Googling, but plan for reality!

Appendix A - Command Prompt Stuff

Most users find the graphical user interface (GUI) the easiest and safest option, as it shows you the choices you have and is relatively intuitive. However, for a number of tasks it is faster and more powerful to use the old fashioned command prompt method.

The GUI is a bit like a drinks vending machine, you see the choices and click on the button for what you want. The command prompt is more like asking to speak to the sommelier (wine waiter) in an upper class restaurant: typically an intimidating step but one that ultimately allows you a far better choice of what you can get to suit your meal, particularly if you have some knowledge of wine.

You start by opening a terminal window (Applications → Accessories → Terminal Window) which offers you a basic way of typing in commands. You end up with something looking like this:



```
paul@paul-ubuntu:~$ pwd
/home/paul
paul@paul-ubuntu:~$ ls
archttpsrv.conf     Desktop                  Music     set_env~   win
Backup              Documents                News      Swdev
backup_scripts      get-iplayer-current.deb  Pictures  Templates
clear_pan_files.sh  googleearth              Public    Videos
clear_thumbnails    google-earth             set_env   VM
paul@paul-ubuntu:~$ su test
Password:
test@paul-ubuntu:/home/paul$ exit
exit
paul@paul-ubuntu:~$ 
```

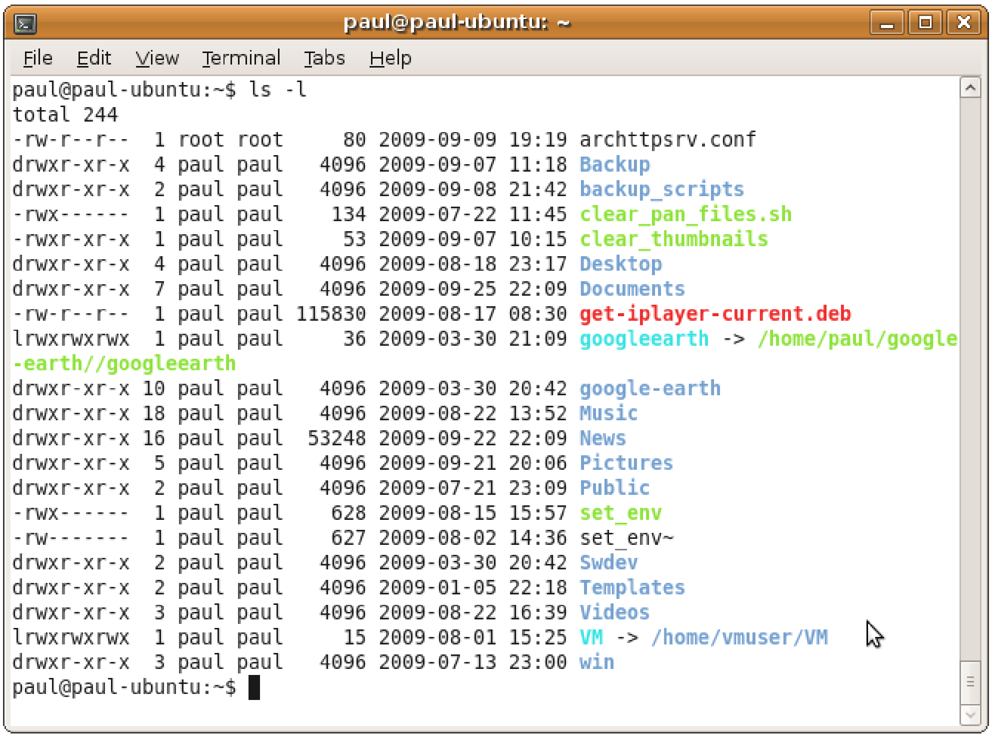
This example requires a little explanation. The text **paul@paul-ubuntu:~$** shows the user who is logged in, in this case starting as “paul”, and the computer's name “paul-ubuntu” and where they currently are in the file system, with “~” representing their home directory. To prove this, I typed in the command 'pwd' and pressed the return key and it printed out the working directory as /home/paul which agrees.

Next I entered the command 'ls' and that listed the files and folders in my home directory. The colours tell you something about the type of file with the name (e.g. blue is a directory/folder).

Then I used the 'su' command to switch user to become “test”, at this point the command prompt changed to **test@paul-ubuntu:/home/paul$** showing the new user name, and also the location changed from ~ to /home/paul as that is not the test user's home directory. Finally the exit command returned me back to myself.

Take great care, as you can do a lot without any warnings this way!

To find out more about the files, I can then use the “long” format of the list files command '`ls -l`' as shown here:



```
paul@paul-ubuntu:~$ ls -l
total 244
-rw-r--r--  1 root root     80 2009-09-09 19:19 archttpsrv.conf
drwxr-xr-x  4 paul paul   4096 2009-09-07 11:18 Backup
drwxr-xr-x  2 paul paul   4096 2009-09-08 21:42 backup_scripts
-rwx------  1 paul paul    134 2009-07-22 11:45 clear_pan_files.sh
-rwxr-xr-x  1 paul paul     53 2009-09-07 10:15 clear_thumbnails
drwxr-xr-x  4 paul paul   4096 2009-08-18 23:17 Desktop
drwxr-xr-x  7 paul paul   4096 2009-09-25 22:09 Documents
-rw-r--r--  1 paul paul 115830 2009-08-17 08:30 get-iplayer-current.deb
lrwxrwxrwx  1 paul paul     36 2009-03-30 21:09 googleearth -> /home/paul/google
-earth//googleearth
drwxr-xr-x 10 paul paul   4096 2009-03-30 20:42 google-earth
drwxr-xr-x 18 paul paul   4096 2009-08-22 13:52 Music
drwxr-xr-x 16 paul paul  53248 2009-09-22 22:09 News
drwxr-xr-x  5 paul paul   4096 2009-09-21 20:06 Pictures
drwxr-xr-x  2 paul paul   4096 2009-07-21 23:09 Public
-rwx------  1 paul paul    628 2009-08-15 15:57 set_env
-rw-------  1 paul paul    627 2009-08-02 14:36 set_env~
drwxr-xr-x  2 paul paul   4096 2009-03-30 20:42 Swdev
drwxr-xr-x  2 paul paul   4096 2009-01-05 22:18 Templates
drwxr-xr-x  3 paul paul   4096 2009-08-22 16:39 Videos
lrwxrwxrwx  1 paul paul     15 2009-08-01 15:25 VM -> /home/vmuser/VM
drwxr-xr-x  3 paul paul   4096 2009-07-13 23:00 win
paul@paul-ubuntu:~$ 
```

Interpreting this page of gibberish requires some knowledge, but it shows one of the more important command examples you might want to understand:

Starting from the left of each file's listing, we have the file type. In this example a regular file (i.e. document, program, etc) starts with '-' while a directory (folder) starts with 'd' and a symbolic link starts with 'l'.

The next 9 characters show the permissions settings as read/write/execute for each of user/group/other. For example, the script file set_env is readable, writeable and executable by the owner, but not by anyone else, while the script file clear_thumbnails is also readable and executable by both group and other.

The next value is the number of hard links. For regular files this is normally 1, but for directories is it normally 2 or more. Hence it is possible to deduce that 'Public' has no sub-directories (2 hard links only), while 'Pictures' has 3 sub-directories (due to 5 hard links).

Then follows the owner and group. Most of these are my files, except for the archttpsrv.conf file which is owner & group as root.

Next is the file size in bytes. However, for directories and links this is not a very useful number.

Then there is the date/time of modification, which is often useful. This is given in local time, but the file system in LINUX stores this in UTC so if you checked before and after the daylight saving shift the values would change, but they are showing the **same time**!

Finally there is the file name. As shown here, when the file is a symbolic link, it shows both the link's name and what it links to. For example the 'VM' link is pointing to /home/users/VM.

NOTE: The permission and ownership of symbolic links are not very significant, as what you can do depends on the permissions/ownership of what they point to.

A lot of commands allow 'wildcards' which are ways of specifying multiple files, where '?' allows for any match to that character location and '*' allows for all matches[64]. Also note you can give combinations such as “sp*test” to select any file name beginning with “sp” and ending with “test” such as: speed_test, speed_last_test, special.document.test, etc.

The '~' at the start of a path/name is converted to your home directory (e.g. user paul would have ~ expanded to /home/paul). Remember that LINUX file systems are case-sensitive, so upper and lower cases are distinct!

It is also a good practice to edit your .bashrc file to include aliases so that **rm**, **cp** and **mv** are changed to include '-i' to prompt you if data could be lost (remove file, overwrite file). You can un-alias the command with the backslash if you really want to delete lots of stuff without prompting (e.g. “\rm sp*test”). Also note the -R recurs option allows you to apply some commands all the way through a directory tree from the starting point, which can be useful (e.g. to delete a directory and all files and sub-directories) but is also dangerous if used without due care.

Following is a short table of some common commands and what they do:

Command	**Function**
alias	Without any options, lists those commands currently being re-mapped to alternative commands. Otherwise you can use it to map something to something else, for example, using: alias rm='rm -i' so you are prompted for each delete (a sensible addition to the .bashrc file).
blkid	Lists the unique identifier (UUID) for each disk in the system. This UUID is a safer way of mounting a disk than the name '/dev/sda1' etc as it won't change if you add more disks (whereas 'sda' could become 'sdb' if a disk was added to a lower number in the disk controller's ports).
cd *newlocation*	Change directory to the new location
chgrp *groupname files*	Change the file(s) group (typically you have to be root to do this).

[64] Under DOS/Windows you sometimes had to use *.* for all files, while LINUX usually treats * as everything (though not always the hidden files starting with '.' it seems).

Command	Function
chmod *options files*	Change access mode of files (permissions) using the options. Typical values for options are of the form {users} {+/-} {settings} with: Users: u=owner, g=group, o=other, a=all of them Action: + add settings, - remove Settings: r=read access, w=write access, x=execute permission, X=execute for directories only. As an example, to make myfile private you could use this: chmod go-rwx myfile Where your group and all others have read, write and execute permissions removed. To remove execute permission on a whole directory, but keep directory access, and to make files read-only to others, you could do this: chmod -R a-x+X,go-w mydirectory
chown *user files*	Change the file(s) owner (typically you have to be root), can also change owner and group simultaneously with: chown *newowner*:*newgroup files*
cp *source destination*	Copy file source to destination. May be aliased to 'cp -i' which will prompt you before overwriting an existing file.
dd	To use the quote from Wikipedia: It is jokingly said to stand for "data destroyer" or "delete data", since, being used for low-level operations on hard disks, a small mistake, such as reversing the *if* (input file) and *of* (output file) parameters, may accidentally render the entire disk unusable. Operates like a file copy action but for specified block size, and possibly number of blocks, and can operate on storage devices as 'files'. Can be used for disk copy, data wipe, and similar tasks, so it requires great care!
df -h	Disk free, the '-h' reports this in human friendly format. Typically only take note of areas that start with '/' such as /dev/sda1 and so on. There is a much friendlier method for disks using: System → Administration → System Monitor (“File Systems” tab) And for more detailed usage: Applications → Accessories → Disk Usage Monitor

Command	Function
grep [options] 'pattern' [files]	A program to search for text in files. As with a lot of LINUX utilities, it can be complicated but simple examples are: grep 'frog' animals.txt to print out an lines of text in the file animals.txt or a more complicated (but useful) case: get_iplayer | grep -i 'frog' to list any of the BBC's iPlayer program titles with 'frog' included anywhere, and case-insensitive (the "-i" option) So it would find "Frogs in history" or "Tomato seeds look like frog spawn", etc.
kill *processid*	Stop execution of any process by its process ID number (as listed by **top** or **ps**). Often you have to be root and this is a dangerous thing to do!
ln -s *existing newlink*	Create symbolic link 'newlink' pointing to 'existing' file/directory.
ifconfig	Interface Configuration. Lists the network connections and can be used to bring them up/down and to change IP address, etc.
ls ls -ltr	The ls command lists the contents of a directory to show files and sub-directories. You can restrict it to show certain files (e.g. "ls a*" for files starting with 'a') and the options (e.g. "-ltr") allow you to adjust how it shows things. Some common options are: -a Show all files (includes hidden ones starting with '.') -l Long format, shows type, permissions, hard links, owner, group, size and filename -t Sort by time (default is by name) -r Reverse sort direction -h Human readable file sizes (e.g. big files shown in kB, MB, GB, etc, rather than bytes).
man *program*	Display the system's documentation for 'command' but not always as helpful as one might want. As an example, try 'man ls' and so on for help. Press the 'q' key to quit.
mkdir *directory*	Make a new directory.
mount umount	Commands to mount and unmount any of the file system(s) in use.

Command	Function
mv *current new*	Move file, effectively a rename but can move across the file system.
ntpq -p	Network time protocol query, print results. Shows you the status of clock adjustment to other time servers (the offset is the difference in clock times in milliseconds).
ping -c 4 *ipaddress*	Program to see if a computer address is alive. The "-c 4" tells it to try 4 times, and *ipaddress* could be either a numerical value or a web address (ping is an easy way to find the IP address from its name, in fact). Some computers disable ping responses though, so they won't respond even though they are reachable.
ps -A	List all processes (can be quite verbose).
pwd	Print working directory (i.e. where you currently are in the file system)
rm *filename*	Remove the file *filename*.
rmdir *directory*	Remove the existing directory which must already be empty. Or else you can use 'rm -R' on the directory to do both.
sudo *command*	Run the command with root privileges. ***Take great care!***

su
su *user*
	Switch User to become someone else (as far as computer access is concerned), default is root, or you can be another user if you know their password. To quote:
	`We trust you have received the usual lecture from the local System Administrator. It usually boils down to these three things:`
	`#1) Respect the privacy of others.`
	`#2) Think before you type.`
	`#3) With great power comes great responsibility.`

sync	Synchronise the storage devices to the current in-memory state of the file system. Typically takes a few seconds to have the disks updated.
tail -n 10 *file*	Display the last 10 lines (from '-n 10') of the file. Typically used to see the last system messages with: tail -n 10 /var/log/syslog

Command	Function
tar	Tape Archive, originally used to combine all files on disk for backup to tape, has become a common LINUX took for creating a single file of a directory tree that can be compressed and copied to CD/DVD/external disk, etc. Needs other options (not shown) and can be tricky to use. Quite a few backup utilities provide a GUI to use 'tar' for the actual backup. For the full complex details try here: http://www.gnu.org/software/tar/manual/tar.html
top	Table of Processes, shows you what is running and its use of memory, CPU time, etc, updating one per second or so. Normally sorted by maximum CPU use. Press 'q' to quit. Much easier to use is System → Administration → System Monitor (the “Processes” and “Resources” tabs)
which *program*	Finds out where *program* is located (as if you were to run it).
who	Lists those logged in to the system. You, and others, may appear several times, as each terminal is considered a separate log-in case, along with the normal GUI display you started with.

Appendix B – Transmission BT client configuration

Updating

The version supplied with Ubuntu 8.10 "Intrepid Ibex" has some minor bugs, and lacks some features added since its release, so it is best to arrange your system to get the latest stable release of Transmission. This page has the details:

http://forum.transmissionbt.com/viewtopic.php?f=13&t=5604

Briefly the steps are first to add the sources to the list file by editing it with this:

```
gksudo gedit /etc/apt/sources.list &
```

You will be promoted for your password, then add these two lines to the end of the file:

```
deb http://ppa.launchpad.net/transmissionbt/ubuntu intrepid main
deb-src http://ppa.launchpad.net/transmissionbt/ubuntu intrepid main
```

Save the file, then exit the editor. Then enter the following commands, remembering that copy (drag mouse with left-button down over text, then right-click and 'copy') and paste (right-click in terminal window with mouse) is safest!

```
gpg --keyserver keyserver.ubuntu.com --recv 976b5901365c5ca1
```

```
gpg --export --armor 976b5901365c5ca1 | sudo apt-key add -
```

Finally run the update of the package manager with:

```
sudo apt-get update
```

You should then have the "orange star" top-right showing updates are available for your PC, follow the usual to update your system to the latest Transmission client (version 1.72 or later).

Configuring

There are some setting features in the updated Transmission client that should be checked and set to suit. These are found with Edit → Preferences and the following are suggested

•Torrents: Change default destination to a dedicated directory, not the Desktop, for output. For example, create a /home/paul/Torrents directory for user 'paul', etc, and use that.

•Peers: Enable block list, change to "encryption required" for connections.

•Speed: enable speed limit mode 8:30 to 21:00 with down=20k & up = 15k for typical Virgin Media 10Mbps account[65].

•Network: Incoming post checked / enabled in router for forwarding.

Please note that while the block list and encryption provide some protection against snooping on your activities, ***you are not anonymous*** by any useful definition of that term!

[65] The Virgin policy is set out here http://allyours.virginmedia.com/html/internet/traffic.html While users do not like such measures, at least Virgin are open about the limits they apply!

Appendix C - Glossary of terms

Term	Meaning
ADSL	Asymmetric Digital Subscriber Line – a data-over-phone-wires system that has a higher speed for downloads (Internet to your PC) than uploads.
ASCII	American Standard Code for Information Interchange - a mapping between 'numbers' (which a computer works with) and their equivalent characters for printing and so on. For example, numbers 65-90 represent the upper case letters 'A'-'Z' and so on. Originally this was very US-centric (e.g. no '£' symbol or accented characters) but variations have been used to extent it to other languages.
bit	A binary 'digit'. This can represent 0 & 1, or false & true, etc, and is the way in which information is represented in a computer. Most computers work with groups of 8-bits (byte) or 16, 32, or 64, depending on the CPU type.
byte	A group of 8 bits which can represent different things, for example, the numbers 0-255 or an ASCII character, etc. This is the basic unit of data for files and most storage systems.
cache	In computer usage 'cache' is a fast local store of some information. For example, a web browser normally keeps visited web pages and images on your local disk (up to a limited time and/or disk space limit) so when you go back to a previous page, etc, it need not re-load everything from across the internet.
CD	Compact Disk (originally for digital audio, but later data as well).
codec	From 'encoder-decoder', a device or (more likely) software that encodes and decodes data, most commonly for playing audio and/or video data. They occasionally need to be up updated as new formats come out.
cookie	When you visit a web site, they often leave a small 'cookie' file on your system to allow them to keep your preferences and see when you return. However, this can also be used to track you web site history beyond the original site. I prefer to set the web browser to delete all cookies on exiting, so each day is a new dawn.

Term	Meaning
compression	The process of reducing the amount of storage needed for a data set. This comes in two distinct types: ● Lossless compression: Here the file size is reduced by removing redundancy in the data, and on decompression the resulting file is ***exactly the same*** as the original (examples include ZIP, gzip). ● Lossy compression: Here some data is 'thrown away' in order to make the file smaller. Typically used for audio and video data where human perception is insensitive to some types of quality reduction (examples are JPEG, MP3). Morse code was an early example of lossless compression in practice, where common letters were assigned the shorter dot/dash codes to reduce the average time to send a typical English message.
CPU	Central Processing Unit, the 'thinking' bit of a computer.
daemon	A program that runs in the background. From Greek mythology, some of whom handled tasks that the gods could not be bothered with (not the satanic association of Christian theology). The Windows equivalent is a 'service'.
default	The settings that will be used unless you explicitly change them (i.e. what happens if you do nothing special).
DHCP	Dynamic Host Configuration Protocol - a system where your PC asks for network settings such as its IP address and DNS servers automatically (typically from your modem/router, and normally it gets its settings ultimately from your ISP).
DNS	Domain Name Server – the Internet's “phone book” that maps a human readable address such as www.sat.dundee.ac.uk in to a computer used IP number such as 134.36.22.54 If hijacked, or its content store is 'poisoned', you can be silently re-directed to a fraudulent phishing site!
DRM	Digital Rights Management, see TURDS.
DVD	Digital Versatile Disk, designed as a data disk originally for high quality video replay.
encryption	The process of 'hiding meaning' so only those with the correct key can reveal the real contents.
firewall	A device (software on the PC, or hardware on the network switches, etc) that serves to limit the access of one system to another for security. From the way a firewall limits the damaging spread of fire in a car, building, etc, should it start.

Term	Meaning
FTP	File Transfer Protocol – a fast way of performing file transfer (but generally not other stuff, so not used for web browsing).
HDD	Hard disk Drive, a common example of a non-volatile mass storage device.
GUI	Graphical User Interface – the visible 'window' bit of a program that make it easier for a human to use.
hash	(1) The '#' character. (2) A cryptographic function that takes an arbitrary block of data (e.g. a file) and returns a fixed-size result (e.g. string of numbers, or letters, etc), the *hash value*, such that an accidental or intentional change to the data will almost certainly change the hash value. Often used to provide a simple way of comparing or indexing data, as a match of hash values is a near perfect guarantee that the blocks of data are identical. The considerable difficulty of crafting different blocks of data with the same hash value makes it useful as a digital signature, etc.
HTTP	Hypertext Transfer Protocol – the method generally used for web site access. The secure version (https) has encryption so your traffic is not easily spied upon.
IP	Internet Protocol – related to the internet addressing and data transfer systems, but usually it is simply meaning the computers numerical address (often written as a set of 4 numbers like 192.168.1.100).
iso	In computer usage this normally means an 'image' of a CD or DVD disk, from the International Standard Organisation that defined the formats. In storage terms an image is a byte-by-byte copy of the storage device's sectors, rather than a file-based copy, so it is independent of any type of file system used on the storage device.
ISP	Internet Service Provider – the company the provides the link from your home to the global internet system (e.g. by cable, ADSL, dial-up modem).
java javascript	Java is a computer language intended to make software that can run anywhere safely. It, and the scripted version, has become a common tool to make web pages more interactive, but of course the aim of high security has not quite been met.
JPEG	Joint Photographic Experts Group - but usually an image compression method defined by that standards body.

Term	Meaning
malware	A play on the terms hardware & software, used to indicate software that is malicious in its intentions. Sometimes this is hidden in a image/video file that exploits security flaws in the player, in other cases they use names such as somemovie.avi.exe where it looks (in Windows) as if it is an audio-video file, but the .exe part of the name makes Windows execute it rather than play it as a video, thus allowing it to comprise the system.
modem	From modulator-demodulator, a device that converts data in to something suitable for transmission over a different network (e.g. phone lines).
MP3	Music compression algorithm, also known as “MPEG-1 Audio Layer 3”. Now synonymous with computerised music players.
NAT	Network Address Translation - typically a system where you have one IP address facing the Internet (as dictated by your ISP) and possibly multiple PCs behind that using different private addresses (usually of the form 192.168.1.xxx). This system helped prevent the world running out of IP addresses with the older IP v4 system (v6 is not in wide use yet, and offers more addresses than we can possibly ever use), and also makes it slightly harder to randomly attack a PC behind the router as normally only data coming back in response to an outward request will be forwarded to the PC(s).
NTP	Network Time Protocol - a system of time message exchanges that allows computers to reliably synchronise to each other and, ultimately, to the stratum-0 clocks at the top of the tree (which are typically atomic clocks or radio time receivers like GPS).
OS	Operating System, the most basic layer of software on a multi-task computer.
P2P	Per to peer, usually implies file-sharing, but strictly speaking just any system of computers without a central control.
PC	Personal Computer. Any home or office computer, but more generally it implies one using the Intel / AMD x86 type of CPU.
PDF	Portable Document Format. A standard for documents created by Adobe for easy distribution over different platforms (LINUX, Apple Mac, Windows, etc) that aread on screen as it is printed.
phishing	“In the field of computer security, phishing is the criminally fraudulent process of attempting to acquire sensitive information such as usernames, passwords and credit card details by masquerading as a trustworthy entity in an electronic communication.”

Term	Meaning
PostScript	A language for talking to laser printers, usually the one type that “just works”.
proxy	A 'proxy server' is used as an intermediary for requests from clients seeking access to other servers. Typically used for anonymity, for security or censorship reasons (e.g. blocking bad web sites, filtering for viruses, etc), or to bypass restrictions (e.g. when an IP address is used for geographic filtering of services).
RAID	Redundant Array of Inexpensive Disks - a system of using multiple disks to provide various combinations of: more space, higher access speed, redundancy against disk failure (but not necessarily all at once).
RAM	Random Access Memory, fast electronic memory (normally volatile).
script	A script is a list of instructions, rather like a play, where the 'actors' are other programs on the computer. Typically scripts are used to automate basic tasks to save you entering a long list of commands.
server	A computer that provides a service to other computers or users (rather than supporting direct user-interaction). Examples include: ● File server: lots of disk space (typically RAID) that provides a centrally accessible area for users to store their data (e.g. for sharing to ease the problem of backing-up). ● Mail server: stores and forwards email messages. ● Time server: provides time (usually via NTP), normally that is ultimately derived from world-wide atomic clock(s). ● Print server: computer with a printer attached that manages the correct ordering of print jobs, usually from multiple users. Some printers now have this capability built in.
SI	From the French “Le Système International d'Unités”, the international metric system of scientific units.
spam	Unsolicited email, typically the bulk of what you will get in the real world.
switch router	Both a switch (also known as 'network hub') and router provide network connections between several computers, but at home a router usually also provides NAT from an external network (e.g. the internet).
TIFF	Tagged Image File Format, a complex format allowing images with various extra stuff included.
TURDS	Technology Users' Rights Denial Systems – a more honest acronym for DRM systems that limit what you can do with copyright works.

Term	Meaning
UPS	Uninterrupted Power Supply – a device that conditions the AC power to your equipment and has a battery (or other energy storage system) so that in the event of power failure it can continue to power your equipment for a short time. In the case of a small computer supply, it will normally issue a command to initiate an orderly system shut-down once there is less than a certain time (e.g. 5 minutes) of battery power remaining. In larger systems it is intended to hold up for enough time to allow a back-up generator to start up.
URL	Uniform Resource Identifier – really a fancy name for a web address.
VCR	Video Cassette Recorder. Tape in a cassette was more or less rendered obsolete by DVD or HDD based recorders, but the acronym remains in use for video recording in general.
VPN	Virtual Private Network. An encrypted (usually) service from one computer to another that 'tunnels' all of your network activity (file transfers, web browsing, DNS lookup, etc) is such a why that you appear to be effectively at another server's network location.
XSS	Cross-site scripting. A security attack based on tricking you in to running a script from a site other than the one you believe you are visiting.

Appendix D - Memory & storage units

Term	Computer Memory Use	SI or Disk Usage
kB	Kilobyte= 1,024 bytes (2^{10}). Sometimes you see kiB, MiB, etc, used to show they are binary (base-2) versions, but not always, so it can be confusing. For disk storage the decimal SI notation (next column) makes things look bigger, by almost 10% for TB-sized disks, so it is now used!	1,000 (10^3) Note for network speeds it is usually in bits, not bytes, with 1 byte = 8 bits. Hence a 2Mbit/sec link could, in theory at least, achieve 250kB/sec (but more likely 30-80% of this in practice).
MB	Megabyte = 1024kB = 1,048,576 (2^{20})	1,000,000 = 10^6 (million)
GB	Gigabyte = 1024MB = 1,073,741,824 (2^{30})	1,000,000,000 = 10^9 (US billion)
TB	Terabyte = 1024GB = 1,099,511,627,776	1,000,000,000,000 = 10^{12} (US trillion)

Appendix E – Automatic Back-up

I have a script that allows a reasonably automated process that appears to work well, but is not so nice to configure/understand. When this script is installed, ***and you have the suitable backup HDD plugged in*** (see the previous example, page 49), it will automatically back up changes to your home area each time the PC is shut down.

More information is posted here: http://www.sat.dundee.ac.uk/~psc/Ubuntu/ubuntu-stuff.html

That page has this guide, and the backup scripts and matching up-to-date documentation.

Notes