Contents

Contributing authors viii

Commenting authors x

Preface xiii

Acknowledgments xiv

Chapter 1: Introduction and Overview 1
Duncan Langford

Chapter 2: Internet Technology 13
David Beckett

Chapter 3: What is New or Unique about Internet Activities? 47
John Weckert

Chapter 4: Privacy and Security 65
Herman T. Tavani

Chapter 5: Law and the Internet 96
John Mawhood and Daniel Tysver

Chapter 6: The Internet and Varieties of Moral Wrongdoing 127
Jeroen van den Hoven

Chapter 7: Information Integrity 158
Richard A. Spinello

Chapter 8: Democratic Values and the Internet 181
Deborah G. Johnson

Chapter 9: Computer Professionals and YOUR Responsibilities 200
Donald Gotterbarn

Chapter 10: Conclusions 220
Duncan Langford

Appendix 1: Glossary of Abbreviations and Acronyms 241

Appendix 2: Internet Use and the Data Protection Act 1998 243
C. N. M. Pounder

Bibliography 267

Index 277

Contributing authors

DAVID BECKETT has worked for the Advanced Development Unit at the Computing Laboratory, University of Kent at Canterbury, UK since 1990. He researches into Internet technology issues such as searching the Web, and description using metadata. David currently works on the UK National Mirror Service project, a large Internet archive of free software and materials.

DON GOTTERBARN is the Director of the Software Engineering Ethics Research Institute at East Tennessee State University. He teaches several graduate courses in the software engineering program. He also teaches computer ethics and has developed ethics materials for the Software Engineering Institute. His research has appeared in more than a dozen professional journals and he has written several encyclopedia articles. He has done funded research on performance prediction for a distributed Ada closure, object-oriented testing, and software engineering education. He also worked as computer consultant developing systems for industry, the Saudi and US navies.

DEBORAH G. JOHNSON is Professor and Director of the Program in Philosophy, Science, and Technology in the School of Public Policy of Georgia Institute of Technology. She is the author or editor of four books, including *Computer Ethics* (1994) and *Computers, Ethics, and Social Values* (co-edited with Helen Nissenbaum, 1995). Professor Johnson co-edits the new, international journal, *Ethics and Information Technology* published by Kluwer, and she currently serves as President-elect of the Society for Philosophy and Technology.

JOHN MAWHOOD is a partner in the London law firm Tarlo Lyons (www.tarlo-lyons.com). For more than 10 years he has advised clients with businesses in, or who are customers of, the IT and telecommunications sector. He has provided advice globally and nationally on, for example, internet and electronic banking and commerce, Year 2000 exposure, systems purchasing and ERP, software patenting and licensing, and EC competition law. John was the architect of the UK government's Action 2000 Pledge 2000 initiative.

DR CHRIS POUNDER has worked in the field of data protection since the early 1980s. In 1984 he was appointed by the then Greater London Council to lead a team which applied the 1984 Act to its processing; there he developed a number of data protection training courses and began providing consultancies and helpline advice on privacy and security matters. In 1990 he launched *Data*

Protection News, and from 1988 until 1999 provided these services from within a large IT company. He now works in the London offices of Masons, a leading international law firm.

DR RICHARD A. SPINELLO is an Associate Research Professor in the Carroll School of Management at Boston College, USA. He has written two books on computer ethics, *Ethical Aspects of Information Technology* and *Case Studies in Information and Computer Ethics*, along with a book on knowledge management entitled *Corporate Instinct*. He has also written numerous articles and scholarly papers on business ethics and the ethical issues associated with information technology.

HERMAN TAVANI is Chair of the Philosophy Department and Director of the Liberal Studies Program at Rivier College, USA. The author of several publications in computer ethics, including a comprehensive bibliographic work in the field, he is currently Associate Editor of *Computers and Society* and Book Review Editor of *Ethics and Information Technology*. He has also worked in the computer industry as software technical writer and software publications supervisor.

DAN TYSVER is a partner with the Beck and Tysver, P.L.L.P., an intellectual property law firm in Hopkins, Minnesota. A graduate of Carleton College and Harvard Law School, Mr Tysver serves as an adjunct professor of law at the University of Minnesota Law School. He is a frequent lecturer and author on legal issues relating to the Internet and to computer technologies. Mr Tysver is also the author of the BitLaw Web site (www.bitlaw.com), an Internet resource on technology law.

JEROEN VAN DEN HOVEN is Professor of Philosophy and Information Technology at the Department of Philosophy of Erasmus University, Rotterdam, The Netherlands. He is founding chair of the international conference on moral philosophy and information technology Computer Ethics Philosophical Enquiry (CEPE) and editor of *Ethics and Information Technology* (Kluwer Academic Publishers).

JOHN WECKERT has qualifications and wide teaching experience in both information technology and philosophy. He is a senior lecturer in Information Technology at Charles Sturt University, co-Director of the Australian Institute of Computer Ethics, and Chair of the Ethics Task Force of the Australian Computer Society. His recent research and publications are in computer ethics and in knowledge-based systems. He is co-author of *Computer and Information Ethics* published by Greenwood Press in 1997.

Commenting authors

Australia

Melissa de Zwart BA(Hons), LLB (Hons), LLM (Melb) is a lecturer in the Law Faculty, Monash University. Melissa teaches and researches in the area of Internet law, including an online course on the Law of the Internet. She has published numerous articles on Internet-related issues, including copyright, trademarks, privacy, pornography and electronic commerce. She has recently acted as an adviser to the Victorian Parliament on the Y2K issue. Prior to joining the academic staff at Monash, Melissa was the Legal Manager at the Commonwealth Scientific and Industry Research Organisation, where she practised in intellectual property, licensing and technology law.

Botswana

Dr Sunday O. Ojo holds degrees from the University of Ibadan and Glasgow University. His areas of speciality include database systems, object-oriented systems, information systems, social informatics, computer science and information systems education, internet use in business and education, and IT in socio-economic developments of developing countries. He has published extensively in these areas. He has published a book on computer science and also co-edited proceedings of a conference on IT in developing countries. Dr Ojo is currently Head of the Department of Computer Science, University of Botswana.

Brazil

Alberto Levy Macedo was born in Rio de Janeiro, Brazil, and is a professional computing engineer with interests that include the internet, computer graphics, virtual reality and human interactions and relations. An academic and consultant, he has being working with the Internet since 1994, developing university courses and producing virtual worlds for the Web. Alberto is currently working at the Brazilian national TV Network, Globo, in the R&D Laboratory.

Germany

Professor Dr Herbert Kubicek is Professor for Applied Computer Science and Head of the Information Management and Telecommunications Research Group at the University of Bremen, Germany. He has been editor/co-editor of *Jahrbuch Telekommunikation und Gesellschaft* (*Telecommunications and Society Year - book*) since 1995, and was a member of the Enquete Commission of Deutsche Bundestag 'The Future of Media in Economy and Society, Germany's Way into the Information Society'in 1997–1998. He is also a member of the Management Committee COSTA1, 'Government and Democracy in the Information Age'.

Hong Kong

Robert Davison is Assistant Professor of Information Systems at the City University of Hong Kong. His research interests range from IT/IS in developing countries, through the application of group support systems in organisational contexts, to ethical and professional issues in IS. His work has been published in CACM, DSS and Information and Management, as well as at HICSS. He is also active in Web-enabling course materials at City University. Further details of his current work can be found at: http://www.cityu.edu.hk/is/staff/rd1.htm

Japan

Shinji Yamane is currently researching Internet security as a PhD candidate in the Graduate School of Information Science, Tohoku University, Japan. Born in Brazil and educated in Japan, he graduated in 1992 from the International Christian University in Tokyo. His special interests include the hacker ethic and Internet culture. A member of the ACM, Shinji was also a foundation member of CPSR (Computer Professionals for Social Responsibility), Japan Chapter.

Russia

Irina Alexeyeva is a senior researcher at the Institute of Philosophy, Russian Academy of Sciences, where she has worked since graduating from Lomonosovîs Moscow State University in 1983. She also teaches philosophy courses at several Moscow universities, including the Moscow Aviation Institute. Her book *Human Knowledge and its Computer Image* (in Russian) was published

in 1993. Her research work is concentrated in two areas, the epistemological content of the computer revolution and the ethical dimension of technology.

Singapore

Lim Kin Chew works in Singapore, as Manager of Temasek Polytechnic's Centre for IT in Education and Learning. He leads a team of instructional designers, programmers and computer graphic designers who help him put up IT-based educational courseware on the Internet. He uses the Internet very widely and has been involved with the Internet set-up facilities in his present workplace. He also helps disabled people to use the Internet.

Sweden

Dr Viiveke Fåk is an associate professor at Linköping University, Linköping, Sweden. She received her PhD from that university in 1978 on a thesis treating cryptography as a mechanism for computer and network security. She has since then been a teacher, lecturer, consultant and researcher in computer security and cryptography. She has served as chairman for IFIP WG 11:4 on crypto management 1985–1989, and has taken part in other international scientific work, primarily as referee and as a member of the programme committee for conferences. She has also served as an expert on Government committees. Her general interests are in the effects of and possibilities created by modern information technology, with scientific interest concentrated on authentication of computer users and of data.

United States

Terrell Ward Bynum is Director of the Research Center on Computing and Society at Southern Connecticut State University in the USA. He is Professor of Philosophy there and also Visiting Professor at the Centre for Computing and Social Responsibility at De Montfort University in the UK. Professor Bynum is Past Chair of the Committee on Professional Ethics of the Association for Computing Machinery, Past Chair of the Committee on Philosophy and Computing of the American Philosophical Association, and a lifetime member of Computer Professionals for Social Responsibility.

2 Internet Technology

DAVID BECKETT

Introduction

This chapter, a mainly technical description of the development and background of the Internet, initially may seem less relevant to the theme of the book; but appropriate technical knowledge is actually directly relevant to an examination of Internet ethics. Providing a necessary understanding of the physical Internet gives a solid foundation for later detailed discussion on Internet use. There are two further aspects of this chapter that should also be explained. The first is that computer scientists are notoriously prone to use acronyms, which are inevitably included here in some quantity; the second is that there are additional references in this chapter to World Wide Web locations, or URLs (defined below). These additional references are given as superscript numbers and are listed at the end of the chapter.

The Internet is not a superhighway and was not designed to survive a nuclear war (Hafner, 1996, p. 10). This chapter provides a background to Internet technology, from the cables in the ground through to describing a portal. Along the way some of the history of the Internet and the Web (World Wide Web, or WWW) will be outlined and how it was built from free software.

The ARPANET and early networking

In 1966, the Advanced Research Projects Agency (ARPA) of the US Department of Defense was funding several research laboratories around the country with expensive computing machinery for researchers. The ARPA managers were concerned that these facilities were not being used cost effectively, since they were only available to people working locally. It was thought that there should be some way of providing access nationally to the ARPA scientists. There were two main applications that they wanted to be able to perform between the computers – to copy files and to operate the computers remotely (remote access). These needs were communicated to the managers who started to consider how to create such a network.

At this time, the phone system had been used to connect remote computers via dedicated trunk lines leased from AT&T.[1] Special devices called modems connected the phone cables and the computer terminals. These remote terminals were operated the same as local terminals; there were no computer to computer links. Since phone lines were expensive to lease, they could not be used to fully connect every site with every other site, but lines could be connected between near sites, with a few long distance lines. This network shape, or topology, demanded a special kind of communications network that knew how to direct messages along the correct lines to their destinations.

It was decided that a new computer networking technique called *packet switching* would be used for the network. A packet is a short message, usually part of a larger message. On this network, where the typical use was for terminal access, most messages were small containing key presses and thus made good use of the network.

Packets contain:

- source and destination addresses
- some control information
- data, usually in varying amounts.

Packets are transmitted from their source to destination via intermediate computers called *routers*. At each one, a decision is made where to send (or route) each incoming packet. With simple packets, this is quite rapid, and the packet can be sent on the best outgoing link towards the final destination. This made good use of expensive trunk lines, and hence was appropriate for this network.

In 1968, ARPA put out the Request for Quotation for the new network – the ARPANET to

> improve and increase computer research productivity through resource sharing. Technical needs in scientific and military environments were cited as a justification. (Heart, 1978, III: 34)

that is, not explicitly for surviving a nuclear attack. Bolt Beranek and Newman (BBN)[2] won the bid to manufacture the equipment to build the network interface devices, called interface message processors (IMPs). Bolt Beranek and Newman also designed the way that IMPs connected to the computers, and the host-to-host protocol that they used to communicate. *Host* is another term used for a computer on a network.

ARPANET sites were mostly at universities, where they were run day-to-day by graduate students. These students, to their great surprise, were left to determine the applications for the ARPANET and how they worked. They ended up designing and operating the network services but most of them 'expected that a professional crew would show up eventually to take over the problems we were dealing with' (Reynolds, 1987). This uncertainty led to the various documents being labelled rather cautiously Request for Comments (RFC) (Braden, 1999), since it was not clear if these were official documents. This informal tradition inspired an open discussion model for creating common standards by consensus, with no barriers, secret or proprietary content.

The conventions that computers use when networking are called protocols, which can be thought of 'in terms of diplomats exchanging handshakes and information' (Salus, 1995, p. 42). Protocols are operated by networking software to provide the network services that applications use. Most of the applications use the client/server model, in which there are two participants. A request is made for a particular service from the client (usually the user) to an entity that can perform it – the server. The server then usually responds with the result, or the conversation continues between the client and server until one of the participants ends it. This model is widely used in computer networking at various levels.

The two applications that were needed for the ARPANET on top of the basic Bolt Beranek and Newman host-to-host protocol were:

1. remote login – remote access to a computer on the network via the telecommunication network protocol, telnet.

2. file transfer – copying of files from host to host via the File Transfer Protocol (FTP).

These applications were also used as a basis for other services, which is a very common technique used throughout networking software called layering. This is often done by creating a hierarchy of protocols, with protocols for application 'above'ones for bare networking, or 'low level'protocols.

The major application not originally thought of was that of passing messages between people. The ARPANET was not intended as a message system but for resource sharing. However, as the network grew the designers and users found ways to communicate electronically and wanted better ways. Messaging systems already existed between users on one computer, but not over networks. The first

There was another e-mail syntax in use for many years that involved UUCP 'bang paths'; examining the path header of Usenet articles will show what this looks like. UUCP and Usenet are described later.

personal message sent over the ARPANET was by Lenny Kleinrock in 1973; however, Ray Tomlinson at Bolt Beranek and Newman wrote the first true e-mail programs, and introduced the @ sign as the punctuation for e-mail addresses, separating the user on the left from the site or computer on the right (Hafner, 1996, p. 192). Mail was originally called network mail, or netmail, and later electronic mail, e-mail or email.

The Internet and the Internet protocols

The ARPANET host-to-host protocol design evolved several times, but by the mid 1970s it was showing its age, and a better protocol was needed. Developments had continued, and by this stage there were several other packet-switched research networks in use. However, these were using different formats, so some way to inter-network between these was required. A larger 'network of networks' is generally called an internet, so the new network was called the Internet. It was to be based on a foundation of a new Internet Protocol (IP), with other layered protocols above that. (Today, an *internet* implies a network that uses IP, as well as connecting multiple networks). To help in understanding how the Internet protocols operate, two examples may be used to illustrate different areas. The first example you may already have heard of.

An information superhighway

Online, use of the phrase Information Superhighway is taken to be evidence that the writer using it is pretty unskilled.

This concept is that the Internet was funded federally by the Government (the US government) and that main electronic communication routes, like America's interstate highways, were built without commercial support. In this sense, the metaphor is true, but it provides no help in explaining the later development of the Internet, after the government withdrew its funding (discussed later in this chapter) and has no relevance to how the network actually works. In other words, the concept of an information superhighway is actually very little use in explaining the current global electronic network. **[SINGAPORE 1, BRAZIL 1, AFRICA 1, BRAZIL 2]**

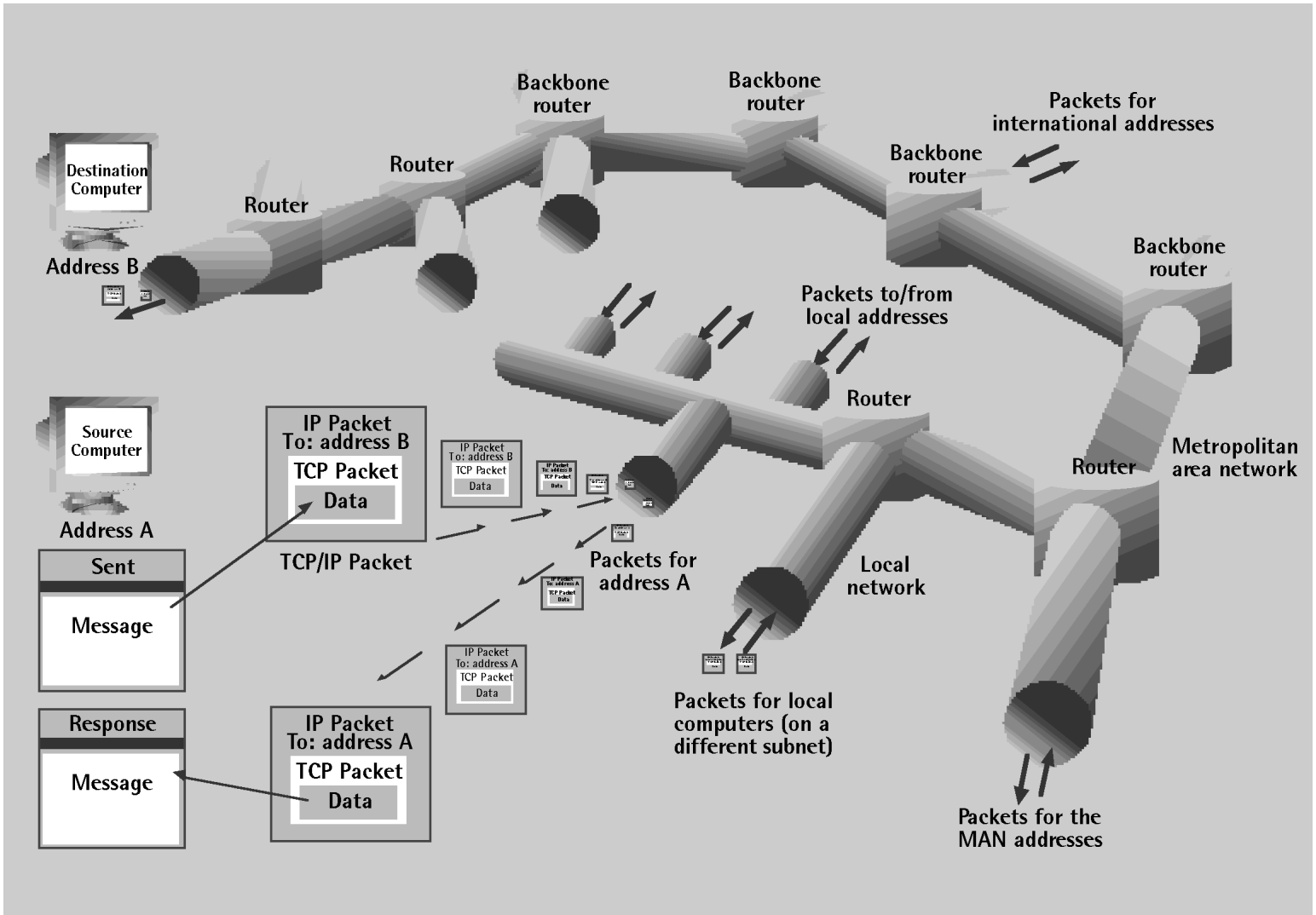

FIGURE 2.1: Digital Pipelines

Digital pipelines

This is much more appropriate analogy, since most people are probably fairly familiar at a basic level with what a pipeline is – a 'thing'that transports 'content' such as water, natural gas, or gasoline/petroleum. So let us consider communications between computers or computer networking being represented by wires (digital pipelines) transferring messages.

Internet protocols standardise the lowest level elements in the IP packets (basically, the shape of messages in a pipe). Messages contain the source and destination addresses of the packet, and some data. These packets are routed by looking at the destination address written on the message; and at each hop in the journey, a router passes the message on. No guarantee is given that an IPpacket will reach its destination intact; failure may happen for several reasons.

For example:

- the packet becomes corrupted (message gathers dirt)
- the packet is dropped, due to congestion (pipe is full)
- the destination address cannot be found (wrong address on message)
- the destination is not listening (pipe disconnected).

Above the IP layer are further protocols, which provide an enhanced service beyond the raw IP. The important ones are:

- The User Datagram Protocol (UDP) which provides single-shot messages (thrown into the pipe) that give no guarantee that the packet arrives at all, since any IPpacket can be dropped without notice. The destination can check any arriving UDPpacket to see if it has become corrupted in transit, but there is no mechanism for correcting the error if there has been corruption.
- The Transmission Control Protocol/Internet Protocol (TCP/IP) which provides guaranteed reliable, error free message transmission between the source and destination. The message can be longer than the maximum size allowed for IP packets, since TCP/IP initially splits up long messages into IP packets, and then reassembles them at their destination. The source and destination also collaborate, to check that the packets are all present, in the right order and that the data has not been corrupted in transit. Should a problem occur, it is corrected by the destination requesting more packets from the source. TCP/IP is used today for most Internet applications, since it handles all of the problems listed above. Its use means a message between applications

is guaranteed to arrive either 100 per cent correct, or not at all. TCP/IP is, however, rather slow compared to UDP because of the extra work involved.

IP is concerned with the contents of the pipes.The diameter or width of the pipe represents the capacity for messages that can flow down the pipe; for data communication this is called bandwidth. **[AFRICA 2]** Different sized pipes can also be connected to each other via a device called a switch. Normally this is done because it is more efficient to use large pipes for liquids (or data) travelling long distances rather than to use many small pipes. This is directly analogous to the use of large fresh or waste water pipes between cities.

Devices also exist devices that switch traffic, and can also convert messages to and from different formats. These devices are sometimes known as *gateways*, and they may be used to connect IP networks to non-IP networks.

Early Internet systems and applications

While the computer systems being used by the ARPA laboratories were made by different companies, their networking software obviously needed to be compatible. For efficiency, the software was developed by the users of the network collaboratively, and then distributed to users as *source code*. Source code is the form in which a computer program is written before it is compiled into a machine-specific version running only on a particular computer. This method of distribution meant that the software could be modified easily, improved and tested by many people at once. Such a development method had several advantages, including getting problems (or bugs) fixed quicker, fostering research into improving the network, and of course keeping a sense of community.

Many of the computers used in the ARPANET labs ran an operating system called Unix originally developed by AT&T Labs (formerly Bell Labs,[3] later part of Lucent[4]). Unix was written in the C programming language, like much of the networking software, and from early on was available with full source code. These combinations meant that Unix ran the majority of Internet networking software from very early in the 1970s. BSD Unix, from the University of Berkeley, California, was one of the most popular Unix versions. Its 4.x series of Unix operating systems (or systems based on it) were the main systems running the Internet throughout most of the 1970s and 1980s.

File transfer

The *file transfer protocol* (FTP) was the main protocol for transfering files regularly used in the early years of the net. A protocol called UUCP was also used, since it was designed for dial-up, phone-based access, and particularly in the early years some sites used this to connect to the net. UUCP works by automatically making phone calls at regular intervals, then picking up or dropping files (or in this case, e-mails). The UUCP facility is still in use nearly 30 years later in situations where there is no Internet access available, or when making connections within networks built from dial-up phone connections, such as FidoNet.[5] **[AFRICA 3]**

Remote access

The *telnet* protocol allows people to access remote computers as if they were actually sitting at directly connected terminals. This protocol has been enhanced over the years, but its functionality has basically remained the same, providing terminal-like access to computers, such as those running Unix, designed to handle multiple users.

A related application, called *talk*, was developed later. Talk allows a small number of people to contribute to a typed conversation over a network, sharing a split terminal window in order that all the participants may see each side of the conversation. This application is generally known as *chat* and has developed into more advanced systems such as Internet Relay Chat (IRC).[6] IRC takes place via a server and allows real-time chat between multiple participants. Each user picks a nickname, and selects subject-based channels. On each channel there are topics in current conversations that can be joined by an interested user.

E-mail and mailing lists

E-mail was, as mentioned above, an unexpectedly important application that the original pioneers did not include in the original Internet design. However, it rapidly became the most popular application, and has remained so for many years. The early mail protocols worked on top of FTP, and were quite simple, but the protocols and software have since improved. E-mail headers are an important evolved feature of e-mail. They contain information about the message lying outside the body of the message. Header contents usually include the sender

(From), the recipient (To), the subject (Subject), and so on. They can be seen in some form in most e-mail software, although the full headers may only be available via a special menu option. The original FTP-based mail protocol was replaced by Simple Mail Transfer Protocol (SMTP) which was designed specifically for e-mail. SMTPwas especially designed to be simple, and to allow e-mail to be sent and received easily over a TCP/IP channel. At the present time, the majority of global e-mail is sent using this protocol (or its direct successor, Extended Simple Mail Transfer Protocol, ESMTP).

One of the main original e-mail transfer agents (MTAs) was based on UUCP. A later MTA, called Sendmail[7] was developed by Eric Allman[8] to manage e-mail interactions – essentially, sending, receiving and forwarding messages between the various types of systems. The program Sendmail was made freely available, since e-mail was a key application and people were relying on it. This program is still free today, and is used by many large organisations.

E-mail was used for person-to-person messages, but several other systems evolved for communicating between many people. Ordinary e-mails can of course be sent to many people by listing them on the 'destination'or 'To'header, but this method becomes unmanageable for larger mailings. Software was therefore written to handle *mailing lists* or *listservs* (from 'list servers'), allowing people to broadcast messages to all the list members by sending a single message to the list e-mail address. Joining and leaving lists is done by sending a message to the relevant mail listserv address, usually listserv@somewhere. This is still used on mailing lists today.

Mailing lists were run and maintained on computers scattered around the network, but, for popular lists, this was a slow way to transmit the messages – particularly if there were thousands of participating users. A more efficient way was consequently developed to distribute messages for large audiences. An additional pressure for this development came as, at the time, the network was only available to connected Internet users, while people outside the network dialling in via UUCP wanted to participate in the shared information space.

Usenet

The new system that evolved was called *network news* but is generally known now as *Usenet*. Usenet is based on messages in the e-mail format, with headers and a body. The messages are published or *posted* to one or more particular

If you look at the headers of a news posting, you can see via the Path header, the route that message took to get to you.

subject-specific *newsgroup*. The messages are then copied around the network, from one computer on to its neighbours. Each participating computer looks at the new message, and decides if it had already seen it, taking a copy if it has not.

Some newsgroups are moderated, meaning that each message or article is e-mailed to a human moderator for approval before being published. Most newsgroups, however, are open access. The newsgroups were originally named in a similar manner to the early mailing lists. Later, although, they were reorganised into a hierarchy, with the top level categories based on eight main areas, with others for countries (for example, '.fr', France) and local sites. Below the top level were sub-groups for more specific topics, with more sub-groups added as the demand for them was seen.

Usenet has grown immensely since the original system design. Though there have been several predictions of its imminent collapse, this has yet to happen.

The software to read, post and serve new articles was made freely available, like the Internet software already discussed. The early news articles were copied by UUCP, but, later, since the number of newsgroups and articles had increased vastly, a dedicated news protocol was needed. This was called Network News Transport Protocol (NNTP) and was developed specifically to pass round the news articles more efficiently.

The Internet culture and free software

The Internet was founded on what could be called the open philosophy – systems, and especially software, were developed openly; no proprietary protocols were used in the network, and there were no big restrictions on copying or modifying software. This approach was encouraged as many users based at universities, where there is generally an ethos of collaboration in order to advance common knowledge. The outcome of this original open philosophy included a willingness to share and help others, and to collaborate in advancing the technology.

The openness of the Internet standards process (RFCs) emerged from the people who formed it. Jon Postel[9] (1943–98) who was the RFC editor from 1969 till his death, established and ran the Internet Assigned Numbers Authority (IANA)[10] which maintains the lists of terms and protocol elements needed to keep the network communicating; he was the first individual member of the Internet Society (ISOC)[11] in 1992. Jon and others kept the spirit of fun going by publishing humorous RFCs on 1st April each year (All Fools Day). The main technical development body for RFCs and other Internet standards is the Internet Engineering

Task Force (IETF)[12] which developed from the ad hoc process of the early ARPANET people. The IETF operates through working groups on mailing lists, meeting three times each year; this arrangement derived from the initial ARPANET working group schedule. There is no formal membership – anyone may attend – but the ethos is on good, demonstrable technical work, or in the IETF's own words, to attain 'rough consensus… and on running code' (Carpenter, 1996).

Several new words emerged from the vibrant communications media that ran over the Internet. Usenet users created the rather clumsy term *netiquette*, the (suggested) way to behave in e-mail and newsgroup articles, *flaming* – being deliberately abusive or argumentative and *trolling* – writing outrageous articles to provoke a response just for fun! For more of this kind of thing, see Raymond (1996). The most infamous term that came from Usenet is *spamming* (Python, 1973) which was originally the posting of the same message to multiple newsgroups (also called *cross-posting*) usually to try to sell something through an advertisement. The term spamming was later applied to e-mail messages sent to thousands of victims, most having been selected by copying e-mail addresses from posted news articles.

As the commercial exploitation of the Internet grew, most of the software that was actually running the network still remained non-commercial. (Some examples have been given above – BSD Unix, sendmail.) Established users began to be concerned that, with the new rush of commercialism, the open spirit of sharing would be lost. They wanted to foster the continued development and free distribution of software source code. In 1984, Richard Stallman[13] started the GNU's Not Unix (GNU)[14] project to create a free software version of the Unix operating system, including all its supporting software. To do this, and to protect its freedom, he wrote a new licence called the GNU General Public License (GPL)[15] which protects the freedom for people to copy software, and prevents the creation of derived works that change this state. The GPL was applied to many useful programs, such as the GNU C compiler, that were in use at the time.

In 1985, Stallman founded an organisation, the Free Software Foundation (FSF)[16] to protect and foster the idea of Free Software.[17] Their definition of free software:

> Free software is that which comes with permission for anyone to use, copy, and distribute, either verbatim or with modifications, either gratis or for a fee. In particular, this means that source code must be available. 'If it's not source, it's not software.' (Tower, 1998)

> 'Free software' is a matter of liberty, not price. To understand the concept, you should think of 'free speech', not 'free beer'. (Stallman, 1999)

The FSF is opposed to software patents, but not to people making money selling software.

Much other software already mentioned is free software, such as BSD Unix from Berkeley, although available under a slightly different licence – the BSD licence.[18] This allows commercial binary versions (and in Stallman's terms, is not as free as GPL'ed software). Kirk McKusick,[19] one of the leaders of the Berkeley UNIX project in its heyday, puts it something like this:

> Copyright is designed to protect the intellectual property rights of the people who create something. Copyleft is designed to protect the rights of the users. The Berkeley license is copy central: Take this stuff down to the copier and make as many copies as you want, for whatever you want.

The X Window System (X)[20] is a portable and flexible graphics technology developed from a project at MIT. It can be used to create graphical user interfaces and works well with Unix. This allowed better and quicker software development and made work with graphics more economic. X was released under the MIT licence[21] which was also pretty generous in its terms. Later on the terms were changed to be less free but a major free software version of X, Xfree86[22] forced the change to be reversed by planning to work from the older free versions.

In October 1991, Linus Torvalds[23] a student at the University of Helsinki in Finland released a new operating system he had been developing for fun under the GPLwhich he called Linux.[24] Linux was the first working implementation of an operating system based entirely on free (GPL'ed) software. It was delivered with the full source code of the system available for people to download by FTP, and then experiment. In a surprisingly short time, Linux attracted many developers, who in turn added improvements to the basic system created by Linus. Free software versions of BSD – FreeBSD[25] and OpenBSD[26] were also developed in the early 1990s, but Linux proved far more successful.

The Internet expands

The ARPANET evolved into two distinct entities – the Csnet, a largely academic computer science network, and a separate military network for the US Department of Defense. These networks were constructed around several original long distance networks that formed the main network architecture of the ARPANet; this is commonly called the *network backbone*. Later, the civilian network became the National Science Foundation Network (NSFnet), funded by that US organisation. However, all such bodies normally forbid the use of their funding for commercial or business purposes. With the intensifying commercial nature of the Internet this was increasingly seen as a problem, eventually leading to the US government's withdrawal from operating the Internet backbone. This left participants to purchase connections from telecoms companies, and new commercial operators. IP also began to be used extensively within organisations over their own Local Area Networks (LANs); many such local networks were running on a cheap and relatively easy to install hardware technology called *Ethernet*. This was designed in 1973 by Bob Metcalfe,[27] one of the original Internet pioneers, when he was working for Xerox.[28] Metcalfe later founded the major network company 3Com,[29] and is now a prominent Internet columnist for InfoWorld[30] magazine. **[AFRICA 4]**

Computers use numbers for everything, including the addresses of machines, but humans prefer to use readable names. To accommodate these two preferences on a network, a map giving the relationship between names and the numbers is needed. Originally this was accomplished with one huge file that contained the names and addresses of all the Internet hosts in existence. As the number of machines on the Internet grew, this file was changing so frequently that it became impossible to keep it up to date. A more flexible solution was therefore developed, using hierarchical names for the networked computers with '.'s separating the elements of the names, such as parrot.example.com. These are called *domain names* and the system that makes them available is the Domain Name System (DNS). The DNS distributes the mappings from names to numbers around the network so that organisations are responsible for maintaining their own names – this is known as a *distributed database*. The distribution of responsibility for updating names means that there is no longer a single huge file of information, and details may therefore be kept more up to date. The most popular and complete implementation of the DNS was the freely available BIND[31] which remains the major implementation in use today.

Domain names are structured at the root (or top) into two types – *country* domains (.uk for the United Kingdom, .jp for Japan, for example) taken from the standard ISO two-letter country codes, and *top level* domains which include:

- .com for commercial sites
- .org for non-profit sites
- .net for networks
- .int for international sites
- .gov for US government sites
- .edu for US education sites.

Originally the first three suffixes were intended solely for sites in the USA, but com, org and net gradually began to be used for any type of site, anywhere in the world. If one of the listed top level suffixes is not employed, country suffixes must normally be used for addresses in every country – except the USA. (There is a .us domain for the United States, but it is at present mostly used for state schools and government.) **[BRAZIL 3]**

Expansion of the use of global e-mail led to further developments in the protocols used. The Post Office Protocol (POP) allows people to use e-mail via a server – essentially to pick up and send mail items. This is particularly useful when the computer used varies from day-to-day , or if a user is working off-site and dialling in to read or send mail. A new form of e-mail encoding called the Multipurpose Internet Mail Extensions (MIME) was also introduced; this allowed e-mail to contain messages that were not just plain text. For example, MIME e-mail may be an internationalised message (not in the usual limited character set), multiple versions of the message in different encodings or even an attached binary file (such as a digital photograph) protected from being changed in transit by a special encoding. Attached binary files are now known as *mail attachments*.

BSD Unix was very popular in computer education and research. This popularity led in 1982 to a new company being formed to develop products for this market. Sun Microsystems[32] was founded by Stanford graduates including Scott McNealy[33] and Bill Joy[34] (who also designed some key technologies, such as BSD Unix, NFS, Java and JINI). Sun picked up and used X and developed new hardware specially to exploit it. This hardware sold widely to the Internet community, due to high performance and out-of-the-box Internet functionality. Today Sun Microsystems still provide a large proportion of the hardware behind many major Internet sites.

The World Wide Web (WWW)

In the late 1980s and early 1990s, most of the file transferring on the Internet was done by means of ftp, postings to Usenet, or e-mail. In 1993, researchers at the University of Minnesota created a menu-based interactive system and protocol called *gopher*, which allowed users to find information and files on the Internet. It allowed browsing menus to be created by users, and searches to be submitted using them. For the first time, there was something more sophisticated than a raw list of ftp files. Gopher was used by a client program that displayed menus on a text terminal, and allowed navigation around them. This interface was much more consistent and easy to use across multiple sites than using ftp, because ftp works by entering and exiting each site, which is difficult for beginners to use. (Following the explosive growth of the WWW, gopher was turned off at the University of Minnesota in 1996, and today seems to have disappeared from the net entirely.)

O'Reilly and Associates[35] founded by Tim O'Reilly[36] began publishing Unix, X and related books in 1988, but also covered many free software programs. They commissioned a gentle introductory book about the Internet for novices from Ed Krol.[37] His book was called *The Whole Internet Users Guide and Catalog*[38] (Krol, 1994) and was originally published in 1992. The text covered e-mail, mailing lists, ftp, gopher (and its search system, veronica) although even with this help the net was still rather difficult for a beginner to use. Finally, the book covered a new system called the World Wide Web (WWW)[39] created by a British physicist Tim Berners-Lee[40] who worked for the high energy physics research laboratory CERN[41] in Switzerland.

In 1991 CERN and other labs around the world were communicating electronically, using the network to pass around research papers, images and other materials via e-mail and ftp sites. At CERN there was also a variety of other older information systems, but each of them needed to be accessed by a different program. Berners-Lee thought that it would be better if access to these systems could be made both much easier and essentially seamless. The information space he created to do this was the WWW, or the Web; and the program allowing access was called a *browser*, or a *Web browser*. Browsers used a new protocol called HTTP (HyperText Transfer Protocol)[42] – and delivered documents in a format called HTML (HyperText Markup Language).[43]

HTTP operates over a TCP/IP connection to a server; it is designed to transfer HTML, plain text or any other type of documents. The protocol allows the fetching of one document per connection to an HTTP server (at least initially; later versions allow more). The simplicity of the protocol, which was written using text-based commands, like SMTP and earlier protocols, ensured that HTTP was easy to understand and, importantly, it was easy to write client and server applications. For obvious reasons, HTTP servers are usually called *Web servers*. Unfortunately, the *computers* that run HTTP servers are also sometimes called Web servers!

An HTTP request and reply sequence is also known as *fetching* or *getting* a Web page or a Web request.

HTML is a format that describes hypertext – text that contains links to other documents. Hypertext systems had been used in the research community for several years but had not generally been integrated with networking. HTML is based on an another much more complex electronic document format called SGML(Standard Generalised Markup Language), in which the body of the document contains text surrounded by tags. Tags are description, formatting or other instructions to interpret the document, and are written as '<' followed by a tag name, and ending with '>'. Tags may mark the start or end of parts of the document, as in this example:

```
...some text....
<TITLE>This is the title</TITLE>
...some more text...
```

The first, or *opening* tag indicates the start of the title in the document and the second, or *closing*, tag marks the end of the title. Closing tags have a '/' before the tag name. Tags are not always used in pairs – an example is <P>, which can be used alone to separate paragraphs.

To understand a document, the browser must understand the tags; however in HTML, Berners-Lee added a new rule – that if the browser did not understand a tag, it should ignore it. The impact of this rule became more significant as further tags were added to HTML.

It is of particular importance that Web browsers understand special <A> (or *anchor*) tags in HTML. Such tags point to other HTML documents, and are the links that create the Web of documents, or hypertext. Other documents are iden-

tified by using a specific and unique address, called a Uniform Resource Locator or *URL*[44] (Berners-Lee, 1994). URLs can point to HTTP, ftp, gopher and news servers directly, indicate telnet access to resources, and other services. A URLhas many parts, but all URLs start with the protocol, before a colon :, for example 'http:'or 'ftp:'as in

```
http://www.w3.org/Addressing/
```

Berners-Lee realised that the seamless unified access he wanted could be provided across many types of servers by the use of HTTP gateways that translated the Web requests into the language understood by other services. In this way, the web of hypertext documents could extend well beyond just those documents written in HTML. See Figure 2.2.

The original browser program spawned other versions that ran on almost all computers – mainframes, PCs, Unix systems, Macintoshes and many others. Each of these systems now had equal access to resources previously only available to a select few. The physics community understandably took up this new system rapidly, and began to create Web document trees around the world, using the clients and servers developed by CERN.

New Web browsers, servers and sites

One early adopter of the Web technologies was the National Center for Supercomputing Applications (NCSA)[45] at the University of Illinois. The NCSAgroup developed software for the physicists and others using the supercomputers around the USA. Naturally they wanted to both use the materials at CERN and to share their own work. This led them to work on a new kind of browser for their applications, one involving the display of graphics across the network. The NCSA software development group developed a graphical server, the NCSA httpd,[46] and, in 1993, a graphical client called Mosaic[47] that ran on Unix, Microsoft Windows/Intel and Macintosh systems. The major innovation in the browser was a new <IMG> HTML tag that allowed authors to add pictures to HTML documents. As mentioned above, existing browsers that did not understand this new tag would automatically ignore it. Addition of the tag to a Web page therefore meant old browsers could still read the page, albeit without the benefit of the image. This was the start of backwards compatibility problems and access issues for Web pages.

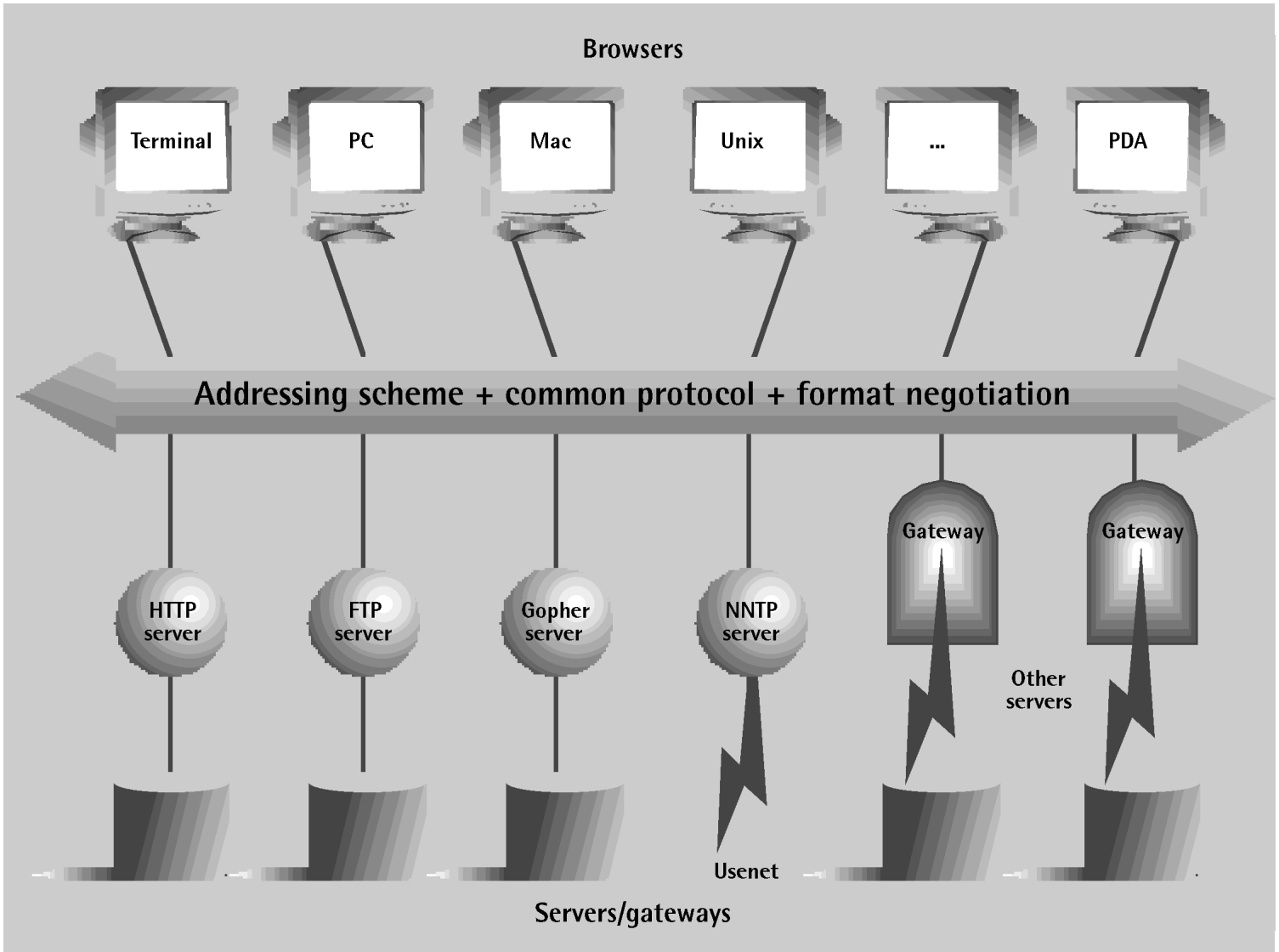

FIGURE 2.2: The Web client/server model

NCSA Mosaic was publicly released and was very quickly picked up by people all over the network, who could now not only use ftp, gopher and the Web in a much friendlier way but view images too – a big change from the terminal-based text Web and systems that were current at the time. Within a year Mosaic had an estimated 2 million users globally. The most popular Web servers at this time were from CERN and NCSA, although there were many others, since writing a server was relatively easy.

NCSA later added a facility to connect their Web server and external programs; called the Common Gateway Interface (CGI),[48] this was used by requesting a special URL. The URL called an external program which actually created an (HTML) page, allowing the generation of *dynamic* Web spaces, rather than the *static* ones which are formed from plain HTML files. Dynamic pages are particularly useful in creating gateways to legacy systems, databases and programs by converting their output to HTML, and thus making them world accessible via the Web. The CGI interface was rapidly added to all servers, and remains the standard today.

The best way to create CGI programs is by using what are called *scripting languages*, computer languages designed to allow programs to be written quickly and, usually, to work with text. The most popular of these is Perl[49] created by Larry Wall,[50] (which is available as free software). Perl is so often used for CGI and other Web tasks that it has been said that 'Perl built the Web'. Perl was designed in the Unix culture, and took lots of its syntax and ideas from other Unix tools; it was thus familiar and useful to many people. Since most CGI programs are written using scripting languages, they are most commonly known as *CGI scripts*.

In the period 1993–95 the Web grew rapidly and many people started to create their own personal Web areas. Berners-Lee created the idea of a person's Web home page – the starting point for people using the Web, containing links to their own documents as well as links to interesting or useful places on the Web. The basic concept evolved into two separate ideas:

- the *home page*, which tended to be a person's public home on the Web
- the browser *bookmarks* (later also called *favorites*) which were more structured Web links, containing the link, the title of the link, the date the bookmark was made, and usually the date it was last visited.

Browsers also stored all the links that were taken during a Web session in a browser *history*. A history usually contained records of several days'activity, to allow people to return to locations they had seen recently. There was also another sort of record, in the form of the 'go back' and 'go forward' buttons which allowed the user to travel back and forwards along the current route they were taking through the Web. The Web document trees, areas or sites that were constructed began to be known as *Web sites* and many innovative Web sites began being created. In April 1994, the Web was becoming so large that two graduates at Stanford University created a directory, classifying their favourite Web sites, which they called Yahoo![51] This directory was the foundation of a multi-billion dollar company. Incidentally, Yahoo! was built, managed and is still run using Perl and BSD Unix.

Coincidentally, it was also in April, 1994, that the core of the National Center for Supercomputing staff who wrote Mosaic (including the key designer Marc Andreessen[52]) left to form a new private company called Netscape.[53] Netscape soon began to design and build a new browser and server, and there was considerable interest in what the products would be like. On 13 October 1994, Netscape released a beta version of their new browser on the Net, free for personal use, and there was an immediate positive reaction. Millions of copies were downloaded. The browser had new and, then, unique features, such as displaying a document as it loaded, the ability to fetch multiple images simultaneously, and support of the JPEG image format. Netscape invented some new HTMLtags to change and improve the look of Web pages, including <CENTER> to centre lines on the page; this was to be the first of many changes. Its impressive abilities made the new browser, Netscape Navigator, look very fast and slick compared to Mosaic. The next week, *Wired*[54] magazine created one of the very first commercial Web sites called HotWired,[55] and displayed the first advertising banner on the Web. The HotWired site was designed to be very graphical, and it worked very well with the new browser.

In December 1994 Netscape shipped its server products, targeted at commercial use of the Net, or e-commerce. For shopping sites, people need to be able to browse, add items to some form of 'shopping basket', and perhaps return several times before purchasing. This means that a commercial server needed to be able to identify individual customers. Some sites use IP addresses to do this, but such identification is of no use if several people come from the same machine or appear to. (This can happen when using caches, see later.) An alternative was needed. To resolve this problem, Netscape added a new feature to their browser,

Some sites still use IP addresses as unique identifiers for people and I recommend you avoid them since it can expose your personal information to any person who shares the computer with the same address.

a feature they called *cookies*. These are small text files that sit on the user's computer, and are sent to Web sites when the user visits. For example, the netscape.com site may register a cookie called *user*, with some value. Whenever you visit any Web page below a netscape.com address, the *user* cookie and the value are sent as part of the HTTP request for every page. In this way, the content of the returned page can be customised for an individual visitor. The issue of cookies does have several privacy and other implications, discussed in other chapters.

Netscape was a major e-commerce user itself, selling its browser and server products online. Customers generally wanted to use their credit cards, and this meant card numbers had to be sent to Netscape on a normal TCP/IP connection. IP packets passing across the Internet can be read by any intermediate computer and, potentially, this meant credit card numbers and personal information could be intercepted. Netscape needed a way to encrypt data used for sending such sensitive information. Encryption is done by complex mathematical transformations of data called cryptography. **[SWEDEN 1]** Netscape created a protocol called SSL (Secure sockets layer), which formed a secure wrapper protocol on top of HTTP; this appears in URLs as the protocol https:.

In February 1995 the popular NCSA Web server was not being maintained (the main author had left to join Netscape) so a group of Webmasters gathered their extensions and bug fixes as *software patches*, – which are text files describing the differences between the old and new versions of files. The group created a new server called Apache[56] (think 'A patch-y') and made their 1.0 release in December 1995. Within a year Apache was the most popular server on the Net; it has remained so ever since, and now runs the majority of Web sites.

The Internet commercialises

In 1993, the US government removed the federal funding for the Internet backbone, and let it be operated by new commercial organisations. These specialist telecommunications companies run high-speed fibre optic cables over long distances, and provide connectivity by various switching centres. These networks are generally known as Metropolitan Area Networks (MANs) when they span cities, or Wide Area Networks (WANs) if the network is over a longer distance. Individual users dial into the Internet from home using their modems to connect

to Internet Service Providers (ISPs), who then connect to the main Internet, sometimes via larger ISPs, and out to the whole of the worldwide network.

There were several organisations running large private online services with their own content and using their own protocols, such as AOL[57] and CompuServe,[58] but these organisations did not connect to the Internet directly until around 1995, and then only through select gateways. It was several years before they provided full Internet accessibility providing Web, e-mail, ftp, telnet and so on, but today they have turned almost completely into major ISPs with the 'extra'of additional Web content. Microsoft[59] also completely ignored the Internet for several years, and instead tried to build its own proprietary online service called MSN.[60] It eventually gave up when Bill Gates[61] re-targeted the company in 1995 to be Internet orientated. Microsoft then licensed the Mosaic browser source code from NCSA to build its own Web browser – Internet Explorer[62] – to compete with Netscape, who had over 80 per cent of the market share at the time.

The US government also gave the operation of the DNS, by open tender, to a private company called Network Solutions Inc.[63] Network Solutions was then solely responsible for the .com and other domains, providing the central database and handling the billing for them. The .com domain, intended for US companies, was used for most sites wherever they were located and hence was the 'place'to be if you were creating a Web site. In 1998 this situation needed reforming, since Network Solutions was in effect running a government-sanctioned monopoly, and there was additionally a need for more top level domains. Many discussions around the world took place on appropriate changes, with the help of bodies such as the Internet Society and IANA. In early 1999, IANA was reformed as The Internet Corporation for Assigned Names and Numbers (ICANN)[64] to oversee the new DNS arrangements. The first five new registrars, responsible for allocation of domain names, were announced in April 1999.

Some pioneer e-commerce companies starting up in this vibrant period included:

- Amazon,[65] who started selling books online in July 1995, using encryption to protect user data, and giving substantial discounts on the store prices
- Geocities,[66] which in May 1995 was one of the first sites offering free Web home pages (supported by advertising)
- Dell[67] who opened an Internet computer store in 1996, selling PCs online directly to the end user and consequently saving money in intermediate costs

- Hotmail,[68] who provide free Web-based e-mail, supported by advertising on their Web pages. Unfortunately HotMail accounts (and later, other free e-mail services) have been taken up enthusiastically by e-mail spammers, since such e-mail accounts can be created easily and without any responsibility
- CNN,[69] who in 1995 created the first major topical news Web site, with continuous updates from correspondents around the world, together with pictures, audio and video clips.

A new kind of service that allowed you to search the Web appeared in the mid 1990s, based on software that 'crawled' the links that made up the Web, or for short, a *Webcrawler*. These software agents (or robots) take the words from Web pages, store them, and create indexes that can be searched. One of the first commercial sites to do this was Lycos,[70] which was founded in June 1995. Later AltaVista[71] created a much larger search database that could be searched much faster. Several other large search companies also formed to do similar work, using advertising on their pages to support the services. These types of systems are now generally known as *search engines.*

Security, privacy and ratings

By now there were malicious people online who took pleasure in cracking computers – breaking in to them. The term hacking, which is discussed in several chapters of this book, is a different thing; a very good programmer is a hacker, a cracker is a thief or vandal. To protect against such electronic attacks, there are various techniques including the *firewall* – a special computer sitting between the Internet and the organisation, with the task of checking access in both directions to try to prevent illegal operations.

Any computer on the Net can read IP packets that are transmitted through it, which are hence not private. This lack of privacy is even worse for packets on a LAN, such as Ethernet, where every computer connected to the cable can see every packet sent. Applications above IP thus need encryption for privacy (like SSL for Web traffic) and the most important application of these is for e-mail. A US programmer called Phil Zimmerman created a program called Pretty Good Privacy (PGP)[72] that allowed individuals to securely encode their e-mail such that even the most powerful military computers could not read the contents. PGP has become the de facto worldwide e-mail encryption standard, despite the US

government trying and failing to prevent people using it outside the USA. The issue of encryption is discussed in more detail in Chapters 4 and 7. **[BRAZIL 4]**

E-mail and other Internet applications are generally built on top of TCP/IP, with no particular knowledge of who may be the person actually requesting the service. This means that e-mail or news can be faked easily, by installing a mail program, and setting it to use any outgoing e-mail address. However, experts can use the route the forged mail takes across the network to work out the likely source machine of the e-mail. This is how most spammers are found.

Some of these problems may be resolved, to provide service such as e-mail that cannot be traced, and e-mail that has a guaranteed known sender. For example, anonymous e-mail can be provided over the Internet, through the use of anonymous e-mail forwarders. These accept e-mail, rewrite the headers to remove the identification of the sender, and then forward the message to the intended recipient; this may also happen in reverse, so both people may communicate anonymously. PGP allows e-mail to be marked with a digital signature – a guarantee that the message was sent by the person with the PGP certificate. The recipient can check this if they already have information about the senders'PGPcertificate.

Advanced Web and Internet technology

The Internet Society, founded in 1992, oversaw the IETF and related bodies as well as encouraging development of the Internet around the world. The IETF concentrates mostly on protocols and lower level standards development. The Web, however, had no such body, and by the mid-1990s there were growing problems. Netscape and Microsoft (mostly) were adding features to HTML. Each browser version produced without consensus caused major problems for Web site authors. In 1994, the World Wide Web Consortium[73] (W3C) was founded at MIT in the US and CERN (later replaced by INRIA, France) to promote common specifications and cooperation. The W3C is an industry consortium, and the founder companies included both Microsoft and Netscape, as well as many other large computer companies. (In 1999 the membership was well over 100 organisations.) The W3C soon set to standardising HTMLand related standards, as well as developing new ones, so Web authors could begin to rely on tags remaining at least mostly consistent. This consensus led to the many new standards, including HTML 2, 3.2 and 4. More recently the W3C has designed style sheets, which

allow the look of Web pages to be separated from the content, and Extensible Markup Language (XML)[74] which will be the basis for future HTML and Web markup, in which there will be no HTMLs.

Netscape as well as extending HTML added *active content* to Web pages by licensing the Java language[75] from Sun in May 1994, and adding it to their browser. Java allows interactive Web pages to be created where the reader can run and use the Java programs running on the page. These programs are called *applets*. Java programs can run on any computer since they run in their own 'virtual computer' and are not compiled for any particular processor.

Netscape also developed a scripting language called JavaScript for use within Web pages in December 1995. This allowed HTML authors to add actions when the user interacted with the Web page by, for example, moving the mouse over an image. Microsoft introduced their own proprietary technologies to perform similar things – ActiveX and Visual Basic Script – which, however, only work on MS Windows systems. These are not used widely in Web pages on the Internet, since they requires browser users to modify their software which is at best inconvenient. Understandably, these technologies are used much more extensively inside Microsoft's Web server, although that is a minority of all Web servers.

The vast increase in users and traffic on the Internet meant that eventually the HTTP protocol dominated bandwidth and use. This led to the development of Web *proxies* or *caches* which fetch Web pages on the user's behalf, and then store copies of popular files so that they are not fetched repeatedly. Such storage can save considerable traffic and money. The most popular proxy is the free software Squid,[76] although there are several large commercial systems. All HTTPaccesses are recorded or logged by the proxies. There are clear privacy issues here, which are discussed in detail in Chapter 4. Logging may also be carried out at the remote sites, and for some Web sites this is a problem, since caches hide or anonymise their customers, and ignorance of a caller's identity may have a real financial penalty.

By the late 1990s schools and children were using the Web significantly. Some caregivers demanded content ratings of Web pages (by the authors or by third parties) so they could select categories of Web page content types suitable for children in their community. The W3C created a format called the Platform for Internet Content Selection (PICS) for these descriptions, but the software that applies these rules, generally called *censorware*, has also to filter Web pages for particular words since PICS is used on very few Web sites. Censorware software

has considerable pitfalls, such as tending to censor innocent terms that happen to be close to a forbidden word, such as might occur in descriptions of travels around Essex or to Scunthorpe, real places in the UK. **[SWEDEN 2]**

Large organisations now use the open Internet and Web technologies internally, rather than older proprietary systems. An internal IP-based network is called an *intranet*, but is generally closed to the outside world. It may be found useful to make some part of the private network visible outside the organisation, and this is generally known as an *extranet*. An extranet may link parts of one organisation or multiple organisations; this will usually need some security since the connections are running over the public Internet. This is done by use of a Virtual Private Network (VPN) in which an encrypted 'tunnel'is formed over a TCP/IP connection; all traffic between the organisations then passes through them, via gateways or proxies.

Broadcasters have begun to use the Web to transmit video and audio as well as Web pages with static pictures. They mostly have used RealAudio[77] for sound, but a variety of video standards are used including RealVideo. A camera pointing at the action and putting the images directly on a Web page is also popular. Such a camera is called a *Webcam*. Musicians (and, later, publishers) started experimenting with digital online music published directly to fans, usually in RealAudio or a new format called MP3[78] which made the data containing songs small enough in size to transmit online. Multimedia content could also be added to Web pages by using technology such as Java or ShockWave, which allows the creation of multimedia 'shows'or performances running inside a Web page.

The free software community expanded with companies like RedHat[79] selling millions of copies of Linux in shrink wrap form, the creation of Web sites such as freshmeat[80] and slashdot[81] keeping the community together, and the new idea of Open Source.[82] This was invented by Eric Raymond[83] (Raymond, 1997) and is free software, but presented in a way that could be understood more easily by the commercial software world. This influenced Netscape, who made their new browser into free software in 1998, codenamed Mozilla.[84]

Several early applications evolved or spawned newer versions. The POPprotocol was joined by variants and by a new one, Internet Message Access Protocol (IMAP), that was more useful for dialling in and manipulating e-mail stored on a server. IRC attracted a new commercial competitor ICQ,[85] a chat system that allows a user to log into a server and be signalled when friends arrive and depart.

Finally, all the big companies such as Yahoo!, the search engines, and the commercial sites realised that they were really providing entrances to the Net – most people started at one specific site when they looked for things online. Large Web operations had become entrances, or portals to the Web. Yahoo! developed a 'my' customisation concept for its portal, adding many services, and it has remained a very popular Web site. The browser manufacturers realised that adding searching systems to the software could direct people to their own portals, and expanded accordingly. Most of the portals today offer some additional content such as news, a Web search service, and customisation.

Future Internet technology

These technologies will continue to grow in use in the next few years. The next version of IP, Internet Protocol version 6 (IPv6), is designed for a much larger Internet, including bigger IP addresses, more data and encryption in the packets. More applications of metadata on the Web are likely, including Dublin Core[86] format written in the Resource Description Framework (RDF)[87] format, an application of XMLfor better Web searching, digital libraries, e-commerce and multimedia applications. Dynamic HTML(DHTML), HTMLtogether with JavaScript and style sheets, offer authors ways to rewrite Web pages on a browser when the user interacts with them. There are more advanced search engines in development, including those such as AltaVista that support language translation, and AskJeeves[88] that allow users to narrow searches by answering questions.

It is quite impossible to keep up with these technologies, since major new applications can appear within weeks or a few months. There will undoubtedly be novel Web systems and expansions which have appeared since this chapter was written in June 1999.

Comments

SINGAPORE

1

In Singapore, the Internet started with its introduction at the National University of Singapore some time in 1991. However, at that time, it was entirely for research and development use. By 1992, the Singapore government formed the TechNet Unit and promoted its use among R&D and educational organisations. In 1994, the first Internet Service Provider was set up by Singapore Telecommunications Ltd. This ISP is known as SingNet. By 1995, it became obvious that the Internet was really taking off in a big way. The Singapore government decided to sell off the TechNet Unit. This new entity became the second Internet Service Provider called Pacific Internet, or PacNet for short. Subsequently, a third Internet Service Provider, CyberWay, was set up. In 1996, the Singapore government announced the set-up of the high speed, broadband, multimedia network called Singapore ONE (One Network for Everyone). This high speed network is based on Internet and ATM (Asynchronous Transfer Mode) technologies. Its pilot project was officially launched by the Singapore Prime Minister on 9 June 1997.

In Singapore, it is the government that takes the lead in introducing and funding the information superhighway.

BRAZIL

1

In Brazil, in the mid 1980s, academic computer networks were linking big universities and research centres in the cities of Rio de Janeiro, Sao Paulo and Porto Alegre to the United States. But these networks were independent and were not working together. At the end of the decade, in 1989, all the efforts became coordinated by the foundation of RNP (Rede Nacional de Pesquisas – National Network of Research) and the creation of the national computer network backbone.

From 1991 to 1993, the work was focused in the development of the backbone and by the end of the third year had 11 of the 23 states connected in speeds from 9.6 to 64kbps. At the same time, the Internet services available were being explained to the academic community, through seminars and training showing their strategic value to the country, while the computer communication addicts started the big BBS boom in the country. The second part of the Internet development in Brazil happened from 1994 to 1996, concerning the need for increasing the backbone speed and reliability due to the great number of institutions connected to the Net. In addition a big change was being noticed: people used to Gopher and FTP started to use the new World Wide Web. In May 1995 the commercial Internet started to work in Brazil bringing new frontiers to be explored.

The current phase started in 1997 and brought infrastructure developments such as increasing the number international connections to five and promoting high speed metropolitan networks (MANs). The first tests with the Internet2 project have just happened (April 1999) and were successful.

Author response

Internet2 is, despite the rather arrogant name, a project to connect mostly American universities via newer, faster network technology. There is no 'second Internet' under construction.

The growth of the Internet on the planet was really amazing. From 213 Internet hosts in August of 1981 to 43.2 million in January of 1999 (source: Network Wizards). Meanwhile, what happened in Latin America? As we all know, big changes take time to arrive and settle in the developing and third world countries. Cultural, political and economical issues are responsible for that. But as the world is being changed and the 'Global Village' is more and more a reality, technology is coming faster and these countries already have an important role in cyberspace.

Take Brazil as an example. There were more than 215 000 Internet hosts as at January 1999 (ibid.) in the country, which makes Brazil the world's 17th largest country in number of hosts and the first in Latin America. The second place is Mexico showing around 112 000 Internet hosts. If we look at South America, Brazil has a big lead. The second place in number of hosts is Argentina, with about 66 000 hosts.

Countries in Latin America have a great Internet growing potential. Internet users in Brazil represent only 2.1 per cent (source: NUA Internet surveys) of the population and in Mexico 0.5 per cent (ibid.). Comparing to Canada (5th in number of hosts in the world) and Denmark (14th place), that have respectively 26 per cent and 22 per cent of Internet users among the population. The Latin American countries have a lot to show. As development countries become more developed and more technological, these numbers should increase. In the next 5 years we should see a Latino country in the world's top 10 countries in number of Internet hosts and users.

BRAZIL

2

When you are looking for a company on the Net, what is the first thing you do? Well, you guess the URL, like www.company.com. If you are looking for a company in a certain country, just after the TLD (Top Level Domain), you add the country domain, as in www.company.com.br.

But there is not an international standard for TLDs and it is becoming more and more difficult to guess. Some countries adopt different number of letters for TLDs acronyms, as in South Africa (co.za, gov.za, school.za) and also some countries adopt their own language to build the TLDs, as in Brazil (psi.br for Internet service providers, lel.br for auctioneers, esp.br for sports) and Mexico (gob.mx for government).

BRAZIL

3

The Internet is not American. The Internet doesn't belong to anybody or any country. Once PGP was released on the Internet it was already part of it – the Internet that knows no borders. How could it be possible to control something like that on the Internet?

BRAZIL

4

Maybe it should be mentioned here that a) credit card numbers are not secret, and knowledge of one is not sufficient proof that you have a right to bill the corresponding account and b) numbers can also be stolen by waiters, hotel receptionists and so on.

SWEDEN

1

There are other pitfalls of such screening. A friend of mine just got an e-mail censored because its subject was 'examination in quality criteria', which in Swedish contained 'kvalitetsexamen'.

SWEDEN

2

AFRICA **1**

Unlike their counterparts in the Western world, the African academic community did not make any known contribution to the Internet development in its early stages. A major role they played within the African context is in the adoption of the Internet in their respective countries. For instance, in South Africa, an Internet user community was first established some years ago within the academic community, who were the only ones with local access to full Internet. The academic network was called UniNet. A number of store and forward dial-up services operated for the non-govermental organisation (NGO) community and computer enthusiasts, and some businesses used CompuServe (before it had full Internet). As more of these services sprung up, UniNet relaxed its membership policies a little, allowing the NGO e-mail service provider (SANHGONet) to become the first non-academic or government system on full Internet in South Africa (Jensen, 1996).

At the continental level of academic network, the AAU (Association of African Universities) has been playing a significant role. The AUU presently links 119 universities in 42 African countries.

For lack of adequate telecommunications infrastructure, less than 1 percent of the world's Internet traffic currently reaches Africa: the telephone network hardly exists. Compare Sweden (with 68 telephone connections per 100 inhabitants), the USA (with 57), and the Netherlands (49) on the one hand to Zimbabwe (with 1.22), Ghana (0.3), and Chad (0.07) on the other. Not all Internet traffic travels via the public telephone network. Very busy routes are served by lines with very large throughput capacity, which are dedicated to Internet traffic and usually leased from the public network. But these backbones of the Internet are found nowhere in Africa except South Africa. To obtain a fast Internet connection in Kenya or Cameron, you need a leased line to the UK, France, or South Africa. Hegener (1995) reported Donald Ekong, the AAU Executive Secretary, as observing that African scientists often hardly know what their colleagues at other universities are doing, and they lack the capability to disseminate their own research results throughout Africa. The extreme shortage of up-to-date academic publications at practically all African universities could be remedied with a robust Internet link, a powerful printer, and a plentiful supply of paper.

AFRICA **2**

Bandwidth is measured in number of bits per second (bps). As the bandwidth capacity is commonly in thousands of bps, the standard units of measurements are in Kbps (Kilo bps), Mbps (Mega bps), Gbps (Giga bps), and so on. Bandwidth usage for Internet ranges from the 9.6Kbps, 14Kbps, 24Kbps, 64Kbps, 128Kbps, 256Kbps, ...2.8Mbps.

As the bandwidth capacity is directly proportional to its costs, in most African counties, bandwidth capacity usage for Internet purposes is often limited to the lower range.

AFRICA **3**

The dial-up (also called store and forward) Internet connection is the common starting point for Internet development in African countries. In fact, it remains so up till present time. Fidonet, Healthnet, Esanet, and other similar NGOs do effectively use the dial-up Internet connections for institutional electronic networking purposes.

AFRICA **4**

The fact that Ethernet is a relatively cheap and easy to install network technology has made it one of the popularly used network protocols in African countries.

Further African Comments

Further comments on this chapter from the African commenting author, Sunday Ojo, were much longer than most other comments in this book. Like those following Chapter 6, from Terrell Ward Bynum, they form an entire essay. However, the issue of Africa and the Internet is relevant not only to discussions on Africa itself, but is equally applicable to consideration of other Third World countries and the Internet. Dr Ojo's commentary is therefore included here in full.

African Commentary

Sunday Ojo on Chapter 2

It is a well-known fact that the utility of the Internet depends to a great extend on the quality of the underlying telecommunication infrastructure. Unlike the economically advanced world, the poor quality of the network still remains a basic impediment to rapid development of the Internet in Africa.

Generally, African countries, except for a few, are experiencing rather slow extension and modernisation of their telecommunication networks, Sub-Saharan Africa's (excluding South Africa) teledensity has continued to remain at less than 1 per 200 inhabitants, most of the telecommunication network is analogue and many sections are highly unreliable, especially during the rainy season. Despite the poor telecommunication infrastructure, in recent times, 94.5% of the 54 African countries presently have some level of Internet development.

As at June 1999, 51 out of 54 African countries have Internet access in the capital cities, 3 have no local Internet access, 7 have only one public access ISP, while 13 have local ISPs or points of presence (POPs) in some secondary towns. Dial-up Internet access remains the commonest in most African countries. As at March 1999, for the whole of Africa, dial-up Internet accounts stood at 428075, international bandwidth at 114454 Kbps, public access ISPs at 319. South Africa is counted among the top 20 countries worldwide for the number of Internet hosts. If South Africa is excluded from the African Internet statistics, the rest of Africa put together is still behind in Internet developments. For example, of these figures, South Africa alone accounts for 46%, 30%, and 78% of the Internet accounts, international bandwidth, and public access ISPs, respectively. The Internet density (population/user) for Africa as a whole is 1819, while that of South Africa alone is 4123 (AIC 1999a, 1999b).

The costs of Internet access vary widely from one African country to another. Monthly subscriptions for dial-up Internet access with ISPs average US$53.49 and range between minimum US$13.97 in South Africa to maximum US$197.74 in Djibouti (AIC, 1999b). This means that in some countries, even if a computer is available, the service is beyond the reach of all but the top elite. Also, because telephone call charges to the service provider are usually the major cost, the absence of a national service effectively cuts off the majority of the population from the Internet. Reflecting the high cost of full Internet-based services, and also because of the overriding importance of electronic mail, the small e-mail-only store and forward systems with dial-up connections to the Internet are generally continuing to attract subscribers.

In recent times, some African public telecommunications operators, PTTs, have started to establish Internet services. For example, PTTs in Benin, Central African Republic, Djibouti, Mauritius, Madagascar, Senegal, Botswana and South Africa have recently brought full Internet services on stream.

The majority of international connections to the rest of the Internet operate on analogue circuits rated at 9.6Kbps, but often pushed to 14Kbps and sometimes to 24Kbps or even 64Kbps. None of the countries outside South Africa had international circuits larger than 64Kbps until

very recently when Tunisia and Egypt upgraded to 128Kbps. By contrast, South Africa has a half dozen international links, most of which are over 256Kbps and some are up to 2.8Mbps. While some Internet circuits in Africa connect to the United Kingdom and France, (as well as one to Italy), the majority connect to the USA where suppliers include AT&T, Global One, UUNET/AlterNet, MCI, NSN, Sprint and BBN. Nevertheless, France Telecom/FCR has more Internet connections into Africa than any other single supplier, largely because of its close ties with Francophone PTTs.

Aside from the marine optical fibre link in South Africa and Djibouti, which has access to the SEA-ME-WEA cable, most of the other international connections are carried via satellite, except for the countries having borders shared with South Africa. For these countries, the lower-cost terrestrial links and historical ties have resulted in all of the ISPs there connecting to the South African infrastructure.

Because of the high cost and low international bandwidth available in many African countries, increasing attention has recently been drawn to the possibility of using satellites for Internet services using VSAT. It offers reasonably high bandwidth (64K–2Mbps) and substantially lower costs than most PTT-supplied international leased circuits. Existence of regulatory barriers has stymied most attempts to use this technology so far, except in a number of countries including Botswana, Ghana, Uganda and Zambia, where the telecoms market has been substantially relaxed. As a result there are a number of countries with VSAT-based Internet Service Providers, for example InfoMail and Starcom in Uganda, and one each in Ghana (NCS) and Zambia (ZamNet).

The recent availability of the higher-powered KU-band satellite footprint in southern Africa, and the prospect of other KU bands being directed at Africa shortly, further improves the potential for VSAT. There are other satellite-based communication systems being planned that are expected to radically improve access from the most remote areas of the continent. However, the costs are unlikely to be within the reach of the average African citizen.

Most countries in Africa have some form of local or internationally-hosted Web server with varying degrees of comprehensiveness, but the quantity of information is generally very limited.

Of particular importance for Internet developments in Africa has been the establishment of national cross-sectoral Internet working groups comprising actual or potential Internet access providers, users, telecommunications operators and government. These groups have been formed in Angola, Ethipia, Gabon, Gambia, Namibia, Sierra Leone, South Africa and Tanzania. On a related note, in East Africa, the East African Internet Association (EAIA) has formally been registered. It is the first regional grouping of Internet Service Providers, collaborating to improve their service, share resources and ultimately to set up an international hub to share leased line costs.

Donor agencies including NGOs and multilateral organisations are playing major roles in Internet development in Africa. For example, in Zimbabwe in 1990, a co-operative of local and international NGOs established MANGO, the first store-and-forward e-mail gateway to the Internet in sub-Saharan Africa outside South Africa. Used as a model for many of the other store-and-forward systems in Africa, the non-profit service grew quickly to the point where there are now over 250 users accessing the service through a single phone line using the highly efficient FIDO protocol. Despite the recent emergence of full Internet services, MANGO continues to operate successfully as a low-cost alternative for those whose primary requirement is e-mail. In November 1995 a collaboration between BellaNet, UNESCO, IDRC, UNECA and ITU, called the African Network Initiative (ANI), made a study on future information-infrastructure building activities in Africa. It identified a substantial number of ICT-related development projects being planned or in process in Africa. These and over 100 other finalised projects were identified during the study, with others added subsequently in preparation for this report, indicating an unexpectedly high level of activity in this area from the international community (Jensen, 1996).

URLs

1. http://www.att.com/
2. http://www.bbn.com/
3. http://www.bell-labs.com/
4. http://www.lucent.com/
5. http://www.fidonet.org/
6. http://www.irc.net/
7. http://www.sendmail.org/
8. http://www.sendmail.com/company/manage.html#CTO
9. http://www.iana.org/postel/
10. http://www.iana.org/
11. http://www.isoc.org/
12. http://www.ietf.org/
13. http://www.gnu.org/people/rms.html
14. http://www.gnu.org/
15. http://www.gnu.org/copyleft/copyleft.html
16. http://www.gnu.org/fsf/
17. http://www.gnu.org/philosophy/free-sw.html
18. http://www.opensource.org/bsd-license.html
19. http://www.mckusick.com/
20. http://www.x.org/
21. http://www.opensource.org/bsd-license.html
22. http://www.xfree86.org/
23. http://www.ssc.com/linux/linus.html
24. http://www.linux.org/
25. http://www.freebsd.org/
26. http://www.openbsd.org/
27. http://www.idg.net/metcalfe/
28. http://www.xerox.com/
29. http://www.3com.com/
30. http://www.infoworld.com/
31. http://www.isc.org/bind.html
32. http://www.sun.com/
33. http://www.sun.com/corporateoverview/ceo/mgt-mcnealy.html
34. http://www.sun.com/corporateoverview/ceo/mgt-joy.html
35. http://www.oreilly.com/
36. http://www.oreilly.com/oreilly/tim_bio.html
37. http://www.uiuc.edu/ph/www/krol
38. http://www.oreilly.com/catalog/twi2/
39. http://www.w3.org/
40. http://www.w3.org/People/Berners-Lee/
41. http://www.cern.ch/
42. http://www.w3.org/Protocols/
43. http://www.w3.org/MarkUp/

44. http://www.w3.org/Addressing/
45. http://www.ncsa.uiuc.edu/
46. http://hoohoo.ncsa.uiuc.edu/
47. http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/
48. http://hoohoo.ncsa.uiuc.edu/cgi/
49. http://www.perl.com/
50. http://www.wall.org/~larry
51. http://www.yahoo.com/
52. http://home.netscape.com/company/about/backgrounder.html#execteam
53. http://www.netscape.com/
54. http://www.wired.com/wired/
55. http://www.hotwired.com/
56. http://www.apache.org/
57. http://www.aol.com/
58. http://www.compuserve.com/
59. http://www.microsoft.com/
60. http://www.msn.com/
61. http://www.microsoft.com/BillGates/
62. http://www.microsoft.com/ie/
63. http://www.networksolutions.com/
64. http://www.icann.org/
65. http://www.amazon.com/
66. http://www.geocities.com/
67. http://www.dell.com/
68. http://www.hotmail.com/
69. http://www.cnn.com/
70. http://www.lycos.com/
71. http://www.altavista.com/
72. http://www.pgp.com/
73. http://www.w3.org/
74. http://www.w3.org/XML/
75. http://java.sun.com/
76. http://squid.nlanr.net/
77. http://www.realaudio.com/
78. http://www.mp3.com/
79. http://www.redhat.com/
80. http://www.freshmeat.net/
81. http://www.slashdot.org/
82. http://www.opensource.org/
83. http://www.tuxedo.org/~esr/
84. http://www.mozilla.org/
85. http://www.icq.com/
86. http://purl.org/dc/
87. http://www.w3.org/RDF/
88. http://www.askjeeves.com/