How To Survive Email

Miloš Vacek *

* University of Hradec Králové, Hradec Králové, Czech Republic
milos.vacek@uhk.cz

Abstract— Email is a good servant but a bad master. Email as a useful tool for communication hides many potential threats. As well as it is fast and straightforward it can easily grow to a state when user must deal with hundreds of email every month, not being able to do other work. This state is often referred to as email overload. This paper follows on a research conducted at UHK in 2012 and extends it to its second phase. The results are used to analyze users’ requirements on electronic communication and to describe formally the process of dealing with email. Another objective is author’s own definition of email overload in terms of obtained and derived findings. A set of improvements is proposed for a new communication tool which shall be designed and shaped in the future work.

Keywords — communication, email, email overload, process analysis, workflow.

I. INTRODUCTION

The structure of work of information users has definitely changed in academic as well as business sector over the past few decades. Most of the agenda is handled on desktop computers, laptops, tablets or even smart phones. With development of new software and mobile applications it is now possible to exchange emails, share documents and use instant messaging almost anywhere due to good coverage of new generation mobile networks. Whereas daily work, consisting of meetings, document processing, teaching, programming, etc., may not have been affected by this too much, communication has changed immensely for good. On one hand it has positive impacts, such as economic and environmental, in a form of cutting costs on printing, postage or maintaining paper archives. But on the other hand, the huge amount of information coming through electronic channels, mostly by email, requires much more attention and time to sort it and deal with it. It is out of discussion that answering emails belongs to everyday duties of information workers but it remains unclear why some employees take this fact obviously and some feel overloaded by email. This paper focuses on analyzing work with email and tries to propose some improvements. There are two major targets: the first is to identify what email overload really is, the second is to describe the formal process of dealing with email which will be used later for reengineering and defining functional requirements of a new tool.

The major input for the analysis was a research conducted at University of Hradec Králové which took place in April 2013. It produced two types of data: computer generated reports from email accounts of respondents and answers to a questionnaire also related to communication, mostly using email. Besides, personal experience shared informally during interviews was also valuable. In this paper, the results are summarized and used to present some conclusions. In author’s previous work it was claimed that the problem called email overload has been around for many years and without any significant change of communication patterns no solution can be expected. To find the solution is author’s long term objective and a subject of his continuous research.

This paper is structured in a following way. The second chapter provides an insight into the history of email related problems, more recent scientific findings in this area as well as author’s previous work. The chapter III brings an overview of research methodology and overview of obtained results which are analyzed in more detail in the following chapter IV. The recommendation and discussion of the most important findings are given in chapter V and, finally, the conclusion of the paper is in chapter VI.

II. RELATED WORK

This chapter gives an overview of literature published on topics related to email overload and email research in general. It includes some recent studies as well as those that go back to roots of this problem. The paper widely recognized as the one that originates the term email overload was published in 1996 by Whittaker et al. It described email as a tool that had become used for multiple purposes it had not been designed for, such as document delivery and archiving, work delegation or task tracking. The authors investigated different email management strategies and their conclusions are valuable yet overlooked even today, for example they suggest grouping emails to threads according to the same subject – feature that some email clients still do not support.

It was noticed that since this paper [1] was published the email research has concentrated into three main areas: theoretical and psychological research, practical research which focuses on daily work with email and technical research which designs new features, tools and add-ons. A few examples from each area are cited below.

A. Theoretical and psychological research

The same methodology as in [1] was used 10 years later by other authors in a paper [2] that revisited the email overload problem by analyzing a new email dataset from a technological company. It was found that some metrics, e.g. archive size or number of folders have changed dramatically whereas others, like inbox size, have remained the same. Authors conclude that email overload problem will continue to grow. Another approach was showed in [3] which concerned the number of email recipients – a variable examined in this

This work was supported by the INDOP No. CZ.1.07/2.2.00/28.0327, financed from EU and Czech Republic funds.
paper too. Experiments on the Enron corpus revealed that with an increasing number of email recipients the strength of relationships between users decreases. It can be assumed that receiving group emails from someone we do not know even within the same organization can be considered as spamming. To prevent this, companies often deploy policies on maximum number of recipients, using work email for private purposes, etc. This was a subject of study [4] which classified emails without examining its content. The purpose was to create high and low priority email queues (HI priority contains working email, LO contains private) so the email server could deal with requests accordingly, thus more efficiently using the bandwidth. The psychological aspects can be found in [5] where authors predict personality of email writers based on word analysis of emails (e.g. neurotics tend to use more negations) or in [6] where it is claimed that in email communication users often discover their alter ego by behaving more aggressively than in a face to face communication.

B. Practical research

Studies in this category often try to propose some solution of a particular problem. This is a case of [7] which introduced three facets of information overload as a large volume of incoming information, inefficient workflow and poor communication quality and conducted training of email users. Then, some hypotheses, such as knowledge of email functions, confirmed that better knowledge after the training support more efficient email communication. The fact that it is important to spread the best practices in using email is also mentioned in [8]. It was claimed that using the best practices has domino effect thus enhancing a better picture of communication and relationships. However, it can be argued that this effect can also work vice versa and bad habits can spread as fast as the good ones, regardless company policies. A different example of practical studies is [9] which focuses on work interruptions caused by checking email too frequently. It verifies and recommends a strategy of creating blocks when employees work on their regular tasks and switching to intervals when incoming emails are dealt with as more efficient. Although the authors propose implementation of this strategy to email clients which would check for incoming emails in predefined intervals the design of such a feature is out scoped.

C. Technical research

The papers in this category introduce real implementations of various features that either enrich user data or make the work with email easier. An example of data enrichment is given in [10] where the described system Acoma gathers context information from various sources based on keywords in machine-generated emails (e.g. account statements, e-tickets, confirmed reservations). The structured information can be used to provide directions to a retrieved address or tracking information for parcel ID. In a different case [11], the email is integrated with social networks (facebook, linkedin, myspace) to find a social context. The practical impact it brings is when a user receives an email from an unknown sender the system automatically retrieves and displays the links based on friend-of-a-friend ontology from social networks. Perhaps a more interesting feature to help email users with better orientation in email workflow is presented in [12]. It approaches categorization of emails using own ontology and visualization of email workflow. Thus, if a sender requests some action or delegates a task to colleagues the system generates appropriate user interface to react, e.g. if some approval is demanded the system provides buttons Approve, Reject and Own answer. Also, grouping of emails is possible for better orientation in the context of email workflow.

It seems that all of the works briefly described above as well as many others somehow contribute to solving consequences of email overload. But the fundamental question – “why do we have to deal with so much work arising from email?” – still remains. One of the likely causes is also described in literature. Hassini [13] in 2004 published a preferred way of communication with students by sending group emails to everybody within each of teacher’s courses, thus generating up to 500 emails in a semester! This was 8 years after the first Whittaker’s email overload study [1] was published and many e-learning environments with discussion tools were commonly used. However, Hassini still insists on sending email attachments as a way to submit students’ works.

The author believes that the solution lies somewhere in a combination of 3 approaches summarized above: psychological aspects (awareness), practical aspects (improving inefficient daily routines) and technical aspects (using advanced tools).

The author’s previous paper [16] focused on finding the causes for a high number of incoming emails to inboxes and attempted to generalize the findings based on results from the case study of UHK employees. It was found that the amount of incoming emails had not been significantly higher at UHK in 2012 compared to a random sample of employees from Enron email corpus in 2001. Several undesired communication patterns were identified and these will be discussed later in this paper once again.

III. METHODOLOGY

This section explains the methodology of the research conducted at UHK in April 2013. It should be mentioned that the process of data collection and processing was almost the same as in the first phase conducted in November 2012 for the previous paper [16]. Therefore only differences will be discussed here.

The biggest change in the process was selecting participants for the research. Whereas in the first case, random selection from all valid email addresses of employees had been made, for the second phase all employees were asked by email for cooperation in order to get relevant answers from as big sample as possible. The total of 420 employees were asked to state whether or not they would be willing to participate, however, the return rate was not much better than in the case where participants had not volunteered. The numbers are in
Table I.

<table>
<thead>
<tr>
<th>TABLE I. REACTIONS TO PARTICIPATE AT RESEARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Will participate</td>
</tr>
<tr>
<td>Will not participate</td>
</tr>
<tr>
<td>Automatic response</td>
</tr>
<tr>
<td>Did not respond</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>

The most of the negative replies were explained with the lack of time or doubts about the sense of the research. Despite their positive statements, 17 colleagues did not attend the research after all because it had not been possible to arrange an interview. Thus, a total of 25 personal interviews were conducted and 19 data reports generated. The rest was not successful due to technical problems or were empty (no message copies on server). The data was fetched for March 2013.

Whereas the first phase had been totally anonymous and reports as well as filled questionnaires mixed up, during the second phase pairs of data reports and questionnaires from each person were bound together under the same identification. By this, more varieties of facts could be analyzed by supporting answers with generated data and vice versa.

Since individual experience was often important, for further analysis examples of participants will be used and such persons will be identified with a code representing his/her answers. The code will be for example P(500, 20%, 12, 4) where numbers in brackets represent these variables:
1. 500 – estimated number of emails received in a month
2. 20% – percentage of important emails out of all received
3. 12 – hours spent weekly on dealing with email agenda
4. 4 – category where participant see him/herself regarding email overload (categories describe how often subject feels not being able to do other work because of dealing with emails and are following: 4 – every day, 3 – 1-2times a week, 2 – 1-3times a month, 1 – less frequently, 0 – never)

IV. RESULTS

In this chapter, the obtained results are summarized for presentation and used for explanation of related facts. Very basic sets of aggregated results are in Tables II and III.

TABLE II. SUMMARY OF SAMPLE RESULTS FROM DATA REPORTS

<table>
<thead>
<tr>
<th>Number of reports</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of emails</td>
<td>2186</td>
</tr>
<tr>
<td>Total number of attachments</td>
<td>1292</td>
</tr>
<tr>
<td>Number of emails with attachment(s)</td>
<td>834</td>
</tr>
</tbody>
</table>

TABLE III. SUMMARY OF NUMERIC ANSWERS FROM QUESTIONNAIRES

| Average estimated number of emails received a month | 246,4 |
| Average % of emails important for the recipient | 61,76% |
| Average % of emails requiring response or other action | 70,56% |
| Average number of weekly hours spent on dealing with emails | 7,42 |

There are several interesting factors, arising from the numbers. There is a significant difference between estimated number of emails received by an average user (246,4) and the average number of emails really retrieved from data reports (115). Of course, there was a delay between the end of March 2013 and each interview and in this meantime unimportant emails were simply deleted. But still, even if all unimportant messages had been deleted about 160 would have remained in mailboxes. This indicates that either messages with important information are deleted or the rate of unimportant messages is much higher (53,3%).

Second observation is that still a high number of emails represent internal communication (63,5%). It was argued before [16] that email is not a good tool for sharing information and documents within organization. Additionally, it was found out that more than a half of interviewed employees had admitted using cloud technologies for sharing documents (such as Google docs, Dropbox, Skydrive) and instant messengers (icq, Skype, jabber) to chat with colleagues. Thus, a combination of cloud storage and instant messaging seems more suitable for internal communication. On the other hand, instant messaging requires both sides being online at the same time which cannot be always assumed.

Time spent weekly on working with emails varies from 1 to 20 hours with average of 7,5 and half hours. This number is slightly smaller than result derived from the previous sample (8,1 hours/week), however, it is still a significant portion of time indicating that dealing with emails and email related agenda is a regular duty of university employees. It should not be strictly separated that there is “usual” work and then “email related work” which is often understood as compulsory overhead. To prove this, some cases were reported, e.g. P(1000, 70%, 20, 0), where high-end values had no impact on email overload and emails were understood as regular work.
However, the number of emails need not to be the only factor, $P(60\%, 80\%, 12, 0)$, though a correlation was confirmed ($\rho=0.78$).

A number which is also quite fascinating is a total size of attachments copied to mailboxes of all recipients of an email with attachment. Let us suppose that an email has attachment of size 500kB and is sent to 30 recipients. Together it would take about 15MB of disk space. From the sample it was found out that this number is actually 3,1GB of attachments in 2426 distinct mailboxes. It certainly can be interpreted as a huge amount of paper being saved by sending documents electronically, on the other hand, this is also responsible for the time spent on emails and thinking that reading often irrelevant documents generates the overhead and causes email overload.

What is considered here as a disadvantage and a reason for taking too much time, that is actually the biggest benefit of email. When users were asked about their likes and dislikes about email their answers were following, Table IV and V.

TABLE IV. WHAT IS IMPORTANT ON SENDING EMAILS

| Possibility to send attachments   | 88% |
| Possibility to send email to multiple recipients | 84% |
| Reliability of message delivery   | 76% |
| Need to confirm oral agreement    | 70% |
| More time to think properly before replying | 61% |

TABLE V. WHAT BOTHERS USERS THE MOST ABOUT SENDING EMAILS

| Formal writing instead of concentration on content | 44,0% |
| Chain replies                                    | 19,2% |
| Missing confirmation that message has been read   | 29,6% |
| Taking too much time                             | 17,9% |
| Sorting email to folders                         | 3,4% |

Except the predefined answers, users were able to give their own positive and negative experiences. Whereas they provided no positive answers a few negative ones were gathered. For instance, what they mentioned:

- double sign-in to webmail client
- missing metadata of message importance
- limits on sending attachments
- internal spam

All of the information retrieved from the filled questionnaires and data reports have been greatly valuable in order to understand better users’ needs to communicate. There are certainly some common patterns and expectations of what email should enable to exchange information efficiently. However, it was realized that there are severe drawbacks as well. All these inputs will be used in future work to propose a tool which would eliminate the problems and raise in front all its benefits.

V. DISCUSSION AND RECOMMENDATION

In this chapter the results described above are observed from a different point of view and used to identify the major requirements on internal as well as external communication.

There are many reasons why the amount of incoming emails have multiplied. Perhaps one of them is the fact that we do not wish to receive information printed on paper. Many types of computer generated reports and forms have transformed in emails, such as newsletters, booking confirmations, invoices, account statements, event reminders, delivery notices, etc. This has a positive impact on costs, resources and the environment as these items used to be printed. It is clear that sorting and archiving of electronic documents is easier than the paper ones. The other side of this benefit are minutes and hours which are taken from paper overload, not mentioned frequently anymore, and moved to email overload which is recognized as a phenomenon of electronic age.

It is obvious that the time spent on emails cannot be the same for all professions and is dependent on job description, importance of a person in company hierarchy or simply achieved results. Though it was proved that number of hours spent dealing with emails depend on the number of received emails there is no evidence that either number of emails or time spent on dealing them have direct effect on feeling email overload.

Information and data transfer is highly ineffective using the standard email protocols. In a typical scenario, whole emails (metadata, content and attachments) are saved on a mail server which is regularly backed up for Disaster&Recovery purposes. Then, emails are downloaded to users’ workstations and mobile devices, important attachments saved on hard drives, important messages archived. Meanwhile, the border between important and unimportant is unclear as much more emails are deleted than is regarded as important. Users obviously build their electronic archives with folder structures and do not mind spending time on this. But, due to limits they delete more than a half of emails or they use other services, e.g. Gmail with its ever-growing storage, even for work purposes.

A huge portion of email communication is used for internal purposes only. Whereas it can be challenging for various reasons to share a common data storage with external users, within the same organization it would be simple to upload a document once to the common storage and provide access to anyone who might be interested in it. The same system might work well for posting internal news, announcements, event invitations, etc.

To learn about the process of dealing with email a process analysis was made and a diagram drawn (see Fig. 1) to identify its strengths and weaknesses. The workflow will be described in following paragraphs, with possible suggestions and improvements in italics.

First reading – user briefly identifies the content of the email, finds out if it is urgent, checks for attachment (without opening them) and decides about further actions. In perfect world, this could be a semi-automatic activity, supported with learning engine that decides about flags based on a defined set of rules. Only in uncertain cases would be required human
interaction.

Decision making – according to results of the previous activity, user decides what to do next with an email. Unimportant emails are deleted, important saved and further processed, based on the content and urgency. It is obvious that decision making is a great candidate on machine processing. A lot of emails could be dealt with automatically and much faster but is not because of lacking input metadata.

Archiving – emails are sorted according to various factors. Unimportant are manually deleted, emails for processing are moved to To-do folders and Done folders when finished. Some users maintain folder structures according to topics. An automated workflow could be assigned to messages for archiving. This could be, e.g. delete immediately, delete after 30 days, index and archive for 5 years, index and archive permanently, archive encrypted.

Execution of attachment workflow – this activity is handled differently by users and in most cases there is no workflow involved as the actual files are treated manually in order to perform the required action. A great improvement of this activity would be automatic recognition of file extensions and offering pre-defined workflow for each type. For example, text files and pdf files would be synced to a reading list in reading devices (tablet, Kindle); spreadsheets saved to a working folder, images resized and uploaded to a gallery, music files tagged and played, etc. Only standard systems such as Document Management System or Digital Asset Management need to be deployed to handle this.

Processing email content – this represents a wide range of activities depending on the content. It seems to be complicated to help users with this, however, using appropriate ontologies for email categorization could be a solution. For example, an email tagged as event invitation would create a new event in user’s diary or at least check if there is a suitable slot.

Moving to queue – a lot has been written about interruptions of work by emails. While experiencing email overload, 37.5% of users claim that they try to answer email immediately and 56% at least on the same day. Email integrated task list could prioritize items and let users deal immediately and 56% at least on the same day.

Creating reminder – majority of email clients do not allow to create a reminder on certain item and users have to create them separately and often in separate tools. Thus, context of the reminder is often forgotten. This feature could be handled more efficiently by setting a reminder directly on emails which would pop up at the time when it needs to be responded or moved to the task list based on its priority.

Email processing – it is apparent from the Fig. 1 that this is an activity which may consist of a set of sub-activities. Each of them could be probably supported in a different way. However, from the research results it seems to be obvious that at least direct communication is required. Before writing an email answer user could be given a choice to chat if the other person is online (even with group chats possible) or call on Skype or other phone number retrieved from integrated contacts. Work with documents could be improved in the same way as attachment workflow already described, but extended by approval process to enhance cooperation.

There is a great amount of ideas how email could be improved and processes around dealing with email agenda made much more efficient. To name the most important features which seems to be missing nowadays and which need to be supported in order to enable deployment of automatic processes, we can stress the following:

- start with internal repository for knowledge transfer
- support direct communication within organization
- bring ontology in email communication
- work with metadata for machine processing
- define automatic workflows for repeating actions

While the problem of email overload persists, it should be clearly identified what stand behind this fact. According to many indicators, such as mining in research data, interviews or online research, there are 3 factors which cause that a person feels overloaded with email and not being able to do other work. These factors are:

1. Not awareness of the fact that email IS regular work, it has only transformed from paper to electronic form
2. Using inefficient tools which do not support advanced filtering and automatic processes
3. Psychological aspects, such as workaholism, denying the option to refuse a task, being afraid of delegating, etc.

To avoid at least the first two points from happening in everyday communication, it is a subject of author’s future work to propose and design a new tool for communication incorporating much of users’ current requirement as well as bringing his own.

VI. CONCLUSION

In this paper results of a research conducted at University of Hradec Králové were analyzed, in order to learn more about the problem email overload among the colleagues at university and to identify the real causes. The methodology used in the similar previous research brought results which were observed from the perspective of user requirements on communication technologies. The results are presented in a clear way and critically discussed with regard to potential impacts on the process. Based on the derived requirements a typical process of dealing with emails was proposed and drawn and this diagram will be used in author’s future work to design a new communication tool. This tool should incorporate much of the users’ needs in order to prevent email overload which is caused not only by psychological aspects but, as was also stated, by technological limitations and not admitting that email as a standard communication channel is a part of daily duties and not something extra which takes more attention. Finally, the paper brings an overview of improvements which might be achieved should the tool be successfully developed and deployed.
REFERENCES


FIGURE 1 - PROCESS OF DEALING WITH EMAIL