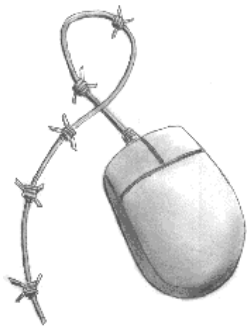
GreenNet CSIR Toolkit Briefing no. 1

An Introduction to the Internet



How it works and what it can do for you

Written by Paul Mobbs for the
GreenNet *Civil Society Internet Rights Project*, 2002.
http://www.internetrights.org.uk/

What is the Internet?

What we know today as the Internet was first devised, in its basic principles, by Paul Baran of the Rand Corporation1 for the US military in 1965. In its early years it was used a way of sending text messages by US academics and government agencies.

A significant leap forward came with the advent of *hypertext* which enabled text, graphics and other media to be linked into single pages and to other related pages. Hypertext was conceived at the CERN research laboratory in Switzerland; it gave us what most people understand as *the Web*.

In the 1980s, when information technology began to transform the corporate environment, the World Wide Web (WWW) began to expand as a global network. Even so, access to the Internet did not become widely available until the mid-1990s, when it was recognised for the versatile communications medium it is, computers became significantly faster and comparatively cheaper, and general use of the 'Net took off.

Increasing media convergence means that that the Internet is a powerful vehicle for communications; not only for passing messages but for all kinds of information and audio-visual material.

Recent technical advances in web page design and graphical capabilities have come at a cost, however. A *digital divide* is increasingly opening up. Large amounts of data now have to be processed when we access the Web; so for full compatibility you need the latest versions of high-speed computers and operating systems. This built-in obsolescence can exclude you if you are on a lower income; you will usually use older, second-hand equipment and out-dated software to run your Internet connection. Groups and organisations wishing to reach the wider public must, therefore, address issues of compatibility and technical standards in the design of their web sites.

How the system works

The Internet is a network of computers and connections that pass packets of data between them, using a variety of *protocols*, or standardised methods of sending information, to make a variety of services available via the Internet. Home users usually connect to the Internet via a phone line to an *Internet Service Provider* (ISP). Your ISP sets up a facility for your computer to connect to their *server*, to enable Internet connections, email and other services.

[1]If you're curious, Paul Baran's original paper's on the Internet prepared for the US military are available at http://www.rand.org/publications/RM/baran.list.html

Many larger offices and organisations act as their own service provider by setting up access to the Internet as part of their network operation.

The server-client relationship

Transactions on the Internet centre around *servers* (computers that transact data over the system) and *clients* (the computers used by people accessing the Internet or other services).

The *server* transacts data over the Internet as part of email, WWW or *File Transfer Protocol* (FTP) operations (see sections below) and interacts with the *client*. The client is not controlled by the server; the server simply organises the transmission of data to and from the client in order to enable communication. Programs or Internet utilities on the *client* machine organise and display information received from the server.

The connection between the client and server is maintained using a standardised communications system called the *Internet Protocol* (IP or *TCP/IP*). The Internet Protocol uses a numeric address, made up of at least four sets of digits, for everything connected to the Internet.

Packets of data are sent under the Internet Protocol; each packet contains its own numeric address, as well as a copy of the address of its source.

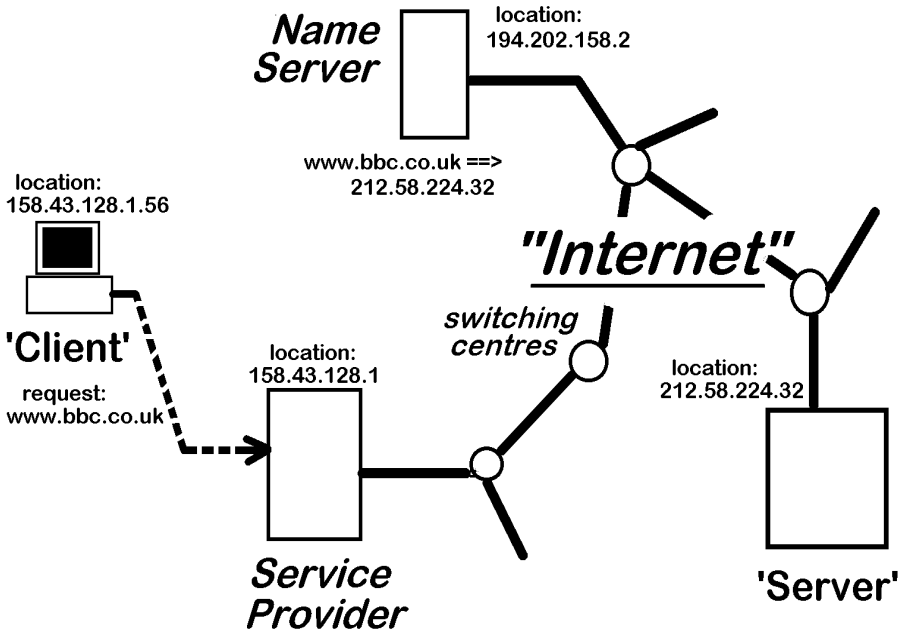
Millions of data packets are shuffled and sorted across the Internet through a large number of *switching centres*. A switching centre usually comprises the computers used by telecommunications companies to run high-capacity data links across the country. The switching centre reads the destination address of the data packet and then routes the packet along a line towards that destination.

The system in practice

To understand how the system works let's consider a simple transaction; we will use the example of you using your home computer to request information from the BBC's web site. This is illustrated in the diagram below.

Whenever you connect to the Internet you are assigned a numeric address that defines your *location*; in our example that address is 158.43.128.1.56 (see diagram).

The first part of the transaction involves you submitting a request for a page, via your Service Provider. To do this you will enter a formal *Universal Resource Locator* address (URL) such as "www.bbc.co.uk".



This URL must be translated into a numeric address. To do this, your computer (i.e. the *client*) first sends the URL address to a *name server*. The name server has its own pre-assigned numeric address (in our example, 194.202.158.2) to which your computer sends your request. The name server sends back to the client a numeric address that corresponds to the location of the server "www.bbc.co.uk", in this case 212.58.224.32.

The system of *domain names* is maintained through databases kept at the name servers, operated by various companies in each country, classified as national *Internet Network Information Centres* (InterNICs).

Your computer then sends the request for data, or an email, or a web page to the numeric address supplied to it by the name server. The BBC server (212.58.224.32) receives this request, processes it, and returns the packets of data to your computer system.

The path that the packets take is not always the same. Packets can be sent via different routes, taking different amounts of time, and hence can arrive in a different order from that in which they were transmitted.

Your computer assembles the packets into the correct order as they arrive, and then passes the data to the program you are using.

Servers process requests on a first-come-first-served basis. So at popular sites there can be a delay while the request is held in a queue for processing. In recent years larger sites have begun to operate network caches - like the cache used by a microprocessor in a computer - which automatically provide copies of very popular files. The time it takes for the system to deliver what you have asked for also depends upon the type of material you are requesting, and where you are requesting it from.

Potential for service disruption

Throughout this system there are vulnerable points that enable the disruption of normal operations. Name servers, for example, and in particular servers that process requests from clients, are susceptible to attack by hackers/crackers. Internet Service Providers can also be the targets of disruptive attacks if they provide services to a person or group against whom someone else has a grudge against.[2]

The Internet's communications media

The explosive growth of the Internet in recent years has been based on the *client-server* model outlined above. The latest generation of high-speed modems and Internet connections reinforce this model; they send data down the line much faster from the server to the client than they do from the client to the server. This is called an *asymmetric* link (the *A* of the ADSL broadband system).

The client-server model is being challenged by the growth of *peer-to-peer* (P2P) networking.[3] A good example of a P2P network was the Napster music-sharing network. P2P enables everyone to host their own content as part of an online *public collective*. Asynchronous links limit the abilities of people to use the Internet to work with each other in this way, and are unpopular in some quarters as a result.
We will now look at some of the key media and features of the Internet.

[2]For more information see GreenNet CSIR Toolkit Briefing no.14, *Keeping Your System Secure*.
[3]*Peer-to-Peer: Harnessing the Power of Disruptive Technologies*, Andy Oram (Ed.), O'Reilly Books 2001, ISBN 059600110X

The World Wide Web (WWW)

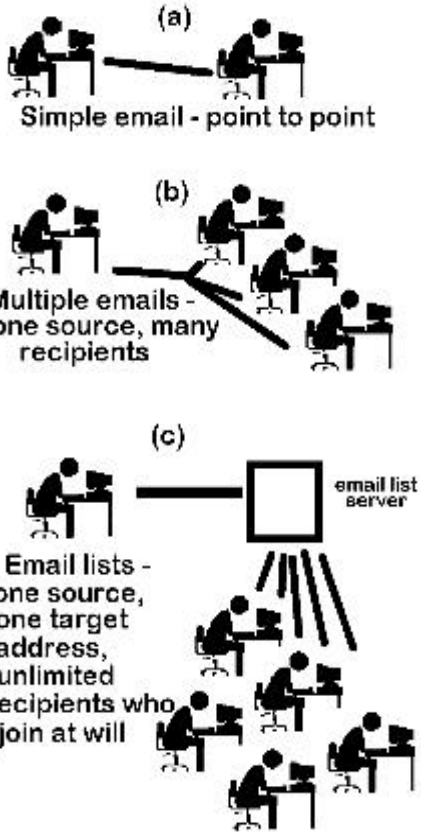
There have been some significant technical developments on the WWW recently:

- Multimedia - As well as controlling text and graphics, you can now use forms, complex animation and even video or sound clips (although these are still fairly ropy because of the restrictions of most people's Internet connection).
- Dynamic scripting - Many WWW pages now incorporate a simple form of computer programming called scripting (see section on scripting languages below). This is usually used for controlling links or the animation of graphics. Scripting can also enable web pages to be programmed to perform dynamic complex functions.
- Plug-ins - Plug-in systems are special proprietary graphics systems that augment the functions of an ordinary web browser, so that it can display complex graphical information according to a pre-defined sequence.

The adoption of *dynamic content* (i.e. content that is not fixed and can change in certain circumstances) has significantly increased the scope of web pages. Changes can be made to the page content depending upon the time of day or date it is viewed, for example. Through request forms and scripting systems to organise the searching and display of information, people reading a page can also search web sites as if they were a database.

E-mail

Messaging or electronic mail was one of the first things the Internet was used for. If you have an email address, anyone else with an Internet connection can contact you, and you can contact them. Email has become a very complex medium over the years and is now a powerful means for group communication, among other things.



(a) Simple point-to-point emails can carry messages. But you can also attach files to them, to send pictures, text documents, complex graphical presentation, audio or video. If it can be made into a computer file, and the file is not so big as to make transmission impractical, you can send it by email.

(b) Multiple emails are where one person can send the same message (with attachments if you wish) to many people (although some email programs start to complain when you list more then seventy or eighty addresses). If a recipient replies to your original message, they can, if they wish, reply to everyone on your original list. This is a simple way of enabling a dialogue and developing a virtual network.

(c) *E-mail lists* are where one person sends an email to an address located in a *list server*. The list server automatically forwards the email to everyone on the list. But unlike multiple emails, you only send the email to the list server address; recipients join (or leave) the lists at their own request in order receive (or not) information. This is a very way of using a virtual network, if only because you don't have to manually manage the list as you do in (b) above. There are a number of free email list services available on the Internet (although you do have to pay the price of having a small advert attached to the emails).

Search engines

Little of the information that is stored on web sites is actually accessed directly. People usually find things through using *search engines* - Internet servers that keep huge classified directories of the contents of millions of web pages.

Success in tracking down information on the Internet depends on how well websites have been indexed and linked to search engines.

All web pages have a *web page title* hidden at their beginning (displayed on the top window of your browser). Titles were used to index web pages when the web was first launched. Today's search engines are much more sophisticated, and examine more of what's actually inside the page - some search engines even specialise in certain file types, such as graphics, video or audio.

There are two main ways to ensure that a website will get a search engine listing:

- Metadata - this is a form of classified keywords and information that identify the page. Metadata can include date of publication, author and status (draft, final, etc.). Metadata is inserted into the head of the web page when it is first written, and a well-drafted metadata description can bring a lot of search requests.
- Site registration - you can complete a form on the web site of most search engines, by which you classify your web site according to certain criteria. Those details will then be added to the search engine's database within a few weeks.

File repositories

Computers can hold large quantities of data. File repositories work by connecting to information servers via the Internet, enabling you access data at will. Since the advent of the WWW, however, storing text and graphics separately has become a thing of the past, so file repositories are less popular than they were.

File storage systems are still regularly used, though, for storing very large files or compressed archives of data, because they are simpler to maintain than a web site.

There are three main types of file storage system:

- File Transfer Protocol (FTP site): FTP was an early means of file storage and retrieval for the Internet. Files are retrieved using a special FTP program (web browsers will now do FTP transfers, too, although, some argue, less reliably). The main advantage of FTP systems is that they enable storage and retrieval of truly huge files - hundreds or thousands of kilobytes in size. These files could be programs, databases, images, video clips or large published documents. Because the FTP system is relatively simple, it is more reliable for large file transfers, and is therefore still in widespread use.
- Gophers: Gopher systems were developed as a refinement of FTP sites. Instead of a file directory, they used text-based menus to guide you through the available files. Although still in use, the WWW have made gophers largely redundant.
- Majordomo systems: Majordomo systems (named after the Spanish/Italian term for a butler) are an email-based form of file retrieval. You email the majordomo for a list of resources, which it emails back. You then use that list to request certain files, and the system loads this information into an email and sends it back. Majordomo systems were popular when most people had email rather than full Internet access, so they, too, have generally been overtaken by the WWW as a tool for information distribution. But majordomos still have one positive advantage over the WWW: security. If necessary, only those with authorisation can retrieve information from a majordomo

system.

File repositories are complex to set up, so they are usually only used by large organisations with the resources to run their own Internet server. But if you need to have access to a large base of information, these systems (especially FTPs) provide a low cost, low maintenance option.

Audio/visual media - and proprietary standards

The built-in obsolescence of IT equipment and programs, promoted by the major corporations, means that the question of proprietary standards has become a major issue. IT companies try hard to encourage us to buy the latest versions of their products. The increasing amounts of data that have to be processed by computers encourages the need for ever-faster processors. Those who do not have the resources to keep up with the latest versions of hardware and software are thus in danger of being excluded from the Internet.

The issue of proprietary standards has become a pressing concern in the area of audio and visual media, and *portable documents*.

Audio signals take up a great deal of space in a file. Video signals take ten times more, even when only using a small projection area. To get around this problem software companies have developed *compression* systems to squash this information into a much smaller space. They give away free copies of the programs you need to read the compressed files; most of these programs are *plug-ins* which integrate seamlessly into your web browser.

The software companies make their money by charging for the programs that take the raw audio or video file and compresses it into their proprietary format for releasing on the Internet. This has two important consequences:

- Audio or video produced for the Web invariably uses the latest compression software - so if you want to access that information you have to use the latest version of the plug-in reading program. This, as we saw above, usually requires that you have the latest versions of equipment, which can exclude a large number of people.
- The programs that create the proprietary files are expensive to purchase - too expensive for most individuals to use them to develop their own online media. There are ways around this, but they are very limited. It is difficult to produce free or low-cost systems for creating proprietary file formats, because of the legal restrictions of copyright laws. Industry moves to patent these proprietary formats would effectively outlaw any low-cost alternatives.

The dominant audio and video standard of recent years has been Real Network's *RealPlayer*, which displays video or audio over the Internet. It is a more complex system than the programs discussed above, and requires not only an encoding program to create the files, but also a special *streaming server* running on the Internet server, to enable you to receive files live.

Audio and video files can be encoded and placed on the 'Net without a broadcast server, but you would have to downloaded them as a file and played separately rather than being displayed automatically.

The other main standards for video (with or without sound) are MPEG and AVI, but their drawback is that they tend to use much larger files (about 10 times bigger) than the *RealPlayer* format.
The significance of audio and visual media on the Internet has not been immediately obvious. This is because most machines connect to the Internet at comparatively slow speeds. But as connection speeds begin to increase and new telecommunications technologies are developed, over the next ten years, audio and video will become vitally important.

Faster connection and processing speeds will enable groups to set up their own virtual radio or TV stations without the restrictive controls imposed on the traditional broadcast media. Indeed, this is already happening; there is a whole range of small stations, including some from the UK, routed through the *Live 365* web site[4]. As *web casting* becomes the norm for society's broadcast media, issues over who defines and controls audio and video standards on the Internet, and hence who controls the form of society's media communications, will become more contentious.

Portable documents

Portable documents formats (PDFs) are another important aspect of proprietary file systems. With the variety of computers, printers and operating systems that proliferate, it can be very difficult to produce a document on one machine in identical form to that produces on another. There is no guarantee that both copies would come out looking exactly the same. Portable document formats - of which the major two are *Postscript* and *Acrobat* - encode documents so that they appear identically irrespective of the computer or printer they are produced on. This makes them very useful for producing leaflets, reports and other documents that can be widely distributed via the Internet.

As with audio and video proprietary file systems, the reader programs are given away free, while the programs creating portable documents are sold commercially.

Some word processors include a facility for generating portable documents, but it is not a full featured version that can be bought from the developer of the system.

It is likely that over the coming years an *open source* (that is, free for distribution and use) system of portable document formatting will be developed. In the mean time the use of portable documents by campaign groups is limited to those who can afford the formatting software.

Complex scripting and data manipulation

The WWW has enabled the development of dynamic pages through the use of *scripting languages*. Scripting languages are very powerful and versatile tools. For example:

- Scripting on a web server can allow the manipulation of databases to provide people with detailed information that they can extract dynamically, according to particular preferences or a geographical location, etc.
- Scripting can provide an interface between different systems, quite often between email and the WWW. You will often see a link on web pages saying "email this page to a friend" - that's scripting.
- In its more advanced forms scripting can be developed into complex data processing systems, for example creating protest letters, or knitting together different protest sites on the web as part of a global action.

The important issue is *where* the scripting is done. Originally all scripting was controlled from the Internet server. This required technicians to set it up, and money to operate and maintain it.
Since the mid-1990s, thanks to the WWW, scripting has begun to jump from the server to the client computer. This means that anyone, independent of the server, can include scripting for certain functions in their web pages.

Scripting provides endless possibilities. The only limits of scripting are essentially the capacity of the computer and software you are using, and your ability to use scripting creatively to solve problems and

[4] http://www.live365.com/cgi-bin/directory.cgi

create opportunities.

Scripting languages

Scripting is a set of programming systems. Rather than being used by the computer's processor directly, they are interpreted by software, based on a scripting language, which instructs the computer what to do. There is a variety of scripting languages, each of them with different features that make them popular. Three of the main ones are:

- The Common Gateway Interface (CGI) - CGI is exclusively a server-side system. It is not strictly a scripting language, but a protocol for addressing different services within Internet servers. CGI is useful for creating documents dynamically according to certain user-elected criteria. CGI can also interface with other software applications to extract information and create collections of information. Search engines use CGI and are a good example of its versatility. Most of the forms that you come across on the Internet are run with CGI.
- Perl (Practical Extraction and Report Language) scripts - CGI is often run from Perl scripts. Perl is a scripting language devised to augment the functionality of servers. Perl can be used to interface the various services of the server with particular Internet services. Using Perl you can:
  - tie web sites and databases together, making complex information simply available without having to write specific web content for the purpose. The Perl script extracts the necessary information from the database, and then composes a web page dynamically, to be sent out via the 'Net.
  - create web site forms that interact with files and email.
  - create email lists and majordomo-like email enquiry systems that can distribute or supply information on demand.

  Using Perl requires some knowledge of programming and Internet servers. Perl is usually a server side system, although Perl scripting systems can be now loaded onto home computers. There are many scripts for Perl available online. An Internet Service Provider may also be able to develop a simple scripting system (although they will make a charge for it) to produce things like online petitions or email pages.
- JavaScript - JavaScript was developed as a subset of the Java programming language for use in web browsers. Originally it was used for simple operations like animating graphics or loading up different web pages according to different user instructions. Today JavaScript is a complex language which can undertake a diverse range of tasks. With JavaScript you can:
  - set up 'virtual web sites' that use JavaScript to tie together different resources on the Web as if they were a single site. JavaScript also allows you to have complex operations taking place within a web page that displays or reorganises the page according to different circumstances.
  - develop action tools for specialised tasks. JavaScript is being used increasingly as a means for developing Internet-based direct action, such as virtual sit-ins and virtual lobbies.
  - develop systems to work with forms and web pages, client-side rather than server-side, making complex web pages easier and cheaper to create and use.

JavaScript can work server-side, but it is usually used as a client-side scripting language. This has important implications. Complex web pages can be developed without the need for complex server scripting or high-powered web servers to process the high volume of scripting requests received by a popular site. Being client-side, you can experiment at home designing your own tools rather than having to use a server.

Most significantly, JavaScript enables *distributed action*. If a server enables the action, then closing or blocking the server closes down the action; because JavaScript is client-side, however, scripts or web pages can be distributed via a server, or via email, and even by post on floppy disk, and the action can still

take place. JavaScript therefore enables complex web-based actions without the risk of being cut off from your server.

The GreenNet Internet Rights Project

GreenNet[5] is the UK member of the Association for Progressive Communications[6] (APC), and is leading the European section of the APC's Civil Society Internet Rights Project[7]. The primary goal of this project is to provide the resources and tools necessary to defend and expand space and opportunities for social campaigning work on the Internet against the emerging threats to civil society's use of the 'Net. This involves developing ways and means of defending threatened material and campaigning, as well as lobbying to ensure a favourable legal situation for free expression on issues of public interest.

Until recently, the social norms of Internet communities, together with a very open architecture based on supporting these norms, regulated the Internet, and was responsible for its openness. The main forces of regulation now, however, are the business sector and government legislation. Corporations and governments are pressing for fundamental changes in legislation and in the architecture of the Internet. Unless challenged, these moves could radically change the nature of the 'Net, making it a place of oppressive controls instead of freedom and openness. It is in this context that APC's Internet Rights project is being developed.

This briefing is one in a series[8] that document different aspects of work and communication across the Internet. Although written from the perspective of the UK, much of its content is applicable to other parts of Europe. There is continuing work on these issues, as part of the European project. If you wish to know more about these briefings, or the European section of the APC Civil Society Internet Rights Project, you should contact GreenNet. You should also check the APC's web site to see if there is already a national APC member in your country who may be able to provide local help, or with whom you may be able to work to develop Internet rights resources for your own country.

Free Documentation License:

For more information about the Civil Society Internet Rights Project, or if you have questions about the briefings, contact ir@gn.apc.org.

[5] GreenNet - http://www.gn.apc.org/
[6] APC - http://www.apc.org/
[7] CSIR Project - http://rights.apc.org/
[8] http://www.internetrights.org.uk/