TECHNOLOGICAL UNDERPINNINGS: COMMUNICATIONS AND NETWORKING

Rudolf Nottrott

National Center for Ecological Analysis and Synthesis, 735 State St., Suite 300,
Santa Barbara, CA 93101

Abstract. At most biological field stations there are few formal provisions for on-going data exchange after individual investigators have returned to their home institutions. However, opportunities for such long distance collaboration and information exchange have recently increased with the development of wide area network technology. Wide area networks have the potential to change the culture of collaborative research in ecology. To familiarize ecologists working at biological field stations with the mechanics of Internet communication and data exchange, this chapter provides a brief review of the history of electronic networking; the architecture, protocols and common client/server applications of the Internet; and basic network security issues.

INTRODUCTION

Ecologists have recognized an increasing need for long-distance collaboration, rapid communication, and increased data access. Biological field stations, which may host hundreds of scientists and research studies over the course of decades, have a clear need for data archiving and access to those data by geographically and temporally dispersed researchers. Once biological field stations and other research institutions have established standards for metadata, protocols, and network software, archiving of long-term data (e.g., relating to the history of a site) for efficient retrieval will be possible. In this chapter, key aspects of network infrastructure such as network functions, hardware, client/server mechanisms, and security are reviewed.

HISTORY

1989 and before - a jungle of networks

Until 1989, there were few opportunities for ecologists to utilize wide area networks for data exchange or collaborative research. The myriad of incompatible networks with variable longevity (Frey and Adams 1989) made it impractical for scientists, both nationally and internationally, to use this technology for information exchange. However, once the potential significance of wide area networks for data exchange between geographically dispersed researchers was recognized, demand for enhanced network functions and increased network access ensued.

A 1989 survey of 18 Long-Term Ecological Research (*www.lternet.edu*) sites (with more than 500 widely-dispersed researchers) indicated three primary needs related to increased network capabilities (Brunt et al. 1990):

- **Local Area Networks** (LANs: Ethernet, Appletalk, PC Networks, etc.) - Resource sharing: files, programs, printers; high-speed links to higher level networks
- **Institutional Networks** (including campus networks) - Access to mainframe computers, Wide Area Networks, and files and printers on other LANs
- **Wide Area Networks** (WANs) - Instantaneous and reliable email; access to supercomputers and national information and software repositories; rapid long-distance transfer of data and information (e.g., graphics and other binary information); electronic infrastructure for long-distance collaboration.

THE INTERNET

Beginning as early as 1969, but accelerating exponentially around 1989 (Rutkowski 1997, Leiner et al.1997), the growth of the Internet, and its current position as a *de facto* global standard, has now made it feasible for widely distributed researchers to utilize wide area technologies. The key elements for the Internet's success were the openness and expandability of the Internet protocols, and their scalability from Local Area Networks (LAN) to global Wide Area Networks (WAN). Although Internet access is not yet truly global, it currently is widely available in the U.S., Europe, Japan, Southwestern Australia, and parts of South America and South Africa (Landweber (*ftp://ftp.cs.wisc.edu/connectivity_table*), and Matrix Information and Directory Services (*http://www.mids.org/mapsale/world/index.html)*).

Definition of the Internet - the foundation of TCP/IP

The Federal Networking Council (Federal Networking Council 1995, *http://www.fnc.gov/*) defines the term Internet as the global information system that:

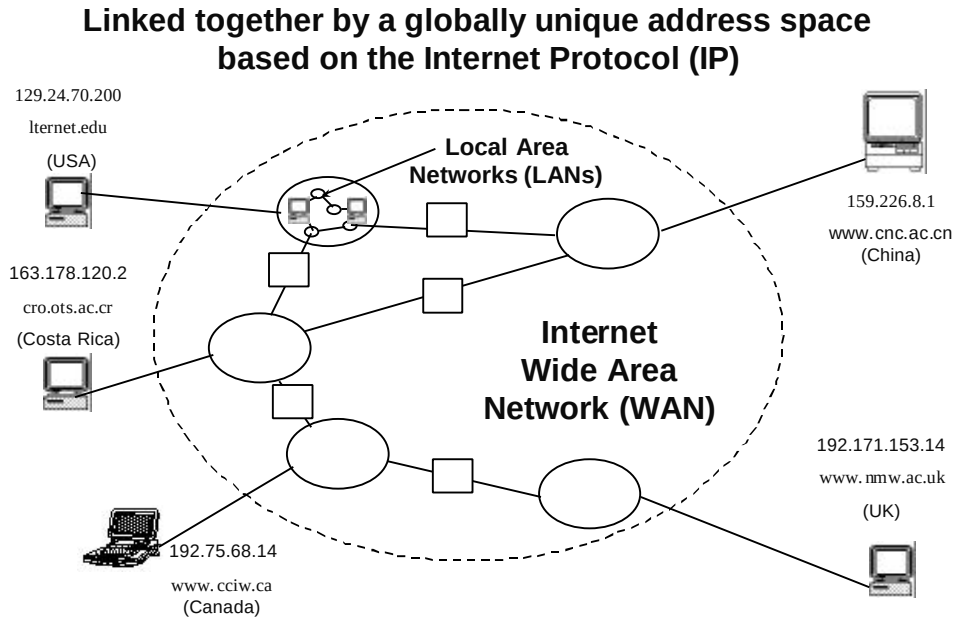
(i) "is logically linked together by a globally unique address space based on the Internet Protocol (IP) or its subsequent extensions/follow-ons;

(ii) is able to support communications using the Transmission Control Protocol/Internet Protocol (TCP/IP) suite or its subsequent extensions/follow-ons, and/or other IP-compatible protocols; and

(iii) provides, uses or makes accessible, either publicly or privately, high level services layered on the communications and related infrastructure described herein."

The global address space of (i) is illustrated in Figure 1 which shows five computers (Internet hosts) on four continents connected via the Internet.

Figure 1. IP global address space.



Each computer is identified by a unique address, called IP number. For better readability, IP numbers are usually shown as four sets of decimal numbers separated by periods (e.g., 128.85.36.9), but they simply represent 32-bit binary numbers, allowing for 2^{32} = 4,294,967,296 computers. In practice the number is smaller, because blocks of IP numbers are reserved for various technical reasons (Hunt 1992). Also, organizations are allocated whole blocks of numbers (usually 256 or 65,536 numbers), which they can use at their discretion. Some argue that present trends indicate a leveling off in the number of host computers at approximately 38 million hosts around the year 2002 (Hilgemeier 1997)**.** However, the future internet may well have 128-bit IP numbers, to avoid the bottleneck of address shortages, and thus keep growing into the foreseeable future.

To further simplify use of the system, IP numbers are commonly represented in the form of hierarchical domain names, e.g., LTERnet.edu instead of 129.24.70.200, and Domain Name Service software is used to facilitate the conversion to numeric IP addresses. The basis of the IP protocol is that all information sent over the network is in small packets (e.g., 1000 characters) complete with destination and sender IP numbers plus other data (e.g., sequence number) needed during and on delivery. The packets may arrive at their destination in arbitrary order, but software on the destination computer can put them back together as needed, using the sequence number. Imagine a colleague in Australia sending you a 300-page story in sequentially numbered, one-page letters, at a rate of one per day. After nearly a year, you compile them to get the full story. Fortunately, the Net is faster than that.

Layers and protocol stacks

Network architects conceptualize, design and implement their network software in what they call “layers.” For the Internet, the layers represent protocols including IP and TCP/IP. In schematic diagrams the software layers resemble stacks of bricks. Hence, they are often called protocol stacks. Before protocol stacks came preinstalled with most computers, one would have to install them before attempting a connection to the Internet (the Trumpet Winsock stack is a well-known example for Windows® 3.1). Figure 2 illustrates the Internet Network layers and how they relate to the Internet definition described above (this is a special case of the ISO/OSI reference model as detailed in Hunt 1992).

Figure 3 illustrates the same layers with an e-mail handling program at the top application layer. The Internet Layer, IP, corresponds to (i), the Host-to-Host Transport Layer, TCP/IP, corresponds to (ii), and the Application Layer, Telnet, SMTP, FTP, HTTP, correspond to (iii). Most network users work at the level of the “high-level services” of the Application Layer, such as Telnet, FTP or HTTP (through Web browsers). By examining some examples of high-level applications, and considering how layers pass data back and forth, the mechanics of the Internet become clearer.

Clients and servers - present-day workhorses of the Internet

The example of e-mail delivery in the previous section illustrates another concept that has found widespread use in software architecture for network-distributed applications - client/server architecture. A client is generally a software program that requests a “service” from another program called server. In the example of e-mail delivery, the client might be a program such as Eudora®, pine, Microsoft Outlook® or the original Unix program called ‘mail’, all available for many different kinds of operating systems and computer platforms. It's server counterpart has historically been a program called ‘sendmail’ running on Unix machines (alternative mail servers are now available). The client and the server communicate using standard keywords and formats, which are called a “protocol.” In the case of e-mail delivery the protocol is called Simple Mail Transfer Protocol (SMTP). Mail delivery using this scheme is not unlike a Telnet session

Figure 2. Internet network layers.

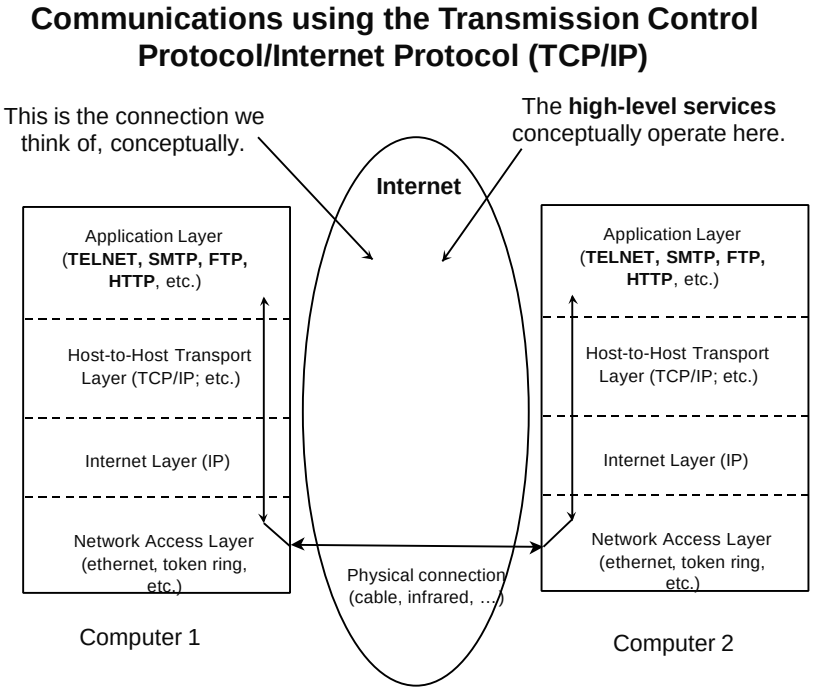
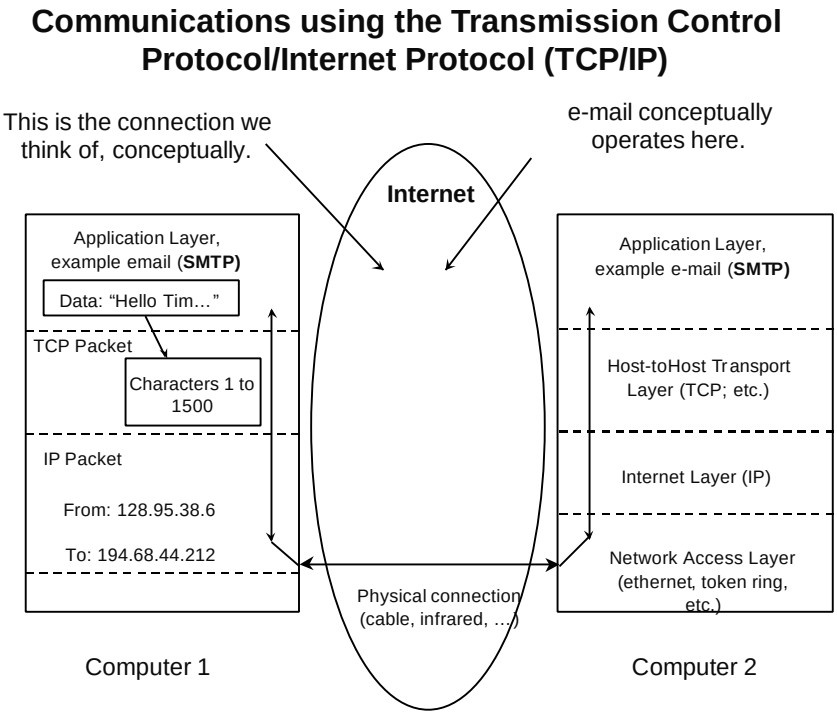


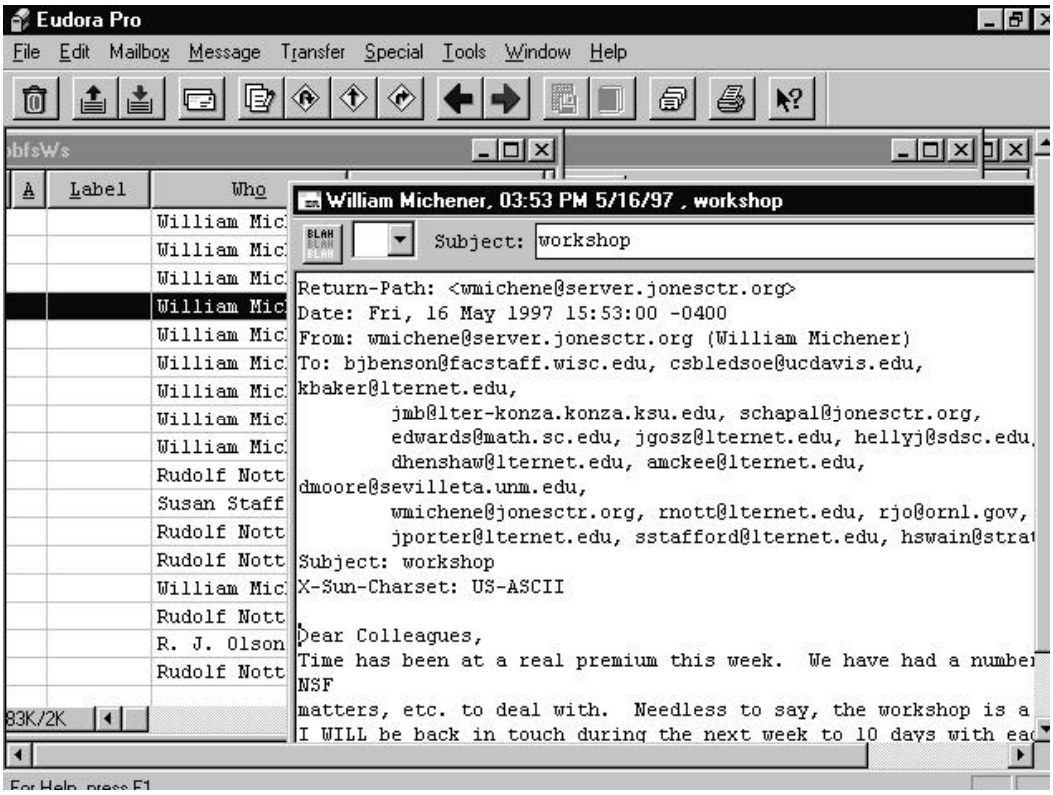
Figure 3. Internet network layers with an e-mail handling program.



(although in practice a software mechanism called “sockets” is used, with Winsock and BSD Unix sockets being most common), and it is indeed possible to ‘talk’ by telnet directly to e-mail servers, as well as many other servers (such as Web, WAIS and News servers).

Naturally, people have come to expect much more user-friendliness, and consequently modern clients hide the protocol exchange behind a façade of windows and menus, as Figure 4 shows for the Eudora® mailer client. With the widespread use of desktop workstations, clients and servers for different application areas are becoming increasingly common. Table 1 gives an overview of the most common types of application protocols. Most Internet protocols are described in Request for Comments (RFC *http://ds.internic.net/rfc/*). A very comprehensive list of Winsock clients, together with reviews and source links, can be found at *http://cws.internet.com*.

Figure 4. Eudora windows.



NETWORK SECURITY

Security of data and other information on wide area networks is a key concern among scientists (Brunt et al. 1996). However, in most circumstances, solutions are available to ensure security of data and information, within reasonable limits. Most of the original Internet applications were developed by engineers and academics with little need for security, and only recently have commercial applications such as bank transactions and online sales necessitated development of extremely secure network applications (including features such as encryption). With the growth of the number of Internet hosts to tens of millions, the atmosphere has changed from that resembling a small town, where few residents lock their doors, to that of a big city, where some doors may need dead-bolts and chains. It is important to keep in mind that network

growth is good, and security consciousness is a small price to pay for the increased services that have come with network growth.

Table 1. The most common client/server pairs at the Internet application layer

Application Layer Protocol	**Common Name**	**Example clients**	**Example servers**
SMTP, Simple Mail Transfer Protocol	e-mail delivery	Eudora®; MS Outlook®; Pegasus®; mail (built-in on most Unix systems)	Sendmail, built-in on most Unix systems
POP, Post Office Protocol	e-mail pickup box	Eudora®; MS Outlook®; Pegasus®	POP3
FTP, File Transfer Protocol	Ftp	ws_ftp; built-in on Unix systems	ftpd - built-in on most Unix systems
NNTP	News	Free Agent®;	Nntpd
TELNET	Telnet	Ewan; Built-in on most Unix systems	Telnetd - built-in on most Unix systems
HTTP	WWW, Web	Netscape Navigator®; MS Internet Explorer®	Httpd from NCSA
	Video conferencing	CU-SeeMe®	Reflector
Gopher	Mostly superseded by the Web		
ODBC, Open DataBase Connectivity	ODBC database Access	MS Access®; Excel®	Oracle®, Ingres®, MS SQL server

Sources for security information

The two most widely used operating systems with built-in TCP/IP capabilities are Unix and Windows NT (Windows 95® was developed with much lesser network capabilities, mostly for use in proprietary LANs). Because Unix is a much older and more mature operating system, several organizations (e.g., Computer Emergency Response Team, *http://www.cert.org* and the Internet Society, *http://www.isoc.org*) have considerable experience with Unix security issues, and a wealth of literature is available on Unix system security.

Recently, numerous books dealing with NT security issues have become available. An easy way to find the latest books is to do an online search at one of the electronic bookstores. For example, a search at *http://www.amazon.com* for "windows and security" retrieved 9 items, including Rutstein (1997) and Dalton et al. (1997). Similarly, a search for "unix and security" returned 7 items including Garfinkel and Spafford (1996). In addition, several newsgroups have been established to discuss Unix-related security issues (comp.security.unix), NT (comp.os.ms-windows.nt.admin.security), and miscellaneous other security issues (all in comp.security).

Simple precautions

Simple measures can help prevent most security problems:

- Choosing safe passwords is the simplest part of network security. Bad passwords cause >80% of all security problems (RFC 1244 1991); change your password at regular intervals!

- Perform the following activities on a regular basis (or choose a good systems administrator who will do it for you): use password software to make sure passwords are "good"; perform password aging; get your software from trusted sources only; keep software updated.
- Keep an eye on your system (unusual files, processes, login activity, etc.).
- Possibly limit access (allow logins only for certain machines, or domains).
- Use programs that help with security checks (e.g., COPS).
- Consider the use of encryption.

Gated Communities - firewalls and intranets

With the exponential expansion of the Internet, some organizations with strict security needs have partially separated themselves from the Internet. They have done this by using a TCP/IP-based LAN internally, called an Intranet, which is connected to the Internet through a separate machine running "firewall" software. Firewalls can effectively protect your institutional network from the outside world and still allow your users access to the Internet. Firewalls obscure the internals of your Intranet from the outside world by refusing to provide name or address information about internal machines, by replacing internal users' login names with aliases (for email), by allowing FTP and other services only to/from the firewall and by allowing telnet or remote log-ins only to/from the firewall.

LITERATURE CITED

Brunt, J., J. Porter, R. Nottrott. 1990. Internet connectivity in the Long-Term Ecological Research Network (LTER): assessment and recommendations. LTER Network Office, University of Washington. Seattle, WA.

Dalton, W., S. Fuller, B. Kolosky, J. Millecan, Nachenberg, C. Goggans. 1997. Windows NT server 4: security, troubleshooting, and optimization. New Riders Publishing

Frey, D., and R. Adams. 1989. !%@:: a directory of electronic mail. O'Reilly & Associates, Cambridge, MA.

Garfinkel, S., and G. Spafford. 1996. Practical UNIX & internet security, 2nd edition. O'Reilly & Associates, Cambridge, MA.

Hilgemeier, M. 1997. Internet growth - host count turning point in June 1997. *http://www.is-bremen.de/~mhi/inetgrow.htm*

Hunt, C. 1992. TCP/IP network administration. O'Reilly & Associates, Cambridge, MA.

Leiner, B., V.G. Cerf, D. Clark, R. Kahn, L. Kleinrock, D. Lynch, J. Postel, S. Roberts, S.Wolff. 1997. Brief history of the Internet. Internet Society. *http://www.isoc.org/internet/history/brief.html*

Porter J., R. Nottrott, D. Richardson. 1996. Ecological databases: new tools and technologies. In Long-Term Ecological Research. Ecological Society of America annual meeting, Providence, RI.

Porter J., K. Baker, R. Nottrott. 1996. Tools for managing ecological data. Eco-Informa '96 Conference. Lake Buena Vista, FL.

RFC 1244. 1991. Request for comment. Site security handbook. *http://ds.internic.net/rfc/rfc1244.txt*

Rutkowski, A.M. 1997. Internet Trends. General Magic, Inc., Sunnyvale, CA. *http://www.genmagic.com/Internet/Trends/*; and *http://www.genmagic.com/Internet/Trends/slide-4html*

Rutstein, C.B. 1997. Windows NT security: a practical guide to securing Windows NT servers and workstations. Computing McGraw-Hill, New York, NY.