TECHNICAL OBSERVATIONS ON ISP BASED FILTERING OF THE INTERNET

www.acs.org.au/ispfiltering

AUSTRALIAN COMPUTER SOCIETY

ICT Professionals Shaping Our Future

About the Australian Computer Society

The ACS (Australian Computer Society) is the recognised professional association for those working in Information and Communications Technology, seeking to raise the standing of ICT professionals and represent their views to government, industry and the community.

A member of the Australian Council of Professions, the ACS is the guardian of professional ethics and standards in the ICT sector, committed to ensuring the beneficial use of ICT for all Australians.

It provides both members and non-members with opportunities for professional development, networking and certification, as well as enabling them to contribute to the growth of their profession.

Visit www.acs.org.au for more information.

AUSTRALIAN COMPUTER SOCIETY

Level 3, 160 Clarence Street
Sydney NSW 2000
Tel: +61 2 9299 3666
Fax: +61 2 9299 3997
www.acs.org.au

Table of Contents

1 Introduction

The Australian Computer Security (ACS) established its Filtering and E-Security Task Force to advise on technical issues, policy, and provide expert commentary on filtering and e-security proposals and issues on an objective and transparent basis.

In particular, the role of the Task Force is to develop pro-active positions around ISP filtering and e-security, taking into account the needs of business and consumer user communities. It will recommend measures to ensure Australia has the highest level of e-security to engender business and consumer confidence in adopting and using on-line business models.

The Task Force recognises there is no silver bullet when it comes to cyber security and solutions to providing a safer and more secure Internet. Addressing this challenge will require an ongoing, multi-faceted approach involving government, industry and social (end user) initiatives to better educate end users with both appropriate technical security and educational awareness measures.

As a first project, the Task Force undertook to examine the key technical issues associated with ISP based filtering and its methodologies, and consider a credible, practical approach based on sound technical considerations.

2 Background

Cyber safety was part of the Federal Government's 2007 election platform. Australian Communications and Media Authority (ACMA)[1,2] has also published three recent studies on Internet content filtering, including "Closed Environment Testing of ISP - Level Internet Content Filtering". Its recommendations and their implications have been widely discussed amongst the ICT community, although the recommendations of these reports will likely be overtaken by the results of the ISP filtering trial being conducted by the Government. A fourth report from ACMA is also being prepared this year.

Key points from the ALP 2007 election policy on cyber security are:

- to provide a clean feed Internet service for all homes, schools and public computers used by Australian children.
- ISPs are to filter out content identified as prohibited by ACMA for this clean feed. The ACMA black list will be made more comprehensive to ensure children are protected from harmful and inappropriate material;
- provide parents, teachers and children with up to date, comprehensive and age appropriate online cyber safety resources and assistance;
- establish a Youth Advisory Group to ensure the Government is kept up to date with issues that affect children online;
- establish a permanent Joint Parliamentary Standing committee to investigate and report on cyber safety in Australia; and
- undertake further research into cyber safety issues to determine where best to target future policy and funding.

Education and information form an integral part of the policy platform stating that Australia children must be taught skills such as:

- how to be responsible cyber citizens;
- understanding how their online actions affect others;
- how to protect their identity online; and
- how to respond to cyber incidents (bullying, privacy etc).

Following on from this, in recent times the Government has stated that it is examining the introduction of ISP level filtering for Refused Classification (RC) material that is on the ACMA blacklist. Under the National Classification Scheme, RC material includes child sexual abuse imagery, bestiality, sexual violence, detailed instruction in crime, violence or drug use and/or material that advocates terrorism.

The Task Force has taken these stated policy initiatives into account when formulating its position on how to best address the issue of ISP filtering of illegal material online and how to best protect children from such material.

[1] *Australian Communication and Media Authority, June 2008, "Closed Environment Testing of ISP-Level Internet Content Filters"*
[2] *Australian Communication and Media Authority, Feb 2008 & April 2009, "Developments in Internet filtering technologies and other measures for promoting online safety" Reports 1 & 2*

3 Key Issues Addressed by the Task Force

3.1 What Are the Goals/Objectives of ISP Filtering?

The Task Force believes that the goals around any successful ISP filtering initiatives must be clearly established particularly the nature of content that is to be filtered and what is considered acceptable in terms of impacts on ISPs, network speeds, business and community from filtering initiatives.

Specifications and regulations should address and define exactly what content is to be filtered, how and where filtering is to occur, the protocols and type of traffic to be filtered, the mechanisms to be used to enforce filtering and the expected effectiveness of filtering.

It is important to note that the Internet is already subject to regulation that prevents domestic Internet content providers from hosting prohibited content (as determined with reference to the National Classification Scheme) defined under the Broadcasting Services Act 1992. The National Classification Scheme applies to publications, films and computer games.

Currently, prohibited content that is found to be hosted by an overseas provider is added to the ACMA blacklist which is provided to filter vendors.

3.2 What Type of Content Should Be Filtered?

The Broadcasting Services Act 1992 provides that online content is prohibited or potentially prohibited in accordance with the National Classification Scheme. This prohibited content comprises:

- content that has been classified or is likely to be classified RC;
- content that has been or is likely to be classified X18+;
- content that has been classified or is likely to be classified R18+ unless it is subject to a restricted access system;
- content that has been classified or is likely to be classified MA15+ and is provided on a commercial basis (i.e. for a fee) unless it is subject to a restricted access system.

The ACMA blacklist is currently provided to filter vendors that participate in the Internet Industry Association's Family Friendly Filter Scheme.

While recent Government statements indicate that ISP level filtering will apply to RC material that is on the ACMA blacklist, there is still a considerable amount of confusion amongst the ICT sector on exactly what content will be filtered.

The definition of material to be filtered must be unequivocal and clear if the initiative is to be successful.

As a guiding principle, the Task Force believes that material which is available off line should be equally available on line.

The Task Force believes that it is important for the Government to determine and clearly articulate the objective of its filtering policy. For instance, is it:

- to avoid inadvertent or unintended viewing of RC or illegal content while surfing the web;
- to prevent, detect, block and prosecute deliverable access, publication or circulation of RC or illegal content;
- to deter both inadvertent and/or deliberate interaction with a wider ambit of RC, illegal or prohibited material using any method of Internet access

In establishing the need for ISP level filtering, the Government has frequently cited the need to ensure children and other vulnerable groups are protected from RC, illegal and undesirable material. The use of filtering to remove access to child pornography has, in particular, been a key element of the ISP filtering debate..

3.3 Where Should Filtering Occur in the Network Architecture?

It will be necessary to clearly establish how and where filtering is to occur within the network architecture. Filtering can occur at multiple levels in the network architecture, for instance: upstream from the ISP (i.e. traffic originating from the ISPs international supplier and backbone service provider); at the ISP; between the ISP and customer; between two customers in the same ISP; or even between a customer and internationally hosted website.

3.4 What Type of Internet Services Should Be Filtered?

A key issue in the filtering debate is the type of Internet services and protocols that can or should be subjected to filtering. The Internet offers a range of services to users including world wide web (WWW), peer to peer (P2P) information sharing applications such as network news, Internet Relay Chat (IRC), email, file transfer protocols, internet telephone and many others.

3.5 Nature of Internet Filtering

The extraordinary diversity of material available on the Internet, much of which is generated dynamically on social networking sites, creates several technical and practical challenges involved in examining the content of each packet.

Potentially every Internet user is a content producer as well as a consumer. The nature of the Internet means that this material can reside anywhere in the world, implying that it is highly likely that the vast bulk of illegal material is generated and resides within jurisdictions that are not subject to Australian laws or regulations, so reducing the capability of Australian authorities in catching the perpetrators.

What's more, perpetrators that are knowingly generating or accessing illegal material, often use techniques such as encrypted protocols and secure networks that make it extremely difficult to assess or examine the material they are transferring.

Even so, the Task Force believes that while it is difficult to prevent those who deliberately generate or go in search of illegal material, much can be done to prevent 'inadvertent exposure' to this material. However, it will come at a cost in terms of a dollar cost for ISPs to implement relevant filters and the potential impacts on the ISP networks, depending on the type and extent of filtering required by the Government.

3.6 How Is Illegal Material Being Distributed?

To effectively answer the question on whether ISP filtering will be effective in reducing inadvertent exposure and whether it will reduce the accessibility of those who deliberately go in search of such material, it is important to know how such material is being distributed.

As an example, one study on ISP filtering of child pornography[3] found five primary means of distribution, being:

- WWW – 96% of offenders used this method;
- Internet Relay Chat (IRC) – 50% of offenders used this method;
- Peer to Peer (P2P) – 43% of offenders used P2P;
- Email – was used by 23% of offenders; and
- File Transfer Protocol (FTP) – was used by 12% of offenders.

While WWW was used by a large majority of offenders, it was found this was usually by inexperienced users of child pornography and that experienced offenders are more likely to use anonymous technologies.

IRC (often used by bot masters to control their bots) can be used with other programs such as mIRC that allows the user to share a section of their hard drive with other users. This allows perpetrators to create and distribute material amongst themselves in an anonymous way.

P2P file sharing software and FTP protocols allow users to transfer large files. FTP was reported to be used strongly by experienced and more highly ICT literate purveyors because it is perceived as being more secure. For instance, FTP servers being accessed are not normally advertised or may not be easily found, and even can be deliberately hidden; hence are unlikely to be stumbled upon by a casual Internet browser.

While this is just one study, it can still be useful to illustrate that filtering of HTTP ("world wide web") traffic alone is not going to solve the underlying problem or significantly impact those who deliberately produce, distribute or go in search of this material.

3.7 What Are the Criteria Behind the Black List?

Currently ACMA investigates complaints about online content and notifies approved filter vendors of overseas hosted content that is found to be prohibited. As part of an investigation, ACMA may apply to the Classification Board to have material classified and it must do so in the case of potentially prohibited material that is hosted or provided from Australia.

Classification decisions by the Classification Board can be reviewed by the Classification Review Board which, in turn, is subject to judicial review under the Administrative Decisions (Judicial Review) Act 1977. Similarly, ACMA take down notices, service cessation notices etc., are subject to judicial review under the Administrative Decisions (Judicial Review) Act 1977.

These processes aside, a significant issue is developing around the transparency of the ACMA black list and the criteria and the process for incorporating sites onto the black list, because of the secrecy of the list itself, following a purported leak of the black list to the media which found it incorporated a number of seemingly legitimate sites.[4]

While the Task Force agrees that the ACMA black list is not suitable for public dissemination, it is equally essential for the Government to establish greater transparency and accountability in the criteria and processes for incorporating sites onto the black list.

[3] *Ememan, M., "A Critical Study of ISP Filtering of Child Pornography", Goetborg University. http://is2.lse.ac.uk/asp/aspecis/20060154.pdf*
[4] *www.smh.com.au/articles/2009/03/19/1237054961100.html*

The Task Force believes that the Government should consider establishing an independent over sighting body and an annual auditing process for the black list to provide an appropriate oversight and ensure the highest public confidence in the black listing process; such a body can also act as an independent central authority to ensure greater transparency around the blacklist and to which appeals and other complaints about the black list can be directed.

4 Technical Issues and Filtering Techniques

Content restrictions can be achieved through essentially four main strategies:[5]

1. blocking of content using various techniques to block access to websites;
2. search result removal by ISPs to remove sites from search results;
3. taking down of web sites and content; and
4. self induced censorship by content producers.

The Federal Government's ISP filtering proposal is essentially aimed at point 1 – the use of technical measures to block access to websites.

The three commonly used techniques to block access to internet sites are:

- packet blocking/dropping at the IP level;
- DNS poisoning; and
- content filtering operating at higher protocol layers - URL blocking using a proxy.

These techniques are primarily used where the authority wanting to block access does not have direct jurisdictional control or does not have control over the web sites. This is the case for the ISP filtering pilot since the sites to be blocked mostly originate from outside of Australia.

These methods use black lists, key words, phrases and signatures, and dynamic content analysis techniques.

4.1 IP Blocking Using IP Packet Filtering/Blocking

Packet filtering is an IP blocking technique whereby the routers or other equipment looks at the headers of each data packet that passes through. There are two types of packet filtering called layer 3 filtering and layer 4 filtering. These differ in the granularity with which they filter and resource consumption on the filtering device.[6]

Layer 3 filtering examines the IP addresses for information on where the data has come from (source) and where it is being sent (destination). It allows development of filtering rules around the destination and source of the data, so that out bound and in bound traffic from a particular site can be blocked. Layer 3 filtering results in all communication traffic to the blocked host being denied including web pages, email, chat, Usenet or any other type of e-communication and uses relatively little resources on the network devices performing the filtering.

[5] *http://opennet.net/about-filtering*
[6] *Dorneseif, M., "Government Mandated Blocking of Foreign Web Content", 2004, http://md.hudora.de/publications/200306-gi-blocking/200306-gi-blocking.pdf*

Layer 4 filtering is similar. It can filter using the destination port number allowing blocking on the basis of a particular service. So while web access to a particular site can be blocked, email, chat and other services run from the blocked host server can still be accessed.

Packet blocking can be implemented at several places within the architecture – for example, at the ISP level or at the backbone service provider and international gateways. Implementation can be carried out in a number of ways including using firewalls (with a list of addresses to be blocked) that are deployed on all connections. Alternatively existing network routing protocols can be employed to redirect traffic for the relevant addresses to be blocked to a "black hole" that discards the packets.

The main problem with packet blocking is the collateral damage that it causes. All of the web content on a particular IP address that is blocked will become inaccessible. This leads to over blocking of legitimate sites and domains using the same IP address.

An example would be a university web site or a commercial website hosting service provider which has many domain names using the one IP address. It is also relatively simple for the same content to be made available through a different IP address (or many different IP addresses) attached to the same server, thus evading the filter.

4.2 Domain Name System Poisoning

The website names, host names, and domain names we are all familiar with are a proxy for numerical IP addresses which are usually written as a series of numbers separated by dots. So when a particular domain name is requested in an Internet search from the Domain Name System (DNS) server hosted by the ISP, this request is then forwarded to a parent DNS server and is eventually converted to the particular IP address that corresponds to that name. The answer is then returned to the original requesting application.

DNS poisoning can take a number of forms:

- name hijacking – a DNS server does not return the IP address of the domain name requested by the domain name server but rather a different IP address of some other web page;
- name "subversion" – a DNS server returns an illegal IP address, leading to a 'could not connect' or similar message from the application when it attempts to access a blocked domain name;
- silence – attempts to access a blocked IP address will result in a DNS server refusing to provide any answer and the original query times out and fails after a period of time.;
- provoked server failure – a DNS server responds immediately with an error code indicating the lookup did not work, so the application fails to attempt to connect to the site after a much shorter period of time.

DNS poisoning can also cause over blocking in that all content within the blocked domain will be un-accessible. So it would not be appropriate for blocking content that is hosted on a site like myspace.com which would also block many legitimate sites and pages. However, with DNS poisoning, the over blocking that occurs is a little different to that which occurs for IP address blocking in that it does not extend to blocking other domains that are hosted on the same machine.

DNS poisoning can result in under blocking because a URL containing an IP address rather than a host name would not be affected.

4.3 URL Blocking Using Proxies

Content filters specifically targeting "world wide web" content can not only be used to block entire web sites but can also be used to block very specific items such as a particular web page or even a single image. If the URL requested is one from a list to be blocked, then the corresponding content is not made available.

There are two main ways of providing content filtering, being web proxies and packet- level firewalls.

Web Proxies

HTTP is the protocol used to transfer web pages from the a web name server to the user's browser and a web proxy mediates the incoming HTTP request and applies filtering or blocking mechanisms. The HTTP proxy can filter out URLs (the ACMA black list for example) by ostensibly mediating communications between the ISP provider and the user. It can block on an individual URL basis because it has full control of the data stream. HTTP proxies can be used to filter out certain specified parts of communications, for example, a specific, nominated part of a web page or even a specific paragraph in a document. All major web browsers support the use of HTTP proxies as they can be made invisible to the web browser.

The two major techniques using HTTP proxies are index and analysis based filtering.

Index based filtering uses lists such as a black list or white lists. Analysis based filtering uses key words, phrases and content profiles (such as the proportion of a web page devoted to images).

HTTP proxies can include image analysis filtering which can examine skin tone to determine if an image contains, for example, nudity; however, image analysis technology is not always particularly accurate and can lead to over and under blocking. For example, image analysis cannot determine the activities being engaged within the image – it may determine that that there is a significant amount of exposed skin but cannot determine if they are they G, PG, R, X or illegal activities so potentially leading to over-blocking or under-blocking.

Other examples of analysis based techniques include file type filtering and reputation filtering based on historical or recurrent behaviours from particular sources.

Packet Level Firewalls

A packet level firewall can implement content filtering using stateful packet inspection or deep packet inspection techniques. Stateful packet inspection typically checks the header portion of a packet, whereas deep packet inspection (DPI) has the ability to look at headers and data protocol structures as well as the actual payload of the message.

The DPI techniques can identify and classify the traffic based on a signature database that includes information extracted from the data part of a packet, allowing finer control than classification based only on header information. A classified packet can be redirected, marked/tagged or blocked. DPI can identify packet flows rather than packet-by-packet analysis, allowing control actions based on accumulated flow information. This will require storing states between packets in flows as well as caching packet contents and additional computational processing.

4.4 Hybrid System

The BT CleanFeed[7] scheme deployed in the UK by British Telecom is a hybrid system and consists of two filtering stages.

The first is a packet dropping mechanism, except the packets are not discarded but are instead routed to a second stage content filtering system.

[7] *Official name for the Project BT Anti-Child-Abuse Initiative*

The first stage examines all the traffic flowing from customers. It checks the destination port number and IP addresses within packets. If the destination is deemed to be harmless, then the traffic is sent to the destination in the normal manner. If the destination is deemed to be suspicious, then the traffic is sent to a second stage filter, which is a web proxy that understands HTTP requests.

This hybrid scheme has some advantages over simple IP address blocking in that it suffers less from over blocking because the second stage web proxy can be as selective as required.

4.5 Content and Site Labelling Based Filtering

Web-based content can be voluntarily labelled and categorised using meta-data embedded within the website content. The Task Force believes that site and content labelling can help to improve the effectiveness of filtering and allow greater control over the content that is being filtered, provided that the label accurately reflects the content and is undertaken by trusted authorities. Legislation making it mandatory to label content accurately would be analogous to existing classification laws and, in respect of all electronic communications, constitutionally well within the Federal Parliament's purview. This raises issues such as who determines whether the content is labelled accurately and how it is being done, as well as potentially requiring additional categories for classification if the mechanism is to be globally useful.

Dublin Core is a well known criteria for labelling web sites and allows web sites to be labelled with the title, creator, subject, description, publisher, contributors, format, language and publication rights.

Content filtering using labels can be done at the ISP level (by HTTP proxy), at the user level (filter programs loaded onto the PC or laptop), at the host web site or can be undertaken by web crawlers (dynamic classification agents that assess and rate content as it produced).

There are number of classification labelling organisations that rate content with minors in mind and these include:

- Internet Content Rating Association RSACi rating system which is voluntary and sites are rated using four categories – language, nudity, sex, violence – with a zero to four sub rating for each category. RSACi is used by a Internet Explorer, CyberPatrol and CompuServe amongst others;
- SafeSurf has categories of age range, profanity, sexual themes, nudity, violence, intolerance, glorifying drug use, other adult themes, gambling. Each category is broken down into multiple subcategories. SafeSurf is used by Internet Explorer and Netscape Navigator;
- Adequate.com is based on the US Television Rating System and includes ratings for children, general audience, parental guidance, parents strongly cautioned, mature audiences only and the like;
- Entertainment Software Rating Board (ESRB) provides rating for web sites and games with categories for early childhood, kids to adults; teens, mature, adults only. Categories are subdivided for the type of content.

4.6 Other Methods of Content Control

In addition to technical filtering mechanism, security mechanisms are available that require user identification before access to data or particular sites is permitted. These methods include:

- encryption – requiring the use of a digital key before access is granted;
- login names and passwords;
- age verification;
- biometrics – use of finger scanning, voice print and other biometric identifiers;
- activity monitoring programs that record the use of certain words, phrases or URLs visited; and
- caching techniques to allow review and analysis of content at a later time.

5 Issues and Design Choices for Filtering

5.1 Content Classification Issues

The fundamental issue with content classification is trusting the labeller to accurately label the content they are creating or hosting.

Even more fundamental is determination of what is an accurate label for a web site because different users, producers and communities have different values and interpretations on what is considered appropriate, offensive and the like.

Labelling and rating are inherently biased and will reflect the ideology of those who create the labels. Further, while a content code of conduct could work well with Australian content producers using relatively static sites, it is unlikely to be effective for offshore content producers or those who deliberately produce illegal material. Nevertheless, penalties for false labelling could be made high enough to deter even those who believe they will not be caught.

Every Internet user is potentially also a content producer, so there may be just too many content producers to influence and/or monitor. Also labelling may not work well with content produced on the fly or for user generated content produced on social networking sites.

This approach would rely on a critical mass of content to be labelled in order to work well and sites and content will require active updating and maintenance to take into account of new technologies and new material.

Even so, taking these challenges into account, there could be a useful role for at least voluntary labelling or rating systems in the development of an overall Internet content management strategy involving the users, communities and the reputation and past behaviours of content producers.

5.2 Criteria Enforcement

The Task Force believes that there should be consistency in the criteria used to classify all media and entertainment content regardless of whether it is available online or offline. In addition, the particular reasons for content being classified as RC (refused classification) or illegal (and so included on the ACMA black list) should be transparent.

For example, as the Australian rating classification for computer games does not have X or R classifications, some computer games end up with an RC classification and so would not be available online from Australian game sites.

A key issue with criteria enforcement is that it does not apply to sites hosted in foreign jurisdictions. ACMA can add these sites to its blacklist if they are found to host RC or illegal material. Also, adult content sites need to be placed behind adult verification systems, although adult verifications systems do not work as well for non commercial content providers.

If filtering is to be acceptable and credible, then we need to have clear and transparent criteria to classify content that applies equally across all media including both online and offline content.

Transparency and credibility should include an independent oversight, a system of checks and balances that incorporates a system of appeals and an independent auditing process.

5.3 What Traffic to Filter

Data is available from the Internet in many different access modes and while web protocol (HTTP) is the most common, non HTTP protocols such as email, discussion groups, chat, news servers and IM, amongst others, are widely used [8]

Filtering of peer to peer traffic, Internet chat rooms and instant messaging from social networking sites remains difficult for ISPs; yet these are some of the primary methods that seasoned purveyors of illegal material use to communicate and exchange files.

Filtering non HTTP protocol applications is generally undertaken by filters built specifically for these applications and can be stand alone or incorporated in larger, more general filtering software. Blocking peer to peer, chat and IM is most effectively done at the organisational or home level by either blocking access to these protocols completely or using blocking based signatures.

Email filtering usually focuses on spam and content of the email by scanning the headers of incoming messages although more intrusive filters can re-route email to password protected folders for manual inspection.

Filtering of news servers such as UseNet, is usually achieved by developing blacklist of unacceptable news groups or using keyword filtering and has to be undertaken at the news group level. The same approach applies to discussion groups, chat rooms and IM services. Access can be controlled by only allowing access to a list of acceptable participants (user id and password) by the host and on the user computer.

File transfer protocols can only be stopped by controlling the sites to which access is allowed to an individual or the site host can control who may access their site using a user id and password. File transfer protocols cannot be readily filtered at the ISP level.

ISP level filtering is not effective for these protocols unless they offer a specific user service that blocks all access to sites using non HTTP protocols using firewalls and proxy servers.

5.4 Encrypted Traffic

On the one hand, encrypted traffic poses a substantial challenge for filtering and there are limits to the type of filtering that can be achieved for encrypted content because encrypted content cannot be easily decrypted.

On the other hand, secure communications are an absolute necessity for growth and trust of the digital economy as a means of protecting financial and other sensitive information from cyber criminals.

SSL, for example, is a security protocol which is a de-facto international standard that allows the exchange of encrypted sensitive information over the Internet. It is used by banks and other businesses to conduct e-commerce transactions and payments and its high level of security induces trust in consumers and businesses to conduct business online.

However, the issue remains that security protocols are beyond the reach of ISP filtering and present a means for those peddling illegal material or Internet fraud to avoid detection.

[8] *Shepherd, M., & Watters, C., "Content Filtering Technologies and Internet Service Providers", 2000 http://users.cs.dal.ca/~shepherd/filtering/ISPweb.htm*

5.5 Filtering and Network Architecture

Internet filtering can occur at a number of places within the network architecture:

- Internet backbone;
- ISPs;
- Institutional level (companies, government, schools, cyber cafes etc); and/or
- Individual computers.

Success of filtering depends on several factors such as where the filters are placed in the Internet architecture, the types of traffic that are to be filtered and how, responsibility for generating, maintaining and implementing lists and profiles used in the filtering process, management of security risks, legal jurisdiction and authority, and the level of individual control required. Filtering at the institutional and individual level also relies on a certain level of ICT literacy and maintenance of filtering software.

In terms of the best place to filter in the ISP network architecture, small ISPs with 1000 to 5000 users would more likely have a relatively simple core network, with a small number of upstream data connections of a maximum of 20-50 Mb/s. In this situation there could be enough performance in the filtering solution to ensure that the filter would not create a bottleneck and significantly affect the performance of the ISP. However, consideration needs to be given to ensuring adequate redundancy so that failure of the filter will not cause service disruption.

ISPs operating with faster backbones would require higher capacity filters if they wish to filter all traffic at their maximum speed. In this case, content filtering may be better located at the subscriber-edge of the network where slower upstream speeds exist before they are aggregated into faster pipes. This implies a requirement for larger numbers of filtering units within a single network.

In general, the larger the ISP (in terms of users and speed) will require more sophisticated and efficient filters, which can be more costly as a unit price. However with large ISPs with a greater number of users, the determining factor may be the cost/user which could be more relevant. Most ISPs have between 3 to 30 upstream providers, often in different states. So there can exist, within and on the ISP backbone, a very large number of interconnections with other providers, online providers and ISPs' PoPs (Points of Presence), with each type of connection having very different connecting speeds.

If filtering is done at the ISP's core, then it is necessary to ensure that filter placement does not lead to a single point of failure. The risk can be remedied with the filtering solution equipment to some extent through load balancing and server clustering, with the filtering functions residing at multiple servers and the load balancer distributing the filtering request across these multiple servers. Using a single load balancing device itself would introduce another single point of failure; this risk needs to be mitigated using redundant load balancing devices, leading to a complex network topology that may only be feasible for large ISPs.

While filtering at the PoP (which is located at the periphery of the ISPs network) would be another option, for a large ISP, this would require installation of dozens of gatekeeper boxes, which can make this option somewhat impractical.

Another concern with filtering at the ISP core is the potential for it to necessitate changes in the ISP's routing (depending on their existing routing, traffic flows, transit & peering arrangements and international gateways), and performance impacts as traffic is forced into sub-optimal paths in order to pass by or through the filtering systems. This could necessitate costly upgrades to backhaul capacity and increased transit costs.

It also needs to be taken into account that many ISPs (smaller ISPs, in particular) purchase their upstream bandwidth from other ISPs that may already be performing content filtering.

In terms of filtering illegal content that is coming from outside of Australia, the intuitive place to deploy the filter would be in the backbone service provider at the international gateways, however there is no clear catalogue of international gateways, and an ISP can connect directly with an international network through a variety of undersea fibre or satellite-based transmission systems.

Also international companies can have Internet based networks with a number of servers distributed through out the world. There can be dedicated communications between two such servers which are permanently connected and may not use the service of a local ISP or a backbone service provider.

5.6 Implementation Issues

There are a number of issues that need to be addressed when it comes to the implementation of ISP based filtering, including:

- sites can easily be renamed and so the names will not match the black or white list;
- language translation (often automated) often produces mistakes and so international sites may not be filtered effectively;
- lists must contain domain names as well as IP addresses to be highly effective;
- not all applications work well with a proxy server and so the performance of the ISP can degrade;
- push technologies (such as RSS) often bypass the proxy server and deliver content directly to the user so circumventing the filtering process;
- not all users access the Internet via an ISP;
- many sites have mirrors and multiple URLs and if these are not included in the black list then the filtering process can be circumvented; and
- proxies can degrade ISP performance particularly during periods of high traffic – they become bottlenecks and can reduce Internet speeds;
- mandating or architecting a network so that all packets pass by a filtering point can create performance problems, duplicated traffic paths and may increase the bandwidth costs for ISPs.

5.7 Addressing P2P and BitTorrent

BitTorrent is a popular P2P data sharing protocol used to transfer large files and uses the other clients who are also downloading the file to effectively act as servers to one another, while simultaneously uploading the parts of the file received to others requesting the file.

When a user selects a file to download, several connections will be made to receive "slices" of the file that combine to create the entire file. While the slices are being downloaded, they are also being uploaded to anyone else that needs the parts being received. Once the entire file is received it is considered polite to keep your client connected to act as a seed: that is, a source that has the entire file available.

In this way BitTorrent relieves the burden from a server, but more significantly, makes it possible for anyone to disseminate a file quickly and easily without requiring expensive servers or a central infrastructure for distribution. If the demand is there, the file will spread from highly distributed points.

In recent times there have been a number of reports of ISPs attempting to contain or limit BitTorrent traffic because it is a prevalent means of distributing film and music illegally, in addition to being used to distribute illegal material.

ISPs can restrict the aggregate volume of BitTorrent traffic by throttling BitTorrent transfers, limiting the available bandwidth, or by seeding the transfer with reset packets which essentially cut off the individual connections between hosts.

While ISP actions against BitTorrent traffic can effectively prevent distribution of large files of illegal material, they will also block legitimate file transfers using P2P transfers such as Internet TV and legitimate video downloads.

Furthermore, encryption of torrents and headers can limit the throttling ISPs are able to perform.

5.8 Circumventing Filters

There are a number of ways of circumventing filters based on whether index or analysis based filtering is being used. While blocking and filtering mechanisms are not easily breached, it is feasible and does happen, although it requires a reasonable level of sophistication and knowledge of ICT and a determination to do so on the part of the content provider and/or the user.

Methods of working around filters are listed below..

Mirroring – mainly used for static websites that don't have interactive elements. Mirrors are duplicate web sites used to reduce the traffic load on servers hosting high traffic web sites. Users seeking to access a blocked site can access the mirror instead and so circumventing the filter; however, this requires users to be informed of the URL or IP address of the mirror.

Additional domain names – content providers often have multiple domain names and multiple URLs that point to the same IP address where the blocked content resides. While IP based filtering is immune to this technique, it is effective against DNS poisoning and HTTP proxy filtering. Additional domain names are often published on the site being blocked.

Changing IP addresses – content providers can change IP addresses on a regular basis and avoid IP address filtering processes. So filtering and black lists based on IP address blocking or even hybrid solutions that use IP based filtering in their first stage will suffer from the problem of dynamically changing IP addresses.

Change of port – changing port numbers used by web server software is used to circumvent layer 4 IP filtering. However, like mirroring and additional domain names, it often requires users to be informed of the new port number through a separate means of communication.

Anonymisers – Users can configure their web browsers to seek content through a proxy (anonymiser) server to access sites rather than accessing the site directly. The ISP filtering software will see only the URL or IP address of the anonymiser and not the URL or the IP address of the web site being requested. (We note that the Australian Federal Privacy Commissioner encourages users to employ this technique: www.privacy.gov.au/topics/technologies/security#6)

Translators and encryptors – these sites are used to translate web page text into different languages or to encrypt text. Typically a user will enter the URL of the website to be translated or encrypted and the translation software will present the translated information within its own web page thereby hiding the URL of the site being blocked from the ISP filtering software.

Encryption is also used with BitTorrent traffic to get around ISP shaping or blocking with Reset packets.

Generally, ISP level filtering of sites that contain dynamic content, such as blogs, chat rooms or other interpersonal communications, is not accurately achieved and there remains a significant challenge around effective blocking of real time content.

In addition to the challenges associated with those who deliberately aim to circumvent filters, the protection of the filtering system itself from malicious attacks and the secrecy of black lists (their URLs and IP addresses) remain paramount if filtering processes based on them are to be effective.

5.9 Over Blocking and Under Blocking

Filtering technologies need to be implemented with care as issues surrounding filtering can be complex, leading to over blocking or under blocking.

Web sites are hosted on web servers that typically have a single IP address. So sites which are hosted on the same server as one which is on a black list (and so consequently will have the same IP address as the site which is on a black list) will also be blocked.

So techniques that filter solely on the basis of IP address will inherently over block sites and exclude many legitimate as well as illegitimate sites, raising concerns and creating significant problems for businesses that rely on Internet commerce and web traffic for their business model.

Analysis based filtering usually depends on two factors namely the amount of resources available to do the analysis and the how long does it take to do the analysis.

Static and dynamic filtering techniques are mostly efficient for text analysis only. Current technologies are not able to accurately target specific categories of content found within websites and cannot effectively analyse dynamic real time content such as found on news groups, chat rooms and instant messaging.

Pictorial, audio and video files are also difficult to accurately analyse (especially in real time) leading to under blocking or over blocking. Similarly, automated text analysis can result in over blocking and under blocking of sites depending on the context and use of certain key words and phrases.

Both over blocking and under blocking have clear implications for business and end users. While over blocking can lead to loss of business, under blocking can lead to a loss of confidence particularly where inappropriate material is viewed by minors.

6 Other Issues

6.1 Improved Control Over Domain Name Registration

ICANN is the Internet Corporation for Assigning Names and Numbers. ICANN is an international body that coordinates the unique Internet identifiers/addresses to insure there is one integrated system that forms the global Internet. Internet addresses have to be unique so that any device on the Internet is able to find any other device that is connected to it.

ICANN coordinates the allocation and assignment of three sets of unique identifiers used by sites on the Internet:

- domain names (DNS);
- Internet Protocol (IP) address and Autonomous System (AS) numbers; and
- protocol port and parameter numbers.

ICANN also coordinates the operation and evolution of the DNS root name server system and coordinates policy development related to these functions.

A key role of ICANN is to accredit Registrars who have responsibility for allocation of domain names and IP addresses. Ultimately, it is the Registrars who have responsibility for allowing web sites to get onto the Internet in the first place by providing them with DNS and IP addresses. Greater professionalism and surveillance amongst Registrars could have a significant impact on reducing access to the Internet by those who produce illegal material.

ICANN, through its accreditation and Registrar Disqualification Procedures can, in turn, impact the behaviour of Registrars who allocate DNS and IP addresses to purveyors of child pornography or other illegal material.

Australia is well represented on ICANN and it provides a strong and legitimate route for the Australian Government to work with ICANN to improve the effectiveness of the Registrar accreditation and disqualification processes to ensure that sites containing illegal material are prevented from being allocated DNS and IP addresses.

The Task Force believes that the Government should explore this pathway as an integrated part of its cyber safety strategy to improve the security and processes around domain name registration. It will be effective not only against illegal sites but also against sites being used for cyber crime such as phishing and other fraudulent activities.

One possible approach is to encourage ICANN not to approve any new domain names that do not include eligibility criteria similar to those required by auDA in the .au space. At a minimum this should require accurate WhoIs information and independently verified identity and address details for domain name registrants.

6.2 ISP Filtering Trial

The current ISP filtering trial being conducted by the Federal Government will provide an insight into how ISPs can best filter traffic, the type of traffic they can effectively filter, types of filtering and the potential impacts of filtering.

The Task Force believes it is appropriate to wait until the results of the trial are available before commenting on its efficacy, its scalability, methodology and potential impacts on services provided by the ISPs.

Outside of this, the Task Force considered that effective filtering will require all ISPs to be part of the process and will need to satisfy at least specified minimum Government requirements. This will likely involve specifying a series of metrics that ISP filtering methodologies should meet and perhaps helping to set a national filtering standard.

Filtering metrics should be transparent and results should be published annually either by the government body given responsibility for implementing and oversighting ISP filtering or by the ISPs themselves as part of their annual reporting processes.

The effectiveness of any ISP filtering process introduced and its impacts on Internet services should be regularly reviewed to ensure that any impacts remain within minimum specified parameters.

Higher levels of filtering could be available to individual users, depending on their requirements. For example, schools, libraries, public net cafes and homes may want all material above PG rating to be filtered from their feed.

The Task Force believes that ISPs should provide options for various levels of filtering so that businesses and end users can choose the level of filtering that suits them best (above that which is mandatory).

7 Awareness and Education of Users

The Task Force believes that education of end users must form an integral part of any effective ISP filtering regulatory regime.

The Task Force applauds recent cyber safety initiatives by ACMA which recently launched www.cybersmart.gov.au to provide up to date information for parents and activities specifically designed for children. In addition, ACMA's Outreach program has been expanded to provide general cyber safety awareness for teachers, parents and students on issues and strategies to minimise online risks. The Task Force applauds this initiative.

While many people understand how to use devices and systems, they do not fully understand how the Internet works, how e-security systems work and where the vulnerabilities in their systems occur.

The Task Force believes the community needs to have a much better overall understanding and working knowledge of the factors associated with threats, computer and network vulnerabilities and how countermeasures work, their limitations and what they can do to adequately protect themselves.

The reality is that, regardless of the level of ISP filtering or filtering system used, it will still need to be combined with appropriate parental guidance and supervision to ensure effective online child protection.

To this end, ISPs and equipment vendors can help people customise the level of control that best suits their needs. This is likely to involve PC based security systems that will compliment any mandatory ISP filtering regime.

However, even with the best ISP level and PC based security systems and education programs in place, it is unrealistic to expect that all illegal material will be caught. A set and forget solution simply does not exist and filters do not replace adequate parental supervision.

People will still open suspicious emails and click on malicious code. Empowering people with adequate knowledge of e-security threats and how they can take responsibility for reducing those threats remains the best defence against Internet security threats. In this context, there is also a role for technology developers and providers to provide more relevant information and to increase transparency on what users can reasonably expect in terms of security in different ICT technologies, so that users can make better informed decisions.

8 The Way Forward

Whilst the broader issues surrounding e-security and cybercrime will be the subject of a subsequent report, at this stage the Task Force makes the following points as a means of progressing the debate on ISP filtering:

1. The Task Force believes that while ISP filtering techniques can be useful in helping to reduce inadvertent exposure to child pornography or other illegal material, filtering alone is unlikely to solve the underlying problem or significantly impact those who deliberately produce, distribute or go in search of this material.

2. Although ISP-level filtering can reduce the likelihood of inadvertent exposure, it cannot completely prevent inadvertent exposure as it is only feasible to filter out those sites and pages that have been identified at an earlier time (due to the dynamic nature of the Internet, it is possible that prohibited content which was previously identified could have moved to another location by the time the original location is added to the black list).

3. The technical assessment of the Task Force is that there is no single mechanism that can filter out or block illegal material on the Internet accurately 100 per cent of the time. A multi faceted approach is needed to address this issue that will involve filtering technologies at the ISP, user and enterprise levels, increased professionalism and tighter controls around domain name registration, education at all levels of society and parental oversight.

4. The technical assessment of the Task Force is that there is no technological substitute for appropriate education and parental supervision of young people who are using the Internet. Education and oversight remains the best method of ensuring that children (and other end users) are aware of online safety and are not viewing inappropriate material or engaging in inappropriate behaviour online.

5. Based on findings of points 1 to 3, the Task Force believes that the policy objective for filtering should be clearly articulated based on whether it is:
   - to avoid inadvertent or unintended viewing of RC or illegal content while surfing the web;
   - to prevent, detect, block and prosecute deliverable access, publication or circulation of RC or illegal content;
   - to deter both inadvertent and/or deliberate interaction with a wider ambit of RC, illegal or prohibited material using any method of Internet access

   It is vitally important to understand the policy for filtering, as the means of realising the filtering to achieve the policy will be quite different in each case.

6. The objectives of any ISP filtering program should be clearly defined including performance standards, clarity around the definition of material to be filtered, reporting processes and filtering processes to be used.

7. The type of traffic to be filtered must be clearly specified. There are many different access modes and protocols to access the Internet including wireless, email, chat, IM and peer to peer protocols. The objectives of the filtering program should clearly state what protocols and traffic are to be filtered.

8. Different filtering processes achieve different results in terms of impacts on speeds, resource use and accuracy of filtering (over blocking and under blocking). In mandating or regulating for ISP level filtering, the government should develop a set of minimum standards that are to be achieved and upon which the efficacy of filtering can be measured.

9. The Task Force believes considerable thought and planning needs to be given to the location of filters within the ISP architecture (depending on the size, speed and level of redundancy) to avoid multiple filtering of feeds, to minimize impact due to ISP failures and to optimize performance due to filter operations.

10. As part of any ISP filtering program developed, the Task Force believes it would be worthwhile to consider implementing a national, voluntary content rating system so that content providers can voluntarily rate the material on their sites. Any rating scheme used should be standardised and should be easy to use so that content developers can self rate the content they develop and verifiers can check the match between the rating and the content.

11. There should be consistency in the criteria and the rules used to classify material online and offline and to make them available.

12. The Government should establish clear, unambiguous guidelines on sites and material that are to be included on the ACMA black list. In addition, there should be an independent and transparent auditing process for the black list and an ability for complaints for those sites included on the black list to be lodged and assessed in a timely manner.

13. The Government should strongly encourage ISPs to provide products that allow users to select/customise their preferred level of filtering (above that which is mandatory).

14. The community needs to better understand the factors associated with threats, computer and network vulnerabilities and how countermeasures work, and what they can do to adequately protect themselves.

15. The Government should actively work with ICANN to improve professionalism and integrate part of its cyber safety strategy to improve the security and processes around domain name registration. Australia is well represented on ICANN and it provides a strong and legitimate route for the Australian Government to work with ICANN to improve the effectiveness of the Registrar accreditation and disqualification processes to ensure that sites containing illegal material are prevented from continuing to be permitted to use DNS and IP addresses.

Task Force Members

Members of the ACS Task Force on E-Security are listed below.

Prof Vijay Varadharajan, Microsoft Chair in Innovation,
Director of Information and Networked Systems Security Research, Macquarie University
Mr Philip Argy, CEO, ArgyStar, Immediate Past President of ACS
Dr Paul Brooks, Director, ISOC-AU, Managing Director, Layer 10
Mr Graham Ingram, General Manager, AusCERT
Mr Alastair MacGibbon, Internet Safety Institute
Ms Holly Raiche, CEO, ISOC-AU