Computer Networks: Prospects for Scientists

Allen Newell and Robert F. Sproull

A fascinating aspect of the computer revolution is the increasing use of computer communication networks to connect together large numbers of computers and interactive terminals. These networks provide communication not only among machines but among their users as well. Computer networks also exemplify some larger themes that characterize the development of computer technology generally: the hierarchical structuring of systems based on an extraordinarily simple base; the power and usefulness arising from the integration of the entire span of computer technology; and an undisciplined, heterogeneous growth that defies simple control or even description. The end result will be the merging of communication and computation into a single technology, though that is some way ahead and unpredictable in its fine grain.

In this article we review the nature of computer networks and preview their development by focusing on the ways computer networks are useful to scientists and to science. The impact of networks extends to all segments of society, and it is doubtful that science is either the greatest beneficiary or the most affected. But the underlying technology is common to all application areas and the higher level considerations are analogous.

Network services are made available to scientists through a collection of tools that have many roles. Electronic mail, for example, is used to send messages over a network. Other tools make remote computers and special computing services accessible to a wide audience. Network tools also augment a scientist's existing computing environment by providing communications to share data in many different forms. In the last two cases, the network is but one part of a scientific computing environment: communications may enrich the environment, but the principal ingredient is the development of software tools useful to scientists.

The term "network" is used in at least three ways. First, a communications network is the interconnected equipment that provides data communications to a set of host computers. Second, because the host computers are all connected together, they are often called a "network of computers." Finally, the users of these host computers represent a network of individuals who may communicate and work together. Usually the distinction between these uses of the term is evident from the context.

Computer Networks

To begin at the bottom, computer technology is the technology of doing things with bits. Bits are so simple that they admit of only three operations: moving from one place to another, keeping in one place unaltered, or changing back and forth between 0 and 1—in short, elementary acts of communication, memory, and transformation. In a digital computer, operations on data bits are controlled by a computer program's instructions, which are themselves represented with bits and saved in an instruction memory. Everything that we call computer technology comes from building hierarchies of processes out of these simple capabilities.

Communication is an elementary computational operation; but to attain communication over large distances and through diverse media requires fitting into the established technology of telephony and electromagnetic broadcasting. The existing technology has two pervasive characteristics: it deals with continuous signals (that is, not with bits), and it connects the sender and receiver by a physical link that not only passes messages but "remembers" between messages the identities of the sender and receiver.

Although it was relatively easy to arrange to send bits rather than continuous signals, it was not until the 1960's that a technically and economically viable way was found, called packet-switching (*1*), to avoid connections. The motivation for avoiding long-duration connection was economic; messages to and from computers are short and bursty and waste communication capacity during idle periods. The solution breaks the communication into small units called packets, each of which contains data to be communicated as well as address and status. Packets are routed by switching computers, which examine a packet's address and forward it through communication channels toward its destination. This scheme dynamically allocates communication capacity, thus reducing the cost of the communication. However, much more than economics was achieved. Packet-switching creates a completely digital technology of communication that joins all communication with the rest of computation, permitting the growth of hierarchy and the integration with other advances in computation.

Computer networks provide a primitive communication operation: the ability to send from one computer to another a datagram, a message consisting of a few hundred or thousand bits of information. This primitive is provided by a network that uses more primitive operations which involve packets flowing between switching computers. It is analogous to the primitive arithmetic operations provided by a single computer, such as addition and multiplication, which are themselves composite, as testified by the structure of a floating-point

Summary. Computer networks are an integral part of the rapid expansion of computing. Their emergence depends both on evolving communication technologies, such as packet-switching and satellites, and on diverse experiments and innovations in the software tools that exploit communications. The tools provide computer users with facilities such as electronic mail, access to remote computers, and electronic bulletin boards. Scientists can both adapt and extend tools to meet the communication needs of their work, and several networks are developing to serve particular scientific communities.

Dr. Newell is the U. A. and Helen Whitaker University Professor of Computer Science at Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213, and Dr. Sproull is an associate professor in the Computer Science Department, Carnegie-Mellon University.

0036-8075/82/0212-0843$01.00/0

multiplication unit. Just as numeric applications of computers build on these arithmetic primitives to perform more complex computations such as solving differential equations or performing statistical analyses, so higher level communication tools are built up from primitive datagram communications, combining them by means of computer instructions, as in any computer program. The tools then become the components by which the services of the network are made available to users, and the services expand as they become integrated with other computer tools.

Network dimensions. There are many different ways to implement the datagram communication primitive. Again, this situation corresponds to the arithmetic primitives of a computer: different computers may use different circuit technologies, different designs for adders or multipliers, or different performance-enhancement techniques. The differences give rise to a range of computers with various prices, sizes, and performances. Similarly, networks may be designed to achieve different performance along several dimensions. Some dimensions of a computer network design are discernible to its users and are important in choosing a network for a particular application. There are three principal dimensions of this sort: scope, performance, and cost.

Perhaps the most important property of a communication system is its scope: what computers (hence, users) does it connect? The great strength of the public telephone system is its scope. Because computer networks do not yet have such universal scope, the utility of a computer network to a scientist is likely to be a function of which other computers are connected to the network.

The issue of scope is related to a distinction often made among computer networks: the local network which connects computers with a cable at most a few kilometers in length, and the geographically distributed or long-haul network, which uses long-distance communications to connect to any computer in the world. The distinction has arisen largely because the networks tend to use different techniques and technologies, but it also affects scope. A local network may link computers of a scientist's laboratory, his department, or his university, but it will not allow him to contact computers used by distant colleagues.

The scope of a network may also be determined by the heterogeneity of the computers it can connect. Some networks are designed to connect only certain computers and are sometimes further restricted to running a particular kind of software or application. These networks are usually private special-purpose ones; an example is the network linking a bank's branch offices, tellers, and teller machines. By contrast, general-purpose networks and commercial public data networks aspire to connect computers of diverse sorts for many purposes.

The performance of a network comprises those aspects that affect the timing and condition of the messages delivered over the network. They may be summarized in three types of quantities: the network's data rate, delay, and reliability.

Data rate is simply a measure of the rate at which the network transmits information from one computer to another; data rates vary from 150 to 10^8 bits per second. A network with a high data rate will cost more than a slower one, especially if it is a geographically distributed network. Not all transmissions require high data rates, however. For example, the rate of information flow from a keyboard to a computer is limited by the typing rate of the user, and transmitting 15 character codes per second is sufficient. But the ability of the user to absorb visually displayed information is much greater, so that the data rate back from the computer to the video display needs to be large. Data rates can be asymmetric in a network.

A second performance dimension of importance to the user is the network's delay—the time it takes from the initiation of a message to the arrival of the beginning of the message at the destination. (A moment's thought will show this is independent of the data rate, which affects only the time between the arrival of the beginning of the message and its end.) If a computer network provides communications between an interactive terminal and a computer, a low network delay is required to permit graceful interaction with the user. If a computer program responds to each keystroke individually in order to change the image on the terminal's screen, a delay of as much as a second can seriously impede the user's progress. The delay is, of course, encountered twice. The network delays the datagram that flows from the terminal to the computer and contains the code describing the keystroke. The return datagram, which contains codes that command the terminal to alter its screen, is likewise delayed.

The final important property of a network's performance is its reliability. No communication network avoids errors entirely; it may garble data or it may lose a datagram altogether; that is, it may fail to deliver the data to the intended recipient. The problem of erroneous data is doubtless familiar to anyone who has used dial-up telephone connections to link a terminal and a computer: transmission is disturbed by electrical noise on the telephone line and by transient malfunctions in telephone-switching equipment. Failure to deliver a datagram is in some ways analogous to a telephone circuit being broken in error, although the reasons for network failure are often more complex.

Much of the unreliability in computer networks may be eliminated by suitable error detection and correction methods. Indeed, all the tools of network communications are able to detect garbled data or lost datagrams and request retransmission. However, if repairing a broken switch computer takes several hours, the host computers connected to that switch will be unable to use the network. This sort of unreliability, virtually unknown in the U.S. telephone system, is not an infrequent occurrence in computer networks.

The increasing use of computer communications networks is in large part due to their competitive costs. For example, the use of a network to connect a terminal to a distant computer costs about $5 per hour of terminal use, irrespective of the distance between the terminal and the computer. By way of comparison, interstate telephone calls cost about $20 per hour, although off-peak discount rates cost only $7 per hour.

A digital network achieves economies by taking advantage of the bursty nature of digital messages. It avoids the assignment of network capacity to a link when no information is flowing by breaking communication into packets and routing each packet as a separate event. Thus, no network resources are consumed while a computer or terminal is idle for a few seconds or minutes, a frequent occurrence. The cost of network communications is still dominated by the basic communications cost of telephony and broadcasting, which has participated only partially in the radical cost reductions in electronics. The effect of the spectacular cost-effectiveness of electronics shows up primarily in the relative inexpense of building up a massive hierarchy of programmed systems.

Network protocols. In general, a communication system treats the messages that flow through it merely as signals; that is, it does not assign the signals any meaning or interpretation. By contrast, computer networks operate by giving an interpretation to at least a portion of the message. The interpretations are couched in terms of communications protocols, or rules of discourse, that any computer connected to the network must observe. For example, each datagram must contain an address that identifies the computer for which the data is destined; all the addresses must be recorded in the same way in each datagram and must observe any addressing rules imposed by the network. Interpretations are processed by the network switching computers and by the host computers. The potential for both increased efficiency and functionality of computer networks resides in the freedom to elaborate the part of the message that is interpreted, because the processing is done by computer.

Several levels of protocols are used by computer networks. The protocol for transmitting datagrams is a low-level protocol. Higher level protocols provide a reliable transmission service by sending unreliable datagrams and expecting return datagrams that acknowledge correct receipt of the original datagram. The sender will retransmit the original datagram if no acknowledgment is received within a suitable period. Still higher level protocols can transmit a stream or sequence of data from one computer to another by breaking a long sequence of data into segments short enough to fit into datagrams and by requiring the receiver to reconstruct the original sequence from the segments. All of these protocols are constructed from the datagram communication primitive. The analogy with the levels of programming systems is precise because the levels are all defined by software systems within the network-switching and host computers.

Adherence to a common set of protocols allows diverse types of computers to communicate. Several protocols have been standardized and are gaining acceptance: for example, the X.25 protocol developed by the telecommunications industry (*2*) and the Internet protocol, now the standard of the Department of Defense (*3*). Higher level protocols have not been standardized as successfully, mainly because they begin to assign interpretations to the data in messages, and different computers use many different data formats. Just as there is no universal computer programming language or spoken language, there is no universal encoding of computer data. Some standards come quite close; for example, the ASCII standard for encoding text characters enjoys wide, but by no means universal, acceptance. The exchange of meaningful data between computers is limited not by network protocols and conventions so much as by the data representations and formats of the computers and their software.

Network tools. The typical user of a computer network is concerned not with its underlying protocols, but with a set of tools that the network, its protocols, and its host computers provide. These tools serve a role similar to that of other user-level services of a computer, such as text editors, compilers, statistical analysis programs, and so forth. These tools present to the user a range of services that still preserve the flexibility of the computer: they can be used in many different ways to address many applications. Perhaps the simplest example is a text editor, which is used to create and change a computer representation of a document, irrespective of the document's intended use. In a similar way, communication tools can be employed by the user in many different ways.

The tools described below are the dominant ones today, but the set is sure to change as the tools are improved, generalized, or particularized and as networks are used for as yet untried functions. The growth and elaboration of tools occurs in a laissez-faire fashion, and the scope for invention is large.

1) Remote computer access. Perhaps the simplest computer communication tool is that which allows one computer to serve as a terminal to another by transmitting character streams over the network. The computer serving as a terminal usually acts as a terminal too; it displays the information received from the remote computer and transmits to the remote computer the keystrokes entered by a user. This tool functions like a dial-up telephone communication to access a remote computer, but its performance is usually superior. The network often allows 9600 or more information bits per second to be presented to the user (960 characters per second); dial-up access is limited to 1200 bits per second (120 characters per second), and only a few years ago the limit was 300 bits per second.

2) Electronic mail. A network can be used to transmit messages to a user of another computer. Messages are stored in the user's mailbox, a memory file usually on a disk that retains a copy of every message sent to him. A computer program is invoked to scan the mailbox file at the user's convenience. The first few characters of a message contain fields that identify the recipient by name and network address, the time the message was sent, and so forth, and that allow the user to categorize incoming messages. Since text messages usually contain between 100 and 10,000 characters, reasonable mail systems require networks with relatively modest data rates. A sample message might appear as follows.

Date: 22 October 1981 1002-EDT (Thursday)
From: Allen Newell at CMU-10A
To: Bob Sproull at CMU-10A
Subject: today?
Message-Id: <22Oct81 100239 AN02@CMU-10A>
Origin: A310AN02 at CMU-10A; 22 Oct 1981 1003-EDT
Bob: As you said, we need a get together on our paper. How about today at 1600:1700? Just tried to call your office/home and failed at both ports, so I understand you may not be around. AN

Electronic mail tools can have distribution lists that allow a single message to be transmitted automatically to a collection of recipients. These are similar to distribution lists on office memoranda. The electronic mail distribution list is considerably more convenient since the copying and physical distribution are performed by computers and networks. Distribution lists are used extensively and may include hundreds of recipients. Other features of many mail systems are a capability to enter the standard editor on the host machine to ease the composing of messages; indexing for past messages; simplified commands to answer received mail and to forward received mail to others; notes to record which messages have been answered; and archiving of old messages.

There are many different ways to implement an electronic mail facility. If all senders and recipients of mail have access to a common computer, the mailboxes all reside on that computer, and network transport of mail messages is not required, although a network may be used to provide remote access to the common computer. The commercial mail system, named Telemail, operates this way. At the other extreme, thousands of personal computers may be connected to the network, presenting different implementation problems (*4*).

The themes of hierarchy and integration of communication and computation can be seen clearly in mail systems. Without the file and editing systems of the host computers, mail would provide a relatively lean and awkward, though still useful, facility. Intracomputer and network mail systems are merged into a single unified mail system in which the user need not be concerned with whether the recipient is on the same computer or on some distant host. As it grows, an electronic mail system gradually becomes adapted to users' actual needs and patterns of activity. However, each mail system is to some degree idiosyncratic, having been designed incrementally in what must be described as a craft-like manner. Commercial systems, though putatively designed more systematically, are in general less capable and less graceful than many of the privately built systems.

3) Bulletin boards. A variant of electronic mail is the bulletin board, which is in effect a mailbox that is publicly accessible and can be read by an entire community of users. Entries are placed on the bulletin board by sending electronic mail messages to a fictitious user (such as "BBOARD") that represents this public mailbox. Users read the bulletin board periodically, as when signing on to the computer. Bulletin boards differ from distribution lists in that the messages are intended to be read by a wide audience that has not been identified by name.

4) Computer conferencing. Another variant of electronic mail is the computer conference. Tools for conferencing enable communications of a collection of people to be recorded, transmitted, and presented to all participants by the computer (*5*). The tools differ from electronic mail tools in that they provide different ways to structure the communication, such as messages, conferences, debates, notebooks, bulletins, and so on. The sharpest difference is the ability to support simultaneous communication, as in a face-to-face conference, so that the latest messages by all participants are displayed simultaneously at each participant's terminal in separate areas of the display.

5) File transfer. A file transfer tool is used to copy from one computer to another a file, which is a stream of data bits. Transfer of a file through the network is similar to transfer of an electronic mail message, but a typical file is much larger than a typical message, and the tools used are somewhat different. The file transfer tool knows nothing of mailboxes, recipients, and distribution lists, but instead must know how to specify the file name to send or retrieve from a remote computer. It may also be required to convert the data representation of a file if it is copied to a computer that uses different representations. The size of files (10 to 100 times as large as mail messages) increases data rate demands. Effective rates of 9600 bits per second (which can require network capacities of 50 kilobits per second) are required or it takes too long to transmit a file, for example, about 1/2 hour to transmit a file the size of a typical scientific review article, if only 1200 baud is available.

6) Resource sharing. A computer network can provide communications between separate parts of a single application program when the parts are executed on separate computers. In general this requires that a process on one computer (for instance, a procedure call in the application program) be able to carry through automatically the whole sequence of sending a message to another host machine to execute a process there and of accepting the answer in return, including the full range of exception and error messages that can occur throughout the communication. These techniques allow the resources of a network to be shared by all computer applications in the network. A shared resource might be a specialized computer or instrument, or it might be a general-purpose computer. Sometimes this organization is called distributed computing because parts of an application are distributed to different computers. In the extreme case of a homogeneous high-performance network of general-purpose computers, the effect of a multiprocessor or multicomputer can be achieved.

How much performance is demanded of the network to make interprocess communication useful depends on the duration of the program executed in the remote host computer. The overhead of going through the network must be amortized over this productive time. In general, however, resource-sharing is much more demanding than remote terminal access, mail, or file transfer; it has not become a standard part of existing long-haul networks, but only of specialized networks. In these applications, it is the programmer who is aware of the network, which may be invisible to the user of the application. The techniques used to distribute computing effectively are the subject of active computer science research (*6*).

7) Embedding nontextual communications. Conventional long-distance communication forms such as voice and facsimile images can be embedded in a computer network by digitizing the signals, transmitting the digitized data in the computer network, and then with the use of some form of transducer reproducing the signal. Network voice transmission of this sort has been demonstrated, in which coding techniques were used to reduce the amount of data that must be transmitted (*7*).

Embedding of voice and image signals offers no new services beyond those offered by current telephone networks. However, the host computers in the network may store or index the data, just as they do mail. Mixed-media messages then become possible, as with vocal commentary on written messages. In general, images are data intensive; a second of high quality speech, for example, requires on the order of 10^5 bits, and a high quality color visual image requires 10^8 bits. Because much higher data-rate networks are required for reasonable service, relatively few systems permit such multimedia messages. This is a topic of active computer science research, much of it devoted to reducing the data-rate requirements by intelligent encoding of the signals.

Status. Commercial long-haul public data networks are now in operation. Tymnet and GTE Telenet both offer services that connect interactive terminals to remote computers or connect computers to each other. Most traffic on these networks is between a terminal and a computer, rather than among computers. Electronic mail is implemented, not by sending mail messages through the network, but by using the network to connect a terminal to a mail computer, where all mail users have accounts and where all mailboxes are stored. These two networks have combined annual revenues of about $100 million and charge about $5 per hour of terminal use. In addition to these commercial long-haul networks, many private networks provide internal communications to industry and government. The expansion of communications facilities in the United States has not kept pace with demand, and new customers may have to wait from 2 to 6 months for the communication or network service they desire.

Noteworthy among existing networks is the ARPANET, operated by the Department of Defense (DOD), which pioneered many of the innovations in packet-switching technology (*1*). This network was established in 1966 under the leadership of L. G. Roberts, then head of the Information Processing Techniques Office of the Advanced Research Projects Agency of the DOD (DARPA), as a basic research effort into computer networks. The host computers were those of the university and industrial research groups that were part of the DARPA research program in computer science. It has become the prototype to show how other computer networks can be constructed and the starting point for further technological development.

The ARPANET consists of an interconnected network of 50-kilobit-per-second lines that are leased from the telephone system and span the continent (Fig. 1). The 88 nodes of this network are the network-switching computers, called interface message processors (IMP's), which store packets and forward them to other nodes as they migrate through the net toward their destinations. The 185 host computers, several of which are attached to each node, are highly diverse; they are located throughout the country and are linked by satellite to Hawaii, England, and Norway. There are also nodes, called terminal interface processors (TIP's), which permit terminals from dial-up telephones to connect into the network. The hosts still represent the full spectrum of university and industrial research groups in computer science, but also now include numerous military organizations with diverse operational missions. The network is essentially saturated, and effective file transfer rates often fall below 5000 bits per second in periods of heavy traffic. The traffic consists largely of mail messages and files, the latter containing programs, data, or text. There is substantial remote access, and several organizations use the net to do all of their computing remotely. The use of the ARPANET has become integral to the operation of many of the organizations connected to it.

Local networks are also beginning to proliferate. These networks are sold rather than operated as services. Most minicomputer manufacturers and some microcomputer manufacturers plan local networks. Many word-processing and office-information system manufacturers are also offering local networks to provide communications between work stations, printers, file storage, and other special devices. A notable example is the Ethernet, jointly offered by Xerox, Digital Equipment Corporation, and Intel (*8*). The local networks operate at data rates from 1 million to 10 million bits per second, much higher than the long-haul networks. These data rates allow file transfer at a rate similar to that between the disk secondary storage of a computer and its central processor. The usage on these local networks tends to be predominantly file transfer rather than mail. They are often used to provide common secondary file storage for all hosts on the net. Resource-sharing is still rare on these nets, though the data rates are high enough to permit useful variants to occur, and a few have recently been occurring.

Scientific Functions

Network tools are used to perform scientific functions. The viewpoint now shifts from a description of the structure of networks and their capabilities to a description of science and its needs. No existing description of scientific activities has analytical merit for such a purpose. It is easy enough to enumerate broad classes of scientific activities: scientists create theories, derive inferences, design experiments, acquire data, analyze data, test theories, write papers,

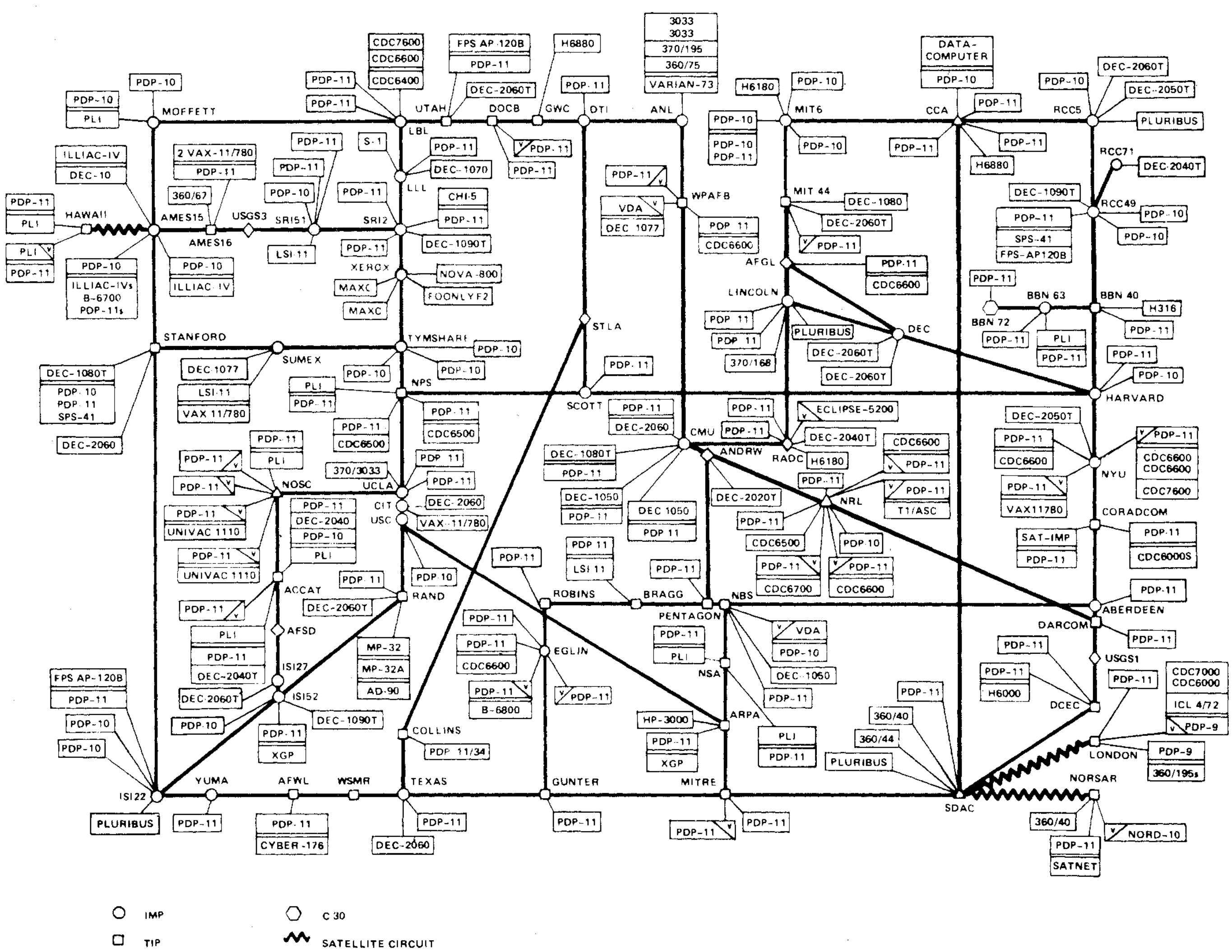

Fig. 1. The ARPANET in December 1980. This map shows the host population of the network according to the best information available, but no claim is made for its accuracy. The configuration of host computers was supplied by the Network Information Center; the names shown are IMP names, not necessarily host names. [Adapted from (*3*)]

give talks, maintain bibliographies, and read papers. But as any scientist knows, much is left out; and what fits under each term is highly diverse and includes activities that can be classified under other terms. More important, communication, the broad function performed by networks, occurs throughout all these scientific activities, in ways both large and small.

The difficulty of analysis here is one that networks share with computers—and precisely because networks are simply a component of an integrated computer technology. That is, wherever information is processed there is an opportunity to apply computers to aid in that processing. At a given historical moment, computers may be unable to supply aid or be able to supply it only partially. The reasons are diverse: the activity has not been analyzed; the cost-benefit rate is not sufficient; the necessary hardware and software have not been developed; legal and regulatory barriers stand in the way; humans prefer to do the processing; even, programmers and engineers cannot be hired. But all of these are contingent, even evanescent, factors that tend to yield to ordinary societal effort.

The spread of computing throughout society can be understood only with an epidemic-like model, in which computing and networking press simultaneously against all the opportunities, with infection taking place all along the boundary according to local conditions and spreading more rapidly around local concentrations. Even the usually trustworthy technique of the engineer, cost-benefit analysis of designs, cannot predict the development of these activities. The continuing improvements of computer technology in increased speed, memory, and reliability coupled with decreased cost, size, and energy demand turn all analyses into fantasy projections.

Perforce, we will be illustrative only. Many scientists already use networks in many different ways. Indeed, for some scientists, especially those whose research is focused on computer models or on computer science, computer tools, including network tools, constitute their entire working environment. Thus, we make no attempt to mention all the important or impressive examples of network use. Application has already gone well beyond that. Instead, we choose our illustrations to reinforce the larger themes of this article: the hierarchical structuring of systems; the integration of techniques of various computer technologies to accomplish scientific functions; and the heterogeneous and laissez-faire growth of systems to meet the needs of the scientist.

Many tools for many functions. One tool is used to provide many functions, and many tools combine to provide a single function. Most users of computers work freely with a dozen or so common tools, using them in combinations that are unique for a broad range of functions. It is this property that leads us to call these computer programs tools. They are like carpentry tools, general-purpose in nature, that can be used to build an enormous range of furniture, houses, barns, and fences. Network tools simply add to the toolbox.

Electronic mail is a good example of the multiple uses of a single tool. A great many purposes are not specific to science: asking a colleague for help, setting up visits, scheduling meetings, announcing meetings, requesting information, communicating with a secretary or administrator, communicating with funding agencies (*9*), and so on. These also are the woof that permeates the scientific day. Mail is used too for more specific functions of scientific communication: to react to papers, to clear up points with students or colleagues, to send in refereeing reports. These uses differ little from the first set, except they serve scientists doing their science rather than their housekeeping.

Scientists have other tools for accomplishing these communication functions: face-to-face conversation, the telephone, departmental memorandum systems, and letters. It is useful to analyze why electronic mail has established a secure role within the scientific communities that have reasonable electronic mail systems. The value of electronic mail came as a surprise to the developers of the ARPANET, who expected the network to be used principally for computer-to-computer communication for resource-sharing but found instead that the dominant use was mail (*9*).

Electronic mail subsumes some of the functions of the postal service and the telephone system but also provides services that these conventional communication methods cannot. Electronic mail is fast; messages are delivered in seconds to minutes, barring failure of the network and the computers used by the sender and receiver. Unlike a telephone call, the arrival of an electronic message need not interrupt the recipient. The mailbox provides the buffering that allows the sender and receiver to work at different times and in different places and still be able to communicate. Typically, a user will read mail once or twice a day at his convenience and process all messages that have arrived since he last read his mail. He will reply to some messages as he reads them; others he may file away according to categories he has established, much the way interoffice memoranda are filed. When a person travels, he may still read and process mail by using network access.

Consistent with its role, electronic mail has come to be used in an informal way. Writers dash off quick notes, usually not taking care to avoid typographical errors and jargon or to phrase the message carefully. Messages have the informality of speech, perhaps because the sender views the message as an alternative to a telephone call. Readers accept this informality and respond in kind. When formal communication is called for, such as in letters of recommendation, job offers, and the like, electronic mail is seldom used. In these cases, the conventional letter is favored, perhaps in part because of its established legal status.

Electronic mail tools are in their infancy. Although a great many people depend on existing systems for day-to-day work, the systems are still primitive. Some important properties of a telephone call or interaction with a secretary are missing from most electronic mail systems. For example, if I am out of town, I might want my mail system to reply automatically to incoming messages with an explanation that I will be unable to read mail for a while. Or I might prefer to allow senders to discover whether I have read the mail that they have sent me and deliver the explanation if they inquire. We expect mail systems to develop both in the direction of improved user interfaces and augmented communications facilities, and in the direction of becoming an active "mail agent," able to perform many routine mail-processing functions such as sorting, setting priorities, and answering simple requests.

The informality of electronic mail and its provision of communication among people with varied schedules and working habits make it particularly attractive to research centers and to dispersed co-working scientists. Indeed, there are those who speculate not only that electronic mail will improve the communication necessary among scientists, but that it may solve problems that appear to hinder scientific communication today (*10*).

It is sometimes expected that objective data will be developed about the trade-offs between such tools as face-to-face conversations, telephones, and electronic mail. Such expectations occur in part because a novel, but relatively expensive, system enters the range of alternatives and requires justified management decisions. It is not clear that such expectations can be satisfied for such changes as the incorporation of computer networks. In fact, there have been a few studies of electronic mail systems and quite a few on computer conferencing. A typical attitudinal finding is that users (here university management) consider them appropriate for factual, but not emotional or personal, information; a typical behavioral finding is that use decreases somewhat after an initial period (*11*). A problem with such studies is that, partly for reasons of scientific tidiness, they almost all work with isolated tools in relatively isolated situations. Since communication is radically affected by whether the desired recipient can be reached through the network, it is difficult to extract anything from this literature if the goal is to evaluate the usefulness or costs and benefits of the tools. In the relevant field situations, such as some large research communities that have both long-haul and local networks, no studies have been made. However, common observation shows that large fractions of daily interaction are made by electronic mail. (For example, each day, each of us receives about 10 to 20 electronic mail messages, about eight telephone calls, and fewer than five letters.)

Instead of working from a tool to its uses, we can start with a scientific activity and ask how network tools are used to aid it. Reading scientific material provides an example. To read a document requires detecting it, acquiring it, and finally reading it. As we noted above, the future computerized reality may be simply to take each component of the activity and make it electronic. We then get something akin to electronic journals, detected by on-line retrieval or advertising, ordered on demand from anywhere (both these functions involve network capabilities directly), read on-line at a display capable of print quality, and with the document capable of active processes to retrieve other material, absorb comments and emendations, and evoke computation-like manipulations. Although hard to improve upon as a long-term projection, this description is unedifying because it does not reveal the specific shape of the difficulties to be overcome. It does indicate the large number of computer tools (including network tools) that enter into a single activity. It follows that the total function will only be possible as envisioned when the entire computer facilitation of the environment has evolved to an appropriate state.

The present reality is substantially less advanced but does represent an interesting way station. Demand printing is the use of file transfer tools to obtain from a central repository a copy of a computer file that represents a document. The user may then scan it on his display, print it, edit it, or do any of the things he does with such files in his computer system. To give just one example, the groups working on the design and implementation of network protocols in the ARPANET maintain a repository that contains a great many working notes, reference documents describing current protocols, comments submitted by committee members, periodic reports of members, and a document that indexes the other documents. The appearance of a new document is announced by sending an electronic message to a distribution list of people who have asked to be kept informed of the group's progress.

As demand printing has become established more generally in computational communities, electronic bulletin boards are used to advertise the existence of documents. The essential means of retrieval is the general-purpose file system, with whatever scheme of file names is available. File transfer over the network is used to read the files. Such systems are not esthetic and their human engineering usually leaves much to be desired. Indeed, they were not designed by anyone, as that term is usually thought of in connection with computer systems. They were cobbled together from the tools available along with the communication of a few conventions that users should follow. However, the systems work quite effectively.

Unification of communication and computation. We have emphasized that communication and computation need to be taken to be an integrated set of tools. The largest part of the benefits, both now and in the future, comes from use of the entire kit of tools in concert. Functions such as data acquisition, data analysis, theoretical calculations, modeling, and simulation are at the heart of science, whether the data are predominantly numerical or textual. The pivotal uses of computers are those that aid these functions. When communication becomes a way of enhancing these functions, it too becomes central to the science. Its use for less central functions, such as retrieving literature, editing papers, and arranging scientific meetings, will be seen as being less central to the scientific enterprise, just as are these functions.

The use of network tools to access specialized resources not available in the local computing environment provides good examples. The MACSYMA system at MIT (*12, 13*) is widely accessed over the ARPANET in this way. MACSYMA is a large, complex program for manipulating mathematical expressions that are used to solve recurrences, differential equations, linear systems, and much more. While the software could in principle be copied and installed at other computer sites, it is more convenient to maintain the system at a single site and provide remote access. Moreover, many users of MACSYMA use it only occasionally in the course of their research, and would find the effort required to import the program to their site excessive.

An interesting and less familiar combination of network tools is used in the ARPANET to allow researchers across the country to fabricate experimental integrated circuits. A system called MOSIS has been developed by the Information Sciences Institute at the University of Southern California to provide fabrication services. Clients communicate with the system through electronic mail, using stereotyped message forms to request that an account be established, to inquire about details of fabrication processes available, to check that a design is acceptable to the system, to enter a design into the fabrication queue, to request the status of a design, and so on. The design is forwarded to MOSIS as an electronic mail message describing in a text form the geometry of the several masks that control integrated-circuit fabrication. Once chips are fabricated and packaged, they are sent by air carrier to the designer.

MOSIS uses the network to allow a great many designers to share access to fabrication. Moreover, the system is able to combine several separate designs onto one chip (a so-called multiproject chip) in order to reduce fabrication cost (*14*). Centralizing fabrication services in this way simplifies interactions with vendors and frees the chip designer from a great many troublesome details. An important advantage is the avoidance of dealing with a human bureaucracy (the alternative organizational technology for managing the same process), which tends to become unresponsive, error prone, and hard to control. With MOSIS, the user communicates directly (and with delays of only a minute or two) with remote files and systems to monitor his progress and check the correctness of the proceedings. The speed and accuracy of the responses far exceeds anything he could get through a human intermediary on a routine basis.

This is another example of network use that becomes an integral part of a larger computational enterprise. The design sent by mail to MOSIS is not prepared by hand, but is produced by computer-aided design tools for preparing mask geometry and for checking the design.

Not all uses of a computer network fit the model of integration with the full set of computational tools. A computer network can offer the services of a "turn-key" or "closed" computer program with a single function; the network is merely a way to distribute the function to clients with interactive terminals. The services are usually designed for clients who have little or no familiarity with computers. An example is the bibliographic search services used by many libraries. A less familiar example is computer conferencing (*5*).

An interesting example on the boundary between the extremes of unifunction turn-key systems and full integration is the PROPHET system, developed by the National Institutes of Health to support the needs of research pharmacologists and others working in chemical and biological interactions (*15*). It provides facilities for maintaining files of chemical structures, experimental results, and laboratory notes and has computational tools for reformatting data, analyzing data, preparing graphical presentations, and so on. The users of the system may share data files and may use simple electronic mail and bulletin board tools to communicate. PROPHET has over 700 users at 40 sites in the United States who access a computer in Massachusetts with interactive graphics terminals connected by a variety of telecommunications equipment; a connection to a public data network is planned. The PROPHET working environment has fostered some intense collaborations, such as experiments undertaken by three geographically separated groups in the pharmacology, crystallographic structure, and animal testing of a single chemical compound (*16*). Thus, PROPHET does provide an integrated communication and computational world. However, it has been developed as an isolated system and does not intercommunicate well with the larger computational world of the scientists who use it. The reason for this development is the same as the reason for other turn-key systems, namely to provide a more polished product that can be learned easily by a user population with limited initial motivations for becoming sophisticated computer users.

Undisciplined growth. To reevoke the epidemic analogy, the entire computer field is growing in a local and undisciplined way, riotously in some aspects, constrained by external factors in others. Coordination and systematic design are certainly attempted, but these serve only to constrain growth locally. Computer networks fully participate in the laissez-faire growth. The development of network tools and their exploitation to provide new functions (with the subsequent fashioning of more integrated tools) is especially open-ended. But even basic network technology, though it involves large-scale investment and commitment, is still mostly undisciplined, because of the interaction of the rapidly evolving technology, the play of the market, and a regulatory apparatus developed without digital communication in view.

One effect of this situation on computer networks is essentially positive. The invention of new network tools and a search for new functions to serve goes on everywhere. It appears that the new tools arise initially in heavily computational research environments, outside the more established channels for computer development. The computer manufacturers almost never lead in the development of new tools (editors, electronic mail, bulletin boards, demand printing—essentially all the things we have discussed in this article), although eventually commercial software tools become available. In other ways manufacturers' developments are highly significant. The explosion of microprocessor-based personal computers, which is quite independent of networking, is of the utmost importance to it; it is creating an immense population of computers, ready to become more useful when networks are provided, thus changing radically the cost-benefit calculations for whether networking is worthwhile.

A second effect is negative. The tools almost never fit together perfectly, are often unreliable, do not work as advertised, and require substantial effort to learn. Although smoother, more integrated systems are developed, such as the PROPHET system, they lag in their functionality. Because of the rapid advance of the field, the lags come to be too great to tolerate; that is, the new tools are potentially too useful to be ignored, and users choose the more advanced, more chaotic world.

We believe that neither of these features will change, even in the medium term, and that users of computer networks will have to find their own balance between them.

A single example of a network tool developing in the ARPANET will have to suffice. It originated under the name FINGER but is also known as WHERE. Anyone on our computer system may give a command such as "WHERE Sproull," which will determine whether a user named Sproull is logged on the system and print out a message if so. If not, the program reads a file created by Sproull to determine whether other information should be revealed, such as when he last read his electronic mail, or whether he has left instructions to permit people to contact him in other ways, such as by telephone. If a user anywhere on the ARPANET gives the command "WHERE Sproull@CMUA," a specialized network protocol is invoked to execute the command on the computer known as "CMUA" and report the results back. If this user does not know where Sproull has computer accounts, he may give the command "WHERE Sproull@NIC" (NIC is the mnemonic name of the Network Information Center). The program contacts the NIC through another specialized protocol to search a database of all known users of the ARPANET and send back any entries that match, thus providing full name, telephone number, postal address, and an electronic mailbox address in the network. A user can update this database by sending mail to a person at the NIC.

FINGER is an example of a function that is not easily provided by any of the conventional network tools. A separate tool and network protocols have grown up to support it. It clearly addresses a real need: to find people amid a community of over 1000 users that is changing all the time; in effect, to provide directory assistance. But it is just as clearly only a beginning: the construction of the network-wide directory should be automated, presumably through the use of network protocols to ask each computer on the network to enumerate its users and the information they wish to reveal in the database. This step has yet to be taken.

Furthermore, the set of functions that FINGER supports is not yet stable. For instance, since there may be several users named Sproull, FINGER prints out the alternatives, on the basis of substrings; for example, all user names starting with "Spr." It thus provides a way of recollecting names of users that have been forgotten, a quite different function. Since FINGER can reveal when you last received mail, FINGER has become a standard way of finding if any mail has arrived without having to log in. Since FINGER can also reveal when mail was last read, it provides an implicit acknowledgment mechanism. You can be reasonably assured that your mail was seen by a recipient without his having to take the trouble to acknowledge it explicitly.

The entire development of FINGER is more like a community project than a deliberate system design. At each step, some members of a computational community design an incremental improvement. One effect of computer networks is that the participants are extended throughout the entire network. Many of these functions are not highly significant in themselves—one would not purchase a turn-key system just to provide them. However, in the aggregate, they make up an environment in which the business of the day (here science) can be carried on with significantly more dispatch. And they accrete at low marginal cost to systems already in existence.

Networks of Scientists

While it is clear that computer networks provide functions that are valuable to scientists, there remains the problem of assembling a group of scientists of similar interests on one network to take advantage of the functions: in effect, to build a network of scientists. Whereas the widespread scope of the telephone network encourages all scientists to participate, the current limited scope of computer networks makes participation more problematic. A network's value to a scientist is a strong function of its scope, as measured by the number of his colleagues who are similarly attached. Although any network may provide terminal access to a particular service that a scientist requires, and therefore justify a network connection, this form of use fails to exploit most of the network's potential functions.

It appears at present that the growth of networks will be piecemeal, with special circumstances and entrepreneuring accounting for each additional group. That is the history so far. The ARPANET was a network for some computer scientists. It was justified by the network itself being computer science research and by computer scientists being willing to spend much time and effort making a computer network work. SUMEX-AIM (Stanford University Medical Experimental Computer–Artificial Intelligence in Medicine) (*10*) is a group of medical scientists distributed around the country who are concerned with sophisticated computer applications to medicine. They form a network that uses primarily terminal access to computer facilities at Stanford (and more recently also at Rutgers), but they use many other tools as well (electronic mail, bulletin boards, and so forth). They were formed as an intellectual spin-off from the computer science research groups in 1974 with NIH support. The central justification was the need for an advanced computer science computational environment. Recently, a group of geneticists formed a network (GENET) to make use of some computer programs for genetic research. This has grown out of SUMEX-AIM, and its structure will be quite similar.

A more general effort is CSNET (*17*) which is a network to connect computer science departments with a level of network technology equivalent to ARPANET. It will create a level of common protocol within existing long-haul carriers (Telenet, Tymnet, and ARPANET). It has only recently been funded by the National Science Foundation and is still being implemented. The justification was to provide the same advantages to the entire computer science community as the ARPANET provides to the community which has access to it. Yet another net, COGNET, has been proposed and is undergoing evaluation by funding agencies (*18*). This network is for cognitive scientists (psychologists, linguists, and artificial intelligence researchers). It will work within CSNET and will concentrate on building up a computational community by building and providing higher level network and computer tools, as well as training and other help. The justification is the need for the dispersed community of cognitive scientists to make use of advanced computer science tools for symbolic modeling and simulation. In all these cases, the justification is usually focused on some central scientific need, but the use of a network, once in place, spreads out to accommodate a larger range of scientific functions through use of whatever tools are in place.

To start a scientific network, then, appears to require several things. First, an initial group of users must be recruited, large enough yet of sufficiently common interests that the group anticipates substantial communication. The initial connection of this group to the network must be subsidized. Once a large enough group is using the network, the incentives to join should be substantial, and the subsidies can decline. Second, the group of users needs to be familiar with computers and to have harnessed them to their scientific work. The justification for the initial subsidy rests in some critical feature of their work that requires network tools in connection with essential scientific activity. It helps if the group has similar computing practices or systems. The more nearly identical the computer systems, the easier it will be to share software for network tools, to share file formats that represent documents, and so on. If the network is used in part to share access to one or more computers for general work, the homogeneity of the computing systems is even more important, simply because it is a burden and an inconvenience to learn to use two or more different computer systems.

Finally, the performance of the network must be selected to permit the kind of communication that is sought. If the network is small and used solely for electronic mail, a low-performance network or even a set of conventions for computer conversations on the dial telephone system is required (*18*). As the network grows or as it becomes used for remote access, greater capacity with low delay will be required. For functions involving file transfer (for example, sharing data, demand printing, joint paper writing), the network bandwidth must be sufficiently high. It is clear from experience with networks such as the ARPANET that performance influences usage a great deal. One can argue that the failure of resource-sharing experiments in the ARPANET can be traced to insufficient bandwidth.

Conclusions

Computers are being used increasingly by scientists because the hardware is becoming less expensive and is being packaged in convenient personal computers, because computer tools increasingly address needs of scientists, and because the general literacy of scientists about computing is increasing. Computer networks play several roles in this overall expansion of scientific computing. One role is to provide user-to-user communication: electronic mail is the principal example. A second role is simply to make accessible to a great many computers and their users the services of special resources. A third role is to augment computer tools that are already useful on single computers with the ability to share and transmit data among users of the tools: paper writing and demand printing are examples. It is important to recognize that the last two roles depend principally on the development of useful computer software, and only secondarily on the ability to communicate among computers.

We have chosen to concentrate on the basic structure of computer networks and the underlying features that seem to govern their character: (i) the hierarchical structure building up from the underlying datagram primitive to network tools to collections of tools that perform scientific functions; (ii) the integration of network tools with computer tools generally in order to obtain the various benefits, so that networks are not to be evaluated as isolated functions; and (iii) the laissez-faire manner in which network (and computer) tools will grow, with attendant luxuriance in opportunities but roughness in execution. We have not been careful to enumerate all or most of the tools and existing scientific networks. They are part of the current scene, which—if anything is clear at all—will change complexion rapidly.

References and Notes

1. L. G. Roberts, *Proc. IEEE* **66**, 1307 (1978).
2. International Consultative Committee for Telegraph and Telephone, *CCITT-X.25-Orange Book* (Geneva, Switzerland, 1977), vol. VIII-2, pp. 70–108.
3. J. Postel, Ed., *DOD Standard Internet Protocol* (NTIS No. ADA079730, Department of Defense, Washington, D.C., January 1980); *ACM Comput. Comm. Rev.* **10**, 2 (1980).
4. A. Birrell, R. Levin, R. Needham, M. D. Schroeder, *Commun. ACM*, in press.
5. S. R. Hiltz and M. Turoff, *The Network Nation* (Addison-Wesley, Reading, Mass., 1978).
6. K. J. Thurber, Ed., *A Pragmatic View of Distributed Computing Systems* (IEEE Computer Society, New York, 1980).
7. D. Cohen, "Specifications for the network voice protocol" (ISI/RR-75-39, DDC AD A023506, USC/Information Sciences Institute, Marina del Rey, Calif., March 1976).
8. Intel, Digital Equipment Corp., Xerox Corp., *The Ethernet, A Local Area Network: Data Link Layer and Physical Layer Specifications*, version 1 (Xerox Corp., Palo Alto, Calif., September 1980).
9. J. C. R. Licklider and A. Vezza, *Proc. IEEE* **66**, 1330 (1978).
10. J. Lederberg, *ibid.*, p. 1314.
11. R. E. Rice and D. Case, *Proc. Am. Soc. Inf. Sci. Annu. Meet.* **18**, 228 (1981).
12. R. Pavelle, M. Rothstein, J. Fitch, *Sci. Am.* **245**, 136 (December 1981).
13. Massachusetts Institute of Technology, "MACSYMA reference manual 9" (Laboratory for Computer Science, Cambridge, Mass., 1977).
14. L. Conway, A. Bell, M. Newell, "MPC79: A large-scale demonstration of a new way to create systems in silicon," *LAMBDA* **1**, 10 (1980).
15. P. A. Castleman *et al.*, in *Proceedings of the 1974 National Computer Conference* (AFIPS, Montvale, N.J., 1974), pp. 457–468.
16. H. Bilofsky, personal communication.
17. L. H. Landweber, "CSNET—a computer science research network," a proposal submitted to the National Science Foundation, Washington, D.C., 1980.
18. D. H. Crocker, E. S. Szurkowski, D. J. Farber, in (*6*), pp. 577–584.

19. We gratefully acknowledge the help and information provided by H. Bilofsky, D. Cohen, I. Durham, D. Farber, R. J. Feldmann, R. Rice, K. Uncapher, and B. Wessler. Sponsored by the Defense Advanced Research Projects Agency (DOD), ARPA order 3597, monitored by the Air Force Avionics Laboratory under contract F33615-78-C-1551.

Computers and Electronics for Individual Services

Ruth M. Davis

Individuals have long used computers to perform personally needed services. Today, the term personal computer generally refers to a computer system that fits on a table or desk top, is in the price range of a consumer product, and is marketed for a wide range of popular, personal applications. Memory size, computing power, and display capability do not definitively characterize the personal computer because they are rapidly changing.

The advent of personal computers can probably best be viewed as another step in the continuing development and diversification of the computer field in terms of both products and applications. Like the minicomputer and the microprocessor before it, the personal computer appears to be ushering in a new era in computing history. Although it was long anticipated by computer professionals, to the new customer to whom computing power was not previously available personal computing opens the door to a variety of new activities. Probably the greatest change brought about by personal computing is that it takes the computer from the exclusive province of scientists, engineers, and company professionals and makes it available to almost any individual in any environment.

Ten years ago, saying that the computer offers to the individual the ability to significantly control his immediate environment might have been considered whimsical. Today, it is an observation that can be supported by many examples. In discussing personal computing in this article I will use some examples from areas that were not previously affected by computers, including entertainment, home management, and mail as an individual means of communication.

Genesis of Individual Computing

There are probably three principal factors that have contributed to the increasing popularity of individual computing devices. One was the provision, beginning in the 1960's, of "carryable" terminals that professionals could use at home to carry on office computing tasks through remote access of computer systems. The second, which was due to dramatic decreases in the cost of electronic equipment, was the advent in the early 1970's of home computer kits with which one could build computer and computer display devices. The third, also originating in the early to middle 1970's, was the availability of electronic games on coin-operated machines or home television sets.

The instructions accompanying a microprocessor kit I bought in 1980 explained the principles and use of semiconductors and programming with pictorials and a minimum of engineering jargon. Written in the middle 1970's, they already identified sample uses of the assembled kit such as temperature control, cooking control, credit verification, portable computing, surveying, traffic control, "smart" toys, and electronic bartending. This material exhibited many of the elements of friendliness that home computer sellers are still trying to project for their products.

The first (cumbersome) carryable remote computer terminals were replaced in the past 15 years by smaller portable keyboard input-output terminals and display input-output terminals. Each type was accompanied by an acoustic coupler through which it could be interconnected

The author is president of the Pymatuning Group, Inc., 1625 I Street, NW, Washington, D.C. 20006.

0036-8075/82/0212-0852$01.00/0