Revisiting Whittaker & Sidner's "Email Overload" Ten Years Later

Danyel Fisher, A.J. Brush, Eric Gleave, Marc A. Smith
Microsoft Research
One Microsoft Way
Redmond, WA
{danyelf, ajbrush, a-ericgl, masmith}@microsoft.com

ABSTRACT

Ten years ago, Whittaker and Sidner [8] published research on *email overload*, coining a term that would drive a research area that continues today. We examine a sample of 600 mailboxes collected at a high-tech company to compare how users organize their email now to 1996. While inboxes are roughly the same size as in 1996, our population's email archives have grown tenfold. We see little evidence of distinct strategies for handling email; most of our users fall into a middle ground. There remains a need for future innovations to help people manage growing archives of email and large inboxes.

Categories and Subject Descriptors

H.1.2.a Human factors, H.4.3.c Electronic mail

General Terms: Measurement, Human Factors.

Keywords

Email, personal information management, email overload.

1. INTRODUCTION

Email is a critical communication tool that enables collaboration. Whittaker and Sidner's 1996 paper on email overload [8] explored how people manage their email and noted that email was not only for communication, but for both task management and maintaining personal archives—that is, email was "overloaded." Whittaker and Sidner described three user strategies for handling email that emerged from a quantitative investigation of 20 users' email archives: no filers, frequent filers and spring cleaners. Their work influenced a decade of email research (e.g. [1, 3, 4, 6, 7]), and structured the way we think about the problem of *email overload*.

Ten years later it seems appropriate to compare the state of email use in 2006 to 1996 and explore whether the three strategies Whittaker and Sidner identified remain in use today. People who may have once cleaned up irregularly ("spring cleaners") may have given up entirely; piling their email in fewer folders and relying on desktop search. Email users may also have evolved new strategies for handling email: many of the tools that

Whittaker and Sidner called for, including flagging and conversation threads, are now available in modern email clients. Email has, of course, remained a popular research topic and many papers have discussed how people make use of email [e.g. 7], suggested alternative designs [e.g. 4], visualized email archives [e.g. 2], and addressed integrating email and task management [e.g. 1]. Our focus in this paper is on modeling email practices through analysis of archives, a kind of archaeology. Understanding the changing state of email use helps us to take stock as we move into the next decade of email research. How does the inbox—and the email archive—look different from a decade ago?

2. METHODS AND DATA

As Whittaker and Sidner did, we collected data from employees at a technology company. We gathered data using the SNARF prototype (Social Network and Relationship Finder) [3], which was designed to assist users when triaging their email. Built as a research prototype, and distributed internally to users at Microsoft, SNARF can also log email behavior: it stores anonymized records of email archives, creating a valuable tool for researching email behavior. Indeed, even if a person did *not* become a long-term user of SNARF, the version of SNARF available at the field site still uploaded a copy of their anonymized data the first time SNARF ran.

We collected data snapshots of email archives from users of SNARF between July 15, 2005 and January 30, 2006. We used the first archive uploaded by SNARF for each user; for the 36 users who had run SNARF on more than one computer, we selected the first archive from the computer with the largest number of messages in order to get a picture of the breadth of the users' email. We excluded from our sample any user who did not have some email older than six months to eliminate recent and short term employees. Our final sample contained 600 participants.

Our study population's job roles are diverse, from managing software development (26%), building software (19%), testing software (12%), to more than 40 other unique roles including administrative assistants, technical writers and consultants. Only 28% of our participants directly manage other people. Seventy-five percent of the participants had email in their archive more than 2 ½ years old, 50% had email more than 4 ½ years old and 25% had email more than 6 ½ years old, suggesting for many people we are analyzing email archives amassed over a considerable time period.

In the 1996 study, participants used NotesMail, an email client for Lotus Notes. All the 2006 participants use recent versions of

Microsoft Outlook (such as Outlook 2003 and Outlook XP) as their email client. Outlook is a many-featured email program with folders that display the number of unread (or total) items they contain, the ability to specify rules to place email in certain folders based on a number of different properties (e.g. sender, recipient, content) and auto-archiving to move messages from the server to local storage. Auto-archiving is particularly valuable at the field site due to limits on the amount of email that can be stored on the server. All incoming and outgoing email is kept by default until the user explicitly removes it. SNARF collected all Outlook email on a participant's computer, regardless of whether the messages were stored on the server or in a local archive, but disregarded messages that were labeled as "Spam" by the server.

Using the data collected we were able to calculate many of the same statistics Whittaker and Sidner presented including: number of inbox items, total stored items, inbox as percentage of mailbox, number of folders, the number of *failed folders* (folders with < 3 items) and the number of messages received daily. As with all archive-based studies, our measurements are impacted by patterns and practices of message deletion. All users likely deleted some percentage of the email they received; that past activity is invisible to our data collection mechanism. Similarly, we have a *snapshot* of each user's archive*;* we cannot know if we collected data from a user shortly before a major cleanup or re-organization.

While Whittaker and Sidner did not present averages across all participants in [8], they presented the average results for three categories of people, for a total of 18 persons. We weighted those three categories together to get averages that approximate 1996 corporate email usage.

3. FINDINGS
As Table 1 shows, a number of interesting differences and similarities emerge from comparing data from 1996 email archives to those from 2006. In this section we compare the size of email archives, the size of email inboxes, the number of folders, and whether users in 2006 employ the email management strategies Whittaker and Sidner identified in 1996. All ordinal counts in the 2006 data are long-tailed "power-law"-like distributions.

3.1 Archives Have Grown Tenfold
The first major difference between email archives in 1996 and 2006 is size. Today, the total messages in our 600 users' email archives are distributed in a power-law curve[1], with the mean at 28,660 email items (median = 15,797). As shown in Table 1, this is more than 10 times the average size of the 1996 archive of 2,482. Initially, we thought archive size might be related to the length of time that someone had been collecting email. We calculated the *archive age,* for each user based on the age of the oldest item in their archive. Surprisingly, archive age and number of total messages in the archive were barely correlated (Spearman's $\rho = 0.185$, $p < 0.001$).

We were interested in understanding what our participants store in their archives. Were archived messages email received from others or did the participants keep mail they sent? Were the stored

---

Table 1. Comparing email statistics from 1996 to 2006. Averages given with medians in ()'s for 2006 data.

	1996	2006
N	18	600
Total Messages	2,482	28,660 (15,797)
Daily # msg. received	49	87 (58)
Inbox Size	1,624	1,150 (512)
Unread Msgs. in Inbox	NA	153 (7)
Inbox as % of total mailbox	53%	8% (2%)
# Folders	47	133 (77)
% of Failed Folders (# folders < 3 items)	39%	16% (14%)
% of Recently Used Folders (containing message received in last 30 days)	NA	38% (35%)

messages recent or old? We found nearly a third (mean = 30%, median = 27%) of the messages in an archive were sent by the participant's themselves. Almost half of the email archive in 2006 is older than 3 months (mean = 43%, median = 44%) while a mean of 13% (median = 9%) is under 1 month old.

Given the prevalence of email in corporate life and availability of cheap storage, the tenfold increase in the size of email archives is perhaps not that surprising. However, the stark contrast to 1996 does highlight the importance of search and organization tools, particularly if we believe people's email archives will continue to grow. While many companies now offer desktop search tools, continued innovation on methods for helping people locate things in their archive seems critical. Promising directions include using automatic labeling or clustering of email based on keywords present in the text [6] and methods that use the correspondence history implicit in our patterns of communication [2]. However, the growth of corporate data retention policies has the potential to dramatically affect archiving and searching behavior.

3.2 Message Flow Is Not Obviously Greater
To see if the increase in archive size was related to an increased message flow since 1996, we estimated the number of messages each participant receives daily by averaging the number of messages in a user's archive received in the last seven days. This approximation is a lower bound for email received because we assume that people delete some fraction of incoming email. For our participants, the mean number of messages received daily was 87 (median = 58). The number of messages received daily is somewhat correlated with size of archive ($\rho = 0.480$, $p < 0.001$).

These lower bound numbers for 2006 are slightly larger than the daily average of 49, which [8] collected in 1996 through self report [Whittaker, personal communication]. It seems reasonable to assume, as we might expect, that incoming email has increased to some degree in the last ten years; however, it appears the amount of email received has not grown tenfold like the size of the archive. Today's larger archives seem at least to some degree to be an effect of a greater dependence—or at least preference—for maintaining email archives.

3.3 Inbox Stayed Same Size

Most surprising is that even with an order of magnitude increase in total archive size the average inbox size has—if anything—shrunk in the last ten years. As Table 1 shows, the average 1996 inbox had well over a thousand messages (1,624); and while the mean for our users is 1,150, half of our participants kept under 512 messages. Since our email archive size has dramatically increased and the size of the inbox has stayed similar or slightly shrunk, in 2006 the inbox is a much smaller percentage of the total mailbox at 8%, compared to 53% in 1996.

We also collected the number of unread items in participant's inboxes. While many of our participants had managed to stay caught up with their inboxes (median = 7 unread messages), the mean was 153 unread messages in the inbox. The percentage of unread messages in the inbox is overall quite low (mean = 16%; median = 2%). The number of unread messages is somewhat correlated to the number of messages in the inbox (ρ= 0.419, p < 0.001).

One might have imagined that with the renewed enthusiasm for the search tools that Whittaker and Sidner advocated, email users would feel more comfortable moving things out of their inbox with the belief they could find them again. However, the large size of the inbox and low percentage of unread messages shows that users today, similar to 1996, retain things in the inbox that they have already read. Without qualitative data we cannot be sure, but it seems likely that the inbox in 2006 is as overloaded as Whittaker and Sidner found it to be in 1996.

The provocative question of why the inbox size has held relatively steady rather than trending upward with the size of our email archives or downward as we made use of new email management tools must be left for future research, however, it seems clear that there is still considerable room for further innovations in technology helping people manage their inboxes. Some promising innovations include assisting with task management [1] and systems that recommend likely folders for messages [5].

3.4 Many More Folders

Another place we see a stark increase is the total number of folders. In 1996, the average number of folders was 47, today the mean is 133 (median = 77), 2.8 times more folders than in 1996. The increase in number of folders is perhaps not surprising given the tenfold increase in archive size, and in fact, the number of total items is somewhat correlated with the number of folders (ρ = 0.423, p < 0.001). Again, one might imagine that the longer someone had been using folders the more folders they might have, but as with number of total messages, archive age and number of folders is only weakly correlated (ρ= 0.181, p < 0.001).

"Failed folders," defined as folders with fewer than 3 items, is a metric that [8] uses to describe folders that were created but not actively used. Some current users may create empty folders to label parts of hierarchies, so this metric seems somewhat obsolete. None the less, the rate of failed folders has declined substantially. In 1996, the average percentage of failed folders was 39%. Today, the mean percentage of failed folders is 16%. As in 1996, the number of folders and number of failed folders are correlated (ρ = 0.836, p <0.001).

Instead of "failed," though, perhaps a better model would be folders that have not been used recently. To estimate how many folders were *recently used*, we calculated the number of folders that contained at least one message received in the month before we collected the participant's archive. Less than half of all folders (mean = 38%), are recently used (mean = 35 recently used folders per person, median = 26). While we have no historical data to compare against, it seems that our participants actively use relatively few of their folders, suggesting that they are keeping substantial archives.

Whittaker and Sidner also hypothesized that large folders were likely to fail because it would be hard to find things in them. To determine how many folders were very large, we set the bar for "very large" at 500 items: fewer than 5% of folders are that large across the entire dataset; per user, a mean of 12% of their folders were very large (median=9%). Perhaps more interesting is that the percentage of very large folders a user has is somewhat negatively correlated to number of folders (ρ = -0.456, p <0.001). This suggests that people with many big folders also have fewer folders overall. This may happen as people come to rely on technologies like desktop search and begin to pile messages into fewer folders.

There are many possible reasons to explain the increase in number of folders and decrease in failed folders compared to 1996. Perhaps the simplest possibility is that users may feel more confident in their use of rules and folders. While both of these features were available in early email clients, they often provided poor feedback to determine when new messages had been filed into folders. The increase in folders may also be partially explainable by local policy. At our field site, participants had a server size quota, and so would routinely copy messages to a local archive. The default archiving behavior creates a local folder hierarchy that mirrors the server folder hierarchy: this could explain a doubling in number of folders. In contrast, the technical and social situation around folders was very different in 1996. In particular, while NotesMail did support a full-text search, there was no support for archiving.

3.5 Same Strategies for Handling Email?

Whittaker and Sidner identified three different user strategies for handling email overload based on two criteria: whether the participant made use of folders and whether he or she cleaned their inbox on a daily basis. Participants that made no use of folders were categorized as *no filers*, participants that used folders and tried to clean their inbox daily were *frequent filers,* and participants who used folders and periodically cleaned their inbox were *spring cleaners*. To explore whether these groups were
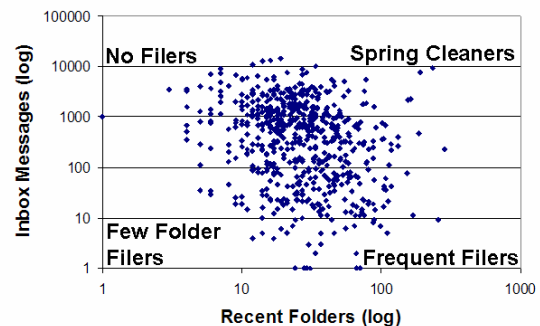

Figure 1: Scatter plot of number of recently used folders (log scale) vs. number of inbox message (log). (N=600)

Table 2. Comparing percentage of participants using different strategies.   Column B uses explicit thresholds of 10 recent folders and 100 inbox messages to divide into groups, while Column C uses 20 recent folders.

	A. 1996	B. 2006 (10, 100)	C. 2006 (20, 100)
No Filers	33% (6)	8% (50)	32% (189)
Spring Cleaners	39% (7)	64% (382)	41% (243)
Frequent Filers	28% (5)	27% (160)	21% (126)
Few Folder Filers	NA	1% (8)	7% (42)

present in our data we first plotted the size of a participant's inbox against the number of recently used folders (that is, having a message dated from within last 30 days) in their archive on log scale axes as shown in Figure 1.

Some users that represent the strategies identified by Whittaker and Sidner stand out. For example, in the upper left corner are several classic no filers: people with fewer than 10 folders and roughly 1000 email messages in their inbox. In the bottom right are frequent filers with many folders and fewer than 10 email messages in their inbox. Lastly, some clear spring cleaners are in the top right with many folders and large inboxes. A fourth strategy, which we term *few folder filers*, also emerges in the bottom left: participants with small inboxes and small number of folders. As is visible from Figure 1, most of our participants are in the middle of the graph with roughly 25 to 75 recent folders and 100 to 1500 messages in their inbox.

We experimented with grouping participants based on their inbox size and number of recently used folders in many different ways, but clear divisions among groups did not emerge. Rather, as Figure 1 shows, there is a continuum of participants from those that maintain very small inboxes to those that have large inboxes and from those with few folders to those with many folders. However, we can pick some explicit values as cutoffs to compare the different strategies. If we use 10 recent folders to divide those who use folders from those who do not and 100 inbox messages to divide filers from non-filers (Table 2), nearly two-thirds of our participants are spring cleaners and we have very few no filers. Shifting that cutoff slightly, to 20 folders, vastly drops the number of spring cleaners, and makes no-filers a full third of the pool.

We believe that the strategies that Whittaker and Sidner describe still function as ideal types. Interviews with participants might allow us to explicitly categorize participants based on how they felt about their folder usage and the size of their inbox. However, our quantitative data does not reflect the clear groupings we might expect if participants were consistently applying one email handling strategy. This effect may be exaggerated by self-selection. We have data only from people that chose to run SNARF, a tool that aims to help people with email, and these may precisely be the people trying numerous strategies to handle their email.

4.   CONCLUDING REMARKS

Our comparison of email archives from 1996 and 2006 shows that some aspects of email have dramatically changed, such as archive size and number of folders, and others, like the average inbox size, have remained more or less the same. Surprisingly, the larger archive size and number of folders only very weakly correlates to the age of the archive, suggesting we may be saving more things

and still attempting to maintain filing systems even with improvements in desktop search tools. Of course, the email landscape has changed radically: spam was nearly unknown in 1996, for example, and the email software prevalent today is far more powerful than its 1996 counterparts.

Like Whittaker & Sidner, our email archives were gathered from participants at one technology company. We believe it is critical to extend this type of email analysis in a broad range of settings to fully capture the variety of email handling behaviors. In particular, we expect that statistics related to size of archive, relative size of inbox to size of archive and daily flow would be quite different for populations of consumers or students. A better understanding of the diversity of email archive structure and usage practices would be a useful guide to the further refinement and improvement of email tools.

Our participants have significantly bigger archives and an increased number of folders; their inboxes are as swollen as a decade ago. Continued research and innovation is critical to help people manage email overload, find email they have archived, and allow email to continue successfully functioning as the crucial communication medium it has become. In fact, as email programs continue to be a catch-all storage and communication medium for many other tasks, like document transfer and staying aware of online content using RSS feeds, it seems likely that email overload will continue to grow.

5.   ACKNOWLEDGMENTS

6.   REFERENCES

[1] Bellotti, V., Ducheneaut, N., Howard, M., and Smith, I. Taking email to task: the design and evaluation of a task management centered email tool. *Proc. CHI 2003.* 345 – 352.

[2] Fisher D., and Dourish, P., Social and temporal structures in everyday collaboration, *Proc. CHI 2004.* 551-558.

[3] Neustaedter, C., Brush, A., Smith, M., and Fisher, D. (2005) The Social Network and Relationship Finder: Social Sorting for Email Triage. *Proc. of the 2005 Conference on Email and Anti-Spam (CEAS).* http://www.ceas.cc/papers-2005/149.pdf

[4] Gruen, D., Rohall, S., Minassian, S., Kerr, B., Moody, P., Stachel, B., Wattenberg, E., and Wilcox, E. Lessons from the reMail prototypes. *Proc. CSCW 2004.* 152-161.

[5] Segal, R., and Kephart, MailCat: An Intelligent Assistant for Organizing E-Mail. Proceedings of the Third International Conference on Autonomous Agents, May 1999.

[6] Surendran, A., Platt, J., and Renshaw, E., "Automatic Discovery of Personal Topics To Organize Email", *Proc. of the 2005 Conference on Email and Anti-Spam (CEAS)*, http://www.ceas.cc/papers-2005/145.pdf

[7] Venolia, G.D., Dabbish, L., Cadiz, J.J., and Gupta, A. Supporting Email Workflow. *Microsoft Technical Report TR-2001-88.* 2001.

[8] Whittaker, S. and Sidner, C. Email Overload: Exploring Personal Information Management of Email. *Proc. CHI 96.* 276-283