FILE COPY

Data Handling Division
DD/87/26
D. Heagerty
October 1987

PRACTICAL EXPERIENCE WITH HIGH LEVEL GATEWAYS FOR MAIL TRANSFER

(Presented at the RARE Networkshop, Valencia, 4-6 May 1987 and
submitted for publication in "Computer Networks & ISDN")

DD-vf

PRACTICAL EXPERIENCE WITH HIGH LEVEL GATEWAYS FOR MAIL TRANSFER

Denise Heagerty, CERN

Geneva, Switzerland

ABSTRACT

Operating a Mail Interchange Service for a large and widely dispersed user community involves a considerable maintenance overhead. The routine tasks include updating of routing tables, checking network connections, processing 'undeliverable' mail, checking log files, analysing statistics, responding to user queries and arranging charging and accounting. Whilst gateways provide mail system connectivity, they suffer from a lack of transparency and do not provide full functionality between mail systems. This gives rise to specific problems such as how a user should quote his address when this depends on the mail system of his correspondent; name clashes can occur when disjoint networks become connected; message loops can be caused by information lost in gateways. In the longer term, a strategy to both reduce the maintenance overhead and improve the service for the users is the adoption of International Standards for Message Handling (currently CCITT X.400 Recommendations).

(Paper presented at the RARE Networkshop, Valencia, 4-6 May 1987 and submitted for publication in 'Computer Networks & ISDN'.)

INTRODUCTION

CERN, the European Laboratory for Particle Physics, operates a number of mail gateways for the HEP (High Energy Physics) Community. As well as providing connectivity between different mail systems in use across the CERN site, the gateways allow mail exchange with collaborating physicists in their home institutes throughout Europe, North America and the rest of the world. Amongst the networks accessible from CERN are EARN/BITNET, DECnet (approx 1000 nodes), Arpanet, CSNET, JANET, EUNET/USENET, EAN and other X.400 networks [12].

The Mail Interchange Service has been operating officially since March 1986, although mail systems have been in use at CERN for many years. This paper describes the tasks involved in maintaining and operating diverse mail systems and their gateways, gives an insight into typical problems arising due to gateways and concludes that a strategy to reduce the maintenance overhead and improve the service in the longer term is by the adoption of International Standards for Message Handling (currently CCITT X.400 Recommendations [3]). Before discussing the operational aspects, a brief outline of the functions of a mail gateway are described.

FUNCTIONS OF A MAIL GATEWAY

The basic function of a Mail Gateway, as the term is used in this paper, is to exchange messages between two networks running different mail protocols [13]. A message can be considered to have three parts: an envelope, header and one or more body parts [4]. The envelope contains routing information such as recipient and originator address. The header part consists of a number of standard fields such as from, to and subject. The header fields available in a mail system give an indication of its functionality since they are used to provide the user with mail system services. The body contains the message contents which users wish to exchange. A mail gateway must be able to recognise and create messages with the appropriate formats for envelope, header and body parts in the mail systems it is gatewaying; its functions can be broken down into:

- interfacing to mail systems
- mapping header fields
- body part conversion
- generating error reports
- generating log data

Interfacing to mail systems

There must be a mechanism so that messages destined for a foreign mail system are passed to the gateway. Similarly, the gateway must be able to insert gatewayed messages into the mail systems. The ease of implementing this interface depends on the particular mail system, but more recent manufacturers' products have started to provide programming interfaces to their mail systems.

Mapping header fields

The degree of functionality a gateway can maintain between mail systems depends on how closely the header fields can be mapped. If the services provided in gatewayed mail systems are similar then provided the gateway software implements suitable header mappings, the user should see little loss in functionality when exchanging mail with a foreign system. However, if one of the mail systems offers a much smaller range of services than the others then the gateway cannot compensate for this and users in the other mail systems will have reduced functionality. This also means that a mail gateway cannot always be made transparent to users.

Body part conversion

If the mail systems being gatewayed support only textual body parts and use matching character sets then no conversion is required. If the character sets do not match, as in the case of ASCII and EBCDIC, or if at least one of the mail systems supports other body types, for example graphical data, then the problem of body part conversion must be considered.

Generating error reports

As with most software, it is often the error conditions which are the most difficult to cope with. For this reason I consider generation of error reports separately from pure message exchange. A gateway must be able to generate error reports which can be understood in the mail systems. Whilst many mail protocols distinguish between an error report and a message, often the only solution for a gateway to convey errors in a comprehensible way is as a message. This can however have side-effects such as the problem of looping messages described in the section on Problems with Mail Gateways.

Generating log data

Detailed log information is essential for resolving user problems, for feedback into gateway improvements, as well as for providing statistics and accounting. At a minimum the gateway should log every message that it handles in terms of who it was from, who it is to, its size and whether the message was successfully accepted by a gatewayed mail system.

GATEWAY MANAGER TASKS

Providing a Mail Interchange Service for a large and widely dispersed user community incurs a maintenance overhead for the operation of the mail systems and gateways. Although effort has been made at CERN to automate tasks wherever possible, the operation still consumes the equivalent of one full-time person, with the work shared amongst experts for the different gateways. The tasks involved are described below.

Updating Routing Tables

Each mail system needs to know where to route mail given its destination [8]. For most of the mail systems at CERN routing is performed according to tables of local hosts and external domains. Each time a new local host or external domain is added to the mail network, all the tables must be updated. The maintenance effort can be reduced by not providing the full information on every host. In this case a default host can be routed to for any recipient address not in the local table. Obviously this has the consequence of increased traffic for the routing nodes and a delay in notifying users of unknown hosts or domains. Some mail networks such as EARN/BITNET distribute regularly complete tables of all sites and their routing data, others such as the RARE MHS Pilot X.400 network distribute only updates. Larger networks such as the ARPA Internet have changed from static tables to domain nameserver protocols [9]. This approach is likely to become more common as the tables continue to grow in size.

Checking Connections

Within the CERN site test messages are automatically sent around the major routes and gateways so that blockages and other problems are quickly detected and resolved. For most of the external connections statistics are available indicating the last time a successful call was established. Regularly checking the connections allows identification of external problems and can feedback into network optimisation, such as determining suitable call frequencies to certain destinations.

Checking Log Files

The log files give a good indication of how the mail networks and gateways are being used. They highlight common user problems which can be useful for identifying facilities and User Guides to be provided, as well as for feedback into future decisions and priorities. Checking why messages fail can identify problems in the system which may otherwise go unnoticed. Log files may be the only indication of a looping message problem. There are also 'undeliverable' messages to be dealt with. These occur when the mail system is neither able to deliver a message to a recipient nor return an error report to the originator. Log files can also identify any illegal traffic or other misuse of the mail gateway facilities.

Analysing statistics

Statistics such as the number of messages exchanged, 'key' destinations, and average size of messages provide useful feedback for adapting the mail service to user demands. The figures can be used for taking decisions such as whether a direct connection should be established to high traffic destinations, whether a leased line is more cost-effective, whether a file transfer service should be provided and estimating when current facilities will become overloaded.

Charging and Accounting

If the Mail Service is funded centrally then charging and accounting is not mandatory. For the moment charging and accounting is only performed for Swiss UUCP hosts making use of CERN facilities, however with the continued growth in use of electronic mail and increasing costs we will need to implement some mechanism internally to encourage use of the service in a cost-effective way.

User Advice

Whilst gateways can be designed to be as functional as possible, they cannot be totally transparent. Users have to learn how to address recipients on foreign mail systems and which functions are not always available through gateways, as well as understanding non-uniform and often cryptic error messages. In order to ease the addressing problem across mail systems the Mail Interchange Service provides a common address syntax for all the recommended mail systems in use at CERN [5]. There is also a CERN-wide Electronic Mail Directory Service. However addressing users external to the CERN site remains a constant source of confusion. User advice includes the provision and updating of User Guides, on-line help, Directory Services, address information services such as lists of hosts and domains as well as responding to individual user queries.

Fire Fighting

In a mail network running a single protocol occasional emergencies arise, however with gateways connected to many external networks the situation is exasperated. Some typical problems that occur in gateways are described in the next section. It is when they result in messages looping or blocking queues that immediate action needs to be taken.

Administrative Issues

Other tasks related to gateway operation which have not been mentioned above are of a more administrative or political nature. Amongst these can be grouped conformance to PTT regulations (e.g. switching of third party traffic), adherence to restrictions imposed by external networks (e.g. use of EARN/BITNET leased lines), naming conventions for sites and domains (there needs to be name management across networks to avoid clashes when disjoint networks are connected), and address mapping (there may be several ways to convert an address in a gateway). For the gateway machines, maintenance contracts and backup procedures must be considered.

PROBLEMS WITH MAIL GATEWAYS

Most of the problems that arise due to mail gateways are associated with addressing. Other problems can be due to incomplete implementations of gateways. A common example is when a gateway masquerades as a user thus causing error messages and replies to be returned to the gateway instead of the original sender. Each problem has its own peculiarities so rather than trying to categorise and discuss the problems on a general level this section describes some typical problems. The aim is to make the reader aware of the difficulties that arise due to the multiplicity of different mail systems and protocols in use in the outside world and over which a gateway manager often has no control.

How to Quote your Address

Each mail system has its own syntax for entering a recipient's mail address [12] so when users of the same mail system exchange addresses they have a common syntax. Within the CERN site we have arranged that all the recommended mail systems accept a common address syntax, but it is not possible to arrange that every mail system that wants to exchange mail with CERN also accepts this syntax. This creates a great deal of user confusion and no single way to quote your address to a colleague outside the CERN environment. Below is an extract from a CERN User Guide [5] giving hints to external collaborators whose mail systems will not accept the CERN common syntax - `user@host.cern`. This subset of possible address formats already shows quite diverse syntax for sending mail to someone at CERN.

SOURCE NETWORK	ALTERNATIVE ADDRESS FORMAT
EARN/BITNET	<code>user%host@cernvax</code> or <code>user%host AT cernvax</code> <code>jnet%"user%host@cernvax"</code> (from VMS)
DECNET	<code>host::user</code> (to a DECnet host at CERN) <code>vxgift::mint%"user@host.cern"</code> (to non-DECnet hosts)
EUNET/USENET	<code>cernvax!host!user</code>
JANET	<code>user@cern.host</code> <code>user%cern.host@ean - relay</code>
ARPA	<code>user%host%cernvax.bitnet@wiscvm.wisc.edu</code>

Address Mapping

A gateway should map the addresses in the header fields so that they conform to the formats used in the destination mail system. This allows functions such as reply to sender and reply to all recipients to be possible for messages which have arrived through a gateway. It also means that users see addresses in a form that they are used to and which they would enter themselves. However when messages are sent from CERN through a gateway to the EARN/BITNET network, we do not know with which mail software it will be read. The destination may prefer the syntax 'user@host.cern' or it may only recognise addresses of the form 'user%host AT cernvax', less sophisticated software may not accept addresses with components longer than 8 characters. There is therefore a one to many mapping and a decision must be taken which will reduce functionality for at least some EARN/BITNET users. Care must also be taken that addresses do not include special characters such as '\$' or '_' since these can have unexpected results on remote systems or simply not be available on the keyboard.

Looping Messages

There is more than one cause for messages to be passed continually between mail systems and never be delivered to a user. One example is described here. Consider a message to a distribution list where replies are to be sent also to the distribution list. It may happen that someone in the list has set their mail system to automatically reply to any message. Within its own environment that reply may be recognised as a report and not redistributed, but once the automatic reply has passed through the gateway it appears as a normal message so will be sent to everyone on the list, causing another automatic reply. Automatic replies will continue to be distributed to everyone on the distribution list until there is manual intervention.

Name Clashes

Many mail systems are using domain oriented addresses of the form user@host.domain where domain may be a list of subdomains. Within a domain there is usually a naming authority to avoid clashes at the subdomain or host level. Name clashes can occur however when joining previously disconnected domains or subdomains. Name clashes can also occur due to address mappings at gateways. Consider a mythical country with ISO code CC. It has three independent mail networks: EARN, Grey Book and X.400. On each network there is a user called ET and a node called SPACE. Because these networks are not connected to each other there is no confusion. However all three networks are connected to CERN. Their addresses in their local formats are respectively ET@SPACE.CC, ET@CC.SPACE, /S=ET/O=SPACE/C=CC but when they have passed through gateways and been converted to the CERN common syntax they all map to ET@SPACE.CC. So now there is one address representing three different people and we do not know which network to send to. This example highlights the need for a common naming authority in each country independent of the existing mail networks. Networks within countries should also be interconnected rather than forcing external sites to maintain the overhead of multiple connections into a country.

EXPERIENCE SUMMARY

Operating a Mail Gateway Service involves much more than just installing the software. The overhead is a function of the number of gateways, the number of connections, the size of the user community and the quality of the mail and gateway software — not only run locally but also at the external sites users are communicating with. Although many of the tasks can be automated there will always be problems due to the fact that gateways cannot provide full functionality between mail systems and cannot be made totally transparent to the users. Detailed log information has been proved essential for resolving user problems, for feedback into gateway improvements, as well as for providing statistics and

accounting. We estimate that an average of 400 messages per day are passing through our main gateway node carrying more than a megabyte of data. The most popular external destinations are sites on the EARN/BITNET network. Traffic figures for the Swiss EARN International node, located at CERN, are closer to 2,000 messages per day carrying 50 megabytes of data [6].

From the user perspective, the overriding problem is how to address 'person x at y'. This has been eased slightly with the introduction of a Directory Service on the CERN site [7]. However until Directory Services are available on a global scale, the problem of representing a foreign address syntax on your home mail system will remain. Other concerns of users are reliability, reachability (who they can exchange mail with), speed and functionality.

CONCLUSIONS

The operation of mail gateways at CERN was always seen as a short to medium term solution to the mail interchange problem [1,2]. After more than one year of providing a Service for a large user community using a variety of mail systems, it is clear based on both the operational aspects and user feedback that a strategy to reduce the maintenance overhead and improve the service in the longer term is the adoption of International Standards for electronic mail [10]. CERN's communications policy, supported by its user community of High Energy Physicists, is to migrate to Message Handling Systems conforming to the CCITT X.400 Recommendations.

As a first step in the migration strategy, we have been running EAN [11], a pilot implementation of X.400 from the University of British Columbia, integrated with the Message Interchange Service. The gateway approach allows sites both within and outside CERN to adopt suitable X.400 products as soon as they become available without losing connectivity to those that can only migrate at a slower pace. In the longer term, adoption of X.400 will provide increased connectivity and functionality with little overhead. Messages can be routed more efficiently since they will no longer need to pass through gateway machines and the maintenance overhead of dealing with problems due to gateways will be reduced.

REFERENCES

- [1] U. Beyschlag, COMICS Final Report, CERN/DD/COMICS – PPR/R4 (1986).
- [2] U. Beyschlag, Going X.400 Example of an Intercept Strategy, Proceedings of the 2nd International Symposium on Computer Message Systems, IFIP, Washington D.C. (1985).
- [3] CCITT, Data Communication Networks Message Handling Systems, Recommendation X.400 – X.430 (1984).
- [4] I. Cunningham, Message – Handling Systems and Protocols, Proceedings of the IEEE Vol 71, No 12 (1983).
- [5] D. Heagerty, A User Guide to Electronic Mail Services at CERN, CERN DD/US/6 (1987).
- [6] D. Heagerty, Electronic Mail Status, September 1986, CERN/DD/COMICS – GEN/G8.
- [7] F. Hemmer, Networking with Oracle, CERN DD/86/11 (1986).
- [8] S. Kille, MHS Use of Directory Service for Routing, Proceedings of IFIP Conference on Message Handling Systems, Munich (1987).
- [9] L. Landweber, D. Jennings & I. Fuchs, Research Computer Networks and Their Interconnection, IEEE Communications Magazine Vol 24, No 6 (1986).
- [10] S. Malde, Message Handling Standards, Computer Communications Vol 9, No 2 (1986).
- [11] G. Neufeld, J. Demco, B. Hilpert & R. Sample, EAN: An X.400 Message System, Proceedings of the 2nd International Symposium on Computer Message Systems, IFIP, Washington D.C. (1985).
- [12] J. Quarterman & J. Hoskins, Notable Computer Networks, Communications of the ACM Vol 29, No 10 (1986).
- [13] D. Redell & J. White, Interconnecting Electronic Mail Systems, IEEE Computer, Sep (1983).