Multiple Email Addresses: A Socio-technical Investigation

Benjamin M. Gross (bgross@uiuc.edu)
University of Illinois Urbana-Champaign

1 Introduction

This research investigates individuals’ everyday use of multiple electronic mail addresses. Email remains the dominant form of online communication. According to research by the Pew Internet & American Life Project, 93% of Internet users have sent or received email. [20] [15] In preliminary interviews I conducted, nearly all participants maintained multiple email addresses. These preliminary interviews, others’ research data, and a cursory analysis of market research data indicate that the use of multiple addresses is both common and mainstream. [13] The phenomenon is not well documented in either the technical or the popular literature, nor are the behaviors well supported within email applications. For example, many users employ workarounds to maintain multiple email addresses, but the process is often frustrating and error prone.

This research agenda arose from a series of preliminary interviews I conducted on the organization, navigation, management and analysis of email by individuals. It became clear that many email addresses had specialized uses and that the functionality available in email applications for maintaining these multiple email addresses was inadequate. My future research will help better inform systems design by examining a common user behavior in email systems.

In order to improve system design, the research seeks to answer the following questions. First, why do people maintain multiple email addresses? Second, what are the socio-technical explanations for this behavior? Third, what are the strategies and tools that people use to manage their own and others’ email addresses? How does user behavior vary across time and space and how do users organize their email collections. Finally, what are the implications of the existing strategies for the design of future email clients and servers?

2 Methodology

In order to answer my research questions, I need to investigate people’s use of multiple email addresses and strategies to deal with them. Data for this research are drawn from three sources: 1) two sets of exploratory interviews; 2) market research interviews and surveys; and 3) my dissertation research—a survey of high-volume email users. In order to understand how people use email addresses, I first conducted a set of pilot interviews to devise a semi-structured interview instrument. I used this instrument in a second set of longer interviews to determine the nuances of the users’ practices and the motivations behind them. I supplemented this information with three sets of market-research surveys that elaborated on these issues. Using the knowledge gained from the previous research I will design a representative survey that will supplement my qualitative data. The survey does not require a population that represents the general population, but rather one that represents users with sustained and substantial use of email.
I first conducted a series of exploratory interviews to develop my interview and survey instruments and to ensure they captured information that is relevant to existing user practices surrounding email addresses and addressing. The information gathered in the exploratory interviews was essential to developing relevant and pointed questions. From 2001 to 2003, I conducted three sets of preliminary interviews with e-mail users.

In the first series of interviews, I informally interviewed approximately two-dozen users from a variety of backgrounds. Although primarily involved in technical fields, the participants were roughly split between male and female users. I interviewed each participant for approximately 20 minutes to an hour. I asked the participants a broad range of questions regarding their behavior with email and the reasons for this behavior.

For the second series of interviews, I interviewed twelve participants in a semi-formal manner for approximately an hour and a half each. There were an equal number of novice and expert computer users. Five participants were male and seven participants were female. All participants had at least five years of experience with email. Their education ranged from a high school diploma to a Ph.D. candidate. Seven participants were employed in the information technology sector, while five participants were in non-technical fields. I asked a broader range of questions than the ones asked in the first series of interviews. I queried the participants about the time, frequency and reasons for email usage. For example, I asked how many addresses they actively maintained and their reasons for maintaining them. This second set of in-depth interviews provided me with a richer set of qualitative data and helped me formulate the questions on my survey instrument.

Next, I supplemented these interviews with three market research surveys and one market research interview. A large market research firm conducted the first survey in 2001 covering about 6000 respondents. The survey is somewhat biased for an older population although this can be corrected. The survey asked many general Internet usage questions. It asked respondents about the number of email addresses they maintained.

Two other market research surveys—one with 65 participants and another with 1390—were conducted by a market research firm for a software company wishing to introduce a product. The first of these two surveys was of 65 participants who were beta testers of the company’s email product. The users were ninety-one percent male and nine percent female. Forty percent of the users were in technical or information technology sectors, twelve percent in education and the rest were in a broad range of industries. The survey focused on questions related to information overload in the inbox.

The second survey for the software company consisted of 1390 respondents to an online request. The participants were sixty-eight percent female and thirty-two percent male. The participants worked in a range of professional service industries with less than 500 people. The survey is skewed towards blue-collar workers with limited computer experience. The questions were primarily focused on users’ workplace behavior.

I interviewed ten participants for a market research study to explore how users spend time with email. Each interview lasted an hour or more. The participants consisted of six males and four females. Most of the participants were academics and professionals in technology-related professions. Most participants processed substantial amounts of email. In particular I asked questions of how much time each task with email typically took for the user.

For my future research, I propose to carry out a survey of participants that send, receive and manage email at medium to high volumes and have at least two active email addresses. The requirement for higher volumes of email will shift the demographic to white collar, technical, scientific and
academic users that are likely to be more white, educated and male than the average population. The goal of the survey will be to quantify motivations for and effects of participants’ use of multiple email addresses and how these fit in the broader socio-technical email infrastructure. The survey questions will be heavily informed by information gained through previous interviews. In addition, I will perform an analysis of the market research studies to compare and correlate the findings with my own data.

3 Investigations of Multiple Email Addresses

In my preliminary research, participants provided a wide array of reasons for maintaining multiple email addresses ranging from the mundane to the sophisticated. The explanations spanned personal, social, institutional, legal and technical spheres. A coarse taxonomy of explanations for maintaining multiple email addresses follows. Some explanations overlap multiple categories in this taxonomy.

There are two high-level clusters in the socio-technical explanations given by users for maintaining multiple email addresses. The first cluster comprises personal and social explanations including: attempting to control the focus of their attention and limit unwanted interruptions. Maintaining multiple email addresses allows users to focus on a single category of communications or tasks without having to process or be interrupted by others. One major source of interruption is Unsolicited Commercial Email (UCE) more commonly known as spam. The first cluster also includes a range of personal reasons such as the separation of professional life from personal life, separation of social groups, the desire to gain status and prestige through affiliation and the desire for privacy. The second cluster includes the technical restrictions evolving out of institutional policy, such as security, archiving and, retention as well as legal and policy restrictions. The second cluster also includes factors such as problems with electronic mail relaying. I expand on each of these clusters further in the text.

3.1 Social Explanations

Users may choose a particular address based on the affiliation, status or degree of prestige it connotes. Typical examples include alumni or professional association addresses. Some addresses have a permanence or continuity that makes them valuable which often overlap with the goals of obtaining affiliation, status and prestige. University alumni email addresses are a pertinent example of the overlap as they typically provide the individual with both the status and prestige of the university connection as well as the benefit of long-term institutional stability and thus the likely permanence of the address. People correlate trust with permanence and continuity. This provides additional incentive for users to keep identifiers static over the long term. Having the same email address for a long time reduces the overhead for people to communicate.

Users select addresses based on the privacy, anonymity or pseudo-anonymity they expect the address will provide. Often, users wish to control the disclosure of identity progressively, for example, providing unknown or untrusted individuals very little information that links them to their “true name.” By using an address that has limited connection to various personal or professional aspects of their lives individuals can achieve some measure of control over the disclosure of their identity.

Issues of role often arise when discussing email addresses. Users manage multiple roles simultaneously. Several explanations users provided for maintaining multiple email addresses touch on the roles that the email addresses connote. In the context of this paper, I define a role as a life role:
for example, doctor, student, scoutmaster, professor, member of a professional organization, etc. These roles may or may not map to specific email addresses. Within the context of email, role and identity are often conflated, particularly since people tie identities to roles that are in turn tied to an email address. No guarantees of transitivity exist for any combination of role, identity or email address, meaning that you cannot rely on an address to map to a particular role or identity or vice versa.

In the literature on identity, there are roughly four ways that identity is conceptualized. This first is identity as a surface or external notion. The second is identity as an internal notion. The third is identity as a combination of external and internal notions. The final category is identity as a combination of external and multiple internal notions that separate the conscious from the unconscious internal identity.

Goffman’s metaphor of the theater is applicable to individuals’ use of multiple addresses. [9] Goffman said, “the individual in ordinary work situations presents himself and his activity to others, the ways in which he guides and controls the impression they form of him and the kinds of things he may and may not do while sustaining his performance before them.”

In this paper, I discuss only a surface and external notion of identity and do not consider internal representations. The theories on identity need to be adapted since not only do people lead multiple public and private lives, but they now also lead portions of their lives online as well as offline. The combinations of public, private, online and offline are often intermixed.

3.2 Email Addresses as Identifiers

A fully specified Internet email address is a rare category of identifier as it is individually and globally unique. Most other identifiers in the world need to be qualified to be unique: for example, a name combined with a physical address or phone number. While most email addresses refer only to a single person, this is neither reliable nor universal. Comparing email addresses as identifiers to other identifiers will be useful in modifying and applying existing theories of naming and identifiers. Email addresses are often overloaded with multiple identities that make analysis problematic.

An identifier is one component of an identity; it may or may not be unique. A unique identity is where a person is individually identifiable. For example, my email address bgross@uiuc.edu is both unique and individually identifiable while sales@somecompany.com is unique but not individually identifiable. My name, Ben Gross, is an identifier that is not unique as there are other individuals with the same name.

For the purposes of this research, it is important to distinguish an email address from an email account. An email address is globally unique and routable that is well documented in email standards. [10] I discuss only Internet email addresses and not those of any other messaging standards or systems. There are no standards precisely defining an email account. I define an email account to be an account on a server that is tied to an individual email spool. An email spool is the location where email is stored in a system after it is delivered and remains there until it is moved or deleted by either the system or the user. Accounts may have multiple associated email addresses. Each account will typically allow the user to both send and receive email. A user may be required to supply a username and password to either send or receive email from an account.
3.3 Technical Explanations

Institutions regularly create policies regulating email. Some of these policies along with governmental policies and regulations are implemented as part of the system infrastructure and enforced through technical mechanisms. [14] Examples of places in the system infrastructure where these implementations occur include: email relay restrictions, firewalls, spam filters, archiving (backups), retention (automated deletion), etc.

Users may choose to or be forced to modify their behavior based on these technical implementations. Users often maintain multiple email addresses to cope with institutional policies as well as technical problems. These additional email addresses become identifiers that the user is then forced to maintain. Standards such as SenderID, the merger of Sender Policy Framework (SPF) and Microsoft Email CallerID, as well as Yahoo! DomainKeys will likely require substantial changes in behavior from many users to cope with restrictions on mail routing. Users may create additional addresses to avoid or work around these policies.

Email relay failures are one of the most common difficulties that users face with email. The typical relay problem occurs when the electronic mail server rejects email sent by users who believe they are authorized to send mail. The failures generally stem from a standard configuration of email relays; where the relay will only accept electronic mail from machines within the same domain or IP address range as the server. Email relay configurations of this type are ubiquitous. Users who are on the same local network as the email relay, who are dialed into the same local network or who are using Virtual Private Network (VPN) software to the local network are typically not affected. Users who travel often need to send email from domains other than their mail server domain and therefore regularly have problems with email relays.

The most common workaround for people who have email relay problems is to send their email using an alternate account such as a Webmail account. The two most common mechanisms provided by system administrators for relaying electronic mail is to implement authentication on the email relay or to implement a VPN. When authentication is implemented on the email relay requires that users supply a login and password to the server to send email. [16] [17] Authentication is relatively simple to deploy although it increases the overhead for both the user and the user maintenance for the system. Secure authentication methods further complicate the process and are supported by fewer email clients. VPNs provide a secure network connection that effectively places the remote user on the local network and sidesteps any email relay restrictions. Problems with VPNs include substantial added complexity for end users as well as the system maintainers. There is often a requirement that end users have additional hardware or software to access the VPN.

A number of major Internet Service Providers (ISPs), require their users to send email using the ISPs provided address. This requirement forces many users to send email using their ISP address even when they would have preferred to use another address. The requirement serves two purposes for the ISP. First, the requirement simplifies its email relay policies. Since all relayed email must have the ISPs address, it makes it easy to distinguish which email should be relayed and which should not. Second, the requirement effectively provides an additional advertising outlet. However, users who wish to send email with an alternate email address must either use a Webmail account or another email relay. The majority of Webmail services require users to send email with the address of service provider. Again, these technical restrictions often cause users to maintain additional email addresses that are effectively additional identities.

Some sophisticated users will run their own email relays on the same computer they send email
from, but ISPs and institutions regularly block this behavior at the network level. Some providers such as Earthlink attempt to prevent users from employing alternate relays hosted by other providers or institutions. This is typically done by blocking all the traffic on email ports that is not destined for that providers’ email relay. Providers typically say both of these measures are to limit the ability for their networks being used to send out spam.

Networking restrictions such as routing, anti-spam software and firewalls may require users to take additional steps to access email. The use of a range of network and content security tools has been increasing for some time. Nearly all enterprise networks include substantial deployments of security technology that can influence email behavior. Anti-spam solutions such as SenderID and Yahoo! DomainKeys will increasingly force users to send email with the same domain in their From: header as their email relay. [21] The standards are still in flux, but they will have a major impact on how email is processed throughout the world as most major providers of email services have agreed to implement a form of these standards. Users may be required to use particular software such as a VPN client, a hardware token or a specific application such as Microsoft Outlook or Groove Network’s Groove Workspace.

Email archiving and retention policies are regularly implemented as part of an institution’s technical infrastructure. Typically, archiving policies for email are a component of an overall backup strategy where a copy of incoming and possibly outgoing emails are stored for a predetermined time. Some archiving implementations have a mechanism for excluding certain messages such as personal messages. Retention policies are implemented to limit the legal liability of stored email by automatically deleting email older than a certain date from system email store. Some retention implementations allow users to protect specified messages from deletion. In order to work around their perceived undesirable effects, some people attempt to circumvent both archiving and retention policies. People may use alternate email accounts, forward or copy email to other accounts and files to circumvent these policies, thus, splitting their email across multiple addresses.

Many systems use email as a simple and lightweight Authentication, Authorization and Accounting (AAA) substitute. [1] The systems require a user to send a message to the system from a specific email address or from an email address in a specific domain. The system then returns information or authorizes resources based on the address. For example, many systems will send users their forgotten password via a response to an email request. System designers make the assumption that only authorized users can both send and receive email from a specific email address. While this assumption is generally accurate, it is only minimally secure and robust compared to other options. On the other hand, these systems are simple, easy to maintain and easy to access. Systems that use email addresses for access control often cause users to maintain additional addresses that they would not otherwise keep.

4 Managing Multiple Email Addresses

Most modern email applications allow users to both check email from multiple accounts and to send email using multiple email addresses. This functionality is often labeled “roles” or “personalities” in the application. Downloading email from multiple accounts is relatively straightforward. Most applications allow the user to specify multiple email accounts to download email rather than requiring a separate application or Web interface for each account.

Selecting which address to use, when there are multiple potential addresses, is more complicated than receiving email from multiple accounts. The application typically changes the From: header in
the email message based on a predefined list of outgoing email addresses. Most applications cannot pair sending and receiving email addresses, meaning that the email application cannot necessarily recognize that the email address used for the outgoing reply to an email should match the address used on the incoming piece of email. The sending and receiving email addresses are not used for organization or retrieval of messages within the clients.

Although many email applications allow users to select from multiple accounts to send or receive email, participants expressed difficulty in managing multiple email addresses. Many participants considered it important to keep certain life roles and associated email addresses separate from other roles. For example, participants consistently reported being embarrassed by mailing a professional contact with a personal address. Participants who experienced a single case of role conflict altered their email usage to prevent the situation from happening again. For example, because of role conflict, one participant now maintains a separate role for her Ebay email to separate her shopping role from her work role. Furthermore, she accesses her Ebay email through a Web interface and her work email through Netscape Communicator. This configuration enforces a visual separation of the two roles so that it is more difficult to make an error.

The most typical mechanism users have for coping with multiple addresses is to forward many addresses to a smaller number of addresses. Users reported that it is cognitively and technically easier to have more receiving addresses than sending addresses. Despite the total number of addresses, no participant surveyed regularly checks more than three accounts. For example, the participant with the largest number of addresses forwards all his email to a single address. Due to forwarding accounts, participants typically have more addresses that receive email than addresses from which they send email.

4.1 Address Books and Identity

Multiple email addresses and associated identities and roles are increasingly mediated through address books. Early address books for electronic mail consisted of alias (or nickname) files. These files provided simple mappings between an easy to remember mnemonic and an email address, sometimes with the addition of the full name. Some email applications with a long history such as: Pine, Eudora and Netscape (Mozilla), still map aliases to an address. With the exception of Pine, each of these applications now has a full-fledged address book. Most applications eventually gained the ability to automatically complete the full email address as the user began to type which effectively replaced aliases for most users. The alias files began to encompass more information and evolved into modern email address books which typically include information such as physical addresses, phone numbers and job titles. These are roughly equivalent to the address books in most Personal Information Management (PIM) applications.

Many email applications still have dedicated address books. However, a system-wide address book application is often considered an essential desktop application and is integrated into most modern desktop environments. Basic forms of system-wide address books are included in Microsoft Windows as well as the GNOME and KDE desktop environments. Both the Apple Address Book for Mac OS X and the Microsoft Entourage address book have greater functionality and fall roughly in the middle of the spectrum. Microsoft Outlook incorporates a far richer set of metadata for entries in the address book, including details of corporate hierarchies, birthdays, family information, and a wide range of labels for addresses and phone numbers. Customer Relationship Management (CRM) applications such as ACT and Goldmine use a database for the address book back end and
include numerous fields related to sales and customer tracking. Address books increasingly contain information that was once solely in the realm of CRM applications.

4.2 Addressing

Researchers have examined the daily rhythms of computer users. [6] [18] [19] Temporal rhythms complicate addressing. Participants send email to different addressees depending on the time of day and location of the recipient. Many participants who have multiple accounts do not have access to all their accounts from every location. There are many reasons for the varying access including security and policy restrictions as well as lack of access to a high-speed connection or the appropriate software. For example, one participant who regularly works from home cannot read her work email at home because of corporate security restrictions. She simply asks her coworkers to send email to her personal address during the days she works at home.

Most modern email applications use an address book to store and retrieve email addresses. Participants rely on many techniques to address messages, including address books, aliases, typing addresses in by hand, relying on the auto-complete feature and replying to previous messages. Most participants in my preliminary study place only their frequently used addresses in an address book. Once an address is entered in the address book, the recipients’ name may “auto complete,” or expand inline after the first few characters are typed. Participants rely heavily on the auto-complete feature. Occasionally, the auto-complete mechanism will select an address that the participant did not expect. Participants reported difficulty in using this feature to send email to recipients who had more than one email address. In general, participants in the preliminary studies rarely used nicknames or aliases. Some participants enter aliases for their most frequently used addresses. Other participants used separate aliases for the same person to distinguish between multiple email addresses for that person. Participants reported little use of directory servers or shared address books outside of institutional electronic mail systems. Participants reported difficulty accessing directory servers outside of the workplace. Many participants reported using Google as a way to locate addresses.

Many users in the preliminary studies reported little reliance on the address book aside from auto-completion. If an address does not auto-complete for a participant, it is common to type the address by hand. Many participants reported that replying to an old message is faster than composing and addressing a new message. Participants often reply to an old message to compose a new message by changing the subject and deleting the body of the message. Several participants keep a separate email folder to store messages that contain contact information. Participants reported that filing the message into this folder and retrieving it is faster and easier than entering information into the address book and searching for it later.

Participants would often create small mailing lists for particular topics or events. Participants reported frustration with creating and managing lists or groups within their email applications, especially with larger lists. The frustration is often caused by the labor-intensive nature of maintaining the lists and the difficulty understanding how the mailing list interface is tied to the address book. Typically these mailing lists would be created by placing all members in either the To:, Cc: or Bcc: headers. More sophisticated users would create lists using the list feature of their email application or use a service such as Yahoo! Groups or Evite.

If the address book is automatically populated, instead of manually populated, with entries taken from existing email messages, there is a higher likelihood that the participant will find unex-
pected auto-complete matches. Many participants reported confusion about how entries appeared in their address book. This confusion is because many configurations Outlook and Outlook Express automatically populate the address book with the sender of any message that the participant replies. In some email clients, such as Outlook and Outlook Express, the recipient’s name is displayed in the interface and the email address is elided, causing additional confusion.

4.3 Retrieval and Organization in Email Collections

Names in email are closely connected with how people locate messages. People typically locate messages through sorting by name or date and then browsing to find the desired item. Participants report that sorting columns is usually faster and easier than searching. A small number of participants rely on the built in search function to locate messages. People typically search for messages by a person’s name, not their email address. One difficulty with searching for an individual is that there is often no way to reference that individual consistently over time, as both their email address and name (or name form) may change. Users reported that trying to locate an individual’s message by sorting is problematic as name forms are not standardized and may have several variations. For example, when first and last names are inverted, this causes blocks of messages from an individual to separate in a sorted list. Nicknames such as Bob and Will have the same effect. The use of initials in names is common for first and middle names. Names may change because of marriage or divorce. Finally, the name field may be missing entirely and just the email address may appear, or sometimes the email address is duplicated in the name field. Email addresses change because of job changes, ISP changes and a host of other reasons. Searching for email addresses is further complicated because some people use additional email addresses for different roles.

Most users categorize email, at least partly, by the sender of the message (e.g. a folder for John Smith). [3] [4] Users can work around problems with name inconsistencies and email address changes by filing all of the messages from an individual into a single folder. Authority control is a common solution in the library science community for mapping multiple entries into one entry for retrieval. [11] Authority control could be a benefit to users if it were available as a feature in email clients. Individuals could have canonical entries containing multiple name variations and addresses automatically constructed from the address book.

For categorization, the burden is on the user to either file all messages from an individual into a single category or to remember the name variants or email addresses to search for that individual later. People’s categories shift over time, and frequently become dated and decline in usefulness. [5] [8] Recategorization is time-consuming because users must move each message to the new category. Often old categories are never fully removed after recategorization leading to “category drift.” [2] [7] [12] Most classification mechanisms in email applications require users to place messages into fixed categories.

References


