How email works

Contents

How email works
What is email?
What makes up an email?
How does email work?
The client
The server
How do email servers work?
What are TCP/IP ports and how are they related to email?
What makes up an email address?
How does an email get transmitted from client to client?
What should I look for in an email server?
How does an email system send attachments?
What types of email servers are there?
Examples of email servers
Frequently-asked questions
What is SMTP used for?
What is POP3 used for?
What is IMAP and how does it differ from POP3?
Is email safe?
Is email just an informal means of communication?
With the gradual move towards off-site web-based email services, do I need a local email server?
What type of internet connection is required?
What is a mail queue?
What are MIME and S/MIME and what’s the difference?
What is the relationship between DNS, MX and A records and who controls them?
Can I use email to send secure information?
Will spammers be able to access my server?

Interesting and useful links

How email works

Electronic mail, or email, has come a long way since the advent of the Internet, with billions of emails being sent and received each day. It is perhaps the Internet’s ‘killer’ application as more people use email services than browse the web. Email has become the way to communicate efficiently, quickly and cheaply.

As the Internet has no distance issues, it is far simpler, cheaper and quicker to send an email halfway round the world than it is to send a traditional letter. It is just as easy to send an email to the person sitting next to you, as it is to send an email to a hundred different people around the world.

Rather than explaining how to use email, which is well documented elsewhere, this document will look at the parts and workings of the email system that are not normally seen by the user and will give an insight into what happens behind the scenes.

What is email?

An email is an electronic message transmitted over a network from one user to another. Email should perhaps be credited with advancing the spread of the Internet more than any other
application or use of the Internet. When the Internet (or ARPANET as it was called) was first introduced it didn't take long before 75% of all network traffic was email related.

An email can be simply a few lines of text sent from one user to another, or include attachments such as pictures or documents. It can be a newsletter sent to subscribers daily detailing what’s happening and when, or an encrypted message sent between government organisations containing classified information. Email has many uses and it is now becoming a mature, feature-rich method of communication that appears complicated, but has an underlying technology that is very simple.

What makes up an email?

The actual email itself is organised into parts called the header and the body. The header contains information about who sent it, when it was sent, and the subject and size of the message. This is the part that is normally displayed to users before they open the full email; it is often displayed in bold text on one line. When the email is opened it displays the body as well as the header – the body contains the actual message sent.

Attachments are often sent with emails and these are normally displayed as icons within the email that relate to the attached document’s type. Alternatively, they may be embedded into the message body.

How does email work?

Email is based on a client-server model.

The client

The client carries out the user’s interactions with the email server. Clients can appear in various forms:

- Application based – these are installed onto users’ machines and include Microsoft Outlook and the freely available Outlook Express and Eudora.
- Web based – these appear in a web browser’s window and include Hotmail, Yahoo and Outlook web client.

Clients vary greatly in functionality, but all provide a basic level of functionality that assists the user. Functions include the ability to create new emails, display and store received emails, hold address lists of contacts, a calendar, journal and other extra functions that help organise the user’s working day. The client is also configured with the account information and names or IP addresses of the email servers with which it will be communicating in order to send and receive emails.

The server

An email server is typically a combination of processes running on a server with a large storage capacity – a list of users and rules, and the capability to receive, send and store emails and attachments. These servers are designed to operate without constant user intervention and should happily process emails for months as sending, receiving and maintenance tasks are carried out at scheduled times.

The client only has to connect to the email server when it sends and checks/receives new email. Sometimes it may be permanently connected to the server to allow access to shared address books or calendar information – this is typical of a LAN-based email server.
How do email servers work?

Most email servers conduct email services by running two separate processes on the same machine. One process is the POP3 (Post Office protocol 3) server which holds emails in a queue and delivers emails to the client when they are requested. The other is the SMTP (simple mail transfer protocol) server that receives outgoing emails from clients and sends and receives email from other SMTP servers. These two processes are linked by an internal mail delivery mechanism that moves mail between the POP3 and SMTP servers. When the client calls the email server to send or check for mail it connects to the server on certain TCP/IP ports (see below): SMTP on port 25 to send email and when it checks for new email, POP3 on port 110. Figure 1 shows the relationship between the clients, servers, POP3, SMTP and the internet.

Figure 1: The relationship between clients, email servers and the internet

What are TCP/IP ports and how are they related to email?

Most email servers run on a webserver platform with email services installed and possibly other services as well. Each server has one or more unique TCP/IP (transmission control protocol/internet protocol) addresses. Attached to all TCP/IP addresses are many ports that range from 0 to 65,535. TCP/IP uses ports to allocate different jobs to different services. The server will listen for a client or application to call it on a port and direct traffic from that port to the required service.

Figure 2 shows an example of a web server with the TCP/IP address of 193.63.50.162. It carries out many different tasks such as website hosting and FTP (file transfer protocol) as well as email. As each request is made from clients requiring information, they are made to a specific port that is allocated for that task. TCP/IP uses different ports for different requests so that it can easily identify what process each request is for.

Figure 2: An example of the ports used for various requests
These TCP/IP ports are used for all kinds of internet jobs: port 80 for serving browsers with web pages and port 21 for file transfers to be routed to the FTP application. Emails usually use port 25 for SMTP and port 110 for POP3.

**What makes up an email address?**

An email address has two parts: the email account (which is the actual user’s account) and the domain name (which is the internet registered name for the organisation). These two parts are separated by the @ (pronounced ‘at’) symbol.

The following example shows a user called Fred Bass with an account on the Techno domain. The email account name will normally be set out in a certain format, the most popular being ‘forename.lastname’ or ‘forename_lastname’. This is followed by the @ symbol to show the end of the email account and the start of the domain name.

Fred.Bass@Techno.com

The domain name is usually the company’s registered address on the Internet and placing ‘www.’ before the domain name should give the organisations’ website. Email addresses never have spaces in them.

**How does an email get transmitted from client to client?**

Here is a simplified version of what takes place in the transmission of an email from one client user to another:

A sender called Bob Doe works as a teacher at Maryscott School in Hampshire. His email address is bob.doe@maryscott.hants.sch.uk. He wishes to send an email to John Far who works at Tami School, Cumbria. Bob has John’s email address on a business card. Bob starts his email client and selects ‘Compose new message’. Bob types John’s email address ‘John_Far@tami.cumbria.sch.uk’ into the ‘To’ box, and chooses not to ‘CC’ (carbon copy) or ‘BCC’ (blind carbon copy) the message to any other users. He types in his subject as ‘Rugby’. He then proceeds to type the rest of the message into the main body: ‘Hi John, can you send me some information on rugby at your school? Thanks, Bob’. As emails are fairly informal, Bob decides this is enough text and clicks the ‘Send’ button. Now the automated processes take over.
The client makes a connection to its local email server ‘mail.maryscott.hants.sch.uk’ using SMTP port 25. The client passes the email’s sender and recipient addresses along with the main body of the message to the email server.

When the server has received the email it looks at the domain part of the recipient's address, the part after the @ symbol – ‘tami.cumbria.sch.uk’. If the domain is local to the server – ‘maryscott.hants.sch.uk’ – it will simply pass the message onto the local POP3 server for delivery. As the recipient’s domain is not local to the email server, it passes it onto the email server that looks after that specific domain.

It does this by asking another server or service called DNS (domain names server) for the IP (internet protocol) address or MX (mail exchange) record of the email server that deals with email for ‘tami.cumbria.sch.uk’. The DNS replies with an IP address for the other domain’s SMTP server. Once the email server has this IP address, it opens up a connection to the recipient’s SMTP server on port 25 and passes the message on.

Once the receiving SMTP server has the email, it checks the domain part of the email address, recognises it as its own local domain and passes it internally to the POP3 server. The POP3 server holds the message in John Far’s mailbox. The POP3 mailbox holds the messages in a queue until the user, in this case John Far, connects to the email server on port 110 to collect mail. He downloads his queued POP3 mailbox from the server to the client and receives Bob’s email.

What should I look for in an email server?

Email servers come with many features that are designed to keep the system running and reduce maintenance time for administrators. A typical email server should include or provide:

- Full support for POP3, SMTP and web-based clients
  Web access to mail (Webmail) is becoming more important and allows users to access their email with only an internet-connected web browser. This is especially useful for offsite access.

- Some form of filtering
  Third party software companies can provide a more comprehensive solution, but the server should have some inbuilt filtering functionality.

- Support for multiple antivirus products
  This allows for unforeseen changes in antivirus software suppliers and the installation of multiple antivirus scanners on an email server.

- Comprehensive support, logging, monitoring and administration features
  Support and administration from remote locations is extremely useful when sites do not have a permanent support presence.

- Multiple domain support
  Most establishments have, or will have, more than one domain.

- Public folders and shared address books for all users to access.

- Auto-responder support for users that are away for a period of time.

- Relay control to prevent the server from being used by unauthorized users, such as spammers.

- Easy backup and restore tasks.

How does an email system send attachments?

SMTP can only transmit text – this creates a problem when it comes to sending images, video and other attachments via email. SMTP gets around this problem by using two different methods.
One method is a program called UUencode. UUencode assumes that the file or attachment contains binary information (1s and 0s). It converts this binary information into text using a simple mathematical equation, almost a type of encryption. Once UUencode has converted an attachment into text, it can travel via SMTP.

The UUencoded file begins with a header that tells the receiving system the original filename. The receiving system's UUdecoding software sees the header and translates the data that follows until it reaches the end marker. This translation results in the file being converted back into its original form.

UUencoding does have problems and there are various formats available, some of which are unable to handle multiple attachments.

An alternative to UUencode is MIME (multi-purpose internet mail extensions). MIME was developed because of the need for a system to translate the array of constantly changing attachment formats. MIME works in a similar way to Uuencode, but creates a header that it wraps around each encoded attachment. This allows it to encode images and sound.

The choice between using UUencode or MIME is normally dictated by the sending email server. Most modern servers use MIME to send their mail, however almost all servers still carry support for UUencode to allow for backwards compatibility.

**What types of email servers are there?**

Email systems come in various formats, but the most common rely on a single server that provides both POP3 and SMTP services. Sometimes, in large organisations, these services are separated onto different servers. The majority of email servers use Windows NT or 2000. When purchasing or obtaining the software that will run the email server, the supplier will normally recommend a certain specification of server components, such as processor, memory and storage capacity, that will be required to support the amount of users.

**Examples of email servers**

- **Windows** – IMail [http://www.ipswitch.com/products/IMail_Server/index.html] and Mail-Gear [http://enterprisesecurity.symantec.com/products/products.cfm?productID=17] are both popular with education establishments and business users alike. They provide good all-round functionality and stability with comprehensive user support. Exchange 2000 [http://www.microsoft.com/exchange/default.mspx] is Microsoft's email and collaborative messaging solution. Although more expensive, it provides all the services of a comprehensive solution, including very good levels of support.

- **Linux** – there are many email solutions for Linux on the market. Some of these are of poor quality and provide little support. As the Linux email market matures, we can expect to see more comprehensive products for this platform, such as Sendmail [http://www.sendmail.org/], a collaboration between Compaq/HP and Sendmail Inc.


An email server is a long-term purchase that may cost more to maintain and run than purchase. Take care when you select a product and make support and reliability your priorities.
Frequently-asked questions

What is SMTP used for?

Simple mail transfer protocol was invented during the 1980s as a standard for the transmission of messages. It is the language that most mail servers use to send messages between each other.

When the SMTP mail system sends a message it uses DNS to convert the domain part of an email address (such as @becta.org.uk) to the TCP/IP network address of the machine that maintains the domain (such as 193.63.50.162). It then connects to that IP address on port 25 and uses very simple commands to communicate the sender’s and recipient’s email addresses and the body of the message.

An example SMTP conversation:

HELO mail.becta.org – says hello to the receiving server and sends the sender’s server’s name.

RCPT TO <joe.bloggs@becta.org.uk> – the recipient’s address.

MAIL FROM <an.other@somewhere.co.uk> – the sender’s address.

What is POP3 used for?

POP3 (post office protocol 3) is used by clients to collect email from email servers. Clients must supply a username and password to the server in order to log into their account or POP3 mailbox. The email server will respond with the number of messages waiting and the client can initiate a ‘deque’ command to download the queued emails. The messages will either be deleted from the email server or marked as read so they are not downloaded again.

POP3 is very good at the simple task of collecting email and can be used as and when required. The POP3 transactions between client and the email server are similar to SMTP.

An example POP3 conversation:

USER JoeB – passes the user or account name to the server.

PASS starfish – passes the password to the server.

STAT – shows how many messages are waiting for collection.

What is IMAP and how does it differ from POP3?

IMAP is the internet message access protocol used for accessing email stored on a server. It is a protocol that allows users to perform certain electronic mail functions on a remote server rather than on their local computer. The fundamental difference between accessing email via IMAP as opposed to via POP3 is that IMAP does not download the messages and store them locally as POP3 does. All message manipulation, such as opening, closing and deleting, is carried out on the server. This makes backup simpler and security tighter since no emails are actually stored locally on the user’s own computer.

Is email safe?

Email can be filtered, censored and monitored by internal or external systems and devices. As email is a flexible tool that can be accessed and used from almost anywhere, it is entirely possible to use it for malicious purposes and it is quite often the cause of virus infections.
By following a few basic rules, it is possible to use email safely and responsibly:

1. Always check all emails for viruses, keeping the antivirus software up to date.
2. Filter email from all sources.
3. Implement an AUP (acceptable use policy) for email.
4. Implement the blocking and filtering of profanities and anti-spam features.
5. Do not allow your server to be used as an open relay.

However, this is by no means a comprehensive list and further sources can be found at the end of this document.

Is email just an informal means of communication?

Email gave early users – and still gives current users – a form of equality. In a paper published in 1978 by the Institute of Electrical and Electronic Engineers, two of the important figures in the creation of the ARPANET, J C R Licklider and Albert Vezza, explained the popularity of email:

‘One of the advantages of the message systems over letter mail was that, in an ARPANET message, one could write tersely and type imperfectly, even to an older person in a superior position and even to a person one did not know very well, and the recipient took no offence. Among the advantages of the network message services over the telephone were the fact that one could proceed immediately to the point without having to engage in small talk first, that the message services produced a preservable record, and that the sender and receiver did not have to be available at the same time.’

Email still presents this equality to users today; it is possible to transcend barriers of race, culture and wealth. Some email users even find it bad netiquette to include more than the basic level of information or content required in each mail. Even so it is important to realise that emails cannot be considered to be private or confidential and are subject to both copyright and libel laws. Emails should never be intimidating, hostile or offensive on the basis of sex, race, colour, religion, national origin, sexual orientation or disability.

With the gradual move towards off-site web-based email services, do I need a local email server?

A local email server provides a single point where all email traffic can be monitored and controlled to prevent the spread of viruses and other potentially harmful content. Should an off-site, third party webmail service be used, this control is severely compromised. Webmail has the benefits of being easy to access, easy to use and convenient. It is quite possible to provide webmail services on a local email server that can be accessed from off-site locations. Webmail does not have the full feature set of an installed client, such as Outlook Express, but users find that it fulfils the basic read, write and reply services they require. If users require these services then, if possible, they should be provided on-site so control over content and malicious actions can be exercised. Use of third party webmail should be discouraged, unless the service is properly controlled and monitored.

What type of internet connection is required?

Bandwidth requirements increase with the installation of any email server. Although it is not possible to predict with certainty the amount of bandwidth required per user, an ISDN 128kbps line would be the absolute minimum, with a substantial increase in bandwidth required for any attachments or prolonged use.

Certain ISPs can provide SMTP and POP3 services that allow the local email server to download and send emails at preset intervals, perhaps every two hours. This can reduce the
amount of bandwidth required by restricting email-related traffic to quieter hours of the day or night. A static TCP/IP address is normally required for SMTP services to be provided. If a firewall or router is being used this may need to be reconfigured to allow traffic to pass through ports 25 and possibly 110, to allow SMTP and POP3 to function.

**What is a mail queue?**

When the sending SMTP server cannot connect to the receiving SMTP server, the message is held in an outbound mail queue. Delivery of messages is retried at set intervals, such as every 15 minutes, for the first few hours. After four hours the user is normally notified of the delay and then, again at preset times, until either the message is delivered or the delivery times-out, in which case the sender will receive notification of failure. These times can be altered manually on the server but the default settings are generally considered acceptable for most circumstances.

If the sending server cannot connect to the receiving SMTP server or find a valid TCP/IP address for the receiving domain's SMTP server, either because of a spelling error or because the receiving server is down or busy, it returns the email to the sender with a permanent failure notice with the reason for failure. Some larger organisations have several SMTP servers and the sending server will try them all in turn before advising of delay or failure.

**What are MIME and S/MIME and what's the difference?**

MIME (multi-purpose internet mail extension) is an extension of the original email protocol to allow clients to transfer file formats other than text over the Internet. The client converts the email and any attachments into MIME format and includes a MIME header in the transmission to inform the client what type of document (audio or image) it is sending.

S/MIME (secure multi-purpose internet mail extension) is a secure version of MIME that prevents interception and forgery. It ensures privacy, tamper detection and verification of sender. Both the sender and receiver need S/MIME compatible clients for them to exchange email securely using this method.

**What is the relationship between DNS, MX and A records and who controls them?**

DNS (domain name service) is the Internet’s ‘directory enquiries’ and translates names into contact numbers. Each domain will have a DNS entry, containing information about where to find the website (the A record), email servers (the MX record) and other relevant information.

An example of A and MX records:

www.becta.org.uk A 142.13.113.34 – all requests for the Becta website will get directed to the webserver with IP address 142.13.113.34

becta.org.uk MX 142.13.113.35 1 – this is the MX record for Becta. All emails will get sent to the email server with IP address 142.13.113.35. If there are multiple servers dealing with email for a particular site they will have a suffix from 1 to 10, depending on their priority to deal with mail requests.

DNS entries are normally controlled by ISPs due to their importance. If the DNS records are configured incorrectly, web and email services may cease to function correctly.

**Can I use email to send secure information?**

Emails are normally sent in plain text, MIME or UUencode format which can be an insecure way to communicate. It is possible to secure emails and this is normally carried out at the client level using a digital ID that incorporates S/MIME or PGP (Pretty Good Privacy). This verifies the sender's authenticity and can also be used to encrypt the message itself.
To send a digitally signed message, you must obtain a digital ID. To send an encrypted message, you require a digital ID for your own use and the digital ID of the intended recipient.

You can obtain digital IDs from independent certification bodies who will verify your identity before issuing them.

**Will spammers be able to access my server?**

Most servers include some anti-relay or anti-spam security measures that prevent the server being used as an open relay. Two popular methods of preventing emails that originate from unauthorised users are making clients authenticate with the server before they are allowed to pass email through, and denying access to clients that are not located on the local network.

**Interesting and useful links**

- The history of email
  [http://www.vicomsoft.us/knowledge/reference/email.history.html]
- All about DNS, A and MX records
  [http://www.merakmailserver.com/Knowledgebase/Merak_Mail_Server/Configuration/DNS.asp]
- How to use email (part of the ICT Advice site)
- Superhighway Safety, DfES
- JANET RFC Guide for Email
  [http://www.ja.net/documents/Mailguide/mailguide-ref.html]