

# **Blocking Content on the Internet: a Technical Perspective**

prepared for the  
National Office for the Information Economy

**Philip McCrea  
Bob Smart  
Mark Andrews**



**CSIRO**

**Mathematical and Information Sciences**

**June 1998**

[blank]

# Contents

<b>EXECUTIVE OVERVIEW .....</b>	<b>5</b>
<b>1 INTRODUCTION.....</b>	<b>9</b>
<b>2 HOW DOES THE INTERNET WORK? .....</b>	<b>11</b>
2.1 THE INTERNET AND THE TELEPHONE NETWORK: A COMPARISON.....	11
2.2 THE INTERNET IS A 'PACKET SWITCHED' NETWORK .....	11
2.3 GETTING PACKETS TO THEIR DESTINATION.....	13
2.4 CONNECTING TO THE INTERNET .....	14
2.5 THE INTERNET IS A HIERARCHY OF MANY NETWORKS.....	14
2.6 SERVICES AVAILABLE TO INTERNET USERS .....	15
<b>3 CONTENT BLOCKING: AN OVERVIEW.....</b>	<b>20</b>
3.1 BROADCASTING.....	20
3.2 PUBLISHING.....	20
3.3 FILMS, VIDEOS AND GAMES .....	21
3.4 THE POSTAL SERVICE .....	21
3.5 ADVERTISING BROCHURES .....	21
3.6 TELEPHONY.....	21
3.7 INTERNET CONTENT BLOCKING.....	22
<b>4 INTERNET BLOCKING AT THE APPLICATION LEVEL .....</b>	<b>25</b>
4.1 WHAT IS APPLICATION LEVEL BLOCKING?.....	25
4.2 NOT EVERYONE ACCESSES THE INTERNET THROUGH AN ISP.....	27
4.3 THE USE OF DIFFERENT PORT NUMBERS CAN BYPASS THE FILTER.....	27
4.4 SITES CAN EASILY BE RE-NAMED TO BYPASS BLOCKING .....	28
4.5 TRANSLATION SERVICES RETURN DIFFERENT MATERIAL TO A USER .....	28
4.6 THE DOMAIN NAME SERVER CAN BE BYPASSED.....	29
4.7 CREATING AND DISTRIBUTING THE BLACK LIST MAY BE PROBLEMATIC.....	29
4.8 THE BLACK LIST IS A VALUABLE COMMODITY IN ITS OWN RIGHT .....	30
4.9 PUSH TECHNOLOGIES BYPASS PROXY FILTERS .....	30
4.10 THERE ARE COSTS INVOLVED WITH EMPLOYING PROXY SERVERS .....	30
4.11 A PROXY SERVER MAY INTRODUCE UNRELIABILITY .....	31
4.12 A PROXY SERVER MAY ADVERSELY AFFECT SOME APPLICATIONS .....	31
4.13 PERFORMANCE IS NOT A MAJOR ISSUE .....	31
<b>5 INTERNET BLOCKING AT THE PACKET-LEVEL.....</b>	<b>32</b>
5.1 ROUTERS: THE BACKBONE OF THE INTERNET.....	32
5.2 A ROUTER CAN BE USED TO IMPLEMENT PACKET BLOCKING.....	32
5.3 WHERE SHOULD PACKET BLOCKING TAKE PLACE?.....	33
5.4 NOT ALL INTERNATIONAL INTERNET TRAFFIC PASSES THROUGH A BSP.....	35
5.5 PACKET-LEVEL BLOCKING IS INDISCRIMINATE .....	36
5.6 PACKET-LEVEL BLOCKING WILL INHIBIT E-COMMERCE INFRASTRUCTURE.....	36
5.7 PACKET-LEVEL BLOCKING MAY AFFECT OTHER SERVICES .....	37
5.8 ROUTERS CAN EASILY BE CIRCUMVENTED .....	37
5.9 SITES CAN EASILY BE RE-NUMBERED TO BYPASS BLOCKING .....	37
5.10 PACKET BLOCKING CAN COMPLICATE FIREWALLS .....	37
5.11 PACKET-LEVEL BLOCKING MUST BE IMPLEMENTED IN HARDWARE .....	38
5.12 IMPLEMENTING PACKET-LEVEL BLOCKING IS COSTLY.....	38
<b>6 CONCLUSION .....</b>	<b>39</b>
<b>APPENDICES .....</b>	<b>41</b>



## ***Acknowledgments***

The authors would like to acknowledge the following people for their invaluable contributions to this report:

- ◆ Luke Carruthers, Secretary, Internet Industry Association
- ◆ Peter Coroneos, CEO, Internet Industry Association, IIA
- ◆ Peter Elford, Cisco Systems
- ◆ Geoff Huston, President, Internet Society in Australia, ISOC-AU
- ◆ Peter Higgs, Access CMC
- ◆ Tom Kennedy, National President, AIMIA
- ◆ Ramin Marzbani, WWW.Consult
- ◆ Rowan Macdonald, Australian Information Industry Association, AIIA
- ◆ Bill Simpson-Young, Carrie Bengston, CSIRO Mathematical and Information Sciences
- ◆ Steven Strange, IT Director, Traveland
- ◆ Michael Ward, Vice President Corporate Relations, OzEmail Ltd

## Executive Overview

This report provides some insights into the technical aspects of blocking material which is delivered over the Internet. It focuses on what can be done to block Content which has *already been identified* by some party as being illegal or offensive. The report does not address the desirability or otherwise of Content blocking on the Internet.

Content blocking on the Internet can take place at two distinct levels:

- ◆ **at the application level** – e.g. blocking a particular web page or ftp site by specifying the URL of the site (or a particular page or file within a site), or by blocking an entire news group.
- ◆ **at the packet level** – this requires routers to examine the IP address of the sender of a packet, and compare it with a supplied ‘black list’.

### ***Application level blocking***

With this type of blocking, ISPs prevent their clients from accessing the Internet directly for some services (such as the World Wide Web and File Transfer) by forcing them to access the Internet through a *proxy server*, which performs blocking and may store (*cache*) frequently accessed material. This requires a client to configure his/her web browser to ‘point to’ the ISP’s proxy server. The proxy server can then compare clients’ requests with a supplied ‘black list’ of web sites, ftp sites, or newsgroups.

Application level Content blocking is carried out in some countries such as Singapore and China.

The effectiveness of application level blocking using proxy servers is limited as a result of the following factors:

#### **Technical issues**

1. **Non-standard port numbers can be used:** Many hypertext links in web pages do not use the standard port numbers that are checked by the ISP’s filter. These requests can bypass the filter and therefore access sites which may be on the black list.
2. **Sites can easily be renamed.** Devotees of a renamed site can discover its new name by using a Search Engine, or from an Internet chat group, news group, or recorded telephone service.
3. **Translation services can return different Content to a user:** A request to an approved Content site operates a translation service, which generates a request to a second site. The information returned to the user is from this second site, but it appears to be the Contents of the site that was originally requested.
4. **The Domain Name server can be bypassed:** A user can access a site by specifying its descriptive domain name, or its equivalent IP address. A black list that checks domain names only can therefore be bypassed unless it also includes the equivalent URLs with IP addresses, which will double the size of the black list.
5. **A proxy server may introduce unreliability:** The policy of forcing users to access the Internet through a single proxy server reduces the reliability of Internet access, as it introduces a single point of failure.
6. **A proxy server may adversely affect some applications:** Some existing applications have problems working through a proxy server.

7. **Push technologies bypass proxy filters:** Push technology delivers Content to a user without being specifically requested by that user. Because a proxy server checks *requests* for Content, Content that is delivered to a site without a specific request will not be blocked by a proxy filter.

#### Non technical issues

8. **Not everyone accesses the Internet through an ISP:**
  - ◆ many organisations operate their own servers, and access the Internet through an Internet Access Provider (IAP), which provides packet routing only to and from the Internet.
  - ◆ most educational institutions operate their own servers and tend to regard blocking as being inconsistent with academic freedom;
  - ◆ many national and multinational organisations provide Internet access to their employees and in so doing bypass local ISPs.
9. **ISPs face additional costs:** Many ISPs and Content hosting organisations do not employ proxy servers at present, and a requirement to do so may be a difficult burden financially for small ISPs in particular. In addition to the hardware cost, there is the ongoing cost of maintaining and administering the proxy server, and supporting their clients that are forced to use it.
10. **ISPs may be placed in a dilemma:** If ISPs or Content providers are asked to adopt the role of a moral arbiter, they will be placed in a difficult position by their clients for either going too far or not going far enough, and may be in an awkward position legally if they block (remove) Content which they host under contract to a client.
11. **Creating and distributing black lists may be problematic:** There are some 600 ISPs and IAPs in Australia, and a much larger number of connected organisations that host Content. The task of updating and distributing the black list in a secure manner to all these organisations would not be trivial.
12. **A black list will be a valuable commodity in its own right:** Black lists should be maintained in a secure environment. A black list will be a prime target for hackers, and having been uncovered will be published on the Internet, thereby creating a 'must see' list for curious Net surfers. This may have the effect of publicising the sites on black lists more widely than if the black lists did not exist.

It should be noted that blocking Content and deleting Content are quite different. Blocking prevents a web page from being accessed, while deletion refers to the removal of a web page after it has been published. The deletion of a web page, or a hierarchy of web pages, can be carried out only by the owner of the web pages, or by the administrator of the site where the material is hosted. After a web page has been deleted, however, it may still exist on:

- ◆ personal computers – where it has already been saved to disc;
- ◆ on proxy servers – till expired or flushed by the ISP; or
- ◆ on mirror sites – till these are updated from the source site.

#### **Packet Level Blocking**

Most Internet Content resides on servers outside Australia. If implemented, packet level blocking should be carried out by the relatively small number of Backbone Service Providers (BSPs) who provide international Internet gateways. Packet Level blocking involves a comparison of an Internet packet's *source* address with a supplied black list of IP addresses. This is implemented by routers operated by the BSP, using a router's Access Control List feature.

The effectiveness of packet level blocking is limited by the following:

### Technical Issues

1. **Packet level blocking is indiscriminate:** The decision to block a site by specifying its IP address means that that entire site would be invisible to Internet users. If the site happened to be a large website hosting organisation in the US – say a large ISP – then the rest of the sites hosted by that ISP could be inaccessible to Australian users.  
  
If the US ISP hosts a large number of commercial websites, then those organisations would be unable to carry out business with Australian companies, thereby disrupting the emerging electronic commerce infrastructure. This would prevent Australian companies from selling to customers associated with the blocked US ISP.
2. **Packet level blocking will inhibit e-commerce infrastructure development:** Any attempt to block packets as they enter Australia may affect traffic that is destined for another country. If the blocked material happens to be an electronic commerce transaction between a client of an ISP that is on Australia's black list and an organisation in another country, then Australia will become regarded as an unreliable supplier of electronic commerce infrastructure.
3. **Packet level blocking may affect other services:** A decision to block a particular site because of an illegal web page means that all other services, such as e-mail, will also be blocked to that site. It may be possible to include the port number (which specifies a particular service) in the filter, but this will have an effect on performance of the router.
4. **Routers can easily be circumvented:** Schemes such as *tunnelling*, where an IP packet is contained inside another IP packet, are commonly used, particularly in the implementation of virtual private networks for distributed organisations. The inside packet is extracted by the receiver to recreate the original message – which may be on a black list. Tunnelling requires both sender and receiver to use cooperating software, but such software may be easily downloaded over the Internet to people who wish to bypass normal packet routing.
5. **Sites can easily be re-numbered to bypass blocking:** In an analogous fashion to application-level blocking, owners of sites that are determined to bypass any blocking filters can easily – and regularly – change their IP number, thereby bypassing the black list.
6. **Routers may need to be upgraded to implement packet level blocking:** A top-of-the-line router from Cisco systems, appropriately configured, can carry out packet level blocking *at normal line speeds*. Older style routers may need to be replaced if they cannot be upgraded to carry out hardware packet blocking. It should be pointed out that the cost of any hardware upgrades for a BSP, however, would be insignificant in comparison with the annual cost of links across the Pacific.
7. **Packet black list blocking can complicate firewalls:** Packet level blocking and firewalls are similar in many respects, but differ significantly in their operation and the amount of computing support required. It may be difficult to integrate and implement a firewall and a packet blocking filter on a single router.

### Non technical Issues

8. **Not all Australian Internet traffic passes through a BSP:** Many multinational organisations have extensive Internet based networks, which may involve the use of dedicated leased lines from Australia to, say, the US. The Australian employees of these organisations would not be subject to BSP packet level blocking.
9. **There are increased operational costs:** There would be a significant cost associated with the creation, maintenance, and distribution of black lists. BSPs would experience the additional operational cost of configuring routers.

### **Summary**

Packet level blocking is too indiscriminate, and its use would create unintended 'holes' all over the emerging global digital infrastructure, which could isolate Australia to a large degree in the emerging digital global infrastructure. It is inconsistent with Australia's desire to become an electronic commerce hub for South East Asia.

Application level blocking is technically possible, but it can easily be circumvented by users in more ways than can packet level blocking. Mandating its use may result in black lists becoming 'hot property', with the result that the black-listed sites may actually become more popular than if they were not black listed at all.

Our conclusion is that Content blocking implemented purely by technological means will be ineffective, and neither of the above approaches should be mandated. Any technology-based solution can be worked around – purely as a result of the sheer pace of technology change on the Internet.

Having said that, we suggest that the following approach be considered to Content blocking, particularly where minors are concerned.

### **ISPs could offer differentiated services**

Where there is market demand, ISPs should be encouraged to offer differentiated services to clients, based on access to the Internet through a proxy server. Two classes of such service should be considered:

- ◆ **A 'clean' service:** the proxy filter includes a list of *permitted URLs only*; all requests outside this list are refused. The 'real' Internet is not actually accessed, and a user cannot escape from the particular prescribed 'universe' that s/he finds him/herself in. Several such proxy-based blocking schemes are currently available, providing access to a universe of thousands of permitted pages<sup>1</sup>.
- ◆ **A 'best effort' service:** the proxy filter blocks a set of known sites, rated according to some prescribed criteria. The result is based on a best-effort approach by an ISP, and cannot be guaranteed. Bess filtering software<sup>2</sup>, for instance, claims to have a black list of 'hundreds of thousands of pages'.

ISPs may incur some costs in setting up services such as these. These could either be passed on to clients in increased fees, or an ISP may see some competitive advantage in providing such environment to clients. Alternatively the Government may consider providing some incentives to ISPs to offer such differentiated services.

---

<sup>1</sup> See <http://www.research.att.com/~lorrie/pubs/tech4kids/>.

<sup>2</sup> See <http://www.n2h2.com>.

# 1 Introduction

This report presents a brief overview of the technical workings of the Internet, with an orientation towards the Content which is delivered over the Internet, and in particular the technical feasibility of blocking such Content. **Issues relating to the desirability of, or policies concerning, Content blocking on the Internet are not addressed in this report.**

The report does not look at technologies that determine the nature of Content that is delivered to a User; instead it is concerned with the blocking of material that has *already been identified* by some party as being either illegal or offensive.

The term *Content* is a generic one that has come to represent any form of information – text, sound, pictures, videos, spreadsheets, and so forth – that is made available to the general public through some delivery channel, including the print media, radio, television, CD-ROMs, or Internet delivery. It is now recognised that all published Content is (or soon will be) produced and delivered by digital means.

The regulation of Internet Content has been the subject of some recent debate. The 1996 Australian Broadcasting Authority report, *Investigation into the Content of On-line services*<sup>3</sup>, identified the main issues. The ABA also contributed to the 1997 UNESCO report, *The Internet and some international regulatory issues relating to content*<sup>4</sup>.

This report differs from the above in that it concentrates on the technical aspects of the Internet that can be used to block Internet delivered Content, particularly by those organisations which collectively comprise the Internet – Internet Access Providers (IAPs) and Internet Service Providers (ISPs).

## ***Illegal and Offensive Content***

By way of background, it is important to highlight the difference between Content that is illegal and Content that may be regarded as offensive.

Illegal Content is material that should not be transmitted/broadcast within a certain jurisdiction, because it contravenes legislation that covers that particular jurisdiction. Material relating to paedophilia clearly falls into this category – regardless of how it is made available.

Offensive Content is material which, although legal to distribute to adults through some media, may be illegal to distribute through particular media such as broadcasting services, or be illegal to distribute to minors.

It should be noted that there is a significant difference in attitude to offensive Content between Internet users and non-users. This is illustrated in Appendix 2, which shows the results of several surveys by market research company, WWW.Consult. Non Internet users are much more concerned about offensive material than are Internet users, although the chief concern of both groups is the cost of Internet access.

## ***ISPs, IAPs, and BSPs***

In this report, we differentiate between an Internet Access Providers (IAP), an Internet Service Providers (ISP) and Backbone Service Providers (BSP).

---

<sup>3</sup> Available from <http://www.dca.gov.au/aba/olcont.htm> .

<sup>4</sup> Available from <http://www.dca.gov.au/aba/unintr.htm> .

An IAP provides an organisation with access to the Internet only. From a technical perspective, an IAP passes Internet packets to and from its client, but has no interest or understanding of the Internet based services used by the client.

An ISP provides a client with Internet connectivity like an IAP, but offers a range of additional services to clients, such as electronic mail (e-mail), world wide web services, file transfer services, network news, telephony and so forth. In the case of the Web, an ISP could provide hosting services for clients. By contrast, a client of an IAP would need to establish a series of servers to operate services such as e-mail, web, and so forth.

A BSP is an IAP that 'wholesales' IP connectivity to other IAPs and ISPs. BSPs operate national networks, and in many cases provide international connectivity.

### ***How this report is organised***

- ◆ Chapter 2 presents a brief overview of the Internet, explaining those technical concepts which are required for the subsequent discussion on Content blocking.
- ◆ Chapter 3 presents an overview of Content blocking in traditional media, and with this as a background introduces the concept of blocking Internet-delivered Content.
- ◆ Chapter 4 considers the first of two methods that can be used to block Internet Content – application level blocking by an ISP using a proxy server.
- ◆ Chapter 5 discusses the second approach to blocking, where routers are used to block Internet packets.
- ◆ Chapter 6 provides a conclusion, and suggests a practical approach that could be taken in relation to Content blocking.

## 2 How does the Internet work?

This chapter provides a brief overview of the workings of the Internet. It is not intended to be exhaustive, and provides appropriate background for the discussion in subsequent chapters on Internet Content blocking.

The term *Internet* tends to be used in a generic sense to denote two quite different concepts:

- ◆ a delivery mechanism;
- ◆ a range of services that are delivered over the Internet.

Strictly speaking the term Internet should be used to denote the delivery mechanism only. This chapter looks firstly at the Internet as a delivery mechanism, and then considers some of the services that can be delivered over the Internet.

### 2.1 The Internet and the telephone network: a comparison

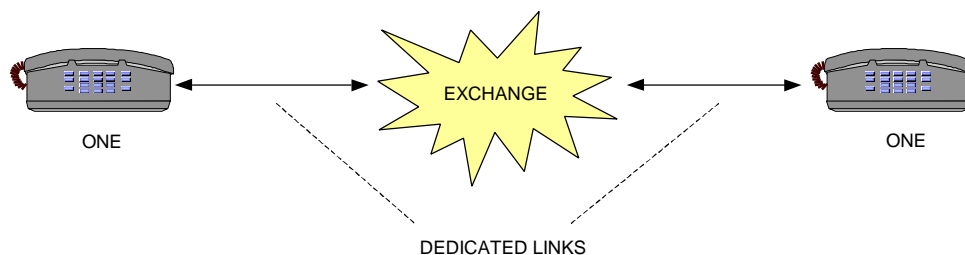
The Internet and the telephone network share many similarities. The telephone network provides a framework for two parties to communicate by voice using specially designed telephone handsets. Similarly the Internet provides a mechanism for transferring digital information from one computer to another.

In the telephone network, the service provider (Telstra, Optus, etc) has no interest in the conversations (i.e the Content) of their networks. In the same way the Internet Access Providers (IAPs), who collectively comprise the Internet, have no involvement with the messages that traverse through their network from one computer to another.

The Internet and the telephone network do differ in one respect, however. In the telephone network, it is technically quite straightforward for a service provider to intercept a conversation and record it. Because of the way the Internet works and is used, equivalent interception methods are difficult to use, and may be quite ineffective, particularly if the messages are encrypted, or if they use an unknown or proprietary protocol.

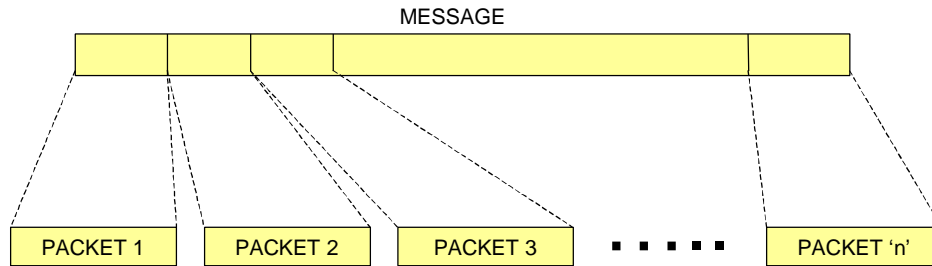
### 2.2 The Internet is a 'packet switched' network

It is important to note one major difference between the Internet and the telephone network. With the telephone network, when a person places a call, the telephone company arranges for the necessary connections to be made at telephone exchanges along the way, so that a physical link is established between the two parties participating in the conversation. This link is maintained for the duration of the call, even if no conversation takes place. As consumers, we are billed for the connect time of a call (assuming long distance), even if we have not spoken a word. This is known as *circuit switched* network and is shown in Figure 1.



**Figure 1:** The telephone systems communication over a dedicated link.

The Internet, on the other hand, is a *packet switched* network: the message being transferred from sender to receiver is broken into small chunks, or *packets*, as indicated in Figure 2. These packets are transmitted independently between sender and receiver. To use an analogy in the transport industry, this is similar to splitting a large consignment of goods – say, wool – into smaller amounts that will fit on a truck or into a container for shipping purposes.

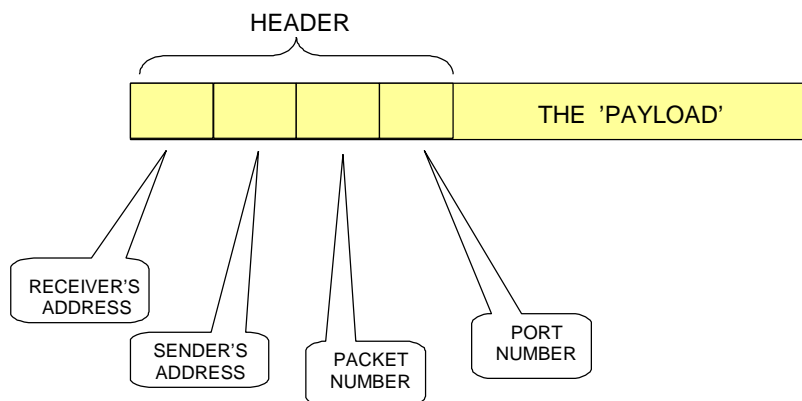


**Figure 2:** An Internet message is split into packets which are transmitted independently.

Each packet is divided into a *header* part and a *payload* part, as shown in Figure 3. The header is used by the Internet *routers* to get the packet to its destination, while the payload contains the actual Content.

The header contains<sup>5</sup>:

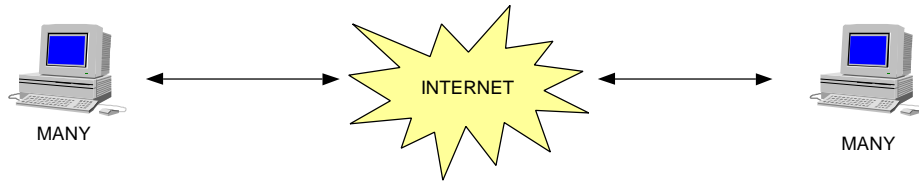
- ◆ the address of the packet's destination (i.e. the receiver's address),
- ◆ the address of the sender,
- ◆ packet numbering information, so the original message can be re-created from its component packets, and
- ◆ other information to identify the application software at the receiver's end that should be used to interpret the packet (the port number).



**Figure 3:** An Internet Packet, showing the Header and Payload.

The advantage of packet switching over circuit switching is that a dedicated line need not be maintained between sender and receiver: a single line can be used by many packets from different senders, and hence many users can be serviced simultaneously over a physical line. The Internet's efficient use of communications infrastructure in this way makes communication inexpensive. This is illustrated in Figure 4.

<sup>5</sup> For the purposes of this discussion, information in the IP and TCP headers is combined.



**Figure 4:** The Internet provides *many-to-many* communication.

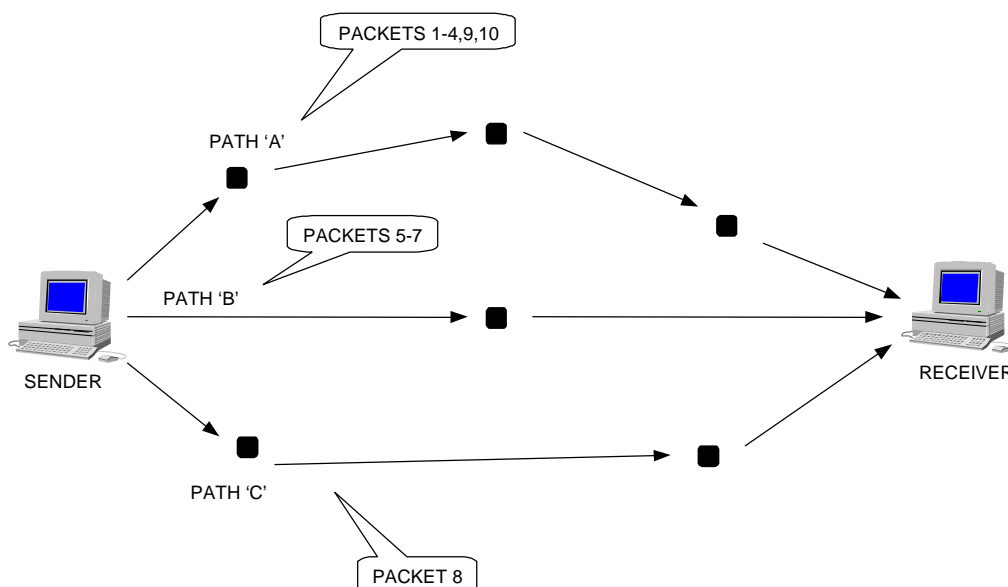
### 2.3 Getting packets to their destination

Since the Internet is shared by everyone, the functionality provided is as simple as possible: the Internet makes the best effort it can to move packets of information from one computer to another. ‘Best effort’ implies that the Internet may be unreliable in that it may:

- ◆ throw packets away – if there is congestion along the route, for instance;
- ◆ deliver packets out of order; or
- ◆ duplicate data during transmission.

As an analogy, in the transportation of physical goods, there is invariably more than one route that a truck could take over the road network, and trucks may arrive out of order at the destination if they take different routes. Similarly Internet packets may take different routes and may arrive out of order at the destination. In addition, Internet packets may fail to arrive at all at the destination (equivalent to the road authority destroying trucks in transit to reduce congestion), or two or more copies of the same packet may arrive.

The Internet software provides a way to detect and recover from these errors. This is the responsibility of the TCP component of the Internet communications protocol, TCP/IP<sup>6</sup>. TCP detects whether any packets that comprise a message have gone missing, and requests re-transmission from the sender. TCP guarantees that the packets that comprise a message arrive intact and re-assembles them in the correct order at the destination.



<sup>6</sup> TCP/IP stands for Transmission Control Protocol/Internet Protocol. Protocols other than TCP can be used with IP, such as UDP. With UDP there is no error correction, and if error correction is required, it must be carried out by the application software that handles the received message.

**Figure 5:** Internet software may send packets of a message along different paths.

Figure 5 illustrates a message that has been broken into 10 Internet packets for transmission between Sender and Receiver. Six packets travel via Path A, three by Path B, and one by Path C. The choice of path will depend on the degree of congestion on each path. The black boxes in each of the paths represent the *routers* that comprise the Internet. The routers read the destination address in the packet header and decide on the next router which will get the packet closer to its destination, taking into account network congestion and other factors.

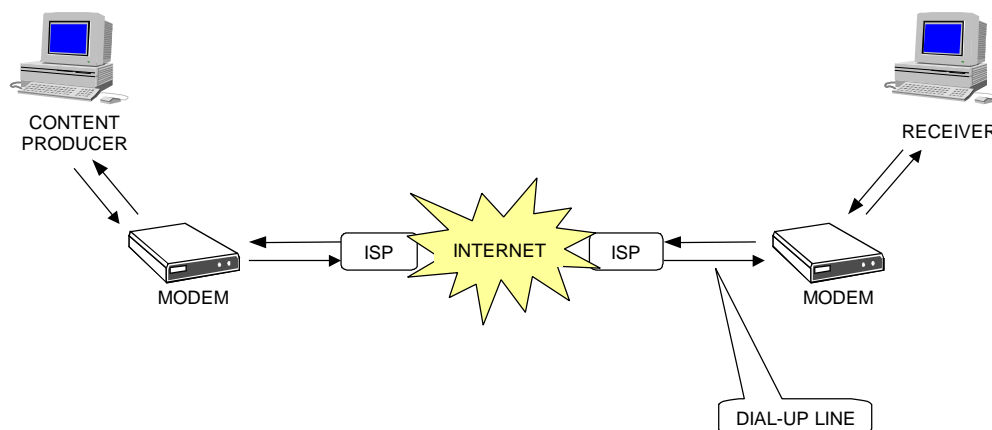
It is unlikely that network congestion would cause the packets in a short message to travel along different paths. A long message, however, such as a continuous radio broadcast<sup>7</sup>, may use a variety of transmission paths over a period of time.

For some applications quality of service is not important. For instance in transmitting a video a small amount of lost data will not be noticed, whereas a loss of some information may be a serious problem if the message being transmitted is a spreadsheet.

## 2.4 Connecting to the Internet

An Internet user requires a PC, a modem and a contract with an Internet Service Provider (ISP) to be able to connect to the Internet, as illustrated in Figure 6. There are over 600 ISPs in Australia, ranging from very large which provide round the clock support (such as OzEmail and Telstra), to very small one person operations. It should be noted that many Australians have access to the Internet through work or an educational institution, and do not use an ISP to access the Internet.

Internet users usually connect to their ISP using a modem and a standard telephone line. Cable modems may also be used. A large Content Producer may have a permanent line to the ISP, which may be a ISDN connection – 64k or higher.



**Figure 6:** The Role of ISPs is connecting users to the Internet.

## 2.5 The Internet is a hierarchy of many networks

No-one 'owns' the Internet<sup>8</sup>. It is not a single network, but rather a large collection of inter-connected networks that use a common set of protocols. It is organised in a hierarchical

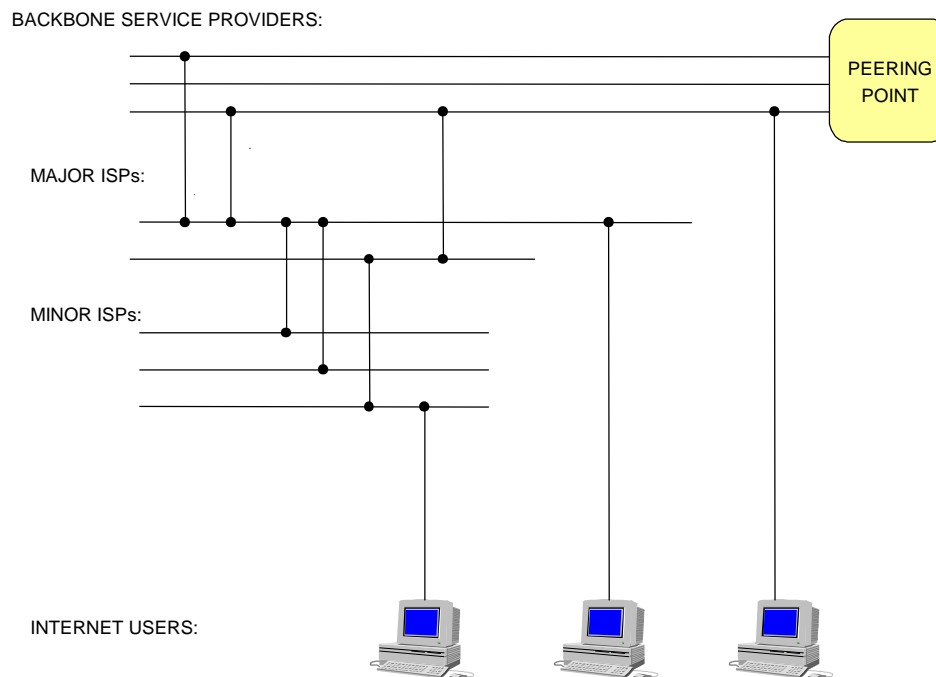
<sup>7</sup> Triple J broadcasts are available over the Internet from <http://www.abc.net.au/triplej/listen.htm>.

fashion, as illustrated in Figure 7, which shows three different levels in a typical Internet hierarchy.

Backbone Service Providers (BSPs) are typically large telecommunications companies who have their own physical infrastructure over which they provide an *IP Service*, which they wholesale to ISPs. Telstra and Optus are BSPs, but in addition to providing wholesale IP connectivity to ISPs, they provide an end user service – *Big Pond* in the case of Telstra, and *Spinnaker* in the case of Optus.

ISPs buy capacity from BSPs, and make it available to Internet users, generally adding some form of value adding along the way. For instance OzEmail has a range of service available to clients only. The larger ISPs have points of presence (POPs) throughout the country, which means that most of their subscribers can access the Internet without having to pay STD telephone costs.

Some of the larger ISPs re-sell IP connectivity to smaller ISPs, and in this capacity they function as a BSP. As a result BSPs and the larger ISPs compete for the custom of smaller ISPs. Many of the smaller ISPs are regionally based.



**Figure 7:** The Internet is a hierarchy of IP networks.

## 2.6 Services available to Internet users

The Internet provides a range of services to users, including some recently introduced services such as telephony. In this report we consider the following services only:

- World Wide Web (WWW)
- Network news – or simply just ‘news’

<sup>8</sup> Much to the annoyance of some of the larger computer companies...

- File transfer
- Electronic mail.

### 2.6.1 The World Wide Web (WWW)

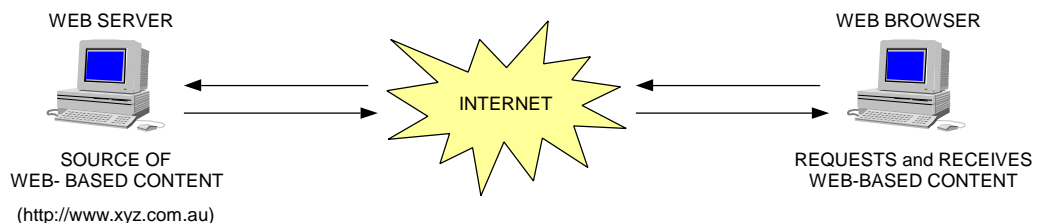
World Wide Web Content is normally referred to as ‘web pages’. An Internet user uses a Web *browser* to access a web page by typing the address details of the computer where the web page is stored. A web address is referred to as a Universal Resource Locator (*URL*), and is typically of the form <http://www.xyz.com.au><sup>9</sup>.

The [www.xyz.com.au](http://www.xyz.com.au) part of the URL is referred to as the *domain name*. A Domain name with the suffix ‘.au’ does not necessarily have to be associated with a computer that is physically located in Australia. The computer could reside anywhere in the world, although it is highly likely that it would be in Australia.

The World Wide Web is implemented in a *client-server* fashion, as illustrated in Figure 8. The software comprises two distinct components:

- ◆ **the client**, known as a *web browser* in the case of the World Wide Web, which runs on a user’s PC; and
- ◆ **the server**, which is generally operated by an ISP, or a Content hosting organisation.

There is generally a *single source* for a web page – i.e. there is a physical server somewhere in the world with a specific address which hosts the web page<sup>10</sup>. However an Internet user may elect to make a local copy of a web page on his/her computer by using the *Save As* feature of a web browser. This saved information can then be accessed subsequently from the user’s hard disc without requiring access to the Internet.



**Figure 8:** A web page is accessed from its source site.

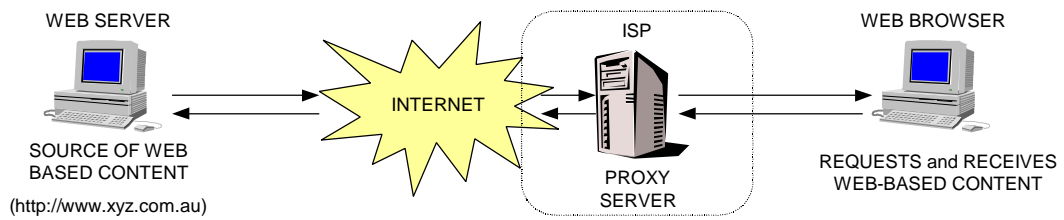
Copies of Web pages (and other Internet Content) can also be stored at sites other than at the source site in proxy servers and mirror servers.

#### Proxy Servers

Many ISPs may store commonly accessed web pages in a *proxy server* at the ISP’s premises. A proxy server is used by an ISP to reduce telecommunications costs, particularly for expensive overseas links, and also to increase download speeds. This is illustrated in Figure 9.

<sup>9</sup> Frequently a file within a particular web site is also specified. For instance <http://www.xyz.com.au/fileA.html/> accesses a specific file called ‘file A’. The suffix *html*, which stands for *hypertext markup language*, indicates that the file should be interpreted by a web browser. The term *http* – hypertext transfer protocol – is the protocol used for accessing a web page. Other protocols may be specified – *ftp* – file transfer protocol – is the protocol used to transfer a file from one computer to another.

<sup>10</sup> The material may be distributed across several geographically separated servers.



**Figure 9:** An ISP may use a proxy server to reduce telecommunications costs.

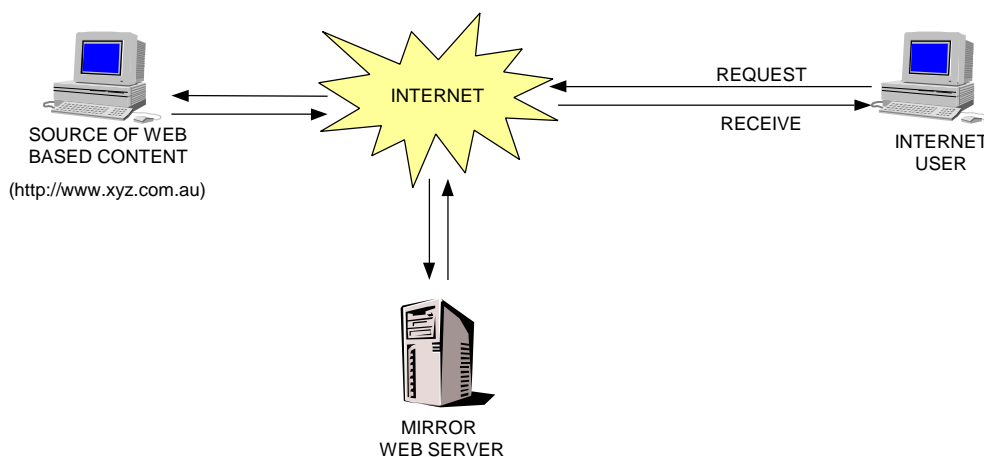
Because proxy servers are commonly used by Australian ISPs, there may be many stored (or *cached*) copies of web pages in proxy servers. The operational policy for proxy servers is determined by the ISP. For instance, a proxy webserver be configured to:

- ◆ cache certain pages, or page hierarchies, but not others;
- ◆ check with the source of the web page each time it is requested to see if the cached copy is still current;
- ◆ delete web pages after a certain period of time, or according to some algorithm such as *least regularly used* (LRU).

### Mirror Servers

Often entire websites, or portions thereof, are *mirrored* on another server in a different geographical location. This is generally done to improve access time, or reduce telecommunications requirements, or both. For instance Netscape have a mirrored site in Melbourne (<http://home.netscape.com/au/index.html>), as does the search engine, AltaVista (<http://www.altavista.yellowpages.com.au/>).

A mirrored site often has a different URL to the source site, so an Internet user can elect to access the source of the Content or the mirrored site. This is shown in Figure 10. A mirrored site is generally updated at regular intervals (eg daily) from the source site.



**Figure 10:** Entire web sites may be ‘mirrored’ in a different geographical location.

It should be noted that the blocking and deletion of a web page (or any other form of Internet Content) are quite different. Blocking prevents a web page from being accessed.

Deletion refers to the removal of a web page after it has been published. The deletion of a web page, or a hierarchy of web pages, can be carried out only by the owner of the web pages, or by the administrator of the site where the material is hosted. After a web page has been deleted, however, it may still exist on:

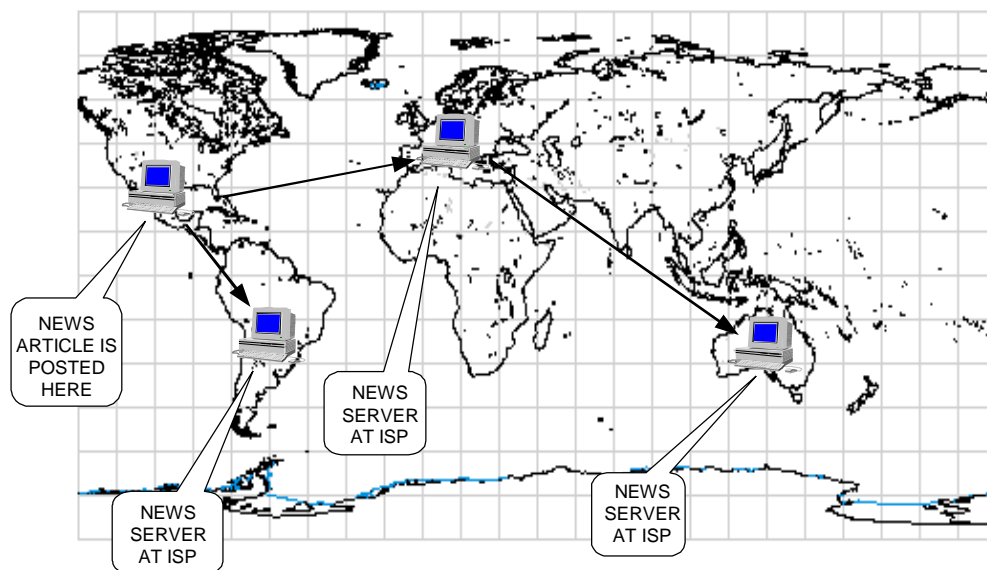
- ◆ personal computers – where it has already been saved to disc;
- ◆ on proxy servers – till expired or flushed by the ISP; or
- ◆ on mirror sites – till these are updated from the source site.

## 2.6.2 Network News

The architecture of the delivery of News is quite different to the web. There is a single source for news, but it is replicated all over the Internet in news servers, which communicate using the *nntp* protocol<sup>11</sup>.

A large number of *newsgroups* have been established for discussion of specific topics<sup>12</sup>. Any Internet user may subscribe to a news group, and having done so will receive all the postings that are made to that group, and may post articles to that group.

Most ISPs have a *news server*, and offer a news service to their clients. After someone posts an article to a newsgroup, that article is distributed the various news servers around the world, and may be read by anyone who has access to that news server, as shown in Figure 11.



**Figure 11:** News articles are distributed in news servers all over the Internet.

Accessing news requires a high bandwidth connection. News feeds currently produce about seven gigabytes of new information per day – well over a million news articles per

<sup>11</sup> Not all newsgroups are worldwide. Some circulate in a restricted area (e.g. the aus newsgroups such as aus.games). Some are not distributed at all but just use the news as a way of implementing discussion forums instead of using mail list servers: Netscape provides discussion forums for its customers in this way. Some worldwide news services are not public: the Clarinet service (<http://www.clarinet.com>) distributes a news hierarchy to its customers using the news (nntp) protocol.

<sup>12</sup> See <http://www.synapse.net/~radio/finding.htm> for useful information on available newsgroups and mailing lists.

day. The task of monitoring news groups for illegal or offensive material is clearly a huge task.

News differs to the Web and to ftp in that there are a large number of Content providers – i.e. there are a large number of Internet users that post to news groups. Individual news postings can be deleted by the person who posted the article, or by anyone masquerading as that person, which is easy to do. Indeed mass e-mail messages (spam mail) are regularly cancelled in this way, and there has been a case of message cancellation being used as a private mechanism for censoring contrary opinion<sup>13 14</sup>.

Some large ISPs receive full feeds from all newsgroups, but smaller ISPs may receive only a subset, or obtain their news from a larger ISP and store it in a caching proxy server.

News articles are often archived in many places, and made available through web sites.

### **2.6.3 File Transfer**

The Internet's file transfer facility enables an Internet user to retrieve a file from another computer using the file transfer protocol (*ftp*) protocol. A file may contain any form of information – a document, a picture, video, or a spreadsheet. It is posted by its owner, and by so doing, the owner is saying “this file is available to the public: anyone can retrieve it”.

Most ISPs operate an ftp server for the benefit of their clients, and generally provide an *anonymous ftp* service, where visitors to the site can retrieve a file from the ISP's public area without the need to identify themselves.

The ftp architecture is similar to the web – i.e. there is a single source of the material, which may be removed by the owner, or by the administrator of the ftp server that hosts the ftp file. A file is specified using a URL of the form <ftp://ftp.xyz.com.au/fileA.jpg>. The *jpg* suffix indicates that the file is a picture that is stored in a particular encoded format.

### **2.6.4 Electronic Mail**

Electronic mail, or *e-mail*, is different to other Internet services – web, news and ftp – in that mail is directed at a particular recipient, or group of recipients: it is not broadcast. As such electronic mail is analogous with physical mail.

E-mail can be distributed from a *list-server*, to which people subscribe. A list-server forms the basis of a closed community, in which registered members can exchange information. It is different to news, in that an individual copy of material is mailed to each registered member, and is not available to Internet users outside that group.

E-mail articles, particularly those generated in a list-server, are often archived in many places, and available through web sites.

The blocking of e-mail is not considered in this report.

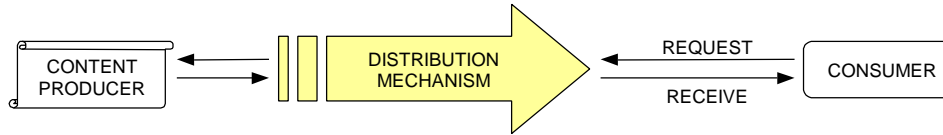
---

<sup>13</sup> The case involved discussions between Armenians and Turks about whether there was a massacre of Armenians in Turkey at the turn of the century.

<sup>14</sup> For moderated newsgroups there has been a move to the use of digital signatures so that only the moderators can approve or cancel news items. It seems likely that all major newsgroups will eventually be at least lightly moderated to keep out spam items.

### 3 Content Blocking: an overview

The basic model for discussing Content regulation and blocking is illustrated in Figure 12.



**Figure 12:** Generic model for consideration of Content blocking.

In this model, Content blocking can be carried out by:

- ◆ the Content Producer,
- ◆ the Distribution mechanism, or
- ◆ the Consumer.

Before considering the Internet as a distribution mechanism, it is instructive to examine where and how Content is regulated in some of the more established distribution systems.

#### 3.1 Broadcasting

Broadcasting refers to the fact that a producer of Content (or an agent of the producer) distributes the Content widely (the *broad* in broadcast) by some means. In broadcasting there are few Content producers and many receivers. Television and Radio fit into this category.

Content regulation is overseen by the Australian Broadcasting Authority (ABA), and administered under an industry code of practice. The ABA does not play an active role in regulation other than to ensure, when requested, that the industry code of practice is adhered to.

Three forms of Content regulation are currently practiced:

- ◆ Blocking of Content by broadcasters (ie Content Producers), according to the industry code of practice.
- ◆ Rating labels on products – eg TV shows use terms such as ‘Coarse language’ etc. This provides consumers with the choice of whether to watch (or listen) to a broadcast.
- ◆ Time of broadcasting, where material deemed not suitable for minors is broadcast after a certain time in the evening.

#### 3.2 Publishing

Publishing refers to the print media, where there are few Content producers (and even fewer owners of the Content producers). Newspapers, journals, magazines and books fit into this category. Content regulation is overseen by the Office of Film and Literature Classification (OFLC), who may classify the Content of printed material, including refusing to classify a publication which makes it illegal to sell, hire or exhibit, or classifying it as legally restricted to adults.

In addition a form of product labelling scheme is used where adult entertainment may be wrapped in opaque material, and so forth.

### **3.3 Films, Videos and Games**

Films and Videos contain the same Content, but have different distribution mechanisms. Content regulation is overseen by the Office of Film and Literature Classification (OFLC), which classifies all Content, including refusing to classify a film or video, thereby making it illegal to sell hire or exhibit, or classifying it as legally restricted to adults. In addition, a well publicised rating system is used for movies that are shown in cinema or available on video, so consumers can make the choice about whether or not to see the film in question.

OFLC classifies and produces a rating system for electronic games.

### **3.4 The Postal Service**

Content blocking is not employed by the postal service. A consumer, however, can take steps to prevent the receipt of unsolicited mail that is specifically addressed to the consumer. The Australian Direct Marketing Association (ADMA)<sup>15</sup> represents the majority of organisations that employ direct marketing involving addressed mail and faxes. Under their industry code of practice, a recipient of unsolicited addressed mail can request the organisation sending the mail – i.e the Distribution mechanism of Figure 12 – to refrain from doing so in the future. If this does not result in a satisfactory response, the matter can be referred to ADMA.

### **3.5 Advertising Brochures**

To the consumer, the ‘junk mail’ industry appears to be increasing in activity. The Australian Distribution Standards Board (ADSB)<sup>16</sup>, however, which is funded by the major distribution companies, has an established code of practice to govern distribution. This includes adherence to the policy of non-delivery of advertising material to a home which displays an ADSB supplied sticker, stating *No Advertising Material*. In this case the consumer exercises choice on whether to receive the material or not, and the blocking is implemented in the Distribution mechanism.

### **3.6 Telephony**

There are industry self regulatory arrangements to set and enforce standards and deal with complaints about the Content of telephone information services (005 and 190 prefixed services) through the Code of Practice relating to Live, Recorded, Data and Fax Services administered by the Telephone Information Services Standards Council (TISSC). TISSC is an independent incorporated body whose determinations (for example disconnection of an information service) are enforced by the telecommunications carriers.

The issue of Calling Number Display (CND) may be construed as a form of Content regulation, since the aggregated information from called parties may be particularly useful to an organisation for marketing purposes. CND is turned ON by default by Telstra, but may be removed at the request of a subscriber.

The foregoing discussion is summarised in Table 1.

---

<sup>15</sup> See [http://www.attgendep.nsw.gov.au/priv\\_fac.html](http://www.attgendep.nsw.gov.au/priv_fac.html)

<sup>16</sup> *ibid.*

Distribution mechanism	Regulator	How regulated?	Where regulated?
Broadcasting	ABA oversees industry code of practice	Blocking: Rating labels on products: Time of broadcast:	Content Producer Consumer Distribution Mechanism
Publishing	OFLC	Blocking (censorship): Rating labels on products:	Content Producer Consumer
Cinema and Videos	OFLC	Blocking (censorship): Rating labels on products:	Content Producer Consumer
Games	OFLC	Rating labels on products	Consumer
Postal Service	Unsolicited addressed mail can be stopped under ADMA code of practice	Blocking	Initiated by Consumer; implemented by Content Producer
Advertising Brochures	ADSB oversees industry code of practice	Blocking – ADSB supplied sticker on Post Box	Initiated by Consumer; implemented by Distribution Mechanism
Telephony	TISCC for 005 and 190 services; Telstra for CND	Blocked by carrier on TISCC recommendation: Blocking - CND facility removed by Telstra	Distribution Mechanism Initiated by Consumer; implemented by Distribution Mechanism

**Table 1:** Content regulation practices in current distribution schemes.

### 3.7 Internet Content Blocking

The Internet is a distribution mechanism over which a range of services can take place. Traditional services that can be delivered over the Internet are summarised in Table 2.

The fact that anyone can be a distributor of Content on the Internet poses a problem for Content regulators. With few-to-many distribution schemes such as broadcasting, the small number of Content producers enables authorities to monitor what is distributed. For instance, it is comparatively straightforward to monitor the programmes broadcast on the five free-to-air TV stations that operate in Australia, as well as the three cable TV companies.

Service	Internet mechanism
Publishing	Publishing on the World Wide Web ( <i>www</i> )
Broadcasting	Multi-casting <sup>17</sup>
Correspondence (written)	Electronic mail ( <i>e-mail</i> )
Correspondence (Voice)	Internet telephony
Document transmission	File Transfer protocol ( <i>ftp</i> )
Meetings	Internet videoconferencing
Discussions	Chat sessions

**Table 2:** Types of services that can be delivered over the Internet.

<sup>17</sup> Other technologies such as streaming audio and video can be used to implement broadcasting. For instance Triple J may be heard over the Internet at <http://www.abc.net.au/triplej/listen.htm> using streaming audio software.

Internet-delivered Content regulation can be implemented (potentially) in the same three areas shown in Figure 12. However, it is not as straightforward to regulate Content on the Internet for the reasons outlined below.

### **3.7.1 Blocking by the Content Producer**

The regulation of Content at the Content producer's end, in advance of that Content being published, is virtually impossible on the Internet, as a result of the fact that every Internet user has the potential to be a Content producer. It is quite impossible for a body such as the Australian Broadcasting Authority, or the Office of Film and Literature Classification to be able to regulate Internet Content at the source for several reasons:

- ◆ There are just too many Content producers to monitor/influence.
- ◆ Internet Content providers cannot be conveniently aggregated into an industry sector for which a voluntary code of practice can be devised.
- ◆ Even if a regulator body were to be established for Internet-delivered Content, it would immediately run into jurisdictional issues since the global nature of the Internet does away with national borders.

### **3.7.2 Blocking by the Consumer**

Two main approaches are being pursued for Content blocking at the consumer end.

#### ***Labels for Web pages***

There has been considerable discussion recently regarding the use of labelling for web pages to enable filtering of URLs to take place at the user (consumer) level. The World Wide Web Consortium (W3C) has produced a specification for *Platform for Internet Content Selection (PICS)*<sup>18</sup>. This requires a Content Producer or third party to label web pages with rating information. The labels on Web pages received by a User are compared against the set of permissible labels that the User has specified.

There are several problems associated with this approach to Content blocking:

- ◆ This method relies on a critical mass of Content to be labelled, in order to work properly. It is not currently standard practice for most Content producers to automatically create labels, although several third party organisations do rate Content, particularly with minors in mind<sup>19</sup>.
- ◆ Whilst some third parties currently label web pages in a voluntary capacity, there would be costs involved if it were to be made mandatory.
- ◆ Different communities have different moral values, and web pages may not be appropriately labelled for all communities.
- ◆ Current web browsers do not handle labels well. To be successful, labels need to be an intuitive and integrated part of standard browsers.

#### ***The use of Filtering Software***

Filtering software, which is installed on a User's computer, examines material as it is received by the User's computer<sup>20</sup>. Such software may be used to:

- ◆ block identified sites – eg specific URLs or newsgroups;

---

<sup>18</sup> See <http://www.w3c.org/PICS/> .

<sup>19</sup> <http://www.research.att.com/~lorrie/pubs/tech4kids/>

<sup>20</sup> Ibid – contains an extensive list of filtering software.

- ◆ carry out on-the-fly filtering of incoming material against a set of pre-determined keywords.

There are problems on-the-fly blocking:

- ◆ Text based blocking does not work properly for graphical Content.
- ◆ Legitimate material may be blocked inadvertently. For instance if material containing the word 'breast' is to be filtered, documents which discuss 'breast cancer' will be blocked. In fact this document itself would be blocked as a result of the inclusion of the word 'breast'.

### **3.7.3 *Blocking in the Distribution Mechanism***

This report is primarily concerned with blocking in the distribution mechanism – i.e what can an ISP or BSP do to block Content that passes through their computers/routers, and at what cost?

The rest of the report addresses this issue.

## 4 Internet Blocking at the Application level

Content blocking on the Internet can occur at either:

- ◆ **the application level** – for instance, blocking a particular web page or ftp file, or stopping a particular news group or news item. This is carried out at the server level.
- ◆ **the packet level** – this is carried out by routers.

This chapter considers application blocking. Packet level blocking is discussed in the following chapter.

### 4.1 What is Application level blocking?

Application level blocking refers to the blocking of Content along the following lines:

- ◆ blocking web pages from a particular location by specifying its URL – <http://www.xyz.com.au> – and all web pages within that domain.
- ◆ blocking a set of web pages – <http://www.zyz.com.au/magazineA/...>
- ◆ blocking a specific web page – <http://www.xyz.com.au/fileA.html>.
- ◆ blocking file access from a particular ftp site by specifying the address of the site, or a particular file (or pattern) within a site – e.g. <ftp://ftp.xyz.com/fileB.jpg><sup>21</sup>.
- ◆ blocking a particular news group.
- ◆ blocking or cancelling a particular news item.

Application level blocking has been mandated by the Singapore Government. The Singapore Broadcasting Association (SBA) has issued an Internet code of conduct. Under the terms of their ‘class licence’ with the SBA, all ISPs and Content providers are required to remove or block Content specified by the SBA<sup>22</sup>. This does not apply to commercial Internet users.

Appendix 2 contains survey results from both Internet users and non-users in Australia, Singapore, and New Zealand in relation to Content blocking on the Internet.

#### 4.1.1 Blocking web pages and ftp files

The most common method employed by an ISP to block web pages and ftp files is illustrated in Figure 13, where the ISP employs a *proxy server*. All clients of the ISP must go through this proxy server to be able to access the Internet ‘proper’. To do so, clients must configure their web browsers and ftp clients to ‘point to’ this proxy server to be able to access web pages and ftp files<sup>23</sup>. Failure to do so will result in blocked access to the Internet, as indicated in Figure 13.

When a user requests a particular web page or ftp file, the following takes place:

- ◆ The proxy server checks to see if the requested URL is on its black list.
- ◆ If the URL is on the black list, the user is informed accordingly that the page or file is unavailable.
- ◆ If the URL is not on the black list, and is in the cache of the proxy server, the requested page or file is sent to the user from the proxy (caching proxy only).

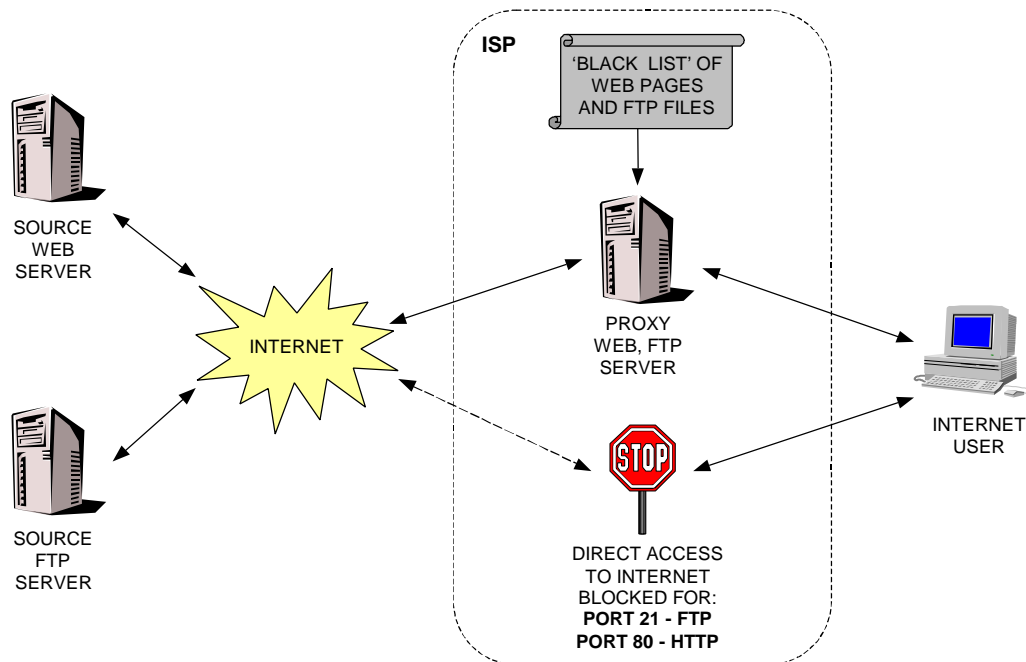
---

<sup>21</sup> The *jpg* suffix to a file indicates that that file is a picture.

<sup>22</sup> See <http://www.sba.gov.sg/netreg/idnote.htm>

<sup>23</sup> Transparent proxy servers, which do not require an Internet user to configure his/her browser to point to the ISP’s proxy server, are a comparatively recent development. Transparent proxy servers break the end-to-end architectural model of the Internet, and are discussed in Appendix 3.

- ◆ If the requested material is not in the proxy server, it is fetched from the source web or ftp site.



**Figure 13:** Blocking web pages and ftp files by an ISP using a proxy filter.

#### 4.1.2 Blocking news

Accessing news is a little different, since a news server receives its news feeds from a subscription list, maintained by the ISP. This is illustrated in Figure 14, which shows that the ISP's newsgroup subscription list is influenced by the supplied newsgroup black list.

The effectiveness of application level blocking using proxy servers is limited by a number of factors which are addressed in the remainder of this chapter.

#### 4.1.3 Blocking e-mail

The blocking of e-mail is not considered in this report, but it is instructive to note that an effective e-mail blocking service, known as the Mail Abuse Protection System (MAPS) has been implemented<sup>24</sup>, primarily to block spam e-mail. This service runs on an ISP's mail server, and detects when mail is received from particular organisations by noting the IP address of the sender of the e-mail. MAPS compares the sender's IP address with a Realtime Blackhole List (RBL). If the sender's address is on the RBL, the mail is blocked, and the sender informed accordingly. The MAPS RBL is designed to create intentional network outages for the purpose of limiting the transport of spam e-mail.

While this blocking service is an application-level blocking service – i.e it blocks a particular service (e-mail) – it contains elements of packet-level blocking which is discussed in the next chapter.

<sup>24</sup> See <http://maps.vix.com/rbl/>

## 4.2 Not everyone accesses the Internet through an ISP

While many home Internet users access the Internet through an ISP, and hence may be forced to go through the ISP's proxy, many Australian Internet users obtain access to the Internet by other means, such as academic institutions and work environments. Many of these organisations operate their own servers, and use an Internet Access Provider (IAP) for connectivity to the Internet only.

To ensure comprehensive blocking coverage, it would be necessary for such Content-hosting organisations to employ proxy servers as well. However this may not be easy to achieve because:

- ◆ tertiary educational institutions may consider that such 'censorship' is inconsistent with academic freedom, and
- ◆ many commercial organisations that provide employees with Internet access may be internationally focused, and have globally-based networks with servers all over the world. This is explored further in the following chapter.

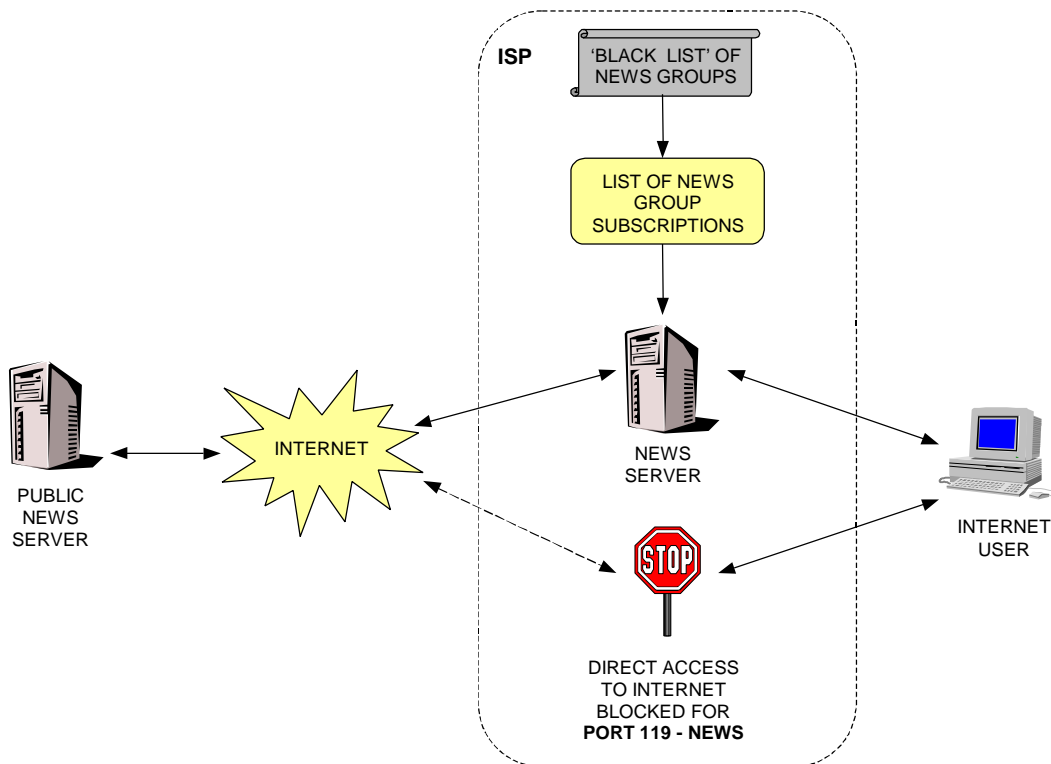


Figure 14: An implementation of newsgroup blocking by an ISP.

## 4.3 The use of different Port Numbers can bypass the filter

The schemes shown in Figures 13 and 14 involve the use of the standard http, ftp and news port numbers – 80, 21, and 119 respectively – to block requests. However these standard ports are not always used. For instance an embedded hypertext URL in a web page may be of the form:

<http://www.xyz.com:100/>

which specifies that port number 100 be used in the webserver where the Content is hosted. The user who requests this page may be unaware that the page is being accessed using a different port number. This page will still be blocked by the proxy server, but if the user re-configures his/her browser to bypass the proxy and go to the Internet 'proper', the request will not be blocked by the ISP.

In other words, there are mechanisms for getting around proxy blocking.

#### 4.4 Sites can easily be re-named to bypass blocking

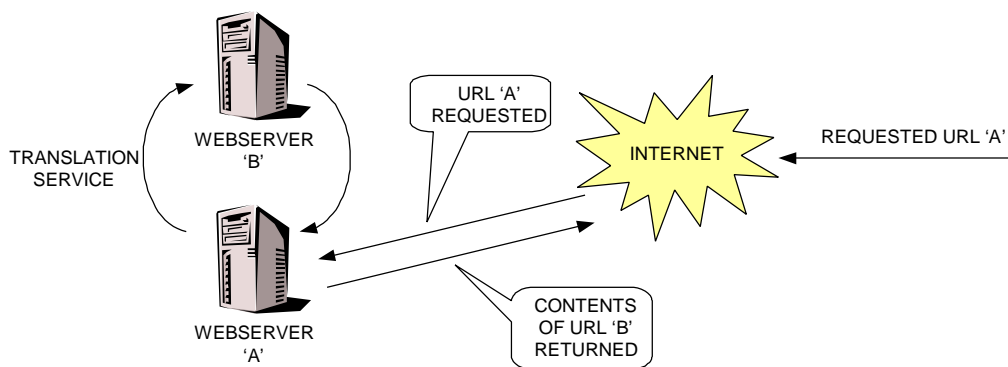
Sites that are determined to bypass any blocking filters can easily re-name their site. Devotees of that site can then discover the new URL by using the standard search engines which are updated at regular intervals, or the new URLs can be disseminated in Internet chat groups, news groups, or recorded telephone services.

Sites could be re-named regularly using a pre-determined algorithm, in an analogous fashion to frequency hopping methods used in radio communications to implement security. This could lead to closed communities that are outside normal regulatory jurisdictions, and which operate in a similar fashion to the pirate radio stations of the 70s.

#### 4.5 Translation services return different material to a user

Services exist on the Internet where a site that has been contacted can automatically access another site, so that the information that is returned to the user is that of a different URL to the one requested. This is illustrated in Figure 15, where a request to webserver A results in a subsequent request to webserver B. The Contents of Webserver B are returned to Webserver A, and in turn transmitted to the person that made the original request.

The Content specified by URL 'B' may then be stored in the proxy server's cache. If URL 'B' happened to be on the black list, not only would it bypass the filter, it would be stored by the ISP for easier access next time.



**Figure 15:** A translation service can return a different URL to the one requested.

Translation services can be particularly useful, and it would be counter-productive for any organisation to try to 'ban' them.

The following are examples of translation services:

- ◆ **Language translation:** the search engine *AltaVista*<sup>25</sup> provides a most useful automatic language translation service – eg English to Spanish. The material returned to the user has a different URL to that which is requested from the search engine.
- ◆ **E-mail-to-ftp (and e-mail-to-web) gateways** exist, such that sending e-mail to a particular site results in an ftp (or http) request to another site, which is then returned to the user. The returned ftp or web site may be on the black list, but the requesting e-mail would not have been blocked by the proxy server.
- ◆ **Word-to-html conversion:** accessing a formatted text document (such as Microsoft Word) can automatically access a html converter, so that the document is delivered to the user as a web page.
- ◆ **A different proxy can be specified using a different port number:** it is possible for a user to specify a different proxy server to the one established by their ISP, and if this proxy server uses a different port number, then requests will not be blocked by the local ISP. Owners of other proxy servers would probably not welcome traffic from non-clients, but organisations such as Anonymizer Inc. provide proxy server facilities, using port 8040, on a commercial basis.<sup>26</sup>
- ◆ **Web-to-voice and e-mail-to-voice translation:** several telecommunications companies have made recent announcements of products that enable web-based Content and e-mail messages to be delivered by voice over GSM.

With appropriate prior knowledge – which may be easily obtained from Internet chat groups, news groups, or recorded telephone messages, etc – it is straightforward for a user to make an approved request, which may return black-listed Content. It is particularly easy if a translation service is used – the user does not need to know the new address.

#### **4.6 The Domain Name Server can be bypassed**

A proxy filter is normally set up to filter out a black list of URLs. A Domain Name Server (DNS), which is normally operated by an ISP, translates a descriptive text URL into its numeric IP address. If the IP address of the source machine is known, a user can specify that address directly in the URL, instead of the more easily remembered domain name for the site.

For instance, the web server at Parliament House in Canberra may accessed by the URL: <http://www.aph.gov.au>, or by its IP number : <http://202.14.81.135/>. Finding an IP address is straightforward using the commonly available *nslookup* program, which queries the domain name service to find the IP number, just as the browser client would.

A black list should therefore contain both the descriptive domain name, as well as its IP address, to block a site effectively.

#### **4.7 Creating and distributing the Black List may be problematic**

If ISPs are required by legislation to implement proxy-based blocking, a recognised authority should compile and distribute the black lists to ISPs. ISPs would otherwise be placed in the difficult position of taking on the role of a moral arbiter by determining which sites should or should not be blocked.<sup>27</sup>

---

<sup>25</sup> <http://babelfish.altavista.digital.com/cgi-bin/translate> .

<sup>26</sup> <http://www.anonymizer.com:8040/>

<sup>27</sup> On the other hand an ISP may see some competitive advantage in implementing a so called *zero tolerance* policy, by employing substantial filtering in order to appear to be more ‘family oriented’.

Apart from the invariable outcry from some clients for going too far, and from others for not going far enough, the ISP may also find itself in the difficult legal position of blocking (removing) material published by their own clients, and in so doing breaking a contractual obligation.

Maintaining a black list at a central site is a simple task, but the distribution of that list in a secure manner to the 600+ ISPs in Australia, and to a larger number of organisations that host Content servers, would not be trivial.

#### **4.8 The Black List is a valuable commodity in its own right**

A black list has the potential to become a most valuable commodity in its own right, and will be a target for hackers. As soon as a black list has been obtained by a hacker, it will be broadcast over the Internet, with the result that it will become public knowledge. This would then be invariably regarded as a 'must see' list – purely by virtue of the fact that it represents a banned set of URLs<sup>28</sup>.

Because of the likely popularity of black lists, it is envisaged that a 'rogue state' may condone – and even support ISPs – who specifically advertise the fact that they host black-listed sites<sup>29</sup>.

#### **4.9 Push technologies bypass proxy filters**

The term *push technology* refers to the delivery of Content that was not specifically requested by an Internet user, but is the result of some previously negotiated arrangement with a Content provider. E-mail is a form of push technology, to the extent that a User receives e-mail without specifically requesting it, although such communication is normally between a set of regular correspondents. Spam mail – i.e. unsolicited e-mail – is an inappropriate use of push technology.

Because a proxy server checks *requests* for Content, Content that is delivered to a user without a specific request will not be blocked by a proxy filter.

#### **4.10 There are costs involved with employing proxy servers**

Many ISPs already use caching proxy servers to reduce their communications costs, but many – particularly smaller ones – do not. A requirement for *all* ISPs to operate proxy servers will place a financial burden on many ISPs, particularly in the following areas:

- ◆ **Contingency plans:** These need to be put in place for the eventuality that the primary proxy will fail. Should no contingency be in place, all access for the ISP's customers would be cut off, resulting in a loss of business. The cost of adding such a contingency in the case of even a medium-sized ISP (say three POPs in Sydney, Melbourne and Brisbane with 1 MB/s of bandwidth between them) can be many times the cost of implementing and managing a single proxy system.
- ◆ **User assistance:** Since access to the Internet via a proxy is mandatory, user support costs may end up being a significant part of implementation costs. The average cost of a support call is estimated to be around \$7.50<sup>30</sup>, and it is believed that around 25% of clients would require a support call to properly configure their browser.

---

<sup>28</sup> In the past, banned books, such as *Lady Chatterly's Lover*, achieved certain notoriety – purely by virtue of the fact they were known to be banned. Obtaining a copy of the book became a challenge, and as a result the book may well have been more widely read than if it had not been banned [personal experience – pmc].

<sup>29</sup> In an analogous fashion to pirate radio stations in the 70s.

<sup>30</sup> Private correspondence with an ISP.

#### **4.11 A proxy server may introduce unreliability**

The policy of forcing users to access the Internet through a proxy server reduces the reliability of Internet access, as it introduces a single point of failure<sup>31</sup>. Without the proxy, an Internet user can access the wider Internet by means of the ISP's routers (see next chapter), and having done so, can then access any server on the Internet. To provide a highly reliable service, while at the same time forcing clients to pass through a proxy server, an ISP would need to take steps to provide a failure-tolerant proxy server.

#### **4.12 A proxy server may adversely affect some applications**

Many Internet applications are client-server in nature, as indicated in Figure 8. The software has two physically separated components – the client part which operates on a user's PC (the browser in the case of the World Wide Web), and the server part which typically hosts Content, or is a front end to an organisation's back end business systems.

Some existing applications have problems working through a proxy server. The early releases of Telstra's *Big Pond* Cable Internet service, for instance, had limited functionality because of the proxy server used, although most of these problems have now been addressed. Many sites reportedly experience difficulties running Microsoft's *Front Page* web authoring tool through a proxy server.

The use of a proxy server, particular a transparent proxy (see Appendix 4), affects any application that collects statistics associated with site visits, such as monitoring the number of hits received by an advertisement.

#### **4.13 Performance is not a major issue**

An Internet user will probably not experience a performance degradation from the use of application level blocking using a proxy server. The initial request by a user for a web page, which is not already cached in the proxy server, will be slower as the page is fetched from its source, and then stored in the proxy cache. Subsequent requests for the same material will be retrieved from the cache, and will hence be faster than the original request.

---

<sup>31</sup> Anecdotal evidence from Internet users.

## 5 Internet blocking at the Packet-Level

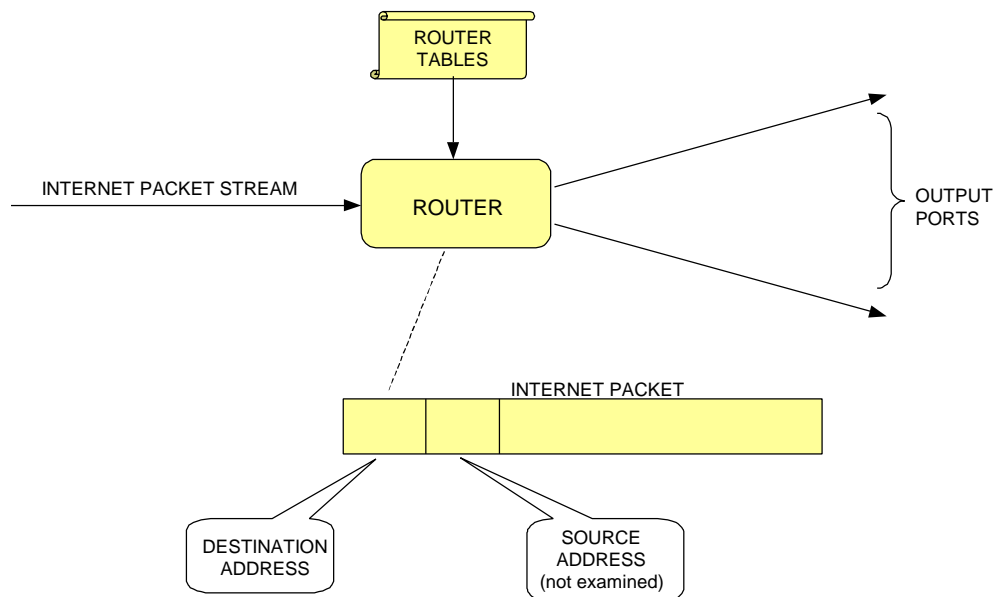
The previous chapter has considered the issue of blocking Internet delivered information at the application level – i.e. blocking specific web pages, ftp sites, or news groups. This chapter looks at blocking Internet packets, on the basis of their *source address* – i.e. where the packets have come from.

We begin with a brief discussion of routers, which comprise the backbone of the Internet.

### 5.1 Routers: the backbone of the Internet

A router is a special purpose computer which examines the *destination address* of a packet, and, using information from its *router tables*, directs that packet towards an output port which will take the packet closer to its destination. This is illustrated in Figure 16.

Although a general purpose computer such as a Sun workstation may be configured with appropriate software to carry out the role of routing, it is normally not fast enough to handle packet streams above a certain size. The actual routing algorithm that is carried out by a router is mostly implemented in hardware to increase speed. The main performance metric of a router is the number of packets it can handle per second. A top-of-the-line router, such as a Cisco 7500 series, can process up 1 million (64 byte) packets per second. Appendix 3 describes briefly how routers operate.

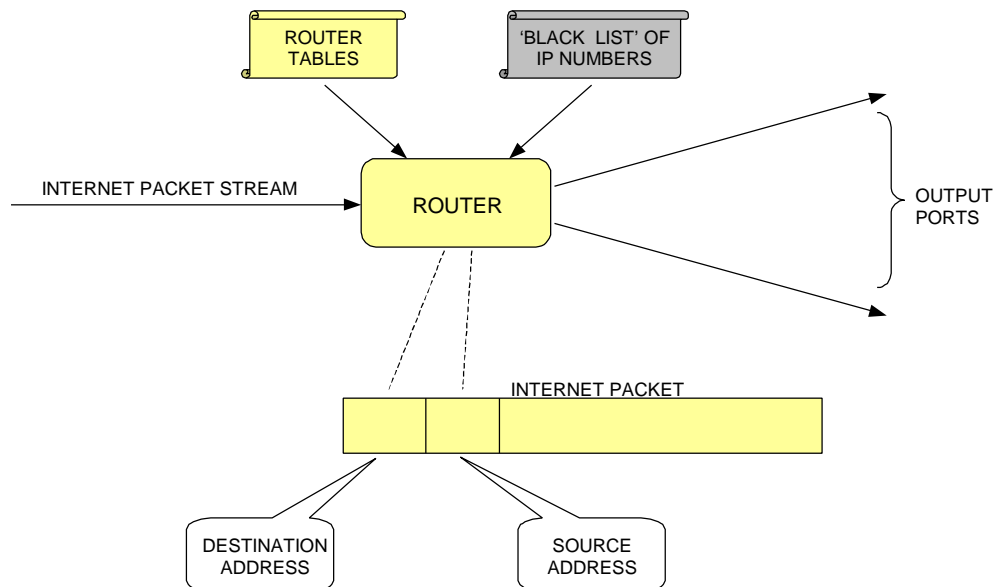


**Figure 16:** A router operating in its ‘normal’ mode examines the destination address only.

### 5.2 A router can be used to implement packet blocking

A router may be configured to carry out packet blocking. To block a packet stream, a router must also examine the *source address*, in addition to the destination address, of each packet, as illustrated in Figure 17, and compare that address against a supplied black list of IP addresses, generally referred to as an Access Control List (ACL).

Blocking may be extended to include the port number, thereby providing the capability of blocking, say, ftp access from a particular site, but not web pages. This requires further decoding of the packet, since the port number is contained in the TCP header which is within the payload part of the packet.



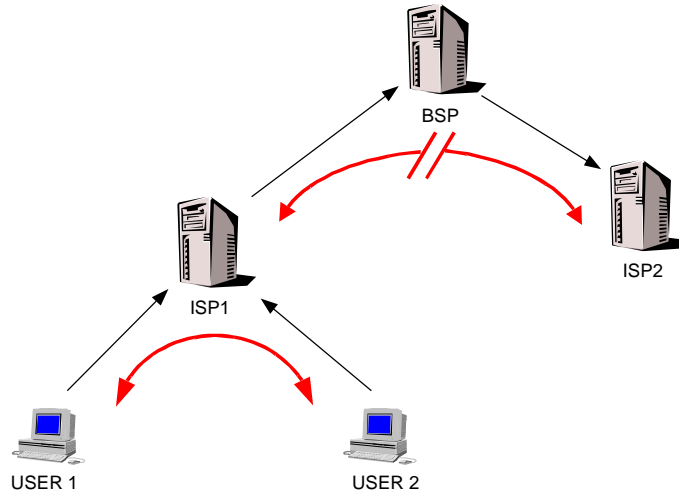
**Figure 17:** Packet level blocking: a packet's source address is compared with a black list.

### 5.3 Where should packet blocking take place?

With reference to Figure 7, packet level blocking may be implemented in several places:

- ◆ **within an organisation:** It is likely that packet black lists will be large and, given the global nature of the Internet, a company would need to filter against the same black list(s) as an entire country. This would place a high processing load on any organisation, and would invariably involve costly router upgrades to be able to handle Content blocking. Because an organisation is in control of its own environment – including servers – proxy filtering would be more appropriate.
- ◆ **at the ISP level:** There are over 600 ISPs in Australia, ranging from very small one-person companies to large international telecommunications companies. For the same reason as above, each of the ISPs would be required to upgrade (or more likely replace) their router hardware to implement packet blocking. The operational and support costs of doing this would place a high burden on the smaller ISPs.
- ◆ **at the BSP level:** All the ISPs in Australia are serviced by a relatively small number of Backbone Service Providers (BSPs). Whilst BSPs may be in a better position to absorb the costs of packet blocking, it would not be an effective place to filter, since a high percentage of Internet traffic may not actually travel through a BSP.

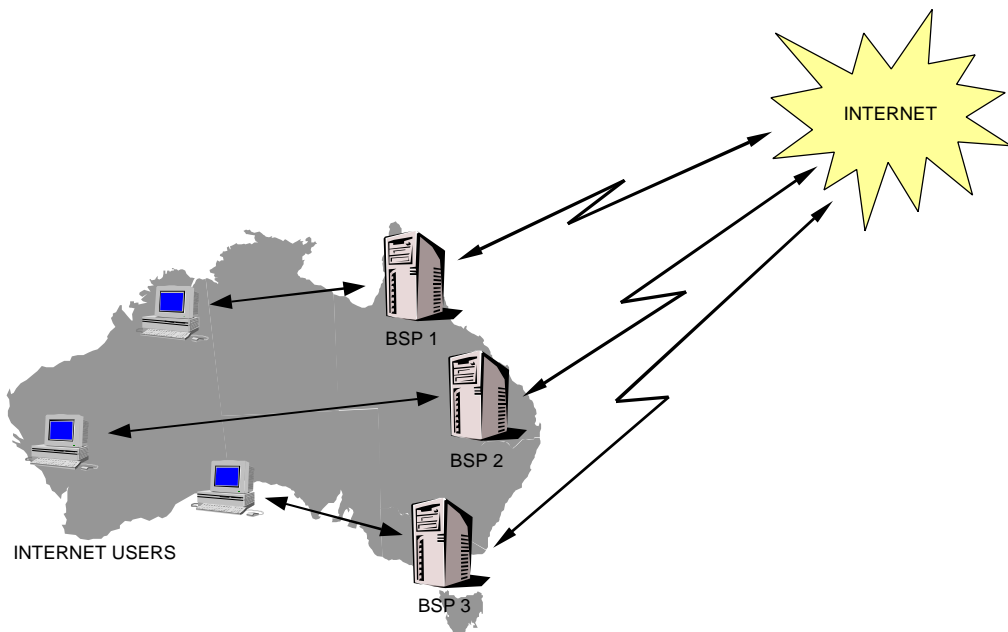
Figure 18 shows a situation where a BSP has implemented packet blocking for packets that pass through the BSP's site. The BSP provides backbone connectivity for several ISPs, two of which are shown. User 1 and User 2 are both clients of ISP 1, but when they exchange a message, no packets actually leave the ISP's premises, and hence do not pass through the BSP's blocking filter.



**Figure 18:** Packet blocking by a BSP will not pick up Content transferred between 2 users of the same ISP.

◆ **at international IP gateways:**

Most Content on the Internet resides on servers *outside* Australia. Because it is outside Australia's jurisdiction, authorities in Australia are unable to ask the hosting organisation to remove offending Content. In fact, due to differences in international regulation, the Content in question may be entirely legal in the jurisdiction in which it is being hosted. Locally hosted Content that is either illegal or considered to be offensive, is more appropriately handled by a direct approach to the ISP or the organisation that hosts the material.



**Figure 19:** A small number of Backbone Service Providers provides Australia with access to the wider Internet.

If packet blocking is to be implemented, the intuitive place to employ it therefore is at the Internet gateways to Australia, which are operated, in the main, by Backbone Service providers (BSPs), as illustrated in Figure 19. Placing blocking filters anywhere other than at the overseas links would be a significant duplication and waste of effort.

Although the number of BSPs that have international links is currently not large, this is set to increase as larger ISPs set up international satellite links. The entire concept of an international gateway may disappear as new infrastructures such as Low Earth Orbiting satellites (LEOs) begin to be deployed as an Internet delivery mechanism.

The placing of blocking filters at international gateways is analogous (to a limited extent) to the situation in the ‘physical’ world, where people arriving at a jurisdiction may be ‘blocked’ by immigration control. Immigration officers are stationed at all the main ports of entry to Australia, and inspect the passports of people arriving in the country. The passport details are compared with a supplied black list, and entry of the arriving person is either permitted or denied. It is interesting to observe that the operational cost of such immigration blocking is borne by Government, not by the airlines or shipping companies. The analogy between the players is shown in Table 3.

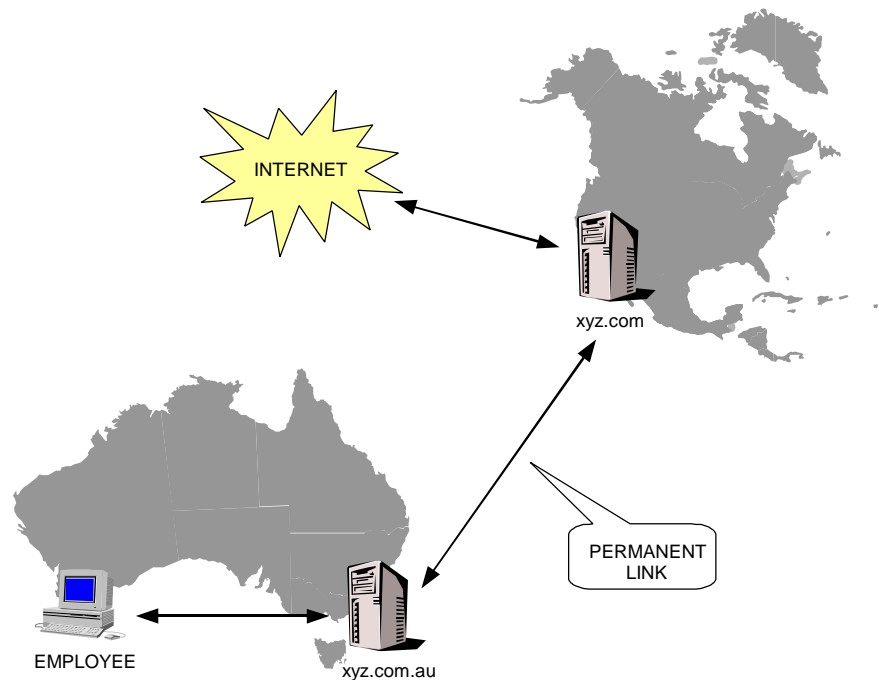
Immigration Control	Packet level blocking
immigration officer	router
passport	packet header
airline (or ship)	Backbone Service Provider

**Table 3:** Packet blocking shares some similarities with immigration control.

The effectiveness of packet level routing is limited by a number of factors which are discussed in the ensuing sections.

#### **5.4 Not all international Internet traffic passes through a BSP**

All multinational corporations have Internet based networks, often with a large number of servers distributed throughout the world. A typical multinational company may have, say, a server in Sydney and another in the US, with a dedicated communications line between the two, as illustrated in Figure 20. The two sites would be permanently connected, and may not use the service of a local ISP or BSP. Packet level blocking by BSPs would not block material on these private IP links into and out of Australia.



**Figure 20:** A multinational organisation with permanent international Internet links would not use a BSP.

### 5.5 Packet-level blocking is indiscriminate

Blocking packets by IP number is a particularly coarse way of filtering, and an unintended consequence may be the blocking of an entire site. If that site – say, OzEmail – happens to host a large amount of Content, then the entire OzEmail site will be unavailable.

If one of OzEmail’s clients places Content on his/her website and this material is considered by some parts of the community or jurisdiction to be offensive, then a packet level block of OzEmail by that community or jurisdiction will result in *all* of OzEmail’s clients – both personal and commercial – being blocked. This will have a detrimental effect on those commercial clients of OzEmail that are using the web as an integral part of their business activity, since they will be ‘invisible’ to those communities or jurisdictions where the block has been applied. If the jurisdiction concerned happens to be an entire country, then Australia’s export performance to that country may be adversely affected.

Similarly, if Australia were to implement blocking of certain overseas-based sites using packet-level blocking, then Australian companies would be unable to carry out business with overseas consumers or companies that are hosted by the blocked site.

The indiscriminate nature of packet level blocking could put entire organisations out of business.

### 5.6 Packet-level blocking will inhibit e-commerce infrastructure

Australia may well appear on the ‘bottom right corner’ of European maps, but geographical isolation is made largely irrelevant by the Internet. Moves are afoot by several multinational telecommunications companies to make Australia a hub for the emerging electronic commerce

infrastructure for South East Asia. This means that information coming into Australia may well be in transit to other countries.

Any attempt to filter packets as they enter Australia may block traffic destined for another country. If the blocked traffic is an electronic commerce transaction between a client of an ISP that is on Australia's black list and an organisation in another country, then Australia will soon become regarded as an unreliable supplier of electronic commerce infrastructure.

### 5.7 Packet-level blocking may affect other services

The application of a packet block on a site affects *all* services to and from that site. If the intention is to block a particular site because of particular web pages that are hosted there, then e-mail communication with that site will also be blocked.

As mentioned above, most routers provide the ability for a port number to be included in the filter. This which would allow, for instance, e-mail communication with a particular site, but not web access. However, many sites use multiple port numbers to hold different parts of their web site, and this will tend to increase the length of black lists.

### 5.8 Routers can easily be circumvented

Schemes such as *tunnelling*, where an IP packet is contained inside another IP packet, are commonly used. The inside packet is extracted by the receiver to recreate the original message – which may be on a black list. Both sender and receiver need to use cooperating software, but such software may be easily downloaded over the Internet.

Tunnelling is frequently used by companies with distributed offices to implement a Virtual Private Network across their organisation, the result being that all employees appear to be on a single network.

### 5.9 Sites can easily be re-numbered to bypass blocking

In an analogous fashion to proxy server blocking, owners of sites that are determined to bypass any blocking filters can easily – and regularly – change their IP number, thereby bypassing the black list. Devotees of that site can then discover the new URL by using the standard search engines which are updated at regular intervals, or the new addresses can be disseminated in Internet chat groups, news groups, or recorded telephone messages.

### 5.10 Packet blocking can complicate firewalls

A firewall is designed to protect a site against unwanted Internet visitors. Packet-level blocking and firewalls are similar in many respects, and both can be implemented on a router. A comparison between firewalls and packet blocking filters is shown in Table 4.

	Firewall	Packet blocking filter
<b>Traffic direction for filtering</b>	Incoming: outgoing traffic may be denied	Outgoing
<b>Filtering method</b>	'permission' list: refuse everything else	'black list': accept everything else
<b>Filter table size</b>	Small – perhaps a few 100 entries	Large – may be 10s of thousands of entries
<b>Computational requirement</b>	Low – can be implemented on a mid-range router	High – needs a high end router

**Table 4:** A comparison between firewalls and packet blocking filters.

It can be seen that firewall filtering and a packet blocking filter are quite distinct activities, which may be difficult, if not impossible, to integrate and implement on a single router.

### **5.11 Packet-level blocking must be implemented in hardware**

As mentioned above, a router is a special purpose computer highly tuned to carrying out a single task – that of reading a packet’s destination address and forwarding that packet to an output port that will take the packet closer to its destination. The program that carries out the routing algorithm is implemented in hardware, to improve the speed of the routers.

Packet blocking must also be implemented in hardware, otherwise throughput performance would be seriously affected. Cisco Systems, the manufacturer of the most commonly used routers used on the Internet, indicate that a Cisco 7500 class router, appropriately configured, is able to carry out source address blocking at line speed<sup>32</sup>. The vast majority of ISPs currently do not have such top-of-the-line routers, however.

The number of entries in a black list should not affect performance significantly if appropriate algorithms are used.

### **5.12 Implementing packet-level blocking is costly**

There are several aspects to the costs associated with packet level blocking.

#### **Hardware costs**

As mentioned above, packet blocking must be implemented in hardware, if performance is not to be affected. Top-of-the-line routers are able to carry out packet blocking in hardware, but earlier router models cannot. If a BSP already has 7500 series routers (or equivalent), then the upgrade cost is minimal. A large black list may require routers to be upgraded with additional memory to handle the larger tables involved.

However if a BSP or ISP does not have top-of-the-line routers, then a major upgrade will be required to handle packet blocking, and this may involve costs of hundreds of thousands of dollars for appropriately configured routers.

It should be noted, however, that the cost of any hardware upgrades will pale into insignificance by comparison with on-going bandwidth leasing costs across the Pacific, which are in the vicinity of \$1.5M per month for a 45Megabit line<sup>33</sup>.

#### **Operational costs**

The greater cost of implementing packet blocking would be the on-going personnel costs associated with the creation, maintenance, and distribution in a secure manner of black lists to those BSPs who would be carrying out the blocking.

The BSPs would experience the costs associated with training staff to become familiar with blocking on the relevant platform, which is not something that is normally required of an ISP, and then testing and implementing an appropriate configuration.

---

<sup>32</sup> Cisco Systems – private communication. Cisco Express Forwarding (CEF) must be enabled.

<sup>33</sup> List price – prices are obviously negotiated.

## 6 Conclusion

The issue of Content blocking is a difficult and, at times, emotional issue. It is beyond doubt that material which is classified as illegal in Australia should not be available to Australian users, whether via the Internet or other distribution means. And most people would agree that minors should be protected, if at all possible, from accessing Content on the Internet – either accidentally or wilfully – that may harm them in some way or another.

Our study of Content blocking relates to Content which is hosted *outside* Australia, but accessed by Australians. Content that is hosted in Australia should not be handled by blocking techniques. If locally hosted material is illegal, then the hosting organisation (which can easily be identified) is required by law to remove it. If the material is offensive, then the hosting organisation can again be contacted directly.

It is technically possible to block Internet-delivered Content at two distinct levels – at the application level, and at the packet level.

**Packet-level blocking**, based on examining the source address on Internet packets, is technically possible and can be carried out without performance penalty using appropriately configured top-of-the-line routers, although it is believed by some that this may not be able to continue scaling. However packet level blocking is too coarse, and if implemented will create unintended ‘holes’ all over the emerging global digital infrastructure. This is inconsistent with Australia’s desire to become an electronic commerce hub for South East Asia.

**Application-level blocking**, based on the use of proxy servers, is technically possible but, as indicated in Chapter 4, it can easily be circumvented in more ways than packet blocking with the result that it would be largely ineffective. Furthermore, users do not have to be particularly experienced Internet users to bypass application-level blocking.

Our conclusion is that Content blocking implemented purely by technological means will be ineffective, and neither of the above approaches should be mandated. Work-arounds will quickly be devised for any technologically-based blocking system and distributed over the Internet itself.

Having said that, we propose two different solutions to the issue of Content blocking – one which can be implemented in the short term, and another for consideration as a development in the longer term.

### ***ISPs could offer differentiated services***

A wide range of filtering software is now available<sup>34</sup>, accompanied by an ever-increasing set of associated URLs which are updated on a regular (sometimes daily) basis. These products fall into several broad categories:

- ◆ They can operate on an ISP’s proxy server, or at the client end.
- ◆ Their filters can either pass and/or block URLs – in other words, they can work with a permitted list, or a black list.

Where there is a market demand, we suggest that ISPs be encouraged to offer differentiated services to clients, and in particular that services for minors be created, based on access to the Internet through a proxy server. Two classes of such differentiated services could be considered:

---

<sup>34</sup> A comprehensive analysis of commonly available filtering software is maintained at <http://www.research.att.com/~lorrie/pubs/tech4kids/>.

- ◆ A **'clean' service**: the filter includes a list of *permitted URLs only*; requests to all URLs outside this list are refused. The 'real' Internet is not actually accessed, and a user cannot escape from the prescribed 'universe' that s/he finds him/herself in. Several such proxy-based filtering schemes are currently available, providing access to a universe of thousands of permitted pages.
- ◆ A **'best effort' service**: the proxy filter blocks a set of known sites, rated according to some prescribed criteria. The result is based on a best-effort approach by an ISP, and cannot be guaranteed. Bess filtering software<sup>35</sup>, for instance, claims to have a black list of 'hundreds of thousands of pages'.

ISPs would incur some costs in setting up services such as these. These could either be passed on to clients in increased fees, or an ISP may see some competitive advantage in providing such environment to clients<sup>36</sup>. Alternatively the Government may consider providing some incentives to ISPs to offer such differentiated services.

To be successful, it is essential that the initial access to the ISP should be to the filtered service. This could subsequently be bypassed by parents, if necessary, with the use of a password.

In addition to the above, individual users can of course acquire and install client-based filtering software as a commodity product.

### ***International Cooperation is needed to determine jurisdiction***

Locally-hosted Content, that is either illegal or considered to be offensive, is best handled by a direct approach to the ISP or the organisation that hosts the material, requesting that the ISP or hosting organisation take appropriate action.

Most Content on the Internet, however, resides on servers *outside* Australia. Because it is outside Australia's jurisdiction, authorities in Australia have no authority to request the hosting organisation to remove illegal or offensive Content. In fact the Content in question may be entirely legal in the jurisdiction in which it is being hosted, as a result of differences in international regulation.

It is proposed that Australia participate in international fora to create the necessary infrastructure, so that organisations which host Content would be able to determine the jurisdiction of the client software making the request. Having determined the jurisdiction, the server can find out whether the requested Content is legal in the client's jurisdiction.

This proposal is expanded in Appendix 5, and is clearly a long term solution. The required infrastructure will not be driven by Content blocking, but will probably be driven by other needs such as taxation, i.e. determining the location of a purchaser so that the amount of sales tax payable in the purchaser's jurisdiction can be worked out.

---

<sup>35</sup> See <http://www.n2h2.com>

<sup>36</sup> AOL and local ISP *iinet* offer filtered services.

## **APPENDICES**

**Appendix 1:** Glossary of terms

**Appendix 2:** Survey of consumer attitude to Internet Content Blocking.

**Appendix 3:** A brief description of the operation of routers.

**Appendix 4:** Transparent Routing – an overview.

**Appendix 5:** Proposal for a framework for an international initiative.

## Appendix 1

# Glossary of Terms

**Application.** An application is a software component that implements some particular use of the network. Typically two sorts of software components talk to each other: a server and a client. More complex arrangements are possible.

**Access Provider.** An Internet access provider (IAP) connects to the global Internet via a Backbone Service Provider (qv) and sells a service where they get their customers' IP packets to the Internet and receive IP packets from the Internet and route them to their customers. Today the customer's IP numbers are normally taken from the access provider, but whether that is the case or not it is the job of the access provider to see that a route to the customer's IP number range is advertised on the Internet. An access provider is traditionally, and still commonly, called an ISP (Internet Service Provider), however that term now carries connotations of also providing additional services.

**Binary.** The decimal system used by (most) humans has digits 0 to 9. The binary system used internally by computers has digits 0 and 1.

**Bit.** Computer store and communicate information as a sequence of "bits". A bit has one of two values normally represented as 0 or 1. Hence numeric information is stored in binary. Communication speeds are normally given in bits/second: A two MB link sends two million bits per second.

**BSP, Backbone Service Provider.** Backbone Service Providers are invariably also access providers. Their customers are access providers which they connect to the global Internet. BSPs typically connect to multiple other BSPs at multiple points giving a redundant, and thus reliable, connection.

**Byte.** A byte consists of 8 bits. Characters (in text) are typically stored one per byte using the ASCII code. Storage, disk or memory, is usually expressed in megabytes (MB). Network usage charging is commonly expressed per megabyte. A byte is sometimes called an *octet* in deference to past computers which stored characters in other than 8 bits.

**Cable modem.** Cable TV networks can be used as an IP link layer (qv) with the use of cable modems.

**Client.** Software, such as a WWW browser or an e-mail user interface, which connects to and uses a service provided by a server (qv).

**DNS, Domain Name System.** Names in the Internet are given a hierarchical structure. So the name "www.widget.com" was created by the owner of name "widget.com" which was created by the owner of the top level "com" domain. The DNS is a total standardised system which supports names of this type for Internet applications.

**E-mail.** The Internet Architecture specifies e-mail in several standards. SMTP (RFC821) specifies how e-mail is moved between computers. RFC822 specifies the format of e-mail messages. These standards have been extended, notably with the MIME (Multipurpose Internet Mail Extensions) standard. There are a number of other mail protocols, most of which are only effective within a single organisation. Nearly all inter-organisational mail passes through the Internet. However it rarely goes directly but instead goes from user to gateway to gateway and then to the destination user.

**Ethernet.** An ethernet is an example of a Link layer that is used to transport IP packets. Ethernets have broadcast and multicast capabilities that are utilised to efficiently implement the Internet over ethernet.

**FTP, File Transfer Protocol.** Original method of information retrieval on the Internet. Anonymous FTP (qv) allows people who aren't registered used on the ftp server to retrieve a limited range of files.

**Gateway.** When users are cut off from services because of incompatible protocols or security restrictions a gateway provides a mechanism that, as the name suggests, lets the user through. Examples include X.400 and other e-mail gateways. Gateways are increasingly a part of firewall systems and seen as having a security rather than a protocol conversion function.<sup>37</sup>

**Headers.** Used in two similar contexts. On packets of information headers are fixed size bits of information in exact positions at the start of the packet. Examples of header information in the IP packet include packet length and destination IP number. The other place it is used is in e-mail where the headers are just lines at the top of a text file, such as 'Subject:', 'To:'.

**Host.** Computers on the Internet are normally classified as 'hosts', meaning end-systems, and 'routers', which fill intermediate positions and pass packets to and fro.

**HTML, HyperText Markup Language.** Format of hypertext documents on the WWW.

**HTTP, HyperText Transfer Protocol.** The protocol, i.e. communication rules, used by the WWW for URLs of the form "http://...".

**Hypertext.** A document with text and graphics in which some areas of the document are 'active'. Clicking on an active area will perform some function, most commonly to display some other hypertext document. Hypertext is now commonly created as HTML.

**Internet Architecture.** Internet standards form an integrated set of documents which make up an overall architecture. When standards are written they can assume that the architecture is in place to build on. An important aspect of the architecture is the presumption that any computer can get any sort of IP packet to any other computer. Many computers that take part in the Internet are behind firewalls and gateways which interfere with the full operation of the Internet Architecture. It is up to the people who design such firewalls and gateways to simulate the correct operation of the Internet Architecture well enough so that the operational integrity of the Internet is not compromised.

**Interface.** A connection of a computer to the Internet through some physical/link layer mechanism such as modem/PPP or ethernet.

**IP, Internet Protocol.** This is the network layer in OSI terms. At this level, computers talk to other computers by sending IP packets addressed by IP numbers.

**IP number (IP address).** In the Internet Architecture each computer has one or more numbers by which it can be reached. The term 'IP address' is common in technical literature. However the word 'address' is heavily overused in communications documents, and the popular term 'IP number' is clearer. It has been the practice in the past for a computer to have a separate number for each interface, so if it had a modem and an ethernet it would have two numbers. Recently it has become common for some computers, particularly routers, to have multiple IP numbers on one interface or to share an IP number between multiple interfaces. These departures from the Internet Architecture seem to cause few problems if done carefully.

**IAP, Internet Access Provider.** See "Access provider".

**ISP, Internet Service Provider.** An ISP provides the packet level connectivity provided by an IAP, but also offers additional services, particularly e-mail service and also WWW hosting services.

---

<sup>37</sup> In the early days of the Internet "gateway" was also used interchangeably with "router". That usage is now extinct.

**Leased line.** A term commonly used for a fixed point-to-point link between geographically separated locations, as distinct from a dialup or other forms of transitory link.

**Modem.** A device that converts an analog link, such as a phone line, and creates a digital link suitable for computer communication.

**Multicast.** Some IP numbers represent multicast destinations. A packet with such a number as its destination may be sent to many computers. Normal destinations are called ‘unicast’ when it is necessary to distinguish.

**Packet (IP packet).** In the Internet, all communication takes place by exchange of IP packets. Each packet has a destination IP number and a source IP number.

**POP, Point of Presence** The larger ISPs have points of presence (POPs) throughout the country, which means that most of their subscribers can access the Internet without having to pay STD telephone costs.

**Protocol.** A protocol is a set of rules that define how entities successfully communicate. It might define the format of information or allowable requests and the meaning of responses.

**Router.** A router is any computer with more than one interface which is configured to forward packets. This means that when a packet comes in on one interface the router will consult its tables and may, as a result, send the packet unchanged out on one of its other interfaces. If it is a multicast packet, it may send the packet out on several other interfaces.

**Server.** A computer providing a ‘service’ on the Internet, which has a process waiting for incoming calls. The caller is the “client”.

**Spam mail.** Unsolicited e-mail distributed in large quantities, normally using third party bulk e-mail address lists.

**Subnet.** When a group of computers are connected together by a particular link layer it forms a subnet of the Internet.

**TCP, Transmission Control Protocol.** The main transport layer protocol on the Internet is TCP. Hence the protocol suite is often called TCP/IP. TCP handles error recovery so that the application level program just sees a continuous stream of data to and from the partner process.

**URL, Uniform Resource Locator.** A text string that describes where a document, or other resource, can be located on the Internet. The format is “type:type-specific-information”. Examples are “http://www.csiro.au/” and “mailto:Bob.Smart@cmis.csiro.au”. For web sites the URL’s type specific information starts “//domain-name” or “//ip-number”.

**WWW, World Wide Web.** The World Wide Web is a distributed hypertext system that has rapidly become a key feature of the Internet.

## Appendix 2

### Survey of Consumer attitude to Internet Content Blocking

These results are a summary of parts of several recent surveys carried out by Sydney based market research company, WWW.Consult<sup>38</sup>, which cover the attitude of Internet users in Australia, New Zealand and Singapore.

#### **Biggest concerns by Internet users with using the Internet:**

	Response time	Cost of Access	Privacy	Indecent material
Australia	26%	20%	16%	3%
New Zealand	18%	25%	10%	5%
Singapore	28%	20%	22%	2%

#### **Concerns of non-Internet users:**

	Response time	Cost of Access	Privacy	Indecent material
Australia	4%	18%	12%	15%

#### **Censorship views of Internet users:**

	Govt should censor	Parents should censor	No-one should censor	No view
Australia	7%	60%	30%	3%
New Zealand	8	65	22	5%
Singapore	10%	60%	25%	5%

#### **Conclusions**

- ◆ Internet users have little concern for indecent material.
- ◆ Non-Internet users are more concerned about indecent material than Internet users, but their chief concern is access cost.

<sup>38</sup> Reproduced with permission.

## Appendix 3

### A brief description of the operation of routers

Cisco routers are the most commonly used in major Internet operations. They are constructed with the same general structure:

- ◆ a central processor;
- ◆ one or more interface processors. These connect the router to external network links.

If the interface processor that receives a packet can understand it well enough, it will just send it directly to the output interface processor. However, if the packet requires special processing, it sends it to the central processor. Sending packets to the central processor has two potentially negative effects:

- ◆ it is a much slower path for the packet compared to going directly to the output processor;
- ◆ the central processor can become overloaded, resulting in packet loss.

In older routers, packets were referred to the central processor if blocking was required. This limited the speed of packet processing in the presence of blocking. In the new 7500 Cisco architecture, the interface processors are more intelligent, and can carry out blocking in hardware without the need to use the central processor.

Packets are classified into *flows*. A flow is a series of packets that has the same source address, source port, destination address and destination port as a previous packet. The interface processors remember the blocking and routing that was applied to a previous packet from the same flow, and treats the incoming packet the same way, without the need to access the central processor nor the Access Control List.

A large percentage of packets (90% in the Internet backbone) can be classified from a pre-existing flow, which speeds up the process of routing considerably. The interface processor can then forward or block most packets, based on this cache of information about how different flows should be handled. As a result, Cisco claim that the 7500 can do packet blocking without any performance degradation.

## Appendix 4

### Transparent Proxy Servers – an overview

With conventional proxy servers, users are forced to use a proxy server because their ISP blocks direct access to the Internet to http services (port 80). This has many negative features, not least of which is that many users don't know how to configure their browser to go through a proxy server.

A recent development is the invention of Transparent Proxy Servers. A transparent proxy server intercepts packets bound for web pages (as detected by the destination port 80). It then looks in its cache and can return the requested page as if it came from the requested destination. Detailed technical information on transparent proxy operation is available at <http://squid.nlanr.net/Squid/FAQ/FAQ-16.html>.

#### **Technical Problems with Transparent Proxying**

While normal proxy operation is a standard part of the Internet architecture, transparent proxy servers step outside the Internet architecture. When using Transparent Proxies the web client receives packets that seem to come from a remote web server but have actually been generated by a closer web proxy. This can easily interfere with the correct operation of the Internet protocols.

#### **Overcoming Deficiencies in Transparent Proxying**

If an ISP blocks port 80 (ie web pages) and clients are therefore forced to use their proxy server, then all web access will go through the proxy server, not just those to port 80. With transparent proxying, however, only connections to web services running on port 80 are caught.

It would not be difficult to enhance transparent proxy servers so that they examine all TCP streams without intercepting them initially. By watching the stream, they can detect if it is an HTTP connection. If so they can cache the page. When the next call to that destination comes in, they can respond from the cache.

#### **Proxy Caching Hardware Requirements**

The requirements are similar for standard or transparent proxy servers. The Australian overseas link to the US is currently in the order of 155Mbs. Since an average packet is about 200 bytes, that equates to around 100K packets per second. An average flow (session) is 10 packets, so this corresponds to about 10,000 web accesses per second. We see in [http://www.cisco.com/warp/public/751/cache/cds\\_wp.htm](http://www.cisco.com/warp/public/751/cache/cds_wp.htm) that a Cisco router cache server can handle 900 simultaneous sessions and a maximum configuration of multiple cache servers can handle 28,800.

Another way to look at it is that the actions of the proxy server are very similar to those of an originating web server and performance should be the same. We see in <http://solo.dc3.com/white/wsperf.html> that a cheap PC can handle about 30 pages per second. So our 155Mbs link could be served by 300 such PCs acting as proxy servers. The logistics of handling such a service are daunting. See section 16.4 in <http://squid.nlanr.net/Squid/FAQ/FAQ-16.html> for an indication of the how to combine PCs with Ciscos in implementing this task.

## Appendix 5

### A proposed framework for International Cooperation

The Internet knows no jurisdiction, and, as a result, material that is perfectly legal in some jurisdictions may be illegal in Australia. This is not a problem with Content that is hosted on Australian servers, as the hosting organisation can be identified relatively easily, and requested to remove the Content.

The fact is, however, that most Internet Content is hosted *outside* Australia. The aim is to provide a mechanism so that service providers outside Australia do not distribute Content, which has been determined as being illegal in Australia, to Australian clients. This requires international cooperation in the form of a treaty to assist in the blocking the delivery of Content from one jurisdiction to another. If such a treaty-based approach is adopted, then an approach like the following be considered.

#### **Desired Final Position**

In our desired final position, there is a group of treaty countries that agree that servers must adhere to the legislative requirements of the jurisdiction of the requesting *client*. Countries that do not cooperate with this treaty may be discriminated against in some way – perhaps by restricting their Internet links. This may just involve blocking sites which seem to break laws of some member countries.

There are two technical requirements for this to work.

1. There has to be a method for a server determine the jurisdiction of the client is that is accessing it. This has two aspects:
  - ◆ finding the jurisdiction in which a particular IP address is physically based; and
  - ◆ creating a way for a proxy service to determine where the relevant client is located. This may done as an extension of the *ident* service.
2. Then there needs to be a simple classification of material that can be done in an automatic way by the Content provider. This would give material a default status according to the different rules of each country in the treaty. Content providers could then seek more liberal ratings from approved PICS rating services.

#### **Getting There in Stages**

- ◆ A mechanism needs to be in place for collecting the information about which IP addresses are located in which countries. Of course we can only do this for Australia and hope that other countries will follow suit.
- ◆ Standards need to be developed to enable a server to find out where its client is physically located.
- ◆ International cooperation that will lead to a treaty of cooperation must be generated.

What Australia can do initially is be a good Internet citizen itself. This means getting Australian servers to respect the laws of other jurisdictions, in so far as they can reasonably determine them. An example might be that Australian on-line casinos should not service customers where the client cannot legally use such a service in their own jurisdiction.

The traditional approach to material coming into the country has been to intercept it at the border and to deal with it without regard to the wishes of the owners of that material. This seems reasonable because the person sending it must know it is being sent to Australia and can

check the Australian legal situation before sending. This approach does not work with the Internet as:

- ◆ the sender has no current way to determine if the destination is in Australia, and
- ◆ it is impossible to examine the material as it enters, so it is only possible to block access to it based on complaint. And even so the block will be circumventable.

Sometimes more than legislation is required. Regulating the Internet can only sensibly be done with a mixture of international cooperation and international standards. These things require time and effort.