E-mail - Friend or Foe?

Dr Stephen Braim

Information Management Dept., Dstl, Malvern, Worcs., WR14 3PS, UK

Abstract

Synchronised Operations fundamentally depend on timely access to accurate information. Indeed, information superiority is one of *the* key enablers for synchronised Operations. However, information access alone is not sufficient - the suitable sharing of a common view of the information across the joint or coalition force is equally important. Thus the timely distribution of information becomes key and the associated *management* of information becomes crucial for the realisation of information superiority.

Modern defence-related IT is increasingly based on COTS. Unfortunately, this can be purchased by enemy and friend alike. This "universality" means that the implementation and method of use of the COTS is the only real differentiator. Thus "we" must be better than the enemy in the use of COTS. Management of information is again a key topic in this. Poor information management directly impacts on the coherence of the decision making process, on effectiveness and hence on the synchronisation of Operations.

E-mail is a key ingredient to many business processes in defence, directly supporting timely information flow. It is very heavily used, and perhaps misused. However, the tools available for the management of e-mail are not well matched to its popularity – "e-mail overload" is an increasingly common cry. Additional functionality is often added to the basic e-mail mechanism, which directly leads to further information overload. Specifically a major problem is commonly associated with attachments to e-mail. Further, the use of e-mail is only likely to grow in the future, given the increased reliance on the Internet and on defence intranets for data exchange, storage and access during times of normal business as well as war. In addition, e-mail increasingly tends to replace formal military messaging, especially for coalition operations. These uses all add to the volume of information to be managed.

This paper focuses on the vital need for the improved management of e-mail, in particular e-mail attachments. It addresses the functionality currently used and delivered by e-mail. It then extracts the associated underlying User requirements. The paper next reviews the tools provided by COTS to manage e-mail and contrasts these with similar tools used in the Internet web environment for e.g. search. It suggests a mechanism to align e-mail and web sharing approaches, allowing the best available tools to be used for each identified User need. The paper then proposes enhanced ways of working and optimal tools for each requirement area. It concludes by addressing their adoption by an ad hoc coalition group, thus enabling synchronised Operations for coalition Forces.

1. Background

Arguably, electronic mail (e-mail) was the earliest widespread functionality offered by networked computers. Nowadays it is certainly one of the most heavily used. Many users complain of "e-mail overload", being faced with tens or even hundreds of new e-mail messages at morning logon. Coping with such a daily message load is onerous in its own

right; finding information held in yesterday's messages is often unmanageable. In contrast, the ability to find text and multimedia information in the modern Intranet and Internet "web" environment is far better supported. The use of search tools such as Google™ and Altavista™ has become normal business for many in their daily work on corporate Intranets, or the Internet. The potential for building a bridge between the e-mail and web technologies, allowing the benefit and broad technical advances of the latter to be applied to the former, is the topic of this discussion paper.

2. Is There Really a Problem with e-Mail?

The feeling of "overload" experienced by many e-mail users is indicative of some level of problem. Whilst the *quantity* of e-mail is often a major contributory factor, the manner in which it is presented and accessed can have a huge impact on a user's efficiency and effectiveness. More appropriate management of the e--mail can lead to a far less overwhelming flow of information. However, one needs to ensure that any management overhead does not at the same time throw away the simplicity of use and benefits that e-mail brings - arguably it is these very benefits that cause the problem with volume. The need is to better manage information access for the overall benefit of the user.

It would seem a fair assertion that e-mail took over from the fax, which in many respects took over from business letters (at least for a time) as the preferred way of working in commerce. The benefit of a fax over a letter was generally one of immediacy; a fax was delivered in minutes, a letter in a day of two. However a fax was found to be relatively limited in capability - adequate for a few pages of typed text, but not for complex documents or images. E-mail offered this advantage, and significantly more (e.g. all document types, forwarding, softcopy cut-and-paste, online and offline storage) at the same speed or faster. The dramatic growth in Internet and commercial WAN¹ provided the necessary "universal" connectivity, allowing e-mail to flourish. The explosive growth in e-mail usage is the end result. It is interesting to recall that there was never a widespread cry of "fax overload", or indeed "letter overload". One wonders why.

The e-mail paradigm is familiar to all, being largely an electronic equivalent of the postal letter service. A letter is composed, sent to one or more recipients for either one-way passage of information (e.g. a holiday postcard) or a two way / many way dialogue (e.g. request for payment, request for information). E-mail can fulfil both of these needs electronically. A letter is generally delivered direct to the recipient, although it can be held in an office mail tray. Similarly, e-mail is kept (on a server) until the recipient is available to collect it. Post, and its electronic equivalent, thus follows a push, read / react paradigm.

One of the strengths of e-mail is its ease of use. This very 'fire and forget' simplicity is, ironically, the most significant factor that has led to the overload problem. Its ease of use has led to a rich set of functionality to be associated with e-mail, and thus different uses to be made of the underlying electronic delivery service. Examples include instant messaging and chat, group calendar and meeting scheduler, and the ubiquitous attachment "postal" service.

It is helpful to categorise the different uses to which e-mail is put. Four distinct areas can be readily identified:-

-

¹ Wide Area Networks

- 1. Use of e-mail to pass basic text information directly (i.e. within the e-mail body) as a message, memo, query, summary, hyperlink / URL information etc
- 2. Use of the e-mail mechanism to send information as an attachment (in almost any multimedia format) to a mail message (which could itself be blank, or simply state for example "see attached")
- 3. Use of e-mail to pass documents as part of a managed workflow process
- 4. Use of e-mail as a meeting organiser / calendar / diary.

In addition, cognisance needs to be taken of the potential for the combination of the four types listed above - for example, types 1 & 2 are often used within the same e-mail construct.

It is suggested that at least part of the problem experienced by e-mail users is that all these differing types of e-mail generally appear, undistinguished, in the same inbox, with the same suite of software tools on offer to manage them. The set of tools is expected to support the differing uses equally well. This is not generally found to be the case, as will be explored later in this paper. This diversity compounds the significant information management problems associated with the volume of information; it is this diversity of uses to which the e-mail carrier is put that leads to many of the *perceived* overload issues. Separation of the different types, allowing each to be treated individually, is proposed as a key step forward; separation of types 1 and 2 in the list above is a particularly important step.

The rich functionality associated with e-mail is recognised (and indeed encouraged) by software vendors. Mail software packages have become increasingly sophisticated, and do offer tools for instance for mail filtering, "organising" and word textual search². However, such tools do not appear to handle the differing types of information listed above equally well. It is proposed in this paper that a more targeted approach, using specific tools for each of the differing e-mail types, will lead to real benefits for the user.

3. Current e-Mail Management Approaches

3.1 Folders

Folders are the most familiar method for organising information on a computer. They are logical extensions of the basic hierarchical file-store structure used by all operating systems, and are familiar as visualised by software such as the Microsoft Windows file manager (Explorer). One can contrast such a hierarchical approach with the manual "linear" card index used in pre-computer days. Such card-files did allow some degree of structure (categories), but manual processes left them relatively unwieldy (e.g. Dewey library classification scheme). Detailed hierarchies really came of age in the computer era. However, whilst hierarchies are fine for storing well-delineated types of data (such as "all reports from the BBC", or "everything on the US space programme"), they tend to become unwieldy once the data volume becomes large and diverse, necessitating more branches, and thus a finer granularity of the description. This is often the major hindrance to information discovery once a fine level of granularity is imposed, allocation decisions become difficult when an information item falls across categories, and even more difficult when trying to recall which category might have been used to allow the item to be found again. E-mail packages such as Microsoft Outlook, Eudora and Pegasus are generally based around a "tree" approach; whilst

² An outline summary of the tools available in a typical e-mail package, for each of the four areas, is listed at Annex A.

this is found to be adequate for straightforward text e-mail messages, it is inadequate for complex messages containing one or more attachments.

Thus, whilst coarse aggregation into folders does seem to help the user; fine aggregation can become a definite hindrance. This difficulty implies that the folder metaphor has been overstretched, and a better approach is needed.

3.2 Web Technologies

As we all know, the Internet hosts the World Wide Web (www), a complex interlinked set of individual information items (generally structured as html pages and/or discrete files) of incredibly diverse information content. Some parts of this "web" are organised hierarchically; most are barely organised at all. The www supports a pull, read / learn paradigm. Its very size presents a real challenge in "finding information". Indeed, given the difficulty experienced by most users in finding an individual item of information on their own computer, their ability to find something specific on the www should be almost non-existent. Remarkably, it is not indeed most users complain of finding too much information on the web (or rather too much data). This ability is due entirely to the existence of sophisticated web indexing tools and search engines, which directly support information discovery. They have evolved significantly over recent years; they spent their lives trawling and crawling the web, indexing all that they find. This "pre-indexing" makes the search task of an individual user almost immediate: one can find information on a specific topic very quickly; the challenge is in finding precisely what is sought.

These search engines are very flexible - they can be used to index and search the Internet, local intranets, or even simple file-stores on a local server or PC. Their speed and sophistication generally outstrip common e-mail inbox functionality (particularly in terms of text and multimedia search) with many further enhancements under development (e.g. natural language processing, semantic web support ³). It is interesting to note that some Web "search" sites, such as Altavista™ and Yahoo™ offer structured hierarchies of categories (e.g. Computing> Internet>Guides>e-mail) as well as their native search engines. The current trend towards Enterprise Information Portals (EIP) has also emphasised a need for common taxonomies across an enterprise.

4. Management Process

The earlier discussion has indicated that at least part of the e-mail "problem" is due to differing functionality being added to an existing mechanism (e-mail). This leads to short term benefit, but arguably has led to the longer-term problem of management complexity and thus overload. It is widely accepted that better processes are essential to support information sharing across an enterprise. Most users have poorly organised information stores on their local PC or server. Local information access and retrieval is often unstructured; widespread information sharing is often not supported. For example, to make an item of information available to a wide (internal / local) audience, the "normal " approach is to send individual copies to multiple recipients. This is not efficient from a system viewpoint. Transmission of the reference to the information item (URL) is potentially more efficient. More simply, announcing to the system that a report on "xxx" is available (thus allowing a profile match to be made to already-subscribed users) is yet more expedient. Users can view (or download)

-

³ See www.semanticweb.org

the report as and when needed [Farrington, 2002]. Inevitably, present COTS software directly supports (and is well matched to) a specific way of working - this relative narrowness, linked to user apathy is the most likely explanation as to why information is not well managed on most PCs.

The need for significantly different processes for information management in general is widely understood – what can actually be done is more of an issue. Some progress can be made in the area of e-mail management, but this needs to be part of a wider change process.

5. Information System Architecture

The process changes intimated above have an impact on the system architecture (and vice versa). The ability to host documents centrally requires not only appropriate storage, but also adequate communication capacity between clients and servers to guarantee user access to information on demand. This is increasingly likely in a (benign) office-style environment with adequate local networking capacity (LAN). It is relatively unlikely in a deployed, operational military environment (either due to bandwidth / laws of physics limitations, or mobility / enemy action limitations). It is interesting to note that whilst over recent years many enterprises have moved from a central mainframe to the local PC for applications and data storage, there is now a reversal of this trend underway. It is recognised that reverting to centralised applications and data-stores offer lower maintenance costs (with LAN & WAN costs falling). Indeed new technology on storage area networks (SANs) and enterprise information portals take this metaphor further forward. E-mail is often "received" centrally onto a mail server and then downloaded to user workstations. It may therefore be easier to adopt some form of centralised management approach.

6. Analysis

6.1 Requirements for e-Mail and Web Functionality

Having discussed the strengths and weaknesses of modern e-mail usage, and discussed the developments in the Internet and www arena, we are now in a position to propose an improved approach to tackle the e-mail management / overload issue. As a start, it will be appropriate to extract the basic user requirements that e-mail is currently used to deliver. Reviewing these requirements against differing "technologies" and system solutions will perhaps yield a more optimal solution, which will lead to a reduction in information (e-mail) overload issues. In other words one should target the user needs and problems directly, rather than continuing to append ad hoc functionality to an existing mechanism (i.e. e-mail). Such an approach may include associated process changes.

The four areas cited earlier in this paper yield the following requirements: -

- 1. The need to pass "simple" text information as memo, query, report, URL etc
- 2. The need to send information (in almost any file format / multimedia)
- 3. The need to pass documents as part of a managed workflow process
- 4. The need for a meeting organiser / calendar / diary.

As indicated earlier, the key step is to separate requirements 1 and 2.

6.2 Solutions

It is believed that requirement 3 can be adequately met by a specific group-working / workflow application, which manages document transfer in a manner either largely transparent to the user or is "well managed" as an integral part of the application.

Requirement 4 is generally successfully delivered as part of the integral functionality of the modern "integrated" e-mail packages, such as Microsoft Outlook 2000. This is a similar example to workflow – although it has added the functionality to e-mail, it does not cause significant information management overhead as it has low information content which is well managed though a complete packaged solution.

Requirement 1 is the original, "simple" purpose of e-mail. It is the core functionality of all e-mail software. Appropriate hierarchy-based management tools (categories, text search tools) are well provided and appear well matched to the e-mail environment. Corporate structure and associated processes could usefully be proposed (best of breed etc) to engender some coherence across an enterprise at the level of common structure for hierarchies, taxonomies and so on.

Requirement 2 is the topic that in many respects has "broken" the simple e-mail approach. The convenience of adding attachments to other e-mail correspondence, or even to a more or less blank e-mail (e.g. "see attached") is not balanced by the effort needed to do something with the hanging attachment upon receipt (e.g. file, index, store, print...). There appear to be two alternatives to this requirement:-

- 1) Encourage users to store the attachment in a manner which will make it easier to find in the future, or
- 2) Adopt a different approach.

As indicated above, the former is poorly satisfied by the hierarchical folder structure offered by most e-mail packages. E-mail search engines do not in general search within attachments (whether stored in the e-mail folder structure, or outside). A better solution would be to put the attachment into some very coarse grained folder structure (e.g. "reports for 2002", or even "all my documents", and then use search / discovery technologies from the web environment (indexing, search) to locate the needed item in the future. Indexes could be automatically generated, thus making the corresponding search virtually instantaneous. This can be contrasted with the typical "find file" utility on a PC, which simply searches all files linearly, on demand. The use of search engines which support searching of metadata tags (which are increasingly being added to www documents) may prove useful, noting that metadata is needed as much for managed information archive as for discovery.

The second option is more radical, but arguably is more robust, and is a better model from the system viewpoint. The approach is to not send attachments at all, but simply a reference to the document, which is held in a central file-store. It can be viewed upon request or copied, printed etc, allowing central storage of a single copy, rather than multiple copies stored locally on each recipient's machine [Farrington 2002]. Note that this approach only works "locally" - there is still likely to be the need to send individually to external, geographically-separated recipients.

7. Applicable COTS and Open Source Software

The overall aim of this paper is to focus on generic solutions to the e-mail management /overload issue; an exhaustive review of available COTS solutions is therefore inappropriate. The following aims to indicate the richness of available solutions. It is recognised to be far from comprehensive and is certainly non-exhaustive.

Tools will be categorised into three topics: "indexers", "search" and "management".

7.1 *Indexers:* (see for example www.robotstxt.org/wc/active.html)

Indexing engines are available at many levels of sophistication and complexity. They range from local / personal freeware and shareware (from those with simple text and html capability to "professional" variants covering a wider range of file types such as .doc and .pdf), through intranet capable engines, to full www indexers of awesome capacity (e.g. Google™ currently claims to have indexed 2,073,418,024 web pages). Some of these indexing engines appear to be able to index e.g. Outlook express inboxes and folders, but few (possibly none) appear to cope with embedded e-mail attachments.

7.2 *Search:* (e.g. www.searchtools.com)

Thousands of search engines are available, from local freeware such as PSE (personal search engine)⁴ or Questagent ⁵, to corporate level search engines (often associated with indexers) such as Google™, or Altavista™. They can, in general, act on a local machine, an intranet, the Internet or any combination. Some also search on metadata tags.

7.3 Management:

There are two areas of interest in this category - conventional e-mail inbox management and attachment management.

E-mail Inbox Management:-

All will be familiar with the e-mail management features offered by such products as Outlook 97, Outlook 2000 or freeware / shareware clients such as Eudora or Pegasus. Typically they all offer folder hierarchies, sort capabilities and text search; Outlook also offers user-definable categories; Outlook 2000 has an "organise wizard" to assist in the process. Mail servers, such as the Microsoft Exchange server, all have store / forward capabilities.

Attachment Management:-

Mail Attender ⁶ manages the attachment at the mail server or via the system administrator function. It offers tools for document retention / deletion, search and archive - see the summary page at Annex B.

Attachment Manager ⁷ is a user tool to manage attachments locally (extraction, storage) - see the summary page at Annex C.

⁴ See www.softcorporation.com/products/pse

⁵ See www.jobjects.com/products/questagent

⁶ Sherpa software - www.sherpasoftware.com/MAExchFeatures.htm or

www.re-soft.com/product/mailattender.htm

⁷ See www.attachmentmanager.com

These latter two software packages appear to be new developments, perhaps aimed at exploiting the gap in the marketplace identified and explored in this paper. Their existence offers exciting new possibilities for the management of attachments at either the corporate or local level, opening up the possibilities proposed in this paper.

8. The Way Ahead for e-Mail Management

The "overload" effects of e-mail are likely to continue to increase, as additional functionality is added to the basic "carrier" mechanism. Recognition needs to be made that simplicity of use (fire and forget) is in danger of being outweighed by the complexity of the information management burden placed on the recipient. A different philosophy and a changed management process are essential to help to solve this problem. The key advance postulated in this paper is to separate out the different types of required functionality, and to offer management tools targeted on each type. In particular, a major benefit will be achieved by the separation of requirement type 2 ("attachments") from the others. It is suggested that attachments should be detached from e-mail at the earliest opportunity and put into a coarsegrained file-store. Such separation can be managed locally (through a tools such as Attachment Manager), or centrally (with tools such as Mail Attender). Freely available wwwstyle index / search tools should be used to manage this information repository. An enhanced second phase would be to have a centrally accessible file-store that should be used for local (and even remote) access. The URL is passed rather than the information (thus potentially conserving bandwidth, and not overloading inboxes), allowing the user to "pull" the item when / if required for online access and browsing. Search capabilities make the item accessible to a wider audience, which can be supplemented by publish / subscribe approaches. This approach clearly needs to embrace system management issues for e.g. low bandwidth environments, emphasising that network-wide system design and management is an essential ingredient.

9. Coalition Operations

Coalition operations are the norm in modern conflicts. Ad hoc groupings of different national elements are brought together, at relatively short notice, to prosecute a mission, be it peace support, humanitarian relief or full-scale warfare. Such ad hoc groupings are a real challenge in terms of timely information exchange. Information systems, processes and procedures all need to be put in place, often from scratch. In the context of this paper, however, such a "clean sheet of paper" approach could prove beneficial. It allows the enhanced ways of working and the use of more optimal tools postulated in this paper to be considered as part of the infrastructure for the ad hoc coalition group. It is suggested that the following steps be undertaken as part of the start up of the coalition force:-

- Clarify information transfer needs, and thus confirm differing requirements for the transfer mechanism (assumed to be e-mail & browser). These can be compared to the four categories discussed in this paper (i.e. basic text, attachments, workflow, diary & organiser).
- Determine level of IT support available to the coalition partners and central staff.
- Clarify necessary processes to underpin the information transfer (such as the level of information sharing, release approaches etc). These should promote centralised document stores, the use of indexing & search tools, and the approach of passing URLs

- rather than documents if the IT systems can support it. The need to publish information and the ability to support user profile subscriptions can be identified.
- Identify software tool support to the requirements and the processes; covering document storage, indexing and search tools, e-mail support, workflow (if needed), calendar & diary support, publish/subscribe mechanisms and attachment management software.

Application of these steps will allow some of the issues discussed in this paper to be "designed-out" of the coalition environment, allowing enhanced information access and retrieval.

10. Conclusions

This paper has addressed the complex issue of "e-mail overload". It has suggested that structuring the user requirements into discrete categories will allow targeted solutions to be derived for each. In particular, the separation of the management of "attachments" from that of the e-mail itself will allow benefits from the use of www index and search technology to be realised.

In summary, the best overall mechanism for each requirement category is:-

	Requirement	Mechanism	
1.	The need to pass "simple"	Linear and / or hierarchical file structure	
	text information as memo,	adequate. Sort, filter & text search tools in e-	
	query, report, URL etc	mail packages generally OK	
2.	The need to send	Phase 1 – send as attachments, BUT strip from	
	information (in almost any	e-mail at user or system boundary (incoming	
	file format / multimedia)	domain boundary) Use coarse grain	
		hierarchical file-store PLUS multimedia index /	
		search tools	
		Phase 2 - pass reference ONLY by e-mail;	
		retrieve via hyperlink. Alternatively use publish	
		/ subscribe mechanism. Use of local / global	
		index & search tools.	
3.	The need to pass documents	Managed adequately via workflow application	
	as part of a managed	layered over e-mail mechanism	
	workflow process		
4.	The need for a meeting	Managed adequately via calendar application	
	organiser / calendar / diary.	layered over e-mail mechanism	

Implementation, particularly of the second category, should be via a phased approach. Immediate action should be taken to set up each user with a "local library", together with tools to allow personal index and search of all received e-mail attachments. E-mail itself should be organised using "best of breed" hierarchies and categories supplemented by regular use of local search tools.

Longer-term action should be taken to explore options for the provision of a "corporate manager" to store attachments centrally on server and to then provide central indexing and search tools. The architectural aspects of this approach need to be fully considered. A change in culture will be necessary to popularise the approach of publishing to central server, as

promoted by work on the Shared Information Environment [Farrington 2002]. Defined processes are needed at system boundaries to cope with incoming and outgoing attachments. Further activities should embrace the concept of only sending references (URLs), rather than documents, heading towards the long-term goal of hyperlinking to server-held copies which are indexed regularly and accessed either directly or via corporately provided search capabilities.

Coalition operations may prove amenable to the early application of these concepts, given their need for timely information flows across ad hoc groups.

11. References

[Farrington 2002] M Farrington, *Shared Information Environment*, Unpublished report QINETIQ/KI/SEB/CR020508 March 2002

Annex A – tool summary

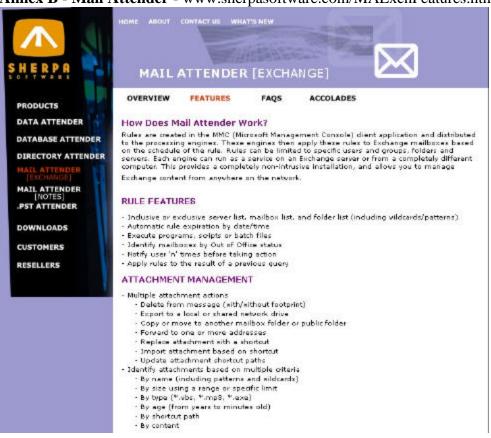
The following Table summarises the functionality available in typical e-mail packages to support the identified data-types.

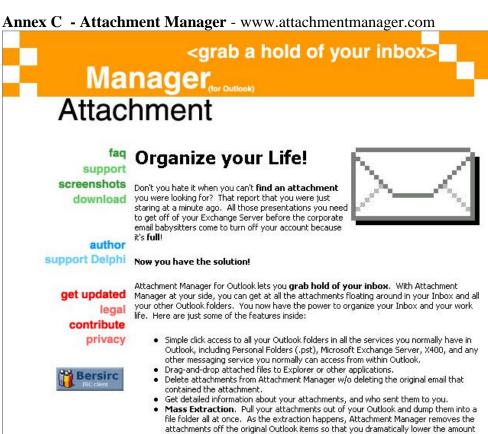
Type	Topic	Functionality	Availability
1	Mail text	Sort, file	Yes
		Filter,	Yes
		Search,	Yes (text)
		Metadata	Limited, yes
	URL		No specific support
2	Attachments	File	Yes
		Sort, filter,	No
		Search,	No
		Metadata	Limited
3	Workflow	Sort, file	Separate application
	documents	Filter,	
		Search	
4	Meetings /	Sort, accept, reject	Integral to e.g.
	diary		Outlook

In general terms, category 1 is well covered (as core email functionality); category 2 is surprisingly badly covered, perhaps explaining many users feeling of overload, given the rapid increase in the use of e-mail for this purpose. Category 3 is often supported by a workflow application either running their own e-mail support or running largely transparently over the e-mail service. Category 4 is sufficiently well handled by integrated functionality.

Mixed uses (especially of types 1 + 2) is very poorly covered, with COTS often exhibiting the worst of both worlds.

Annex B - Mail Attender - www.sherpasoftware.com/MAExchFeatures.htm





of disk space you consume on a Microsoft Exchange server. Extracted attachments can be grouped in folders by sender and Attachment Manager preserves multiple

versions of the same file so you don't lose valuable information.