Support Group Application Note
Number: 265
Issue: 1.0
Author: DW

Acorn

Connecting Acorn Computers to the Internet

This document describes the hardware, software and procedures necessary to connect one or more 32 bit Acorn computer to the Internet. It provides a list of UK Internet Service Providers and useful questions to ask them, and details of considerations to take into account once the system is connected and working,

Applicable Hardware :	All 32 bit Acorn computers running RISC OS 3.1 or greater	Related Application Notes:	234: Peripheral Interfacing via the Serial Port 282: Writing Command Scripts for Acorn InterTalk 283: TCP/IP Addressing, Subnetworking and Interoperability: an Overview 284: Using InterTalk across Gateways

Support Group
Acorn Computers Limited
Acorn House
Vision Park
Histon, Cambridge
CB4 4AE

Table of Contents

Introduction	3
What is there to Gain by Connecting to Internet?	3
Planning, Connection Options and Service Providers	4
Pre-Installation Issues	
Local Infrastructure	5
Dial-up or ISDN?	5
Storage Prediction	6
Installation and Implementation Issues	
Choosing an ISP	7
Addresses and Domains	9
Domain Names and How They Work	9
Email Addresses	10
Choice of Services	10
Maintenance Issues	
Censorship and Auditing	11
Controlling Newsgroup growth	12
Internet Resources: Getting Started	12
Security Considerations	13
How to Find Out More...	14
Appendix A: Contents of the Accompanying Disc	14

Introduction

The Internet is a world-wide network of networks with gateways linking organisations in North and South America, Europe, the Pacific Basin and many other countries. The organisations, ranging from businesses through educational establishments to government and military sites, are administratively independent from one another. There is no central, worldwide, technical control point. Yet, working together, these organisations have created what to a user seems to be a single virtual network which spans the globe.

The networks all use a common suite of networking protocols, TCP/IP (described in some depth in Application Note 283). It is because of this commonality of protocols, network functionality and interoperability that the networks provide what may appear to be a seamless, integrated virtual network, irregardless of the heterogeneity of the underlying computer hardware or physical communications carrier.

The most basic functions provided are electronic mail, access to remote computational and information facilities, and file transfer; many other functions have since been built on top to create a highly multi-functional environment. The networking protocols were first deployed in the late 1960s in the United States by ARPA.

In the UK, Internet activity was started by gateways through JANET, the Joint Academic NETwork, which links all UK Universities. JANET originally used a different addressing protocol to the US Internet (or ARPANet as it then was), which was resolved in the early 1990s by changing from the original X.25 and X.400 "Grey Book" system to JIPS, the JANET-Internet Protocol System.

As private and commercial interest in the Internet has expanded, companies have been set up whose business is to provide easy-to-access connections to the Internet; these companies are referred to as Internet Service Providers (ISPs).

What is there to gain by connecting to Internet?

An institution must first address the question, "What will this institution gain from participating in the Internet community?".

Both commercial and non-profit education and research institutions spend a great deal of high level effort to define their mission and goals. Any introduction of new technology - particularly one which involves new modes and methodologies of communication - should be assessed in light of the institution's own mission and goals as well as the wants and needs of the user community it serves.

Following, and as part of this evaluation, key institution decision makers (at the highest levels of the organisation) will require information not only on the cost of connection, but more importantly on the purpose and scope of participation in the Internet. The decision to participate requires not only the strong commitment of senior administration but also the support and endorsement of the general institutional community.

The deployment of an Internet connection may provide the impetus for the development of a site-wide strategy for the use of information technology in many curriculum subjects, which may otherwise never be accessible. It may be difficult to quantify such benefits, but they must be included in the justification process. Many institutions have already done this and are very likely already connected. An interested institution might well consult with a nearby, connected organisation to see what benefit they have derived from the connection. An institution looking at a connection for the first time must decide if a major reason is simply to participate in a technology that has already proven itself as being important to education; more importantly, it may be a requirement now to compete with peer organisations.

Many universities have also discovered economic efficiencies. Many subscription services have traditionally required a dedicated and expensive access method. More and more of these services are now accessible via the Internet. This trend will undoubtedly continue as more and more commercial companies make their services available. While the subscription fee may not alter, the cost of the dedicated connection may by used to finance an Internet connection; not only will the availability of the particular service be greater but the underlying access medium can be used for multiple functions.

Libraries, many already with automated catalogues, are looking at various new applications to deal with the glut of information, shrinking dollars and limited shelf space. Electronic journals, image-based text, publishing on demand are all issues that are being evaluated for the digital library. Universities are already automating and integrating a variety of activities and providing access to the students and staff via a campus network.

For examples of some of the educational resource material available over Internet, see the "Internet Resources: Getting Started" section later in this document.

Planning, Connection Options, and Service Providers

An institution must anticipate and prepare for four critical phases in the deployment of an Internet connection. The list of issues discussed within this document is not exhaustive, but rather the information provided should alert decision makers to major concerns they should address during the different phases of network deployment.

As each issue is discussed, soft and hard cost items will be identified. Both must be considered when determining the real cost of deploying an Internet connection. The hard cost items include costs for which invoices are created. They include the costs for new circuits or telephone lines, the purchase of modems or routers, network membership subscriptions and upgrades to existing hardware to make it network compatible. Soft costs are harder to quantify but no less important. These costs include training and education of staff, faculty, and students, modifications to support staffing and structure, deployment of new network applications or network services such as FTP servers, centralised electronic mail services, or campus-wide information systems. It should also be recognised that the soft costs involved also result in benefits that can easily be seen as people investment and organisational investment.

The four phases of an Internet Connection deployment are:

1. Initial planning/Pre-Internet installation phase
2. Implementation and Startup phase
3. Full Production/Maintenance phase
4. Evaluation/Upgrade phase

Pre-Installation Issues

Local Infrastructure

The first decision which must be made is whether to connect just a single computer to the Internet, or to implement a proxying Internet gateway system such that Internet access may be available across all Ethernet-connected stations on a local area network (eg one which uses Acorn Access or Level 4 fileserver). Unless the site administrator is confident in their network management abilities, and especially if the staff who will be using the system are unfamiliar with the services the Internet offers, it may be wise to connect a single machine initially for the purpose of evaluation and training.

If a full network connectivity solution is to be implemented, one machine should be designated as an Internet proxy server; it should be ensured that all machines which are intended to be able to use Internet services are on either the same Ethernet network as the machine which will act as the Internet proxy system, or on an Ethernet network which can be routed to by the machine which will act as the Internet proxy system. See Application Note 284 for further details concerning routing across local gateways.

Hard costs which will be incurred at this stage for a network installation involve the purchase of whatever cabling installations, gateways and network cards are necessary to connect the local infrastructure together. If it is planned to use interactive services such as World Wide Web (as opposed to services which can be performed in a batch-mode manner, such as the processing of email and USENET News), there will be a soft cost of assigning a machine as a permanent Internet proxy, removing it from use as a client station; if the only services to be used will comprise email and News, the station can be retained as a client system and set to process net traffic at night.

Regarding the choice of machine to use as an Internet proxy, the following issues should be borne in mind:

The proxy system will need at least 2MB of RAM (4MB is preferable). RISC OS 3.1 or later is also required. If you intend to use an A3000 for this purpose, then you must fit a serial upgrade.

Not all Acorn computers will benefit from being attached to high speed modems. As a result the Risc PC and A7000 best suit the purpose of acting as a server which accesses the Internet, achieving serial port throughput speeds of up to 115200 bits per second (115Kbaud) from the Desktop. The A5000 and previous computers will support speeds up to 9600 bits per second (9.6Kbaud) reliably. Any RISC OS system with 3.1 or later and 2MB may act as a client system, even if it has no local hard disc.

Dial-up or ISDN?

A variety of options exist to facilitate connection to the Internet. Factors besides costs may be used to select the appropriate option, or a series of options. These factors include size (bandwidth) and projected use (traffic) of the connection, nature of the use and purpose of the enterprise driving the effort.

For a site which plans to connect a single machine, or which intends to connect a local area network via a proxy but which plans only to use relatively low-bandwidth services (such as email and carefully selected USENET Newsgroups), an acceptable speed of connection can be obtained using a modem and a dial-up connection operating over a standard telephone line. As on-line costs are obviously an important soft cost centre, we suggest the purchase of a fast modem to gain the benefit of lower call charges. A good basic guide would be to choose a Hayes compatible modem, supporting either the V.32bis (14.4Kbaud maximum) or V.34 (28.8Kbaud maximum) protocol standards.

Regarding suitable cables to connect the modem to the computer, your local Acorn agency will be able to provide you with a suitable cable if one is not supplied with the modem. In the case of the A3010, A3020, A4000, A5000, A7000 and Risc PC a standard cable designed for IBM PC compatibles will work. For earlier models (Archimedes 300, 440, 400/1, 540 and A3000) you will need to purchase a modified cable to ensure that serial data handling between the computer and modem is performed correctly. Details of these cables are supplied in Application Note 234.

ISDN (Integrated Services Digital Network) provides a faster link which allows a site to transfer data to and from the rest of the Internet much faster than by using a modem, making it a useful option for sites intending to provide facilities for bandwidth-intensive Internet use (eg Web browsing) by several simultaneous clients on their local network. ISDN establishes a dedicated line between the site and the Internet service provider. This can be switched if needs be to other sites for other applications which may require ISDN, such as videoconferencing. Typically ISDN offers two B channels, each providing a data transfer rate of 64Kb s^{-1} combining to offer a data transfer rate of 128 Kb s^{-1}. Faster ISDN is also available.

ISDN requires specialist hardware and the rental of a direct line, and this results in it being considerably more expensive than the use of a modem. However, the effect on performance, particularly for intensive use, means that for many users in education it is becoming increasingly attractive.

Rather than a modem, ISDN connections utilise bridges; these are broadly similar in functionality to ordinary Ethernet bridges (which simply link adjacent network segments, allowing pass-through of packets from one segment destined for a machine on another segment), although they must provide their own routing capability.

Your Internet Service Provider will advise you on the cost and availability of ISDN subscription, and both Acorn and your carrier company can advise you on an individual basis regarding the cost and choice of the necessary ISDN bridging hardware you will require to connect the dedicated line to your existing (or future) Ethernet cable installation.

Storage Prediction

Depending on the amount of traffic you predict will be travelling between your site and the Internet at large, you may find it necessary to add extra hard discs to your Internet proxy system. The quantity of email traffic per user will depend on the amount of computer access time available to them, and the number and nature of any collaborative projects they are involved with; a very crude rule-of-thumb figure, which hopefully incorporates some degree of futureproofing, is that circa 10K of textual email per user per day will need to be sent off-site, and as much again can be expected to be received. If you budget for 2 to 2.5 times this when assigning hard disc space, and (especially if you are using a dial-up connection) implement a policy whereby sending off-site of MIME-encoded binaries is the exception rather than the rule, this will give you a workable service once you take into consideration the amount of email traffic which may be distributed internally (which, as no external connection features, can feature MIME encoded inclusions; the only penalty here will be mail storage and, should large messages be picked up by multiple recipients simultaneously, increases in load on the local fileservers and local area network).

Storage allocation for News requires some forethought. Once the system is running, it is probably best to make groups available for site use a few at a time; before a likely group is made generally available, the network administrator should subscribe to the group for a few weeks so that the group may be vetted both for quality of content and the amount of storage space it will consume.

If at any point the volume of articles in a group appears excessive, the best way (unless old messages in the group are being kept online and active for a specific reason) to reduce the volume of a group is to run an expire on it, or if automatic expiry is operative, either to increase the frequency at which expire is run or make expire affect more recently posted messages. News administration is very much of a compromise, however if the system becomes difficult to manage owing to volume issues, more storage can always be added at a later stage.

The hard cost centre involved here is for storage media and possibly further disc interfaces for the Internet proxy.

Installation and Implementation Issues

Choosing an ISP
An Internet Service Provider (ISP) is a commercial company providing Internet access for companies, institutions and individuals. The choice of an Internet Service Provider depends on several factors; the priority of any of these above the others depends entirely on the situation of the establishment in question. A list of known UK ISPs, along with the services they support, is provided on the disc supporting this Application Note.

Broadly speaking, the issues which affect the choice of ISP are:

1). *Whether you intend to connect a single machine or a local area network to the Internet.*
Single-user systems require no internal email forwarding once the message has reached your site, so POP3 (Post Office Protocol version 3) is an acceptable protocol to use for receiving email. Dynamic IP address allocation is also feasible.

If you wish to achieve access to the Internet on a site-wide basis, with each user having their own private account from which they may make email transactions, mail forwarding will be required between the Internet proxy machine and the personal mailbox of the intended recipient, wherever the mailbox may be stored on your local network. In order for received mail to be forwarded, you will specifically require an SMTP (Simple Mail Transfer Protocol) connection for receiving mail; from a user perspective, POP protocols and IMAP allow a user access to a single mailbox on a remote machine, whereas SMTP supports forwarding and more generic access. Ensure that you check this with your provider before subscribing.

All mail is sent using SMTP, regardless of the connection type or IP address allocation method adopted.
See Application Note 282 for details of how to configure Acorn InterTalk for static or dynamic IP address allocation.

2). *Distance to the nearest Point of Presence*
Most ISPs operate on a "Point of Presence" (PoP) system for dial-up connections; a PoP can be considered to be a modem capable of answering incoming telephone connection requests, demodulating the signal and connecting to the ISP's mail and news host systems, which are in turn connected to the Internet backbone. In most parts of the UK, it is reasonable to expect to be able to find a PoP within the range for which local call rates apply. Be aware, however, that some ISPs only have support at some of their PoPs for SLIP **or** PPP (see issue 3 below), not both; this may require you, depending on whether your account with such a provider is configured to be based on SLIP or based on PPP, to ring a PoP other than your nearest one to obtain a connection.

Considering the amount of time for which your system may need to be connected to the Internet, especially if you intend to use World Wide Web, telephone charges will be an important factor in your overall expenditure on Internet connectivity.

3). *Supported Protocols*
For dial-up services, the two protocols which are used to "frame" the IP packets so that they keep coherency when travelling across the modem link are SLIP (Serial Line Interface Protocol) and PPP (Point to Point

Protocol). Some details on these protocols (especially PPP) are provided in Application Notes 234 and 282. InterTalk version 2.x and later has built-in PPP support, whereas earlier versions only support SLIP.
Note also that, if you intend to use ISDN, not all ISPs have the infrastructure to support it.

4). *Mailboxes*
Mailboxes are essentially a POP3 concept. A "mailbox" is effectively an area of filespace on the ISP's mail server where email destined for a user with a particular email address is stored. When the site which owns the address connects to the ISP's mail server with the user ID of the user who owns the mailbox, it is able to download the user's mail.

This is fine for a single-user home-based system. Problems can arise, however, when a single system is to be used by more than one user, as is the usual practice in education. If the ISP only provides for one POP3 mailbox per site, then unless the number of mailboxes can be extended by request the confidentiality of email is severely compromised; received email would be readable by anyone who had access to the machine, and in effect, a system of user accounting and accountability could not be implemented.

If you use SMTP, however, these issues do not arise.

5). *Costs and Services*
The cost of internet connectivity varies depending on the provider. For dial-up connectivity, you will normally be expected to pay an initial amount which relates to the administration costs of setting up your individual account. Thereafter, a monthly service charge can be expected. This is in addition to your telephone call charges. Most ISPs only levy a flat monthly charge, which may be different depending on whether you will be using a single IP address (eg for one machine) or whether you require a Class C network allocation (see Application Note 283 for network class details). However a few ISPs charge per unit of data transfered over their Internet link, and others charge per unit of time for which you are connected to their PoP; be aware of these "extra" charges, especially if you intend to spend a significant amount of time online.

In addition to the basic connection, many ISPs offer a telephone support hotline. Make sure that your ISP is knowlegeable about Acorn systems, and is prepared to offer connections and support to potential users without requiring users to run software proprietary to the ISP. Epithets from ISP staff such as "Oh, you can't use the Internet unless you're running Microsoft Windows" indicate not only a lack of knowledge of Acorn systems, but also a lack of knowledge of the Internet; TCP/IP has been ported to just about every piece of computing hardware which has a serial port or Ethernet interface and enough memory to hold an IP stack, regardless of its operating system or user environment. In fact, the systems which support the majority of the popular Internet sites run UNIX, the Operating System on which TCP/IP was originally developed and deployed.

An increasingly common feature bundled with Internet connectivity is a filespace allowance on the ISP's own server; this enables you to provide data (usually in the form of World Wide Web pages) about your site or organisation which can be accessed by other Internet users without your own system or network being connected to the Internet at the time of access. Although not particularly useful at the current time for the home user, resource of this kind can be valuable to an educational establishment for provision of information about the organisation, displays of students' work, etc.

Once you are happy with the cost and type of service, contacting and sending payment to the service provider will result in your account being set up. Further information will be forwarded to you in the post. You will use this information when configuring your software.

Addresses and Domains

Broadly speaking, Internet addresses are subdivided into blocks with different classes; the concept of IP addressing and subnetworking is described in Application Note 283. An "IP domain" is the list of IP addresses assigned to an organisation for its own use, and tends to be defined in terms of the largest blocks of address space (by class) which belong to that organisation; for example, the Acorn Computer Group's domain can be described as the addresses with the Class B network number 136.170.0.0. This block is subdivided further to create subdomains for Acorn Education, Acorn RISC Technology and Acorn Online Media, which map onto the actual domain as seen by the outside world. Acorn's domain is a subdomain of the domain assigned to corporate entities within the UK.

Similarly ISPs have domains assigned to them, and they in turn sell blocks of addresses to subscribers who wish to connect their networks to the Internet.

Rather than having to memorise lists of octets in dotted decimal format, the DNS (Domain Name Service) exists to translate systematic site names into the appropriate IP addresses; provided that there is a route by which one DNS server can connect to any other DNS server, DNS enables a user connected to a DNS to resolve any registered address in the world. Included as part of any Internet connectivity package will be some form of resolver which interfaces to the ISP's DNS server. Once you have DNS connectivity, the address component acorn.co.uk is mapped onto any address within Acorn's domain.

The body which deals with registering textual names with the DNS is InterNIC; however, in most cases your ISP will handle the registration of your site name for you.

Domain Names and How They Work

In most cases, UK sites have an address which looks like *sitename.name_of_service_provider* (eg should St. Trinian's School have a subscription with Demon Internet Ltd, their site name could be sttrinians.demon.co.uk unless that address was already in use; effectively the contents of the "sttrinians" portion would be at the discretion of the site); however, it is possible, often at the cost of an additional fee to cover mail routing at the ISP, to have any site name you wish provided that the name you choose is not already in use, and has a "tail end" portion which accurately describes the type of organisation operating at your site.

If you wish to take advantage of the capability to completely define your own site name when you go online, your service provider should be able to assist you with the registration process; otherwise, you will need to contact InterNIC directly The form to complete, along with detailed instructions, is supplied on the enclosed disc and should be returned by email to HOSTMASTER@internic.net

Domain names work in a right-to-left order, with each successive subdomain separated from its neighbour by a dot; rightmost, unless you are in the USA (which uses no national code), affiliated to one of a very small number of global organisaions which use ".int" (short for "international") or not using a separately-registered name and attached to a service provider which uses a ".net" address, the rightmost entry in your domain name is the domain covering your country. A fairly comprehensive list of these is supplied as the "countries" file on the enclosed disc. Second from the right is the type of organisation at your site; the available options are:

USA	UK and Elsewhere	Description
com	co	Commercial company
edu	ac	Academic institution (usually higher eduaction)
gov	go or gov	Government organisation
mil	mi	Military site
org	org	Any organisation not covered above (these tend to be used for non-profitmaking organisations)

In addition to these domains, there are changes in progress which are expected to result in the creation of the .sch.uk domain, for use by UK schools.

Domains can be subdivided further into *site.organisation_type.country*, or even *machine.site.organisation_type.country*; what happens to the left hand side of *site* is under the control of the administrator of that site.

Of course, there are a small number of exceptions to this rule. JANET, the academic net for UK Universities, still has a legacy of a "middle-endian" scheme of username@uk.ac.university.machine (eg ma2043@uk.ac.bristol.seqa), but as Grey Book protocols have effectively been obsoleted by JIPS (the JANET-Internet Protocol System), you shouldn't see many of these.

On rare occasions, particularly when sending email to the USA, you may find that you encounter some very long addresses where the elements are separated by "!" instead of "." Such addresses are known as "bang paths", and pre-date DNS. The following definition of a bang path and how it works is extracted from "The New Hacker's Dictionary" (aka "The Jargon File"):

bang path: An old-style UUCP email address specifying hops to get from some assumed-reachable location to the addressee. Each hop is indicated by a !. Thus, for example, the path ...!*bigsite*!*foovax*!*barbox*!*me* directs people to route their mail to machine to *bigsite* (presumably a well-known location accessible to everybody) and from there through the machine *foovax* to the account of user *me* on *barbox*.

For more information on UUCP, see the "Tools" document on the enclosed disc.

Email Addresses and How They Work

An email address uniquely identifies a user (or group of users, if accounts are shared) at a site, and takes the form *username@full_site_domain_name* On InterTalk servers, email usernames mirror the names given to user accounts. To comply with standard practice, all sites should have an account called *postmaster* to be held by the site administrator, and it is also advisable to have the postmaster also hold an account called *root*

Choice of Services

On examining the sheer number of services which may be used over the Internet (most of the widely-used protocols are listed in the "Tools" document on the disc), you may feel that it would be exceedingly difficult to support all these services; fortunately you do not have to, as a number of them are falling into disuse (eg Gopher) and others only tend to be used within particular domains (eg WHOIS is used at very few sites other than those within the US .gov and .mil domains).

The common protocols which are definitely worth supporting on a fully-operational and well-organised system are:

ftp	(basic file transfer)
http	(World Wide Web)
smtp	(email, if you have a fixed IP address; if not, the ISP handles smtp for you)

The other protocol which you may also find useful is IRC (Internet Relay Chat; the most secure way to use this in an educational environment is via KidLink, see "Censorship" section later).

ftp and http servers are available for Acorn systems, and can be used to make resource material and pupil's work readily available over the Internet.

Maintenance Issues

Censorship and Auditing
In an educational environment, it is almost inevitable that some students will attempt to transmit or receive material which is deemed inappropriate for their use. In the case of USENET News, censorship of material may either be provided by the ISP, who may decide that the simplest option is not to carry any newsgroups from USENET-at-large at all, or by the site network administrator. The simplest method of censorship involves only subscribing to groups which carries appropriate material; most groups make the suitability of their material obvious in their title, and if the site administrator has any doubts, a look at the group's FAQ (Frequently Asked Questions list) in the news.answers group will often provide the definitive statement on what is and what is not topical to the group. If a newsgroup is listed as being moderated, there is a very good chance that only messages containing information pertinent to the topics in the group charter will be permitted to exist in the group.

Currently the only effective method of locally moderating a useful group (which may nonetheless have completely off-topic and sometimes offensive messages posted to it by malicious users at other sites) is for the admin to check the list of message subject lines in all subscribed newsgroups by eye and delete problematical messages between their download and any users seeing them; the same applies to private email between users.

Systems have been investigated by other purveyors of Internet access software which purport to be able to filter unsuitable material from an email or News stream; the problem which arises is that the dictionary of unsuitable words against which plaintext messages would have to be checked would be immense, as it is trivial in English to convey the meaning of a word without using the correct spelling (even inserting punctuation characters, say, would fool many checking systems), and as a typical newsfeed would comprise several MB of messages per day, the checking process would thus require a huge amount of computing resource.

Regarding censorship of interactive systems, it is possible within InterTalk to produce lists of World Wide Web sites such that only sites on the list may be accessed, or all sites except those on the list may be accessed. The decision was taken to keep the scope of Web censorship on a site-wide basis since, as new pages are being written and made available over the Internet at the rate of three new pages every ten seconds, a single administrator would be unable to cope with censorship on the scale of anything smaller.

Administrators should be aware, should the latter option of disallowing access be used, that proxy systems and gateways could be configured to fetch files from "forbidden" sites and return them as email, this process being achieved by passing the address of the "forbidden" site as a piece of text in an email message (see the "Email" file on the enclosed disc); the simple way to foil this tactic is to forbid email access to the robot sites described in the information on the enclosed disc.

Multi-user mutually-interactive systems such as IRC (Internet Relay chat) can be problematical regarding the exchange of inappropriate material, however there exist IRC systems such as KidLink which, in addition to having a real-time censoring program to check for the more obvious uses of inappropriate language and eject the issuing user, always have an adult superuser online to ensure the peace is kept.

However, KidLink has limited resources available to carry out effective real-time moderation by a member of staff, and so access is limited.

In the cold light of day, 100% accurate and effective censorship of Internet-based material is impossible to achieve. However, sight must not be lost of the fact that the 100% effective censorship of *any* media, electronic, interactive or otherwise, is fundamentally impossible in a nation which upholds the right to freedom of speech; the Internet is, in this respect, no different from the national press or popular literature. The best compromise with the Internet, as with any other medium, involves a careful choice of which newsgroups are subscribed to and which Web sites are permitted or forbidden access.

Should transgressions occur and unsuitable material be downloaded, it is possible using InterTalk for the site administrator to determine both the source of the information and the user responsible for the downloading by examining the remote access logs; thus the source of the material can be added to the "access forbidden" list and the culprit subjected to the appropriate measures. The access logs can also be used by the administrator to produce profiles of Internet access on a per-account or per-machine basis.

Controlling Newsgroup growth
As newsgroup postings are downloaded to a site proxy, so the volume of disc space consumed by USENET postings on that proxy increases. Unless limits are imposed on the quantity of disc space which News is allowed to fill, and appropriate actions taken once these limits are reached, the space taken by News will expand until the proxy's hard disc is full.

The volume of News stored locally is usually controlled by expiring (ie deleting) postings which are more than a configurable number of days old; unless a group has particularly few postings per unit time, or needs to have "old news" retained for some other specific reason, a reasonable metric would be to expire messages which are more than a week old. If you find that you need to subscribe to a binaries group which carries a large number of new messages per day, you would be wise to expire the messages in this group after a day or two; messages posted to binaries groups tend to be large.

The expiry procedure makes intensive use of system resources, and so is best scheduled to take place sometime during the night; ideally, your expire should be set so that it concludes before any batch downloads start fetching new material. If you are using InterTalk, you should be aware that the system keeps note of the number of the final message downloaded on the previous newsfetch, so only messages more recent than that message will be downloaded.

Of course, the best way to ensure that problems regarding News space do not arise is only to subscribe to low-volume groups.

Internet Resources: Getting Started

Once you have an Internet connection, it can be difficult to know where to begin looking for useful information and resource material; to this end, the "StartHere" directory on the enclosed disc contains several pages, divided by subject, which can be loaded from local disc into a World Wide Web browser and used as a starying point for examining subject-specific resource. The Search Engines page provides links to automatic systems which hold recently-compiled databases of Web-based information, such that the Web can, in effect, be searched by content. The ftp equivalent of these Web search engines is Archie, which can be accessed via a Web front-end as ArchiePlex (http://sunsite.doc.ic.ac.uk/archieplex/) and can then be asked via an easy-to-use form interface to search its library of ftp sites for a file with a given name.

For most freely-available software and documentation on most common hardware platforms, the most useful site to look at is probably sunsite.doc.ic.ac.uk; this is an extremely fast, high bandwidth Sun Microsystems server with over 60 GB of data available online for public consumption. For Acorn-specific information, try http://www.acorn.co.uk/ or ftp.acorn.co.uk (or whichever is your nearest available mirror site) initially.

Regarding USENET News, it is very strongly recommended that anyone who is not already familiar with USENET and its protocols (those relating to the service layer or to the conduct of human discourse) should subscribe to, or be subscribed to, the **news.announce.newusers** newsgroup.

news.answers exists primarily to store the Frequently Asked Questions (FAQ) lists for all the newsgroups which have them; if you wish to know what the intended content of a newsgroup is, and whether the group is moderated, this is the place to look. FAQs tend to be posted on a bi-weekly or monthly basis, so if you wish to keep track of the charters of all the groups on USENET, you will need to keep a longer expiry time on this group (however this can probably be reduced after the first few weeks).

The **control** hierarchy will be automatically dealt with by your news fetcher; basically this hierarchy deals with message deletions, and the deletion and creation of newsgroups. ON NO ACCOUNT should anyone post to this group unless they fully understand the protocols involved.

Security Considerations

As with local area networking, security measures should both attempt to prevent unauthorised access and/or tampering with the system, and allow the administrator to backtrack and account for any misuse which actually occurs. The first measure involves physical security in the area of the modem and link to the outside world (keeping the proxy machine and line in a lockable office, for example); subsequent measures are best implemented by ensuring that all users have individual accounts and passwords, and that the passwords are reasonably secure to minimise the likelihood of one user being able to masquerade as another; using the UNIX metric, a password of 8 alphanumeric characters selected at random, and including at least one punctuation character, can be considered "reasonably secure".

InterTalk features a log scripting system, so that all transactions may be recorded along with their destination, the account details of their originator and their size. Probably the most obvious sign that something untoward may be going on will be large volumes of traffic for a particular user; the administrator would then be wise to investigate the address being accessed.

In the first instance, access to a World Wide Web browser should only be given to trusted individuals, as it is possible to access any part of the IP stack by embedding the appropriate port number in the URL used to access a resource; this is done by specifying the URL as <*protocol*>://<*site*>:<*port*>/

Administrators should watch out for URLs where the port number, if specified, does not equal 80 or 8080, (80 is the recognised standard port for http, 8080 is common at sites where firewalls forbid access to ports <1024) especially if it is low-numbered (<1024) and the protocol specified is not ftp, gopher or http; specifically, mail and telnet URLs with low port numbers should be treated as suspicious.

How to Find Out More...

A number of books have recently been published giving information on the Internet; a range of these are available through most good specialist computer bookshops. However, the de-facto information on the Internet is available via the Internet itself, in the form of RFCs.

An RFC, or "Request For Comments," is a document which was originally a proposal for adoption of a standard detailed in the document. However, following comment and review, many RFCs are now essentially standards documents in themselves, such that packages which implement protocols covered in RFCs are obliged to conform to the RFC's interpretation of how the protocol should behave.

In the UK, RFCs are carried by several Internet sites; two of the most popular sites to ftp RFCs from are:

sunsite.doc.ic.ac.uk	(SunSite Western Europe, located at Imperial College, London; this carries just about all the RFCs which have been written)
ota.ox.ac.uk	(Oxford University Text Archive; this carries all but the most recently released RFCs.)

Both these sites support "anonymous" ftp access, ie the UserID you should use when prompted is anonymous, and the password you should use is your own full email address.

FAQs (lists of Frequently Asked Questions, with answers) to USENET Newsgroups are regularly posted to news.answers; for an answer to a popular question appropriate to a given newsgroup, find the FAQ and see if it has been covered there.

For information on more sophisticsted Internet tools (Archie, Gopher, WAIS, Trickle, Veronica etc) a very useful document to have is the EARN Association "Guide to Network Resource Tools;" the plaintext version of this document is on the enclosed disc, and the fully formatted, paginated PostScript version can be obtained by sending email to LISTSERV@EARNCC.BITNET; the content of the email (which need not have a Subject: field) should be

GET NETTOOLS PS

Appendix A: Contents of the Accompanying Disc

The plaintext files in the Documents directory contain supporting information obtainable via the Internet and referenced in this Application Note. These files are therefore of particular relevance to readers who do not currently have Internet access. The files comprise:

AnetFAQ: The Frequently Asked Questions list, with answers, on the subject of the technical issues surrounding connecting Acorn machines to the Internet. Copyright 1994 Kevin Quinn, freely distributable.

Domains: List of most of the top-level (ie national and organisation-type) domain names.

Email: Details, for sites which only have email gatewaying to the Internet, on how to use all of the Internet's resource tools by email. Copyright "Doctor Bob" Rankin, freely distributable.

InterNIC: The form required to register a domain with InterNIC, with completion instructions.

ProvList: The list of UK Internet Service Providers as of November 1995, and the services they offer to their customers. Copyright 1995 Paola Kathuria, freely distributable via Internet or email, reproduced here with permission.

rfc1359: The current ACM SIGUCCS Networking Taskforce document on Connecting to the Internet. This is designed with colleges / Universities in mind, but has some pertinent comment which would be useful for schools too. Freely distributable.

Tools: Plaintext copy of the EARN Association's "Guide to Network Resource Tools." A comprehensive manual on how to make use of everything from Archie to WWW, including worked examples. Copyright EARN Association, freely distributable.