ENVISIONING A NATIONAL UNIVERSAL E-MAIL SYSTEM

Joel Pliskin

Christopher Kedzie

P-7913

Published 1995 by RAND
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138

OVERVIEW

An important purpose of the Universal E-Mail project is to consider the policy implications of a Universal E-mail system - an entity which does not yet exist. Among the competing interests and the interested observers, conceptions of "Universal E-mail" vary considerably. An important preliminary task for us, therefore, is to elucidate our own vision of this system that we expect will pose profound challenges to system designers as well as analysts and practitioners of public policy.

To envision the characteristics of a future national Universal E-mail system, there are two possible methodological approaches. One can either extrapolate current technology trends forward, or deduce system capabilities from desired objectives. This paper uses the second approach, both because this method will ultimately describe the system that would be of greatest demand and because the analysis will be convergent. A potential risk of this method would be a vision too narrow and exclusionary. Our objective is to imagine a system, encompassing and enabling, which will encourage rather than limit creative applications of wide availability and use. Therefore, we start with a broad conception of e-mail.

The function of "e-mail" or an "e-mail system" is to facilitate asynchronous electronic communications between individuals and groups for its own sake. Here, we use "asynchronous" to mean that the sender of a message and its recipient are not necessarily using the system simultaneously. Within these wide bounds, we explore and catalogue various categories of communication. Then, we examine and describe the services necessary to support these categories of communication. We make no assumptions about whether these services must be provided at the system level, may be provided by the "after-market", or both. We seek, primarily, to encourage a common understanding with a useful vocabulary and framework for further discussion and analysis.

Our point of departure is a future in which a national e-mail system exists and is being used productively by people in their daily lives. We assume that technical challenges which may complicate the creation and development of such a system have been overcome. The system, for example, is not limited to carrying any particular data type. We explicitly assume that the e-mail system treats all data types as digital bit streams regardless of whether the data stream actually represents ASCII, voice, video, images, etc. In fact, we expect that applications residing on this system allow users to combine data types in diverse ways.

Universal access is granted by *fiat* and is presumed to be roughly equivalent to that of today's telephone and postal systems. In this future, most people have access from their homes or places of business and others can access through public terminals and rented e-mailboxes. The universal access assumption circumvents the question of how, or even if, people pay for their access to the system. We do explicitly assume, nevertheless, that business is transacted on the system and that the system (including any value-added services) supports these financial transactions. Our perspective is that of the user; the role of other actors, such as the government claiming a privilege with respect to privacy, is beyond the scope of this vision.

Communication Taxonomy

Our most basic assumptions are that the system exists and that people are using it actively in their daily lives. What are people doing with this system? How has e-mail supplanted and supplemented more traditional forms of communications in daily life?

It is not possible, or even desirable, to list all specific uses or all possible applications that might be present in a future system; this would limit our thinking to those uses that can be imagined today. It is possible, however, to envision several general categories that embrace the spectrum of traditional types and purposes of interactive communication. We can then concentrate on the user functionality and the underlying services necessary to enable and encourage the creation of useful applications within these broad categories.

An alternative approach would have been to define the set of orthogonal dimensions that completely describe an n-dimensional e-mail space. Sender characteristics, recipient descriptions, delivery means, commercial aspects, levels of urgency, degree of authentication, extent of privacy, and various other relevant dimensions geometrically expand the number of cells in the n-dimensional table beyond that which can be easily discussed in this short treatise. Many of the cells would be functionally redundant; unsolicited advertising - independent of specific characteristics of the sender and the recipient — could be considered "junk mail," for example. Conversely, many other cells would be nonsensical and thus irrelevant; inclusion of commercial transactions with zero authentication, for instance, would contribute nothing meaningful to an understanding of the future of e-mail.

While this "dimensional approach" may have certain descriptive utility, it fails to offer the sort of prescriptive insight we seek. Necessary system services derive from the purposes behind the messages rather than from a descriptive mapping of the messages themselves. In the postal world, a special delivery letter is distinguishable from an ordinary letter only in the intention of the sender. Therefore, our approach is use-based. We choose to concentrate on the handful of cells which are most likely to be representative of how a future Universal E-mail system will be used. The downside of this approach is overlap; certain messages could, arguably, fit into more than one cell. Nevertheless, this hazard posed by potential variances of interpretation does not obscure the original objective — to develop a useful framework for further analysis.

Five general categories are listed below. The first three "directed" communication categories constitute a continuum of increasing demands for system surety and security. The notion of "directed" implies that the originator purposefully directs the initial transmission to a specific audience over whose membership he or she has direct control, or at least detailed knowledge. The latter two categories are "undirected" messages, which can be thought of as "launch-and-leave." The sender has minimal control over readership and

may not even be aware of everyone who will have access to the message once it is posted.

Directed

** D-1*

D-1 comprises casual communication for which today's Internet mail represents a useful baseline. There is no requirement for, or promise of, special handling. This is ordinary message traffic such as, but not limited to, personal notes between acquaintances, collaborative efforts between colleagues, and information requests between people who may never have met face-to-face. These messages originate from a single source (e.g. a person, an organization, a corporation, etc.) and are either sent to an individual or are multicast to an originator-defined receiving group (e.g. a group of friends, a project team, an entire company, etc.)

"Best effort" delivery, such as that of current Internet mail protocols, suffices. Neither proof of sending nor proof of reception is essential. Nevertheless, use is predicated on a basic expectation of reliable delivery despite the absence of delivery guarantees.

** D-2*

D-2 can be conceived of as verifiable communications. These differ from the casual communications of D-1 in that at least one of the message attributes is critical and must be ascertained, with a high level of certainty, by the receiver, the sender, or both. Best effort delivery is insufficient for verifiable communications; the system depends on guaranteed delivery. Attributes subject to verification might include the time of dispatch, similar to a USPS postmarking; proof of sending, such as registered mail; proof of reception, frequently referred to as a return receipt; integrity of message contents, assurance – without a contemporary analog – that a message has not been modified either accidentally or intentionally; and authentication of the sender identity.

Examples of D-2 communications are common and diverse. To conclude legally binding contracts between users, the identities of the parties involved, as well as the timing of their communications, must be

verifiable. To file personal income tax returns by midnight of April 15, postmarks must be inspectable and message contents verifiably tamperproof. To serve legal processes, receipt must be provable. To vote in a general election, the identity, citizenship, and age of the voter must all be subject to confirmation.

** D-3*

D-3 encompasses communication in the commercial realm. The marketplace is among the oldest of human institutions and trade is among the most pervasive forms of communication. In this future perspective, e-mail is widely used for the purposes of buying and selling products and services. Whenever money changes hands, security concerns heighten. Demands for guaranteed delivery and information reliability can be more stringent than for either D-1 or D-2.

A commercial transaction can be dissociated into three components. In the e-mail context, these three may take place simultaneously or at different times, and in any order. First, an authenticated agreement between buyer and seller is recorded. This is analogous to the establishment of a contract between the parties and is subject to the same requirements. Second, the buyer transfers payment to seller. Third, the seller delivers goods to the buyer.

The broad D-3 category of commercial communications can be divided into two sub-groups: sale of "soft" and "hard" goods. Soft good transactions are the sale and purchase of electronic data stream commodities. Soft goods include information purchased from private databases, such as stock quotes from Dow Jones, software, videos, text, and anything else that can be delivered to the buyer in digital format. Soft goods sales can be transacted entirely within the e-mail system. Pricing of these items can be connected to direct measures, such as connect time or volume of information delivered, or they may be priced as individual items just like hard goods. Payment schemes may be similar to that of current computer network services like CompuServe or to that of telephone companies for 976 calls. Delivery is completed electronically.

The second type of commercial communication involves the sale and purchase of more traditional hard goods, which cannot be electronically

delivered. In these cases, the e-mail system facilitates the payment mechanism through credit validation and approval codes, or through electronic transfer of funds between the buyer and seller. The system must enable the bank or credit company to validate the serviceability of the account to be debited, and enable the buyer to verify his or her authority to access that account. Further, the system must protect account information and other personal data from unauthorized users. Product delivery can be accomplished by any of the usual delivery mechanisms, such as the post office, FedEx, or UPS.

Undirected

Not all messages are conscientiously guided. Many are simply delivered into various orbits in cyberspace to be snatched (or run into) by receiving parties. There are two general types of "undirected" messages.

** U-1*

U-1 denotes messages posted on virtual walls where anyone who is authorized can access and respond. Messages do not have a directly specified recipient and discourse can be multilateral. Message recipients are self-selected. They actively access posted messages, whether one time, like subscribing to a Bitnet list-serve, or each time, such as reading a Usenet newsgroup. The content of posted communication may be traditional message traffic, remote access to data, or both. Commonly referred to as electronic bulletin boards, this type of communication traditionally facilitates the formation of virtual communities based on shared interests. Other contemporary examples include the many private bulletin board systems currently in use, as well as public ftp, gopher and world wide web sites.

U-1 communication forums may be public, allowing general public access, or private, with restrictions on who is allowed to access the information contained. Both private machines connected to the system and components of the Universal E-mail system itself support posted communication. System operators may accept subscriptions and regulate postings. Participation may be free or for a fee. Some characteristics

of fee-charging for posted communication overlap and complement commercial communication.

** U-2*

U-2 may be conceived of as broadcast communication, the unilateral dissemination of information to passive recipients. Here, the distinction between public service and private interest is particularly important. In a public service mode, these messages may appear when the user engages his e-mail agent. The "emergency broadcast system" is a radio and TV analog of a public broadcast. A most common contemporary e-mail example is the "message of the day" on Unix systems, which a user receives when he or she logs on. System administrators invoke messages of the day for items of sufficient importance or strong general interest to the entire system audience, such as planned system outages.

In the private interest mode, U-2 messages resemble unsolicited mail. Advertising is a dominant use of such broadcasts. Targets are often identified by recipient attributes such as age, gender, occupation, and location, according to the wishes of the sender. When addressing, the sender exercises discretion over the recipients, but the fine line separating D-1 messages from U-2 messages is the sender's knowledge of the recipient. In the latter case, the sender's knowledge is limited to recipient attributes, not specific identities.

Although the differences between these message types may seem subtle, ramifications diverge significantly in the system service requirements (as well as in the policy realm). Broad-ranging implications extend from the specification of fields within the directory databases; to privacy issues, for which some users demand the capability to actively block broadcasts in which they have no interest; to practicality concerns, which preclude general public access to broadcast to-all-subscribers messages.

Classification of Services

Within this future vision, the categories above demonstrate the variety of ways in which people are employing asynchronous electronic media for myriad purposes. To attain this vision, certain questions are consequential. What are the service prerequisites? Through what underlying service provisions does Universal E-mail gain public acceptance and wide use?

A general class of services facilitate the usage and universality of the e-mail system. Additionally, the five categories of communication illuminate the need for specific services. Some of these services are built into the e-mail system itself. Others may be offered as value-added services by a new class of e-mail based businesses. Our task in sketching this horizon is to identify the key distinctive features. No attempt at clairvoyance is undertaken here to discern at what system level the services are maintained or to distinguish the future basic services from the enhanced. That partitioning will result from deliberate policy decisions, market forces, and historical accident.

* *Addressing Schemes*

A scheme that allows any user to address messages to any other user is a most fundamental characteristic for all categories of communication in a universal system. Future addressing must be consistent, simple, transparent and intuitive. The Internet addressing model, called the *Domain Name System*, meets these basic requirements reasonably well up to the host computer (all the information to the right of the "@" sign in the address). User names (to the left of the "@" sign), however, are more arbitrary.

Future addresses must be reciprocal, i.e. message recipients should be able to reply to senders via the message origination information. Further, future addressing schemes should support dynamic binding. Today's statically bound addresses direct messages to particular computers which are physically connected with a certain entity (e.g., a business, university, etc.) and are in a particular geographic location. Smarter addresses stay with and follow the user.

Probably the best contemporary analog is telephone's "700 numbers" which eliminate geographic area codes and ring wherever the owner resides.

** Delivery Guarantees*

The current Internet delivery model is "best effort." This means that, while the expectation is high that the system will reliably transfer messages from originators to recipients, there is no guarantee of success. Nor is there any promise that the originator will be notified in case of failure.

For D-1, U-1, and U-2 communication, best effort can continue to be the standard. However, this paradigm will clearly not be sufficient to support D-2 or D-3 communications; guaranteed delivery is essential. Guaranteed delivery entails more than surety of receipt. The integrity of the message contents must also be part of the guarantee. Neither inadvertent technical mishaps nor deliberate tampering can alter the body or attributes of the message.

Guaranteed delivery could be provided as an extra user-invokable service which might cost an additional fee payable by the sender or the receiver. Alternatively, delivery guarantee mechanisms could be automatically invoked for D-3 commercial applications. In this case, the cost could be incorporated into the transaction fee.

** Annotation Services*

Postmarking, registering mail, and providing return receipts are user services inherent to the guaranteed delivery model. They are functionally analogous to the existing post office services, and have the same general purposes. As these post office services are legally admissible evidence in court, so too must e-mail services be constituted so as to be admissible.

Other annotation services include message tracking and content verification. Via message tracking, a sender can know, in real time, the current status of messages anywhere along the path from dispatch to receipt; even, for example, that a message has been delivered to the recipient's mailbox but has not yet been accessed. Content verification allows proof that the message contains what the sender claims it contains.

** Billing and Payment Mechanisms*

The e-mail system must enable billing for D-3 applications and must also facilitate active payment mechanisms. To support soft goods transactions, commercial service providers such as CompuServe and Prodigy have developed methods for billing users based on direct measures, such as connect time or volume of information delivered. These providers can do so because they are centralized and can keep track of all of their users. An analog of this method will be necessary in the distributed computing environment of a future e-mail system.

For active payment, the preferred solution is a "double blind" on-line mechanism. Account information is transmitted directly to a bank or credit company, which then sends an approval code to the merchant. This provides a loosely coupled goods and service transaction mechanism. Each party is protected from fraud, and the seller does not need and does not have access to the buyer's actual account information. (It is noted, however, that the current model of purchasing with a credit card number by phone offers no such safeguards.)

** Identification and Anonymity*

Much of the communications encompassed by the all categories, D-2 and D-3 in particular, require that one or both parties be able to verify aspects of the identity of the other. In addition to the variety of examples provided earlier, this menu of uses also includes civil transactions, such as applying for a social security number, renewing a driver's license, registering to vote or obtaining an official birth certificate. In some of these circumstances, positive sender identification alone is not sufficient; additional specific characteristics, such as citizenship and age must also be verifiable.

Other transactions, conversely, are facilitated by anonymity or pseudonymity. Certain recreational uses of D-1 and U-1, and electronic cash payments in D-3 are representative examples. Within the system, designers could obstruct tracing the identities or origin by having removable "From:" lines. Alternatively, enhanced services might offer anonymous electronic P.O. boxes for which the mapping of box number to user identity is suitably protected or encrypted. If users are allowed multiple electronic identities, then many may wish to elect a mixed

strategy, with a public e-mail address where all unsolicited messages will go, and multiple private addresses which they can share with family, friends and business associates.

If it is desired to ensure the ability of individuals to remain anonymous, address conventions which give no hint of the real identity must be designed, and databases containing information that maps address to identity must be suitably protected or encrypted.

** User Privacy*

While verification of potentially sensitive user information is necessary to complete certain transactions, the protection of privacy is essential to guard the information from outside parties. (Primacy of privacy is postulated here with respect to other system users, not with respect to supra-system actors, such as the government, which may also have considerable interests.) The system must be able to discriminate and provide personal data only to those who have the right, the reason and the permission to obtain it. It must also be able to filter the specific information which is intrinsic to the transaction at hand. These privacy provisions pertain not only to the correspondents' personal data, but in many circumstances apply also to the message contents and attributes.

In the paper world, postal envelopes are to be opened only by the addressee and only the recipient and mail carrier are to know what is written on the envelope. The presumption of privacy in this country exists due to U.S. laws, postal tradition and the difficulty and high cost of tampering with mail undetected. Similar privacy protections are a greater challenge in the e-mail system where perfect copies may be easily made without disturbing the original. Nevertheless, people will be reluctant to use any system which does not preserve their privacy rights. Therefore, cryptography and other such protections must be high priorities for system designers.

** Directory Services*

The telephone system offers a directory service baseline. Anyone can obtain basic information about any other subscriber (who has not explicitly chosen to be unlisted) by knowing only the subscriber's name and area code, which is a proxy for geographic location. However, in

the context of an e-mail system, geography is no longer the only, or even the best, criterion by which one person would search for another. E-mail addresses, as a means of communication, are independent of the physical location of the individual. Within virtual communities, it may be desirable to list people's coordinates in a multi-dimensional database with extensive cross-references, including such non-traditional data as profession, organizational membership, personal interests, etc. As in the case of the current Yellow Pages, users could determine by which attributes they would like to be listed.

Again, privacy concerns pose a counterpoint. In the future, family, friends, colleagues and non-acquaintances should each have differing levels of access to personal data. Thus, the identity of the person initiating the query should determine the amount and type of information which the directory service will make available. This discussion is further complicated by the possibility of multiple electronic identities and anonymity, which differs subtly, but substantively, from unlisted addresses. Unlisting refers to a being who prefers not to be found. Anonymous messages, like political leaks, are *ad hoc* and content specific; a public persona attempts to hide his or her role as an originator of a certain transmission. Listed identities can send anonymous messages and unlisted identities can send non-anonymous messages.

Synthesis of the global directory database is also an issue to be resolved. Today, it is the responsibility of each local phone company to maintain its own database. Since each database is self-contained and is primarily available in the form of a book or through directory assistance, this is not a major contemporary problem. Once the global directory becomes available on-line to each subscriber, maintenance and synchronization become important technical considerations. A system modeled along the lines of the current domain name or gopher servers may be appropriate, i.e. each piece is separately maintained, yet the collection appears to the user as a seamless whole. The search for arbitrary listings ought to be no more difficult than an analogous search through the yellow pages, the white pages, or directory assistance.

** Interfaces*

The user interface is key to the public exploitation and acceptance of a Universal E-mail system. Usage will depend on the availability of applications and the quality of the interfaces. Compelling applications must be sufficiently available to encourage people to overcome innate conservatism and to try something new (and technical). Consistent and simple user interfaces must be of low enough complexity that people can learn one without referring extensively to manuals, and then feel comfortable using it. The frequent references here to contemporary communication systems underlines the importance of familiarity.

Given a heavily used system, market forces will drive the development of compelling applications (although something that regularly saves one a trip to the post office could be sufficiently compelling). The market will also deliver system enhancements which cannot yet be foreseen. It is further likely that the market will produce many user interfaces which will compete to become the standard, much as has happened in the personal computer marketplace.

If applications designers are to produce usable applications, however, system designers must first provide a good standardized Applications Programmer Interface (API) within which application designers can work. Similarly, for policy makers and policy analysts to get to work, they must first arrive at a compatible and comprehensive understanding of the types of use and services of future Universal E-mail.