Institut für Informationsverarbeitung und Computergestützte neue Medien
(IICM)
Technische Universität Graz

A Mail Archive Server on Top of Hyper-G

Diplomarbeit in Telematik

Alfons Schmid

Begutachter: o.Univ.-Prof. Dr.phil. Dr.h.c. Hermann Maurer
Betreuer: D.I. Dr.techn. Frank Kappe

30.11.1995

Abstract

The increasing employment of Internet services, especially that of Electronic Mail, confronts many users with the problem of processing, storing and eventually retrieving large amounts of e-mail. Many tools have been introduced to facilitate these jobs. However, most existing solutions are limited in their utility, because they apply archives - mostly files and directories - which are rigid in their structure once they are created, and they scarcely provide functions to manage large numbers of stored articles. Moreover, most of these tools are unable to properly process multimedia e-mail, the importance of which is steadily growing.

This thesis describes the design and implementation of an archive server which does not employ files for storage, but Hyper-G, a Hypermedia Information System. By this special approach, the server overcomes the weaknesses of conventional solutions, and in addition, it offers a lot of advanced functions which are new in the field of mail processing.

Table of Contents

List of Figures

Many thanks to D.I. Keith Andrews for preparing Figures 6.1 - 6.3.

Introduction 1

From the beginning, the Internet has fascinated the people who were involved with it, but never before has it been as popular as it is today. This is easy to comprehend, regarding the services and possibilities the Internet provides. Mainly aimed at communication, information exchange and information retrieval, these services give the users access to resources which have been unknown in such quantity, variety and availability.
One special Internet feature is Electronic Mail, a service that partly replaces telephone calls and conventional mail, because it is cheaper, more comfortable and faster than its alternatives. The application of e-mail speeds up many processes, with the effect that the users can do the same tasks within a shorter period of time. Consequently, their amount of work increases, but besides, they are now faced with the problem of how to organize the flood of e-mail messages, how to determine their relevance, in what order to read them, and how to store them with respect to easy retrieval for later reference.
Many tools have been introduced to assist e-mail processing: Mail Filters and Mail Archive Servers presort and archive the incoming messages according to user preferences, and sometimes they automatically take actions in response to certain keywords in message headers. However, designed especially for private ends and applying as archive a filesystem, most of these tools are useful only to a quite rudimentary level - they sort the messages and store them in files. Thus they are unable to meet advanced user-demands like flexible archive-structures, visualization of relations between messages, support of Hypermedia e-mail or powerful search mechanisms, just to mention a few.

In this thesis, a new approach in the design of a Mail Archive Server is taken. The underlying archive is no longer a filesystem, but a special Hypermedia Information System - Hyper-G. As this system has been designed for convenient manipulation of large amounts of information, it provides a lot of functions which are also very useful in the field of mail archiving and support the implementation even of advanced features like those mentioned above. Thus, the decision to choose Hyper-G for mail storage and build a Mail Archive Server on top if it is an obvious one, and it is confirmed by the resulting tool, which offers some innovative features.

This thesis is structured as follows: Chapter 2 gives a short survey of the Internet's history, introduces the services provided by the Internet, and, with special focus on e-mail, investigates how the users are influenced by this new kind of communication.
Chapter 3 then describes the revolution of text- and document design that has taken place in recent years and introduces Hypertext, text that virtually goes into the third dimension. The discussion always keeps an eye on Electronic Mail, which has to keep up with this development.
Chapter 4 gives an introduction to e-mail architectures and standards, describes common features of existing Mail Archive Servers and points out their weaknesses.
In Chapter 5, design goals are specified for a Mail Archive Server which should offer new and improved features for a better usability.
Chapter 6 introduces Hyper-G, which will serve as the archive for the new Mail Archive Server, and its UNIX client Harmony. It describes their features and how they support the implementation of the design goals developed in Chapter 5.
Chapter 7 describes an implementation which was designed to provide Harmony with a set of functions to create an e-mail system exclusively for Hyper-G. Although an ambitious project with interesting new features, it had to fail because of its entire detachment from the standard mail-system.
A second implementation, less daring but more widely usable, is detailed in Chapter 8. Based on the standard e-mail system, it meets the goals from Chapter 5 and still provides most of the innovations of the project in Chapter 7.
Finally, Chapter 9 lists a number of improvements that would make the new Mail Archive Server even better usable by adding some features and, especially, a better User Interface.

The Internet and its Services 2

2.1 - A Very Brief History of the Internet

Technologic Development
[3, 8, 26, 27, 60, 66]

Back in the early 1960's, when the political situation in the world was much more strained than it is today, the US government tackled with the problem of how its authorities were able to keep up communication after a potential nuclear war and tried to find a solution to that. In 1964 their efforts brought forth the model of a digital network; in order to meet their requests it should have no central control, and it should be able to operate even with parts of it destroyed.

This network should be materialized by applying the following principles: Firstly, it was assumed to be unreliable at all times and thus had to be designed in such a way as to transcend its own unreliability. This should be achieved by making all nodes in the network equal with respect to their status and priority to send, receive and pass on messages. Secondly, the messages, when sent across the network, should be split into packets, each separately addressed and numbered, which are then shifted from one node to another towards their destination. The particular route a package would take should be of no importance. Consequently, packets may take different routes to reach their destination. This involves the advantage that messages can reach their destination even when part of the network had been destroyed - they take their way across the remaining network.

Several teams of researchers in the USA and in Europe tried hard to physically implement that kind of net, and in 1969 the first network went in action. Installed in UCLA, it consisted of four nodes and was named *ARPANET*, after its Pentagon sponsor DARPA (Defense Advanced Research Projects Agency). The computers representing these four nodes were able to transfer data on high-speed transmission lines and they could be remotely programmed, giving researchers the opportunity to use computing power by long distance. Around 1975 the ARPANET already connected about one hundred nodes. The growth of the net encouraged changes and improvements in its design and architecture.

New technologies were explored, as for example communication by way of ground mobile packet radio (this net was called PRNET), or communication across the atlantic using point-to-point satellite connection (SATNET). The experiments were very successful and paved the way for ARPANET's global expansion.
Due to its decentralised design it was in fact very easy for the net to expand and machines of arbitrary architecture could be connected, as long as their communication followed a common packet-switching protocol. Originally, this protocol was the *Network Control Protocol* (NCP) [45, 54], but around 1983 it was superseded by a higher level protocol, which is the standard still today - *TCP/IP* [65]. The *Transfer Control Protocol* (TCP) is responsible for converting the messages into a stream of packets at the source and reassembling the original message at the destination; the *Internet Protocol* (IP) does the addressing and routing of the packets to their destination across the network.
At this time the term *Internet* was introduced, referring to the global interconnection of networks.
In 1984 the NSFNet, inspired and sponsored by the National Science Foundation, was introduced as a major backbone, again augmenting the ARPANET. It brought along many technological innovations and was continuously improved, so that finally in 1989 the original ARPANET expired, as a victim of its own success. In April 1995, the NSFNet backbone was retired, too. A fully commercial system of backbones has been put up, and the key networks, which had made all that possible, are gone, with no visible effects for the user.

In the course of the years, due to the open design technology and the TCP/IP protocol-software being public domain, many official institutions like government agencies, universities and commercial organisations connected to the net, but also minor networks joined in. So, in early 1995 the number of host computers had increased to about five million worldwide [27].
Trying to take a look at Internet's future is a very difficult and uncertain thing to do. Today's Internet has only little resemblance to the US government's original plans of a post nuclear war communication network - which is a happy irony. Somehow its development has always resisted planning, and presumably, just this will be the case in the years to come.

Services Provided by the Internet

The Internet provides its users mainly four services.
The *File Transfer Protocol* (ftp) allows the Internet users to access remote machines and retrieve text or programs. Countless sites all over the world can be searched for all kinds of information. Not only lots of software, documentation or computer-related material, but also electronic journals and magazines, even whole books are available via ftp. The amount of information available is so vast that Internet tools like WAIS [9, 39] or Gopher [16] have been developed to facilitate the search and exploration of these archives.
The introduction of the *World-Wide Web* (WWW, W3, The Web) [18, 19] and networked hyperdocuments has expanded this Internet feature. It provides an easier and much more comfortable way of representing information, and it has multiplied Internet's usability by introducing an interface that anyone could use. The Web and its features will be described in more detail in Section 3.2.

Long-distance Computing, one of the original intentions of the Internet, is especially popular among scientists, giving many researchers the opportunity to maintain remote accounts and use the power of super-computers, which may be continents away. This probably makes the Internet the most important scientific instrument today. It makes available enormous amounts of information of all kind, enabling rapid exchange of data, and thus speeds up the pace of scientific research. Also by long-distance computing, libraries offer digital search catalogues or CD-ROM archives, and large amounts of software are accessible.

Electronic Mail is another service which was introduced by the Internet. This feature will be expanded in Section 2.2.

Finally, the Internet provides *newsgroups*, a countless number of discussion-groups, which cover virtually every topic one can think of. This service constitutes a world of its own, featuring news, debate and argument, and it is generally known as USENET, which it is not a physical network, but rather a set of social conventions. The users, a gossipy and news-hungry crowd, come from every walk of life, which makes discussions vivid and colorful. The general tone is casual, independent of the position the author of a contribution or the recipient of an answer may hold, and this produces an air of fellowship. The opportunity for people from all over the world to "come together“ and discuss the

same topics eliminates borders, spans geographic distances and points out that people are the same, everywhere.
The number of newsgroups is increasing every day, as does the number of users contributing. Today, there are millions of USENET users, but trying to find exact figures is pointless, since they would be outdated tomorrow.

The Internet, a World of its Own

What makes the Internet very special and interesting apart from the services it provides is its anarchism. When connecting to the net the restrictions are of technical nature only. There is no social, political or ethnological control, and that is what makes the Internet Population so manifold - as already insinuated earlier. On the net one has the chance to get in contact with millions of people all over the world, chat, discuss and exchange thoughts with them like they were neighbours. This expands one's opportunities and perspectives, while at the same time superseding the limits of time and space.

However, in recent time several interest groups have formed among the users, all having their claims. Some want the net for educational and scientific purposes only, the government wants it fully regulated, others want a better financial basis. But tendencies to leave the net as it is - an institution without institutionalisation - are at least as strong. However, it is obvious that the Internet-World needs some kind of control. This control should not be too rigorous, leaving the users as much freedom as possible, but it should prevent the abuse of the net and violation of local laws.

In May 1995, the *Web Society* [49], an international non-profit organisation, has been founded out of the concern that the rate, at which the Internet is growing, requires accompanying measures.
The Web Society intends to be a strong representation of net users and non-profit net developers. - When anybody gets stuck on the net, he can contact the Web Society and get help.

The major goals of the Web Society are:
- to advocate the development of standards and open systems to make the net accessible as widely as it seems desirable;

- to support the development of tools and techniques to make information, which often is still hard to find, easier accessible;
- to keep the net open to an extent permissible by local laws, while at the same time supporting all moves to protect persons against the violation of their privacy;

The Web Society's goals do constitute some kind of regulation of the Internet, but they still keep limited the restrictions of the user's liberty of action.

2.2 - Electronic Mail, a New Kind of Communication

Already at the very beginning of the ARPANET, when only a few computers were linked together, the users, who were mainly scientists then, took advantage of the possibility to send messages to one another along the network. Just a few years after ARPANET's introduction it became obvious that the users were much more fascinated by this service, called *Electronic Mail* (e-mail), than by the opportunity to remotely program or access computers [66]. Still today, thirty years later, e-mail is one of Internet's most popular services, its importance and availability magnifying along with the net.

The Features of E-Mail

With e-mail, each user is given a world-unique Internet-address, similar to a phone number or a conventional address. When composing a message, the recipient's address, along with some other information, is put in its header, which constitutes a kind of envelope. The underlying layers, Mail Transport Agent and TCP/IP, use this address for routing the message from node to node along the net towards its destination.
At the destination, a file-server puts the message into the recipient's private mailbox - which is in fact a file. On demand, the contents of the mailbox is presented to the recipient, who can further decide, what is to be done with the messages. (More detailed information about e-mail and its delivery will follow in Chapter 4.1.)

Groups of recipients can be combined to kind of interest groups which are given a name (called *mail alias*, in a larger scale it is called *mailing list*). Sending a message with the alias as recipient distributes a copy of the message to each member of associated group.

In many situations, e-mail is increasingly replacing the usual communication media like telephone, fax or conventional mail. One reason is the speed at which messages are transmitted. Taking no more than just a few seconds, a message is delivered to its destination, no matter if it is sent just a few blocks or half around the world. A second reason is that people, when working on computer terminals, do not have to leave their working environment for depositing a letter or sending a fax (which is in the meantime - thanks to Internet - possible via computers, too) and they do not even have to move an arm to make a telephone call. Instead, they activate a mail handling program and send an e-mail message. Finally, e-mail is cheaper than all of its alternatives.
As comfortable as this may sound, some drawback are at hand as well. The high speed message delivery accelerates many tasks with the effect that people can do the same jobs within a shorter period of time. However, this does not mean that they may leave their offices earlier; instead, they have to pick the next job, and in all, this again speeds up their already fast-moving lives.
Another drawback is increasing net-traffic, which has the effect that transmission-speed drops, sometimes even to a degree that messages are delayed for several hour even if sender and recipient are within the same town. While still faster than conventional mail, a fax would deliver the information much faster.
What might happen in the future is that companies discover e-mail as a means for advertisement. They would post advertisement-messages to numerous people within a certain region, which might jam the net as well as the receivers' mailboxes. The receivers then would have the problem of picking out the important messages, and if the situation aggravated, really urgent messages would have to be sent by conventional media again.

Still, as electronic mail today is as new as the telephone was some decades ago, it offers a lot of new possibilities, and it brings new dimensions to correspondence, no matter whether private or official.

E-Mail and Cooperative Work

In 1977, the following paper was published:

"The ARPANET TELNET Protocol: Its Purpose, Principles, Implementation, and Impact on Host Operating System Design"
by Davidson, Hathaway, Mimno, Postel, Thomas and Walden;
Fifth Data Communications Symposium, Snowbird, UT;
September 27-29, 1977.

What is special about this paper is that the authors never had a meeting or made a telephone call about this paper. From the first ideas to the final version, everything was done by e-mail. This may not seem so special today, but it surely was at that time.
A very different example of cooperation via e-mail was a poem, cowritten by two people who never met in real live but got in contact over the net [37].

E-mail has almost become irreplaceable in the coordination of teamwork projects, regardless whether a team works within a single building or in geographically distant locations. No matter if it is a quick brainstorming, the announcement of a new program-version's release for testing, the necessity of design changes or just a message like "I'm not in the house today", it all can be done much easier and faster using e-mail. Usually this does not replace regular team meetings, but certain decisions can be made this way, which speeds things up; moreover, the overall level of information about what is done in the team is much higher.

Before the introduction of e-mail it had been impossible to form teams of experts who may live and work in different countries, but cooperate on a particular project in order to produce the best possible results. Just a few months ago (Fall 1994), an organisation for assembling such teams has been founded in Dublin, Ireland, and it is very well imaginable, that teamwork of that kind will become increasingly important in the years to come.

The Influence on Private Correspondence

Electronic mail brings people together. The speed at which messages are delivered gives correspondence a feeling of immediacy and brings virtually all net-users within convenient reach, no matter where in the world they may live. Both reasons may contribute to the phenomenon that people who did not bother to keep contact by writing letters or talking on the telephone revitalize their friendship as soon as they find out that they could correspond by e-mail. What may also contribute is that, even today, writing and receiving e-mail is somehow new and exciting, and people take every chance to communicate over this medium.
But for sure, it is very handy to write a few lines on the computer and post them by e-mail. When typing a message, one often does not choose one's words as carefully as when writing with pen and paper, and the messages may be very short. So, writing an e-mail message may take less time, and usually it does not take long until an answer arrives. Time may just be the point why people prefer e-mail to hand-written letters; it is a matter of convenience and meets the people's indolence. However, as it is always the case when typing a letter rather than writing it by hand, it lacks the personal feeling. But it is likely that e-mail will never fully replace mail written by hand - just like the existence of books in electronic form will not render printed books obsolete.

There is another phenomenon of private correspondence: The writing style on the Internet is somewhat casual, yet polite, motivated by the impression that all people on the net are fellows. "Talking" to net-users is like talking to friends, and one can always expect the others to be cooperative. This is even true when contacting somebody for the very first time. Yet, at the same time the net provides anonymity, a secure distance to the others. Users are not judged by their appearances or behaviour but only by their opinions and writing styles.
For some people - considering human nature, it may be many - this makes it easier to freely express their opinions and to become acquainted with other users. More than once, these acquaintances led to close friendships - and even to some marriages! [37]
It sounds paradoxical that the net-users are all neighbours and companions of some kind, while that feeling of fellowship is caused just by the distance and anonymity the net provides.

Advances in Document Design 3

Years ago, when e-mail and communication over the Internet was still new, the messages that the users sent one another consisted of plain text only. This was sufficient then, since the facilities to create documents on a computer were limited to text editors. Their function followed that of conventional typewriters and thus provided everything one had become used to. The only thing that was new, of course, was the way of transmission. However, as the potentials of text- and of dataprocessing on computers in general has increased - textprocessing with actual layouting facilities have appeared, images and audio documents can be included in texts, etc. - the request for transmitting such documents by e-mail, too, approached.
This chapter will concentrate on the revolution in document-design, solutions to the above request will be described in Chapter 4.

3.1 - From Plain Text to Hypertext

Over the last thirty years, in which e-mail spreaded along with the Internet, the power of computers grew exponentially, while at the same time the prices dropped, and by the end of the 80's computers had made their appearance in most households. The massive presence of computers and the new demands proposed by the users encouraged, maybe even forced software houses to produce more and still better products. (An additional factor surely was and still is that it is a very paying market!) The early text editors, which did not provide much more convenience than a mechanical typewriter, have been developed to powerful wordprocessors, allowing multiple fonts, page layout and insertion of images into the text, to mention only a few features. Designing texts in a way that was formerly reserved to experts with expensive machinery only is now possible for everyone. Image processing software today allows the user to create and manipulate pictures of very high qualitiy at very high speeds, doing things not even the best photolaboratory was able to do without computers. Just to mention an example, photographs, once digitized and brought into the computer's memory, can easily be processed by changing colors, removing or adding

details, distorting etc. These are potentials which became available through the very application of computers. The printing facilities' quality has been similarly improved, so the results from textural and graphic design are not restricted to computer-displays. With the resources of today's computers, audio documents at CD-quality as well as synchronized movies can be generated and reproduced, and they can even be combined with other types of documents to form multimedia documents.

These new possibilities changed the concept of information presentation, not only on a computer, but also in printed form. This again stimulated the development of new applications, but, as the more important effect, the users' claims with regard to information presentation have grown. Everybody has made the experience that the design of a document or an advertisement which was new and innovative some years ago seems outdated and cheap today, because new methods and techniques for doing the same have been developed and are common now. As it has always been the case, design is interpreted as a measure of quality.

But the improvements described above are not enough: Document design has been augmented by a "third dimension".

The Introduction of Hypertext

The first model of a hypertext, a text that could link various blocks of text, was described by Vannevar Bush in his article *As We May Think* [24] - although he called the concept *non sequential writing*. In this article Bush described *memex*, a conceptual machine that could store vast amounts of information and gave the user the ability to generate information trails, links between related texts and illustrations. The trails could be stored and used for future reference. Bush believed that using this associative method was not only practical, but also closer to the way the mind ordered information.

In 1981, Ted Nelson picked up Bush's idea in his book *Literary Machines* [53] and first coined the concept of *Hypertext*. He developed a system called *Xanadu* [2], in which users could create hypertexts - documents consisting of linked nodes.

> "A link is simply a connection between parts of text or other material. It is put in by a human. Links are made by individuals as pathways for the reader's exploration; thus they are part of the actual document, part of a writing."
>
> – T. Nelson

A node is a discrete unit of text, graphics, sound or whatever. Within the nodes there may be links to other nodes. This constitutes the basic structure of hypertext as Nelson described it. But in his project Xanadu Nelson's ideas went a step further; he envisioned an entire *docuverse* of interconnected, networked hypertext, a system which would replace print publishing.

The concept of hypertext has changed since Bush and Nelson proposed the idea. While their first models could have easily been accomplished on paper, today's hypertext seems restricted to computers. The notion of what hypertext is, what it can and what it cannot, has blurred. Hypertext surely assists the representation of large amounts of information, but not all printed texts will benefit from being turned into hypertext documents (so called hyperdocuments). And for sure, hypertext will not change the way humans think.

In this context another term has been defined - *Hypermedia*. There is no exact definition for that, but hypermedia refers to the ability to combine various media, such as text, images, movies, sounds etc., to a single document. This is not really different from Hypertext, but is used as the more specific term. [55]

3.2 - Networked Hypertext: *The Docuverse*

As expanded earlier, far more than five million host-computers connected to the Internet today, providing information part of which each Internet user can retrieve. Considering the Net with all its connected nodes, it represents an almost unlimitedly vast amount of information. With no special tools at hand it is impossible to seriously take use of this resource, because there is no indication of existence and location of specific items. Nelson's concept of networked hypertexts, he called it a *docuverse*, seems to perfectly match the problem of the Internet with its vast information-space. The nodes introduced by Nelson are still units of text, images, etc., and the links, whatever physical realisation they might have had in Nelson's first ideas, are now references identifying a specific

node on a defined host-computer. These links may span the world, that is why they are then called hyperlinks, and a single document may contain links to nodes spread over the whole Internet. This allows the association of related documents by establishing links (which, for example, offers the possibility of creating indices and making the topics accessible via hyperlinks), or to combine nodes of different sources and authors in a single document, just to mention two examples.

The World-Wide Web - The Breakthrough of Hypertext

Hypertexts had already been around for some time, but especially with the introduction of the *World-Wide Web* (WWW, W3, The Web) [4, 18, 19] this new type of document gained Internet-wide significance.

The World-Wide Web is a large-scale networked hypertext information system started by CERN, the European Laboratory for Particle Physics in Geneva, Switzerland. It introduces two special schemes, the *Universal Resource Locator* (URL) and the *HyperText Markup Language* (HTML).

URLs [10] help locate documents on the net and facilitates link establishment. They consist of three parts; the first one indicates the type of access to the remote document. Figure 3.1 shows the supported types.

```
http      hypertext transfer protocol
hyperg    Hyper-G transfer protocol
gopher    Gopher transfer protocol
mailto    e-mail address
ftp       file transfer protocol
news      newsgroups
wais      WAIS transfer protocol
```

Figure 3.1: Protocol Types Supported by WWW

The second part of the URL is the host's address and the third part is an index to the actual document. Figure 3.2 shows some examples.

```
ftp://ftp.univie.ac.at/mac/info-mac
http://www.w3.org/pub/WWW/Addressing/Addressing.html
```

Figure 3.2: Some Example URLs

HTML [17], the second scheme introduced by the Web, is a hypertext document description format, a mark-up language derived from SGML [28]. Its purpose is to improve computer mediated communication on the Internet by representing information in terms of available hypermedia technology.
Being a document description language like RTF or PostScript, HTML provides authors with tools to express their ideas at a fairly high level, so that the readers can use tools on their computers to navigate and display these ideas. Although the author's and the reader's tools may not be the same, there is a high degree of confidence that the idea will get through in-tact.
To go a bit more into detail, HTML is an SGML format [28, 38]. Special tags which are kind of commands in angle-brackets are used to give a document its layout. Examples of such tags are `<P>` for a new paragraph, `<H1>` to `<H4>` for highlighting phrases; there are tags for quoting, numbered lists etc. (See Figure 3.3 for more examples.) The exact way in which the formatted document is presented to the reader, e.g. with regard to character sets or colors, is determined by the reader's HTML parser, which displays the documents according to the formatting information.
What contrasts HTML from other SGML derivatives is that it is specially designed for describing networked hypermedia documents and thus provides a convenient way for the definition of hyperlinks. The hyperlinks supported by HTML are *embedded links*, links which are inserted into text documents using tags as described above (`<A HREF...>` in the example). Creating such links is easy, but their utility is limited, since they may emanate from texts documents only. So, in the WWW it is impossible for images or movies to contain links pointing to some other document.

For more information on HTML or SGML, please turn to the references suggested above.

```
<HTML>
<HEAD>
<TITLE>The Spanish Wine Page</TITLE>
</HEAD>
<BODY BGCOLOR="#FFDD90">
<CENTER>
<IMG SRC="es_logo.gif"  ALT = "Wines of Spain">
</CENTER>
<BR>
Welcome to the <u>Spanish Wine Page</u>. Spain is a world class producer of wine
both in quality and quantity. Better known are the quality reds from <b>Rioja</b
and <b>Ribera del Duero</b>, reds and "cavas" (sparkling wines) from
<b>Pened&eacute;s</b>,
fine whites from <b>Rueda</b>, and of course the "sherries" from <b>Jerez</b>. B
there's
more, lots more. The aim of The Spanish Wine Page is to help both newcomers
and knowlegable wine drinkers find and enjoy not only the known products, but
also to become acquainted with wines from more than 40 recognized wine
<A HREF="es_regn.html">regions</A> in Spain.
<P>


... stuff deleted ...

<H2>More About Spain</H2>
<DL>
<DT> <IMG SRC="es_bball.gif" ALT="*"><A HREF="es_gen.html">About Spain</a></DT>
<DT> <IMG SRC="es_bball.gif" ALT="*"><A HREF="es_food.html">Spanish Cuisine</a><
<DT> <IMG SRC="es_bball.gif" ALT="*"><A HREF="es_other.html">Other Things of
Interest</a></DT>
</DL>

... stuff deleted ...

</BODY>
</HTML>
```

Figure 3.3: Example HTML Text
Taken from http://www.eunet.es/InterStand/vino/es_vino.html

The introduction of HTML made the creation and presentation of hyperdocuments possible for everyone. Documents may now contain remote parts, transferred upon the user's demand, or they contain parts that were created by somebody else and are included by a reference to their location; and from these remote parts further links may lead even deeper into hyperspace.
Web-clients like *Netscape* or *Mosaic* simplify the use of hyperdocuments and support navigation through hyperspace.

Offering a new and exciting component, hyperlinks also entail a crucial drawback: The user, starting at some deliberate point in the information-space, is tempted to follow one hyperlink after another and thus gets deeper and deeper into hyperspace. But as a peculiarity of WWW, the user is presented just the current document and gets no feedback about the document's environment. So, already after following a few links the danger of loosing control over the current location is very high. This phenomenon of disorientation is called "*Getting Lost in Hyperspace*" [32], and in the WWW there is no remedy to it; the user has to keep control over his actions - and his curiosity.

Fundamentals of E-Mail 4

As pointed out earlier, Electronic-Mail has become a very important instrument of communication - in some respects it is more and more replacing conventional mail. No matter if it is private correspondence, project coordination, an announcement or a subscription to an electronic magazine, it can all be carried out, and in fact is carried out, by e-mail. As a consequence, someone taking massive advantage of this system daily has to cope with a flood of e-mail consisting of private mail, communication and discussion within an office or company, mail received from mailing lists and much, much more, depending only on the user's degree of involvement. Some part of the received mail is important, some is not, one part can be handled immediately, another one must be delayed. Mail has to be archived - putting together related topics, sorted, for example, by author or date, to be referenced later.
Supposed, an e-mail user receives some hundred mail messages every day and has to read and answer them, sometimes doing some research first, this can be a very time consuming job, already starting at the question of how to sort and in what order to read the mail. In fact, this is a scenario frequently encountered in real life, but since it is a problem in the field of data processing, several tools to its facilitation have been developed. These tools will be described below, after a short introduction to formats and standards of Electronic-Mail.

4.1 - E-Mail Standards and Formats

When a user wants to send or receive e-mail, he or she first must be assigned an e-mail address[46], as mentioned in Section 2.2. This address (Figure 4.1) consists of two parts, separated by the *at*-sign "@". The first part is the user's log-name, the second part is the host computer's Internet (IP) address. The *IP address* usually consists of a world unique 32-bit number, split into four parts (Figure 4.2.). Such numbers are difficult to memorize, and that is why a *Domain Name System* has been introduced, permitting the assignment of *Domain Names* to IP addresses; the names may then be used instead of the addresses. Domain Names (Figure 4.2.) are divided into levels (typically three to five, divided by dots)

which constitute a domain hierarchy (Figure 4.3). Since the Internet Protocol operates on IP addresses only, a *Name Server* is responsible for mapping the domain names back to IP addresses.

```
user@host
aschmid@fiicmhp01.tu-graz.ac.at
```

Figure 4.1: Format of E-Mail Addresses

IP Address	Domain Name
18.72.2.1	mit.edu
18.26.0.36	lcs.mit.edu
129.27.153.10	fiicmds04.tu-graz.ac.at

Figure 4.2: IP Addresses and Corresponding Domain Names

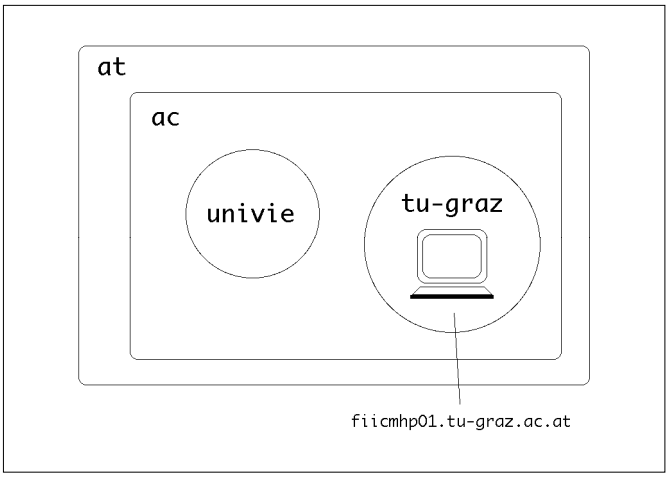

Figure 4.3: Domain Name Hierarchy

When an e-mail message is composed, the recipient's address must be put into the message header. The underlying services (Mail Transport Agent and TCP/IP) refer to this address when the message, maybe split into several packets, is routed through the Internet towards its destination.

Upon its arrival, the message is stored by the host computer and made available for the user. The exact way of storing, delivering and accessing e-mail, however, is different for various types of architecture [48]; the most important ones will be described below, before finally regarding the current e-mail standards.

Architectural Issues

The most common type is in *Timesharing Systems* (Figure 4.4). For each user, incoming mail is stored by the system and made accessible upon login by executing the *Mail User Agent* (MUA, a mail reading program). Transfer between timesharing systems is done by a *Mail Transfer Agent* (MTA). Since these systems always remain busy, mail can be delivered at any time. On UNIX-systems, e-mail works like this.

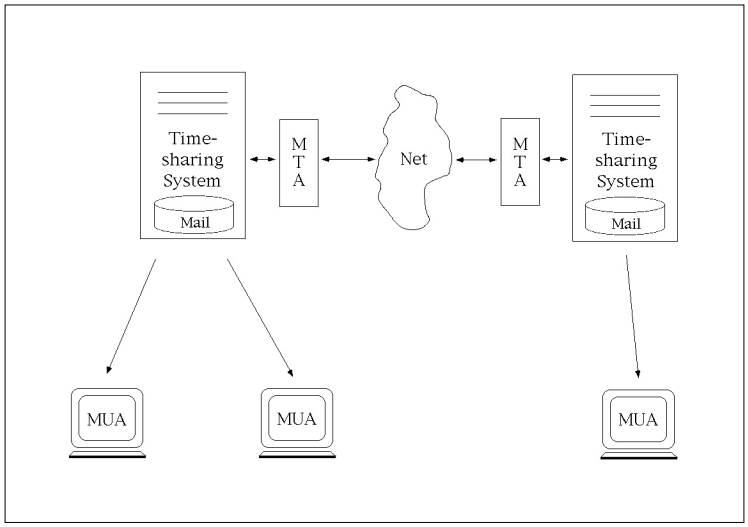

Figure 4.4: Mail Delivery in Timesharing Systems

A second method is used for networks of personal computers (Figure 4.5). Here, all messages are stored on a *central file server*. The MUA runs on the user's PC and transfers the mail from the file server using a *Remote File Transfer Protocol*. For mail exchange between file servers a special protocol is used.

A third type of architecture is the use of a *Mail Access Protocol* (Figure 4.6), which is becoming increasingly popular. In this case, mail is delivered to a mail

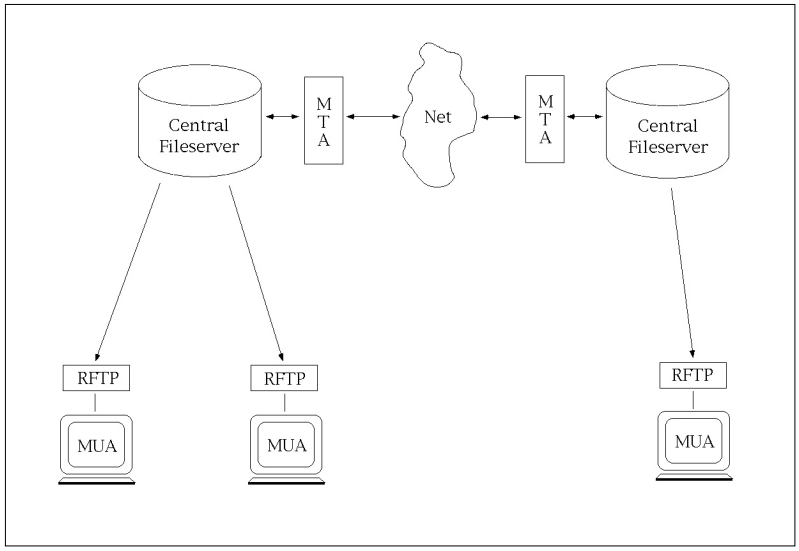

Figure 4.5: Mail Delivery with *Remote File Transfer Protocol (RFTP)*