Experience with Electronic Mail in Andrew

James H. Morris
Carnegie Mellon University
Pittsburgh, PA 15213

The use of networks of computers for human communication is surprising and fascinating. Electronic mail (email) becomes an important tool of any organization whose members spend a lot of time using computers. This paper reports on the state of email activity at Carnegie Mellon where thousands of students, faculty, and staff have been using the Andrew system during its development over the last few years. There is some analysis of why and how email works and what it will take for it to become more widespread.

The Andrew System

Carnegie Mellon is a relatively small university of 8,000 people located on a physically compact campus, with a long-standing record of excellence in computer science research. In 1982 Carnegie Mellon and IBM created the Information Technology Center to design and develop a system to support the university's needs. The system has been named *Andrew*, after two benefactors of the university, Andrew Carnegie and Andrew Mellon. It is one of the largest distributed personal computing systems of its kind. As part of the NSF EXPRES project Andrew is being extended to support interchange of multi-media mail. Some earlier papers[1,2,3] describe the system in general.

Computers

There are nearly 3,000 different computers at Carnegie Mellon today: mainframes, mini-computers, personal computers and workstations. The most interesting of them are the 1,000 or so workstations. They run UNIX and have such advanced features as multi-million-instruction per second processors, bit-map displays, virtual memory, megabytes of real memory, a mouse, and a local area network (LAN) connection. These machines cost about $5,000 today. The primary machine that Andrew runs on is the IBM RT, a 32-bit RISC Technology machine. It also runs on SUN workstations and DEC VAXstations.

The Network[4,5]

Carnegie Mellon uses a number of LAN technologies that provide high bandwidth computer-to-computer communications. Various departments, at their own initiative, have installed Ethernets. Most of this work was not coordinated centrally. During 1985-87 we re-wired the entire campus, both for data communications and future phone services, using the IBM cabling system and token ring. Today there is a network outlet in every room of the university, including dormitories, totalling 11,000.

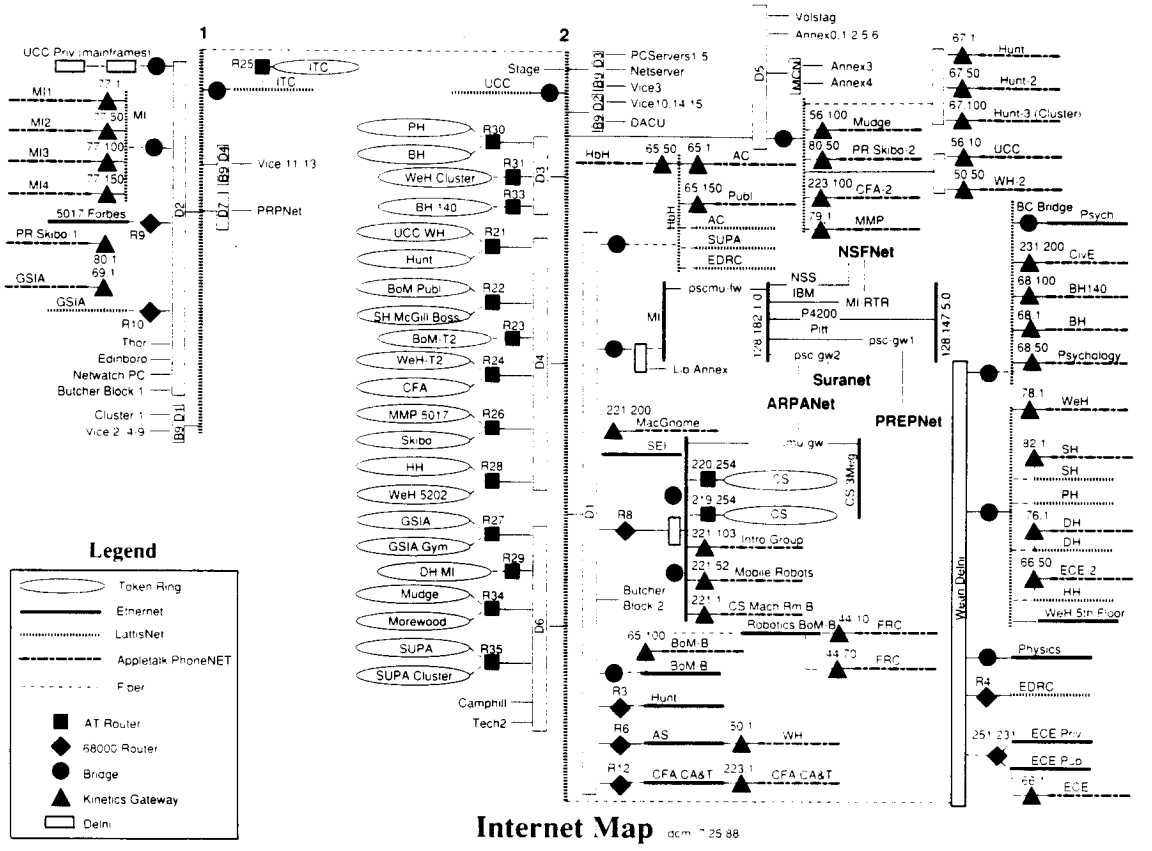


Figure 1. The Carnegie Mellon Network

The Carnegie Mellon internet -- an assemblage of LANs interconnected with routing computers -- is shown in Figure 1. The backbone consists of the two vertical sections under the numbers 1 and 2. Most departmental LANs are connected directly to this net through a fiber cable and a routing computer. The backbone serves both the function of a switch net as well as the main resource sharing network. We recently broke this backbone into two sections to reduce the

volume of traffic.

This network allows one to move information between any two campus computers at the rate of about 50K bytes per second. It has allowed us to build many services that use high-bandwidth communications -- notably a file system and a multi-media message system.

The File System[6,7]

An academic environment requires a large amount of information sharing. A high degree of user mobility between dormitories, faculty offices, libraries, and laboratories is also essential. Thus it was important to choose the central communications model carefully. We chose to emulate a time-sharing file system. The file system provides a common name space for files. Users may thus access files in a uniform manner regardless of the specific workstations at which they are logged in, or at which the files were created originally. The files are actually stored on about 20 file servers -- workstation-class machines named Vice2, Vice3, etc. in Figure 1.

Most other shared facilities, such as mail and printing, can be built on top of the file system, obviating the need for machine or location-specific information. One requests a service by leaving a request file in a designated directory. Software in each workstation makes these facilities appear as a transparent extension of that workstation's operating system.

Graphics Software[8]

Andrew supports documents that include multi-font text, raster images, animations, spreadsheets, equations, and geometric drawings. This is implemented with a basic user interface toolkit that runs on workstations. It is open-ended in the sense that new kinds of graphical applications can be added easily. A document preparation system has been created using this software; this paper was written using it. All the other Andrew applications use the toolkit, too.

The Message System

The Andrew Message System,[9,10] which makes heavy use of the file system and graphics software, supports a large email community. Besides person-to-person and person-to-group delivery it also provides *bulletin boards* (bbs): directories of messages that are centrally stored. Sending something to a bb is called *posting*.

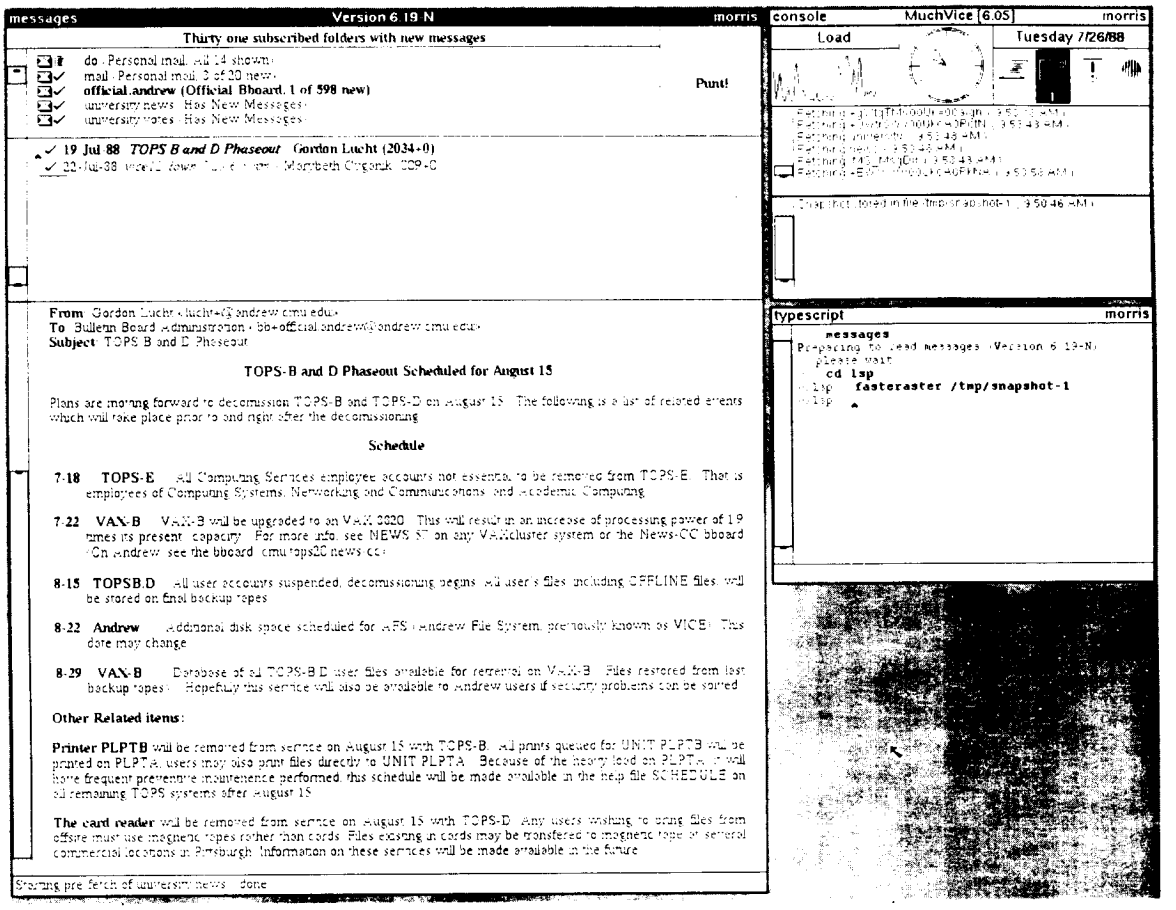


Figure 2. The Message subsystem on Andrew

Figure 2 shows the Message system running under the Andrew window manager. It is the window on the left. The uppermost panel lists all the folders that contain new messages. Folders serve many purposes and have different access rights. My *do* folder is a collection of messages I should act on and can be read and written only by me. My *mail* folder can be added to by anyone but read only by me and my secretary. An *official* folder is a bb that can be written only by authorized people, but read by anyone. Other folders like *graffiti* can be read or written by anyone. The second panel lists all the messages in the

selected folder, and the third panel displays the selected message.

The upper window on the right is a monitor, Console, that indicates various things, including the presence of new mail. The lower window on the right is a command processor. It is important that the system be integrated so that one can copy things from other windows to and from the Message window.

Mail Transport

Local mail moves around the system as follows:

1. A user invokes a mail sending operation from a document editor.
2. A preliminary screening of the addressees is done using a data base called the White Pages. If any address is implausible or ambiguous based on local knowledge the user is helped to correct it immediately.
3. The message is stored in a personal out-basket directory.
4. A background process on the sender's workstation examines each addressee for the message. If an addressee is an Andrew user, the message is written directly to his or her mailbox in the file system; thus delivery between Andrew users can be instantaneous. Otherwise, it is put into a queuing directory to be examined by a post office machine.
5. The indicator in Console tells users if they have mail waiting in their mail boxes. When they invoke the the *read mail* operation, the message is moved from the mailbox to their mail folder.

A post office machine's major duty is importing and exporting mail from Andrew. It reads the queuing directory mentioned above and uses the UNIX sendmail program to direct mail to machines on the networks CMU participates in. It has the mail address *andrew.cmu.edu* for incoming mail and distributes mail to mail boxes, using the White Pages to identify directories. A secondary job for post office machines is to serve as a fall back delivery system for local mail. If a file server is down, preventing quick delivery, a mail queue on an up server is found, and the post office machine handles it later. If things get really bad, e.g. the workstation can't reach any file server at all, mail is queued on the sender's

workstation and the background process retries delivery later.

External Networks

The value of a mail system increases with the number of people it can reach. Therefore, the connections of our mail system with the outside world are crucial. We are active participants in several academic and research networks. Figure 1 shows some of the physical connections.

The *ARPA/NSF* network is one of the most developed networks. Mail travels easily to sites on these nets without complicated addressing rules. There are over 11,000 ARPA sites listed in our directory.

BITNet is a cooperative network started in 1981 that links universities and research centers around the world. A university can join by agreeing to forward mail in its local vicinity. Andrew users send and receive mail through a local VAX/VMS computer.

CSNet is a very good mail-drop service that facilitates communication between networks. No address list is locally maintained for CSNet, but one can send a dubious name and address as the first line of a message to "fixaddr@relay.cs.net" and receive back a reply that is a rewriting of the address into a shorter and possibly correct form.

UUCP stands for UNIX-to-UNIX copy. Like BITNet, it is cooperative, but less formally organized. To join the UUCP system one need only find someone who belongs and get him to agree to forward mail to you. Partly for this reason, UUCP addresses are relative rather than absolute; i.e one must specify the detailed route a message should follow. The Andrew mailer does not have a database of UUCP sites; there are many thousands of them.

Usenet is defined as all sites receiving the newsgroup (bb) net.announce. A Usenet link between two sites is one that net.announce is sent over. Thus, Usenet is a network for wide information dispersal, not a network for mail.

The inter-university network system is growing rapidly and is often confusing. People press these networks to the limits of their capacity, partly because they are free. An important facilitating function is the maintenance of directories for various places and people. The chaotic nature of the networks' growth makes this a hard task. At Carnegie Mellon the question, "Does anybody have an address for ...?" is frequently seen on the local bbs.

Basic Statistics

The Andrew system runs on about 500 advanced workstations at Carnegie Mellon. The message system also runs on about 500 IBM PC's. About 6,000 people have accounts on Andrew, but that does not necessarily mean they are active users. We don't keep statistics on the total volume of mail processed, partly because much of it is passed directly through the file system without passing through any particular computer. However, we have kept track of the amount of use of bbs. There are over 1700 bbs on Andrew. Most are from external places: the ARPANet, Usenet, and Dow Jones -- the source of stories for the Wall Street Journal and other Dow Jones data base services. Several hundred bbs are locally generated. In a typical month there were approximately 60,000 different messages placed on various bbs, about 88 per hour on a 24 hour basis. Table 1 shows the major categories of bb posts. The annual volume of messages is estimated to be 10^{10} bytes.

Bulletin Board Type	Monthly Posts
Usenet	23,300
Dow Jones	16,000
ARPANet	6,700
Andrew Advisor	1,200
All Others	12,800

Table 1. Major Sources of Bulletin Board posts in a typical month

#	Bulletin Board	#	Bulletin Board
779	official.andrew	120	andrew.x11
532	university.news	119	andrew.dev|null
435	andrew	119	andrew.networks
422	andrew.hints	115	org.cs.general
412	cmu.market	114	andy.raster
307	official.andy	114	graffiti
256	ext.nn.alt.sex	113	andrew.ms
248	announce	110	magazines.misc.jgk
231	andrew.games	105	announce.music
212	andrew.gripes	103	andrew.prog.c
203	andrew.picture	102	andrew.bugs
197	cmu.market.apartments	102	andrew.quovadix
163	cmu.mac	102	hobbies.cd
153	andrew.advice	101	andrew.library-info
151	andrew.whatsnew	100	academic.cs.15-12.ann
148	org.cs.market	99	cmu.tops20.general
134	andrew.next	98	andrew.kudos
127	tops20.david-letterman	98	andrew.wanted
126	tops20.job-postings	96	academic.cs.15-212.discuss
125	andrew.quotes	96	magazines.humor.ckk

Table 2. Most-read bulletin boards, # indicates number of readers in a month

A more interesting set of numbers describes readership. During a typical month of the school year there were about 1,500 active readers -- people who read some bb at least once. The average number of bb's read per person was 16 and the median was 7. Table 2 lists the most popular bb's. For example, 779 people read the official.andrew bb sometime during the month. Official.andrew and official.andy contain announcements from the staff about the state of the Andrew system and recent changes. Andrew.hints, andrew.gripes, and andrew.whatsnew are open forums about how to use the system with slightly different emotional slants. Andrew.picture contains various raster images. The market, and wanted bb's are for local advertisements. University.news contains announcements by the university's public relations department. Cmu.mac contains discussion of Macintoshes. Tops20.david-letterman is devoted to discussions of a late night television program. Andrew.dev|null and graffiti are intended as places people can write whatever they want without fear of criticism. The magazines are digests of posts from other bb's that the individual maintainers, jgk and ckk, think might interest readers. Academic bb's contain class discussions.

Multi-media messages

One of the most novel features of the Message system is the facility to include things other than text in messages. This is relatively new feature and has not been fully exploited by the user community. The following examples were taken from messages sent by some of the pioneer users of the system.

One popular use of the raster image facility is to distribute cartoons to the campus community. One student uses MacPaint to produce cartoons in a continuing series called "Nerd@Andrew". Figure 3 shows two samples. We also experimented with scanning cartoons from a daily newspaper until fear of copyright suits took over.

For the system to be used effectively for technical discussion, the facility to include mathematical formulae is essential. Figure 4 shows a message I sent to several places requesting some tutorial information about a mathematical subject. The formulae in the message are manipulated by a special editor that understands them. For example, the size of parentheses, integral signs, etc. are scaled automatically.

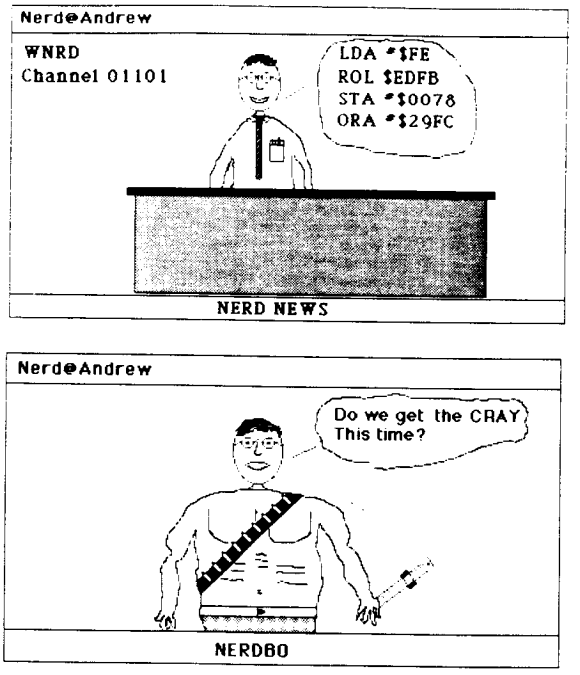


Figure 3. Cartoons distributed by a student

From: Jim Morris <jhm+@andrew.cmu.edu>
To: bb+academic.gsia@andrew.cmu.edu, bb+academic.epp@andrew.cmu.edu,
bb+academic.stats@andrew.cmu.edu, bb+andrew@andrew.cmu.edu
Subject: Understand Lagrangians?
CC: Andrew Palay <ajp+@andrew.cmu.edu>,
"Maria G. Wadlow" <maria+@andrew.cmu.edu>

I would appreciate a little help in understanding the following excerpt from "Moral Hazard and observability" by Bengt Holstrom in *The Bell Journal of Economics*

> Let $f(x,a)$ be a probability density function for x; f_a and f_{aa} are its first and second partial derivatives. G and U are utility functions. The optimization problem is to choose a function $s(x)$ in $[c, d+x]$ and a value a so as to maximize

$$\int G(x - s(x)) f(x,a) dx \qquad (1)$$

subject to

$$\int [U(s(x)) - V(a)] f(x,a) dx \geq \bar{H} \qquad (2)$$

and

$$\int U(s(x)) f_a(x,a) dx = V'(a) \qquad (3)$$

Let lambda be the multiplier for (2) and mu be the multiplier for (3). Pointwise optimization of the Lagrangian yields the following characterization of an optimal sharing rule.

$$\frac{G'(x - s(x))}{U'(s(x))} = \lambda + \mu \cdot \frac{f_a(x,a)}{f(x,a)} \qquad (4)$$

Figure 4. A mathematical query

From John Howard <jhh+@andrew.cmu.edu>
To Jim Morris <jhm+@andrew.cmu.edu>
Barry Silverman <barry+@andrew.cmu.edu>
Jonathan Rosenberg <jr+@andrew.cmu.edu>
[illegible] Walt Schmidt
Subject EXPRES Deployment Configuration
CC Carolyn D. Council <cdc+@andrew.cmu.edu>
Paul G. Crumley <pgc+@andrew.cmu.edu>
Mike Kazar <kazar+@andrew.cmu.edu>
[illegible] John Howard

Folks

Here is the penultimate pass on the EXPRES deployment configuration. If I don't hear differently from anybody within the next day or so I'll be placing an order for five. Jim mentioned more, but didn't specify more than 5 deployment sites. Ordering 5 now gets us into the queue and should leave some money for more later if needed.

Machine/Feat	Description	List
[illegible]	Desktop RT APC	$[illegible]
	Includes 70MB disk, 4MB [illegible]	$0.00
#[illegible]	Mouse	$195.00
#4[illegible]	Extended monochrome display adapter	$1450.00
#[illegible]4	4MB memory expansion	$3800.00
#4[illegible]	Streaming Tape adapter	$430.00
#[illegible]	SCSI adapter	$1080.00
#[illegible]	SCSI cable	$95.00
#[illegible]	Ethernet adapter	$850.00
[illegible]-001	Moniterm 19" display	$1395.00
[illegible]	300MB Disk	$6500.00
[illegible]-001	Streaming tape drive	$1795.00
5799-[illegible]	IBM ACIS 4.3 for the IBM RT PC	$1715.00
	TOTAL	$31905.00
	LESS 40%	$19143.00

Remarks

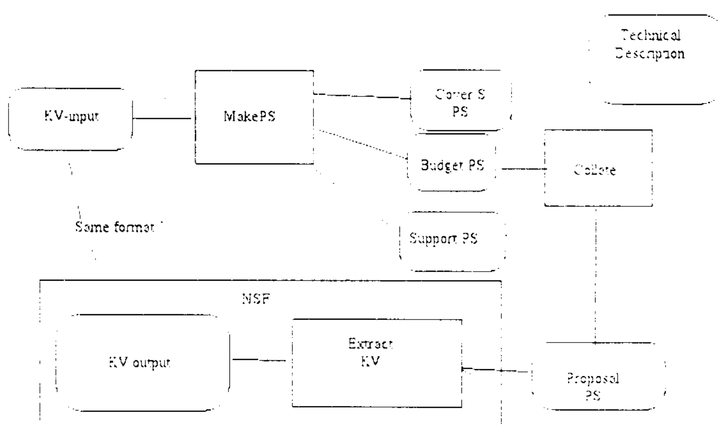
This machine is configured to be a complete standalone workstation with the potential for later expansion to be a file server if desired. It is a configuration which is fully supported by 4.3 and provides enough disk space for a standard standalone installation.

Figure 5. A table of numbers

From Jim Morris <jhm+@andrew.cmu.edu>
To Chao-Ping Yang <cy0C+@andrew.cmu.edu>
Subject Re: PS-EXPRES
CC Barry Silverman <barry+@andrew.cmu.edu>

Don't worry about the file being too big unless we're talking about additional megabytes.

I don't see any great value in storing them as PS field-value pairs. Here is a picture of what I am thinking about.



The square boxes are programs you are going to write. There is no strict requirement that the KV input and output be the same format since entirely seperate groups of people will deal with them, but it might simplify the job.

Where do you want the Laser printer. It may take some time for a workstation, but we could certainly get you an entry badge for the ITC where there are lots of spare machines.

Figure 6. A technical discussion with a drawing

Presenting information in tabular form is an important feature. Andrew has a very general table construct that subsumes the functions of a spread sheet. It is more general because virtually any kind of object -- text, drawings, rasters, etc. -- can be included in a table -- not just numbers and labels. Figure 5 shows part of recent discussion about workstations equipment.

Including drawings in messages has been a useful tool for programmers discussing technical problems. Figure 6 shows a message I recently sent to an assistant outlining the design of a program.

One of the flashier kinds of objects one can include in documents is animated stick drawings. Figure 7 shows an animated version of the Carnegie Mellon logo. Only two steps of the animation sequence are shown. The animation facility has been used to create a variety of demonstrations of various concepts.

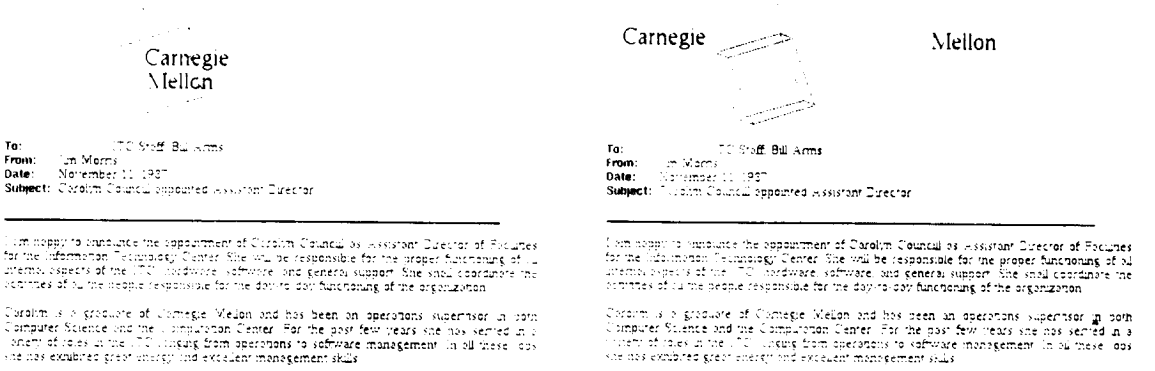


Figure 7. An animated letterhead

The availability of these facilities in Andrew has not yet caused an explosion of usage. There are several plausible reasons. First, it is a new feature and it takes time for its availability to spread. Second, one cannot send multi-media messages outside Andrew, so the audience is limited. Anyone intending a message for non-Andrew users will refrain from using non-text elements in their communication. Third, it takes a certain amount of skill to produce non-text communications and people are not generally trained to do it. The average person might be able to draw a picture on a piece of paper, but today's computer drawing tools take further learning.

Significant Usage by Groups

Non-classroom Discussions

There are about twenty different bb's devoted to particular university classes. They provide an excellent medium for class discussions. Several instructors actually require their students to use the system. The discussions can take longer and be more thoughtful since the system allows more time and requires all communication to be written.

In one undergraduate required course bbs were established by the teacher for perfunctory purposes like exchanging information about assignments, class meetings, schedules of tests, etc. At the beginning of the course, many of the posts were short, to-the-point, and informational -- written by by the professor and teaching assistants. By the middle of the course, students were the main posters and were helping one another with particular problems, discussing course content, and even criticizing the course. Several posts by students were made every day and some students indicated that they saw reading the class bb as an integral part of doing their class work since that is where they got real help, advice, and sympathy from their classmates.

In one graduate course, the professor set up a course bb and required students to make several posts during the semester on their reading. For the most part, these were students who had little prior experience with email, and several were somewhat skeptical about computer technology. Students did the required posting, but soon the bb took on a life of its own. Students in the course reported that this was where they actually learned things about their classmates and did a lot of their intellectual stretching. This is evidenced by a couple of things: the students continued to post on the bb through the following semester; and the students created a new, independent bb of their own to discuss and organize details of various research projects. Several of the students said this was one of the most valuable parts of the course, and several who don't much like Andrew in general said they would use the system because of this bb.

The Comments program[11]

Each year several sections of Freshman writing are taught using a system called Comments. It uses the Andrew editor with a facility that allows someone to annotate the text without altering its layout: a small triangular mark is placed in the text and the comment is place in an separate window. Figure 9 shows how it looks. The upper panel contains the paper and the lower one holds the comments. This allows the reviewer of a paper to comment at length on a point without the usual limitations of margins, etc. The mail system and file system are used by students to submit their writing to the instructor and their fellow students for comments. Thus a student or a teacher in the class may receive several samples of writing each day and can annotate each easily and return them without leaving his or her office.

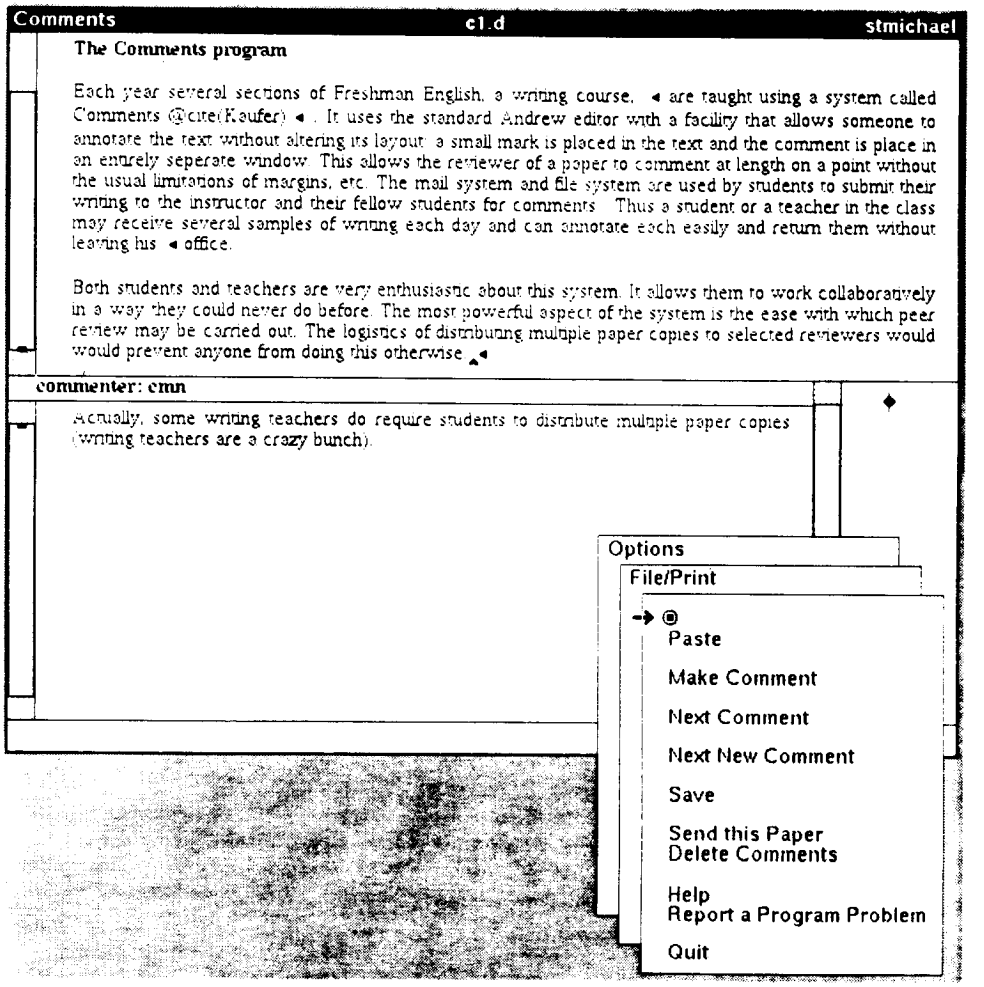


Figure 9. This paper annotated by the Comments system

Both students and teachers are very enthusiastic about this system. It allows them to work collaboratively in a way they could never do before. The most powerful aspect of the system is the ease with which peer review may be carried out. Some teachers began to encourage more frequent submission of drafts because the system is so much more convenient. The logistics of distributing multiple paper copies to selected reviewers would discourage anyone from doing this otherwise. The program has also been used by other classes and even individuals as a way of soliciting help on scholarly papers.

The Advisor System[12,13]

As the statistics in Table 1 indicated, some of the most widely posted-to bbs are the Advisor bbs. They are used by the staff that gives assistance to users. A user with a question just sends it to *advisor*. The mail is actually directed to an *in-box* bb devoted to the particular day of the week it happens to be. The advisor on duty that day, an undergraduate, reads this bb, and attempts to provide advice by return mail. If the question is too hard he cross-posts, i.e. forwards, the message to a subject specific bb, called a *help box*; there are about 35 such help boxes. These bb's are read by various experts who help resolve the question. Figure 10 shows the communication paths. Often the subject bbs will become forums for discussion of the problem and related issues. Most of these bb's are readable by a wide audience, including the managers of the system, so the quality of the advice as well as the major problems with the system are available for all to see.

The striking feature of this system, in contrast to a typical office system, is its openness and ability to accommodate structureless workgroups. To be sure, the daily in boxes have assigned readers and their duties are clearly defined. Similarly, their supervisors have a fairly clear role in monitoring the communication and resolving difficulties. However, the help boxes are read by a more-or-less voluntary group, people who have knowledge and are willing to share it and others who are generally curious. An undergraduate who once read a long technical argument by some experts on a bb remarked that it had been one of the most edifying experiences of his computer science education!

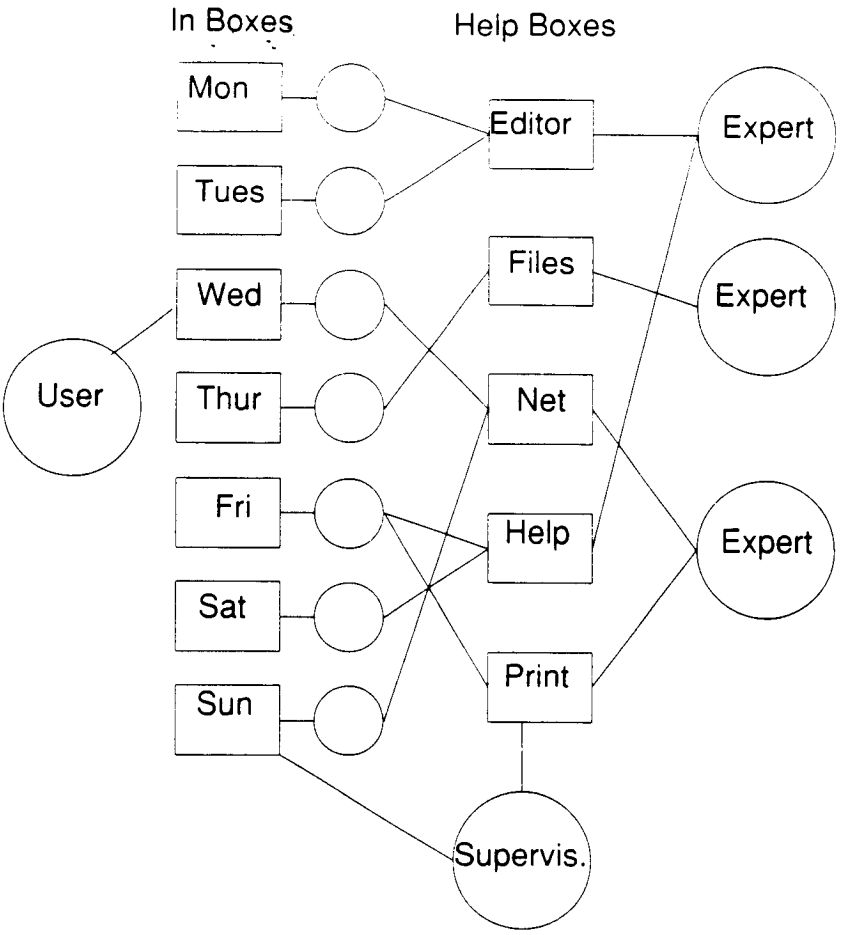


Figure 10. Advisor Communication Paths

Software Distribution

I recently sent out a short message offering some free software for a simple, somewhat specialized task. The message was sent to about four national distribution lists representing groups that use and share various sorts of document preparation software. Within a week, over a hundred people had responded, requesting the software; and I sent it back to them by electronic mail. I have since had several conversations with the users of the software and made improvements to it and sent it out again. Some people enhanced the software

and sent it back. This phenomenon is not uncommon. Development coordination of the M.I.T.'s X window manager depends heavily on the network.

General Characteristics of Electronic Mail

What does it take for an email community to flourish?

First, all the correspondents must be sitting in front of a computer or terminal a significant percentage of their working day. Often they are computer programmers or workers in a highly automated office and work with computers extensively. At a minimum they must look at their mail every day.

Second, they must have reliable connections. It is nice if these connections are high-bandwidth, but not necessary. Many mail communities, especially the UUCP group, manage using low speed links.

Once these conditions are met one can expect the use of email to grow. A typical user may receive 10 to 20 message a day and send 4 or 5. Many of the message he or she sends may be relays or excerpts of received messages.

When it works email displays some striking advantages over conventional mail or the telephone:

Speed with asynchrony. The system can deliver information as quickly as the telephone, but does not require correspondents to synchronize themselves to communicate -- the game of telephone tag. A conversation can proceed at whatever rate the participants desire, from interactions every few minutes to messages exchanged weekly or monthly.

Plasticity of Information. Since all the information that arrives electronically is processable by computer it is very easy to store, retrieve, and modify. Thus one can forward excerpts of messages to others, incorporate contributions from several sources into a report, and generally process information very easily. A request for some information may be forwarded to someone else. A question of policy may cause him to copy a part of a message to several other people and ask for opinions.

Wide, Accurate Distribution. It almost as easy to send a message to many people as one in a mail system because the sender need not make the copies. This makes it very natural to expand conversations and encourage group discussions. One can use either distribution lists which pin-point mail to multiple people or bbs which keep all messages in a central place. Each has different computational and social properties. One can reach hundreds of people around the world with a message and have them reply just as easily.

There are two striking phenomena that occur in a well-developed email community:

Unstructured communication paths. As Sproull and Keisler[14] pointed out, electronic mail tends to break down organizational structure and status. There seem to be two reasons for this. First, the power and convenience implied by the foregoing points make communication something anyone can do. The traditional role of the manager as the only person with access to information is diminished. The second reason is that electronic communication provides very few social clues to the correspondents about their relative status; one doesn't go through layers of secretaries or see any titles on letterheads.

Intensive querying. "Don't search, ask!" has become the motto of many email users. Given a question of contemporary interest it is often more effective to send the question to a distribution list or post it on a bb than to search libraries, data bases, or magazines. There are a lot of people on the networks who have the knowledge, time, and attitude to answer questions by people they've never met. Obviously, the best questions to ask revolve around computer use and products, but other kinds of questions find answers, too. On colleague of mine claims to read nothing less than a year old because it all can be picked up in conversations.

Universal Email

Although email is wonderful when it works, it is still a rather limited phenomenon. The community reported on here is a high-tech, academic group that has free access to email and a predilection to use new things. Although commercial email

seems to be coming of age, with an estimated 1.5 million mailboxes[15], it still tends to flourish only inside companies with extensive computer networks.

What will have to happen for email to become a universal communication service like real mail and the telephone?

Better Transport

The Integrated Services Digital Network (ISDN) will help, but will take a long time to spread. Fortunately, it is not crucial. As examples like UUCP and Facsimile (FAX) show, the existing telephone system is serviceable.

There are more than 15 million personal computers in the world. Why aren't they the basis for a viable email community? One product that might help is a *mail modem:* A conventional modem for a personal computer that has enough storage to hold a day's worth of messages, a red light to tell you when there is mail, and (the hard part) a mail protocol to speak to personal computers. The need for this product arises from the simple fact that if a PC is turned on it is busy doing something besides waiting for mail.

In the long run, however, the equivalent of Post Offices must appear. One needs a place to hold the mail reliably and help locate people. Currently, there are commercial email services that serve these needs -- MCI, Sprint/GTE, CompuServe, Western Union, etc. A current difficulty is linking these systems together. The equivalent of the telephone information system is nowhere to be seen. Getting connected to one of these services at all often involves one in technical problems beyond one's ability. The current divestiture muddle in the U.S. seems to have the Bell Operating Companies off balance. One might expect leadership from the PT&T's; e.g. Minitel.

Exchange Standards

A glaring defect of email is the fact that single-font ASCII text is the only widespread standard. Except for special communities, like Andrew users, no one can exchange electronic messages with pictures. Email must catch up with

electronic publishing. Exactly how it will happen is not clear.

The most significant technical/political challenge facing the email world is to develop a way to exchange documents that allow different kinds of information to be included while preserving their plasticity. There are several candidates. The *Office Document Architecture* (ODA), an emerging CCITT standard, allows one to specify the form and content of a multi-media document containing structure, multi-font text, rasters, and geometric graphics. In principle ODA allows revisable documents to be exchanged, edited with different systems and returned to their original authors for use in the original system. The American Association of Publishers has developed a set of conventions based up the Standardized General Markup Language (SGML) that specify the general meaning and structure of documents without controlling form. MicroSoft's Rich Text Format which links their word-processing products for different personal computers is likely to be an important *de facto* standard. People also exchange documents in the form of TeX, Scribe, troff, and the input languages of other document processors.

As part of the EXPRES project, Carnegie Mellon and the University of Michigan are developing software to support the use of mail by scientists. The major thrust is to promote the interchange of documents created by different multi-media document preparation systems. Our first goal is to link the Andrew system with Diamond, a comparable document editor, through ODA.

A less complicated approach is to transmit documents in non-revisable form. This approach is represented by *de facto* standards like PostScript, a language for the representation of printable documents. They specify the precise appearance of a page without commitment to a resolution, leaving print quality to the device. Text and geometrically described graphics can be represented economically, so that the average page requires 4K bytes of storage.

A near-term activity of the EXPRES project aims to support the transmission of NSF proposals in PostScript. The general idea is that people can create proposals using their favorite document preparation system which is probably capable of producing PostScript output for printing purposes. Then, with some extra effort and software, they can integrate the standard NSF forms with their

proposal to produce a completely digital document. It can then be transmitted to the NSF and on to reviewers who have the proper printers. The fact that a Postscript document is not editable can be an advantage. Unlike most email, a proposal is a formal document and most authors like to exert control over its appearance as well as content. However, by itself this approach is not adequate for the full capabilities needed by email.

Facsimile

The CCITT Group 3 and Group 4 specifications are a successful standard for transmitting images with fixed rasters in the range 200 to 400 dots per inch. The average page requires 40K bytes.

While FAX precludes revision and is relatively bulky, one can't argue with success. Today there are 3 million FAX machines worldwide, most of them in Europe and Japan. It has no trouble with image and does not require tricky computing protocols. A reasonable case can be made that email will be swamped by FAX in the commercial world. Then the question is how one can add some of email's features to FAX. On simple approach is to create a store-and-forward service; at least one U.S. company does. This provides asynchronous communication and wide distribution, but no plasticity.

The growth of FAX should teach us something: stop waiting for everyone in the world to become computer users, typing their communication into word processors and creating pictures with drawing programs. That may happen someday, but there will be a long transition from the world of paper. Today's businesses, even ones with extensive email systems, receive much of their communication on paper.

Email systems should accommodate paper and FAX as input. Advances in storage technology have made it easier to cope with image documents. Storage costs will continue to plummet so that storing all one's paper in FAX form is not crazy. We estimate a full page's image can be saved on an optical disk for about 5 cents today.

Character recognition technology has advanced to the point where it can find well

over 90% of the words on a printed page. While this is not good enough for one to throw away the original, it is more than adequate to feed text to a retrieval system for the documents.

Retrieval and Filtering Tools

The amount of new information now available from mail and bbs is overwhelming. Some people eventually give up reading mail simply because there is too much of it. The traditional tools for dealing with overload -- assistants and secretaries -- can be employed, but it also seems appropriate to try such computer tools as data base systems and information retrieval on the mail problem. Malone's Information Lens[16] is an example, but the extensive work in library retrieval systems should also be brought to bear on this problem.

Acknowledgments

The development of the Andrew system was supported by the IBM Corporation in a joint effort with Carnegie Mellon. The National Science Foundation, under the EXPRES project (grant number ASC-8617695) has funded enhancements to Andrew and work on document standards.

As the long lists of authors in references 1 through 10 imply the Andrew system has resulted from the efforts of many people. The Message system was originally designed and developed by Jonathan Rosenberg, Nathaniel Borenstein, and Craig Everhart.

Parts of this paper were derived from other writing and advice of Jaap Akkerhuis, Nathaniel Borenstein, Craig Everhart, Christine Haas, David Kaufer, John Leong, Dan Muchnok, John Myers, Christine Neuwirth, David Nichols, Jonathan Rosenberg, Bob Sidebotham, and Chris Thyberg.

References

1. James H. Morris, Mahadev Satyanarayanan, Michael Conner, John H. Howard, David S. H. Rosenthal, F. Donald Smith, "Andrew: A Distributed

Personal Computing Environment", *Communications of the ACM*, March 1986.

2. James H. Morris, John Leong, David Nichols, Michael J. West, Mahadev Satyanarayanan, "Andrew: Carnegie Mellon's Computing System", *IFIP Proceedings*, September 1986.

3. James H. Morris, " 'Make or Take' Decisions in Andrew", *Proceedings of the USENIX Technical Conference*, February, 1988.

4. John Leong, "Nuts-and-bolts guide to Ethernet installation and interconnection", Data Communications, September, 1985,**14**,10

5. John Leong, "Data Communications at CMU", Report CMU-ITC-85-043, Carnegie Mellon University, July, 1985,

6. John. H. Howard, "An Overview of the Andrew File System", *Proceedings of the USENIX Technical Conference*, February, 1988.

7. John H. Howard, Michael L. Kazar, Sherri G. Menees, David A. Nichols, Mahadev Satyanarayanan, Robert N. Sidebotham, and Michael J. West, "Scale and Performance in a Distributed File System", in *Proceedings of the Eleventh Symposium on Operating Systems Principles*, November, 1987. *To appear in ACM Transactions on Computer Systems.*

8. Andrew J. Palay, Wilfred J. Hansen, Mark Sherman, Maria G. Wadlow, Thomas P. Neuendorffer, Zalman Stern, Miles Bader, and Thom Peters, "The Andrew Toolkit: an Overview", *Proceedings of the USENIX Technical Conference*, February, 1988.

9. Jonathan Rosenberg, *et al.*, "An Overview of the Andrew Message System", *Proceedings of SIGCOMM '87 Workshop*, Frontiers in Computer Communications Technology, Stowe, Vermont, August, 1987.

10. Nathaniel S .Borenstein, Craig F. Everhart, Jonathan Rosenberg, and Adam Stoller, "A Multi-media Message System for Andrew", *Proceedings of the USENIX Technical Conference*, February, 1988.

11. Christine M. Neuwirth, David S. Kaufer, Gary Keim, and Terri-Lynn Gillespie, "The Comments program: Computer support for response to writing". Center for Educational Computing in English, Technical Report CECE-TR-3, English Department, Carnegie Mellon University, Jan., 1988

12. Nathaniel S. Borenstein and Chris A. Thyberg, "Cooperative Work in the Andrew Message System", *Proceedings of the Second Conference on Computer Supported Cooperative Work*, September 1988

13. Chris A. Thyberg, "Advisor -- An Electronic Mail Consulting Service", *Proceedings of the XVI AACM SIGUCCS Conference*, to appear.

14. Lee S. Sproull and Sara Kiesler, "Reducing social context cues: Electronic mail in organizational communication". *Management Science*, **32**, 1492-1512, 1986

15. J. Tannebaum, "Speed, Cost And Cachet Aid Growth Of Electronic Mail", *The Wall Street Journal*, July 27, 1988.

16. Thomas Malone, *et al.*, "Intelligent Information-Sharing Systems", *Communications of the ACM*, May, 1987.