Procedural Protections for Internet Expression

Dawn C. Nunziato, B.A., M.A., J.D., University of Virginia,
Professor of Law,
The George Washington University Law School

Forthcoming, International Review of Law, Computers, and Technology (2013)

Keywords

Internet censorship, filtering, free speech

Abstract

Early hopes that the Internet would create greater opportunities for freedom of expression have been dampened by the reality that it is increasingly a tool of state censorship and repression. Popular and academic attention alike have generally focused on the repressive tactics of authoritarian and semi-authoritarian states like North Korea, the People's Republic of China, and Saudi Arabia. However, Western democracies have also moved towards greater government control over their own citizens' Internet use. This paper explores the impact of the United Kingdom's Internet Watch Foundation and the Australian government's proposed ISP filtering on freedom of expression on the Internet in those countries. The paper then analyzes how these policies conflict both with traditional Anglo-American jurisprudence regarding prior restraints and with the principles set forth in the United Nations' International Covenant on Civil and Political Rights.

I. Introduction

It was long assumed that the Internet would bring about unprecedented opportunities for free expression. In recent years, however, the Internet has increasingly become a tool of censorship, as scores of countries around the world have imposed nationwide filtering regimes to block their citizens' access to various types of Internet speech that these countries deem harmful. Instead of trending toward greater freedom, the Internet is now trending toward greater censorship and control, as many countries – including many democracies -- are seeking to exercise greater and greater control over and through this powerful medium. Today, more than forty countries – in addition to the usual suspects like China, Saudi Arabia, and North Korea – have implemented nationwide technical filtering of speech on the Internet, and this number is growing (see Murdoch and Anderson [2008] for a general discussion of how nation-by-nation filtering is implemented). Among democracies, the United Kingdom and Australia are leading

the way in restricting access to harmful content. The U.K. established a comprehensive system for filtering and blocking harmful Internet content, through the mechanisms set in motion by an entity known as the Internet Watch Foundation. Australia is also taking a leading role in filtering a variety of harmful Internet content. The Australian government announced that it intends to introduce 'mandatory ISP-level filtering' of certain content and that it will require all Australian ISPs to block websites that contain content that deals with 'matters of sex, drug misuse or addiction, crime, cruelty, violence or revolting or abhorrent phenomena in such a way that they offend against the standards of morality, decency and propriety generally accepted by reasonable adults' (Deibert, Palfrey, Rohozinski, and Zittrain 2010). Nationwide Internet filtering has become a powerful tool for many governments – dictatorships and democracies alike – to control the content that their citizens are able to access. Given the extent and increasing effectiveness of efforts to censor Internet speech throughout the world, protectors of Internet free speech can no longer rest comfortably on the assurance given by Internet pioneer John Gilmore in a *Time International* article two decades ago (December 6, 1993) that 'the Net interprets censorship as damage and routes around it.' Although free speech advocates broadly denounce such censorship, it is likely that many countries – having seized upon these powerful tools of control – will continue to restrict Internet content to prohibit their citizens from accessing speech that they deem to be harmful. The question is, what can and should be done to reverse this trend and to preserve the Internet as a forum for free expression?

It is commonly understood – and understandable – that different countries around the world implement different definitions of what speech is protected and what speech is unprotected, online as well as offline. Given, for example, Europe's horrific experiences with the Holocaust, it is not surprising that some European countries would consider racial and religious hate speech to be unprotected. While there is substantial divergence on the *substantive* contours of free speech protections – which categories of speech are protected and which are not – there is a developing convergence among nations regarding *procedural* protections for speech (see, e.g., Krotozynski [2006]; Sedler [2006]; Farrior [1996]). Such procedural protections are inherent in and flow from widely-shared concepts of fundamental due process, and have been embodied in the widely-adopted International Covenant on Civil and Political Rights (the ICCPR). A second important source of procedural protections for Internet speech arises from the Anglo-American tradition's hostility toward prior restraints on speech and (relative) preference for subsequent punishment as a means of restricting expression. Over the past four hundred years, Anglo-American jurisprudence has developed a presumption against the legality of any prior censorship or prior restraints on expression and has imposed a set of procedural safeguards that must be in place before any system of prior restraint can be legally imposed. The nationwide filtering systems that have become more pervasive in the past few years embody illegal prior restraints on speech – restrictions on speech imposed prior to a judicial determination of the speech's illegality that fail to accord these important procedural protections for speech embodied in the International Covenant on Civil and Political Rights and in Anglo-American free speech

jurisprudence. Such nationwide filtering systems should either be jettisoned or at least revised to as to accord these fundamental procedural protections on speech.

In this Article, I focus in Part II on the nationwide filtering systems established by two liberal democracies within the Anglo-American legal tradition – the United Kingdom and Australia. In Part III, I analyze the procedural protections on free speech that have been articulated both under the International Covenant on Civil and Political Rights as construed by the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, and within the Anglo-American legal tradition (especially with respect to prior restraints on speech). In Part IV, I compare the procedural protections provided under the U.K. and Australia's Internet filtering systems with the protections required within Anglo-American jurisprudence and the ICCPR, and find these nationwide filtering systems to be lacking. I propose modifications to these systems and suggest that certain procedural safeguards be implemented within any nationwide system of Internet filtering – especially those that are imposed within liberal democracies.

II. Case Study 1: The United Kingdom's Nationwide Filtering System

> *Journey with us to a state where an unaccountable panel of censors vets 95 per cent of citizens' domestic internet connections. The content coming into each home is checked against a mysterious blacklist by a group overseen by nobody, which keeps secret the list of censored URLs not just from citizens, but from internet service providers themselves. And until recently, few in that country even knew the body existed. Are we in China? Iran? Saudi Arabia? No - the United Kingdom . . . .* (Wired Magazine, *May 20, 2009)*

One liberal democracy that has had extensive experience with nationwide Internet filtering is the United Kingdom. The UK has implemented a nationwide filtering system that now affects the vast majority -- over 98% -- of Internet subscribers in the UK. The United Kingdom's Internet service providers, at the insistence of the government, implement an Internet filtering system to block access to websites containing certain types of content that have been deemed "potentially illegal" by an organization known as the Internet Watch Foundation. Below I describe the evolution of the Internet Watch Foundation, the development of the UK's nationwide filtering system, and some of the difficulties posed by this system.

Back in 1996, as concerns arose about illegal content being hosted and facilitated by ISPs, the United Kingdom's Department of Trade and Industry facilitated discussions between the Metropolitan Police, the Home Office, and a group of ISPs, with an eye toward addressing these concerns. These discussions resulted in an agreement in which a nonprofit, charitable entity known as the Internet Watch Foundation – a private, charitable organization initially charged with policing the Internet for child pornography – was formed. As its website notes, the IWF was created 'to fulfill an independent role in receiving, assessing and tracing public complaints about child sexual abuse content on the internet and to support the development of website rating systems.' In 1999, after three years of the IWF's operation, the U.K. government

and its Department of Trade and Industry evaluated and ultimately endorsed the operations of the IWF. The IWF is currently responsible for policing and facilitating the filtering or blocking of at least two types of content: (1) indecent images of children under 18 hosted anywhere in the world (child pornography); (2) criminally obscene content ("extreme pornography") hosted in the UK. or uploaded by someone in the UK. The IWF monitors and polices the Internet's content by operating a hotline reporting system, through which UK. Internet users alert the IWF to such potentially illegal content that they have come across on the Internet. The IWF then employs a handful of analysts trained by police to review flagged websites to analyze whether they contain 'potentially illegal content.' If an IWF analyst determines that a website is 'potentially illegal,' she will include the URL for that website on the IWF blacklist. The blacklist, which contains between 500 and 800 sites at any given time and is updated twice daily, is maintained in secret. Approximately 98% of all domestic broadband connections are filtered in compliance with the IWF blacklist (Wei 2011), and the UK government is actively working to secure 100% compliance.

Depending upon the method of implementation employed by an Internet user's ISP, the user may or may not receive any indication that the website he or she has requested has been blocked in compliance with the IWF blacklist or the reason for such blocking. Some ISPs, such as Demon, return a HTTP '403 Forbidden' error message, which provides some indication to the requesting Internet user that the requested site has been blocked because it is 'forbidden.'[i] In contrast, some of the largest ISPs, such as British Telecom and Virgin Media, simply return a generic HTTP '404 Not Found' error message when a user requests access to a site that is on the IWF blacklist. This error message does not give the requesting user any indication that the requested page has been blocked or the reason why the page has been blocked.

In any case, neither the IWF nor the complying ISPs provides any notice to the content provider of the blocked website that its website has been blocked, or the reasons for such blocking. In the words of commentator Edwards (2006, 175), the U.K.'s implementation of the IWF blacklist 'could be the most perfectly invisible censorship mechanism ever invented.' Nor does the filtering system provide for a method of appeal for an independent judicial determination of whether the blocked content is in fact illegal under U.K. law. The IWF website indicates that 'any party with a legitimate association with the [blacklisted] content . . . who believes they are being prevented from accessing legal content may appeal against the accuracy of an assessment' (Internet Watch Foundation 2012a). The appeal procedure provided by the IWF, however, does not involve judicial review. Rather, the contemplated appeal (as shown at the IWF website) merely involves a second look by the IWF itself with no input from or representation of the affected content provider or end user -- and following that, a review by a police agency, whose assessment is final (Internet Watch Foundation 2012b). Further, as discussed above, it is unclear in many cases how a party would learn that the content she was seeking, or seeking to make available, was subject to the IWF's blacklist, since the IWF does not provide notice to the content provider at issue that its site has been blocked, and ISPs merely provide Internet users with a generic 404/File not found or 403/Forbidden error message when a requested website is on the IWF blacklist.

In essence, the UK's nationwide filtering system based on the IWF blacklist operates as an opaque, non-transparent system that does not provide end-users or content providers with meaningful notice that a "potentially illegal" website has been blocked, nor does it provide any method for affected parties to secure an independent judicial determination that "potentially" illegal content that has been blocked is actually illegal under U.K. law. The system operates so as to reposit ultimate authority over Internet content in an unaccountable, nontransparent body. Not surprisingly, this has led to instances of overblocking. Although it is difficult to secure meaningful data regarding how many sites on the IWF's blacklist are actually – not just potentially – illegal, in at least some cases, the IWF's blocked sites have become widely enough known to subject the IWF's discretion to some public scrutiny, as I discuss below.

In December 2008, acting on a hotline notification from an Internet user, the IWF placed on its blacklist a Wikipedia article discussing an album by the popular German rock band The Scorpions. The cover of the Scorpion's 1976 album Virgin Killer depicted a pre-pubescent girl unclothed, but with her genitalia obscured from view. The IWF determined that the cover art was 'potentially illegal,' and placed the Wikipedia article on its blacklist, without notifying the content provider of its decision to blacklist the website. As a result, the vast majority of U.K. Internet users were unable to access this content and were not informed why the content was blocked. In addition, as an unintended result of placing this Wikipedia article on the IWF blacklist, many U.K. users were unable to edit *any* Wikipedia pages. As Heverly (2011) explains, 'When one page on the Wikipedia site was blocked [because it was on the IWF blacklist], the resulting filtering of content forced Wikipedia to deny anonymous editing privileges to thousands of UK users. [A]ll users were forced to either register and log in to the Wikipedia site, or were precluded from editing and contributing to Wikipedia's development.' Surprised by their sudden inability to edit Wikipedia pages, several determined U.K. Internet users investigated and ultimately traced the source of the problem to the IWF blacklist. However, some learned that following up on the causes of such censorship is not for the faint-hearted. One user who contacted his ISP to complain about the blocking was accused by his ISP of seeking to use the Internet to view illegal images. His ISP subsequently threatened to report him to the police and to monitor his Internet use, all as a result of the user's complaining about his inability to edit Wikipedia pages after the Virgin Killer page was placed on the IWF's blacklist.

As a result of the general public awareness and outcry regarding the consequences of the IWF's blocking of the Virgin Killer page, the IWF – apparently for the first time in its history – re-assessed its actions and ultimately removed the Wikipedia article from its blacklist (while still maintaining that the image of the album cover was potentially illegal). Were it not for the general public awareness surrounding the inadvertent blocking of users' ability to edit unrelated Wikipedia pages, it is doubtful that the IWF's actions in connection with this website would have come under public scrutiny or that this page would have been unblocked. As Mike Godwin, General Counsel for Wikimedia, explained to *Wired Magazine* on May 20, 2009, 'When we first protested the block, [the IWF's] response was, "We've now conducted an appeals process on your behalf and you've lost the appeal." When I asked who exactly represented the Wikimedia Foundation's side in that appeals process, they were silent. It was only after the fact of their

blacklist and its effect on UK citizens were publicised that the IWF appears to have felt compelled to relent. If we had not been able to publicise what the IWF had done, I don't doubt that the block would be in place today.'

In another incident of (massive) overblocking, in January 2009 U.K. Internet users learned that all 85 billion pages of the Wayback Machine – the application that archives the Internet's content -- had been blocked, apparently because the archive contained one or more URLs that were on the IWF's blacklist.

Most recently, in November 2011, the implementation of the IWF blacklist blocked U.K. subscribers of the ISP Virgin Media from accessing any files from the popular file-hosting service Fileserve– one of the ten most popular file-sharing sites on the Internet – which allows users to store and share files in the cloud. IWF apparently intended to blacklist only certain Fileserve URLs, but the effect of its placing these URLs on its blacklist was to block any and all files from being downloaded by U.K. users from Fileserve.

As discussed above, because of the secretive, nontransparent, and overbroad manner in which the IWF blacklist is populated – and because of the lack of meaningful notice provided to blocked content providers and would-be end users – it is very difficult to scrutinize the operation of this nationwide filtering system in a comprehensive manner. However, the examples of overblocking above suggest that there are serious concerns with the implementation of this well-intentioned system.

II. B: Case Study 2: Australia's Proposed Nationwide Filtering System

> *Australia's decision to impose mandatory Internet censorship through technology filtering puts the country at the forefront of the spread of this practice from authoritarian regimes such as China and Iran to Western democratic nations. (Travalione 2009)*
>
> *First, China. Next: The Great Firewall of . . . Australia? (*Time, *June 16, 2010)*

Like the United Kingdom, Australia also has a nationwide Internet filtering system, and there is a strong possibility that the government will move to expand this system. Unlike the U.K., the Australian system will be formally mandated by the government, instead of through the informal pressure that the U.K. government has placed on its nation's ISPs to adopt the IWF blacklist. The recent developments in Australia herald 'the first time that a Western democracy will require, through formal statute, ISPs to block users from accessing certain materials online [via a system in which] the criteria by which sites will be designated for blocking remain opaque and uncertain' (Bambauer 2009-10, 495). These developments toward Internet censorship resulted from a change in Australia's government, from the Liberal Party to the Labor Party, the latter which depended upon the support of the conservative Family First Party. Following the

Australian federal election of 2007, the Labor Party formed its first government in more than a decade. Labor's failure to win a majority of seats in either house of Parliament meant, however, that the newly formed government had to rely on minor parties. In particular, the Labor Party found itself in the position of being forced to appeal to the socially conservative Family First Party's lone senator. In doing so, the Labor Party platform announced the goal of filtering websites to all public Internet points accessible by children.

The censorship regime in place in 2008 when the Labor government took office was – and remains at time of writing – a largely voluntary one overseen by a regulatory agency known as the Australian Communications and Media Authority (ACMA). The Broadcast Services Amendment Act of 1999 granted ACMA the power to operate a system requiring Australian Internet service providers to remove objectionable content upon receipt of a notice. Like the U.K.'s Internet Watch Foundation, Australia's ACMA receives complaints from Internet users suggesting that certain websites are illegal. Unlike the IWF, ACMA may, on its own authority, initiate investigations into whether content on the Internet is illegal. And, unlike the IWF, ACMA is an official arm of the Australian state.

The ACMA determines a website's legality by reference to the classification system established by the Australian government, which governs all media both online and offline. Potentially prohibited material includes content that falls within one of the following categories: RC (refused classification); X18 (non-violent, sexually explicit activity between consenting adults); R18 (likely to disturb minors); and MA15+ (restricted audiences) (Australian Communications and Media Authority 2012a). If ACMA determines in response to a user complaint that the material is prohibited – and if such potentially illegal content is hosted within Australia – ACMA sends a takedown notice to the ISP or content provider. If the potentially illegal content is hosted outside Australia, ACMA will notify software filtering companies to add the website to their blacklists. Australians may then use such software to block illegal sites.

The successive Labor governments of Kevin Rudd and Julia Gillard have sought to replace individual software filtering with a mandatory ISP filtering program run by ACMA. The initial plan called for blocking illegal websites on ACMA's blacklist, but the breadth of proposed filtering has been progressively expanded since Labor took office. Stephen Conroy, Minister for Broadband, Communications, and the Digital Economy since 2008, recommended that Australian ISPs block access to approximately 10,000 websites that contain allegedly harmful content. Conroy's ministerial website explains that:

> ISP filtering is a key component of the Australian Government's cybersafety plan. . . . The government has announced that it will introduce legislative amendments to require all ISPs in Australia to use ISP-level filtering to block overseas hosted

Refused Classification (RC) material on the ACMA RC Content List. (Department of Broadband, Communications, and the Digital Economy 2012)

Australia's proposed nationwide filtering system revolves around ACMA's procedures for determining whether a website's content is illegal. That process is neither transparent nor available to the public or the press. Consequently, ACMA's decision to include a website on its blacklist carries with it significant potential for abuse and in fact already has led to minor scandal in Australia. In March 2009, Wikileaks published a list it claimed was the then-current ACMA blacklist of websites. Of the 2,400 URLs on the blacklist, nearly two-thirds apparently contained only material to which adults had a legal right to access in Australia. These included 'online gambling sites, YouTube links, regular [not child] porn and fetish sites, and websites of a tour operator, a Queensland boarding kennel, and a Queensland dentist' (MacBean 2009). The Australian government and ACMA were highly critical of the leak and denied its authenticity – but then promptly added Wikileaks itself to the ACMA blacklist. In a similar development, the Australian government apparently censored 90% of an official account of a meeting with ISPs and business figures about censorship before releasing it to the media. The government claimed the release of the uncensored version could have set off 'premature unnecessary debate.' Equally troubling is the lack of an appeals process for websites erroneously included on the ACMA blacklist.

Australian public opinion is almost uniformly opposed to ISP filtering. As *Time* reported in an article on June 16, 2010, in a poll of 90,000 Australians, 90% responded they opposed mandatory government filtering of the Internet (*Time*, June 16, 2010). Similarly, international opinion is unimpressed by Australian efforts. Reporters Without Borders announced in its 2012 Enemies of the Internet report of the 'worst violators of freedom of expression on the Net' that Australia is now a country that is 'under surveillance,' while U.S. Secretary of State Hillary Clinton decried the Australian government's mandatory filtering initiative in speeches during both 2010 and 2011 (Reporters Without Borders 2012; Clinton 2010; Clinton 2011).

Despite widespread public criticism of the plans for nationwide filtering in Australia, some form of ISP filtering seems likely to occur, whether government-mandated under the Labor plans or government-sponsored and run through a private entity like the IWF. Australia's center-right Coalition led by the Liberal Party is not in principle opposed to an ISP filtering scheme. Instead, its opposition has been grounded in practical concerns with Labor plans, including the ease by which the proposed filter might be bypassed and the expense of the system.

Furthermore, despite the government's delay in implementing a mandatory policy, all of the largest ISPs in Australia – including Telstra, Primus, and Optus – have voluntarily imposed filtering of websites since July 1, 2011 (as reported by *The Australian*, June 23, 2011). As of

now, the URLs blocked are exclusively those of child abuse websites identified by ACMA working jointly with INTERPOL; Australian ISPs also track and report attempts to access these sites. However, the Australian government is also undertaking a review of its Refused Classification categories. It seems likely that the government will try to convince ISPs to voluntarily adopt an expanded blacklist in place of the more limited child abuse one provided by INTERPOL. The Labor Party is in favor of mandatory nationwide filtering and the current center-right Coalition led by the Liberal Party is not opposed to such a system.

III. Procedural Protections for Speech

Nationwide Internet filtering systems like the system implemented "voluntarily" by ISPs in the U.K. and the system that the government intends to impose mandatorily on ISPs in Australia fail to incorporate the requisite procedural protections for speech. International treaties and documents of international law provide both substantive *and* procedural protections for speech that must be respected in the context of nationwide Internet filtering systems. In particular, the International Covenant on Civil and Political Rights (ICCPR), which has been adopted by 167 parties and is considered a binding international law treaty, provides that:[ii]

> 1. Everyone shall have the right to hold opinions without interference.
> 2. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.
> 3. The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. *It may therefore be subject to certain restrictions*, but these shall only be such as are provided by law and are necessary:
> (a) For respect of the rights or reputations of others;
> (b) For the protection of national security or of public order (ordre public), or of public health or morals. (UN General Assembly 1966, 178)

The ICCPR not only has a substantive dimension of which categories of speech to protect and which may be restricted, but also has important procedural dimensions, which require that "sensitive tools" be implemented to distinguish between protected and unprotected speech (see, e.g., *Bantam Books v. Sullivan* [1963]). As free speech theorist Monaghan (1970, 518) explains, 'procedural guarantees play an equally large role in protecting freedom of speech; indeed, they assume an importance fully as great as the validity of the substantive rule of law to be applied.' While there is substantial variation from country to country in the substantive protections for speech, there is more widespread agreement regarding the procedures that are essential to ensure protection for legally-protected categories of speech. These procedural requirements were recently set forth in the Report of the Special Rapporteur on the Promotion and Protection of the

Right to Freedom of Opinion and Expression. While recognizing that countries enjoy some discretion to restrict speech that constitutes child pornography, hate speech, defamation, incitement to genocide, discrimination, hostility or violence, La Rue explains that:

> Any [such] limitation to the right to freedom of expression must pass the following [multi]-part, cumulative test:
>
> It must be provided by law, which is clear and accessible to everyone (principles of predictability and transparency);
>
> It must pursue one of the purposes set out in article 19, paragraph 3, of the [International Covenant on Civil and Political Rights], namely (i) to protect the rights or reputations of others, or (ii) to protect national security or of public order, or of public health or morals (principle of legitimacy); and
>
> It must be proven as necessary and the least restrictive means required to achieve the purported aim (principles of necessity and proportionality).
>
> Moreover, any legislation restricting the right to freedom of expression must be applied by a body which is independent of any political, commercial, or other unwarranted influences in a manner that is neither arbitrary nor discriminatory, and with adequate safeguards against abuse, including the possibility of challenge and remedy against its abusive application. (2011, 8)

Such internationally-recognized procedures and sensitive tools for protecting free speech are as important as the substantive protections themselves; as Justice Frankfurter explained in his concurrence in *Malinski v. New York* (1945, 414), '[t]he history of … freedom is, in no small part, the history of procedure.'

In addition, within the context of Anglo-American jurisprudence, courts have established procedural protections for free speech -- a powerful 'body of procedural law which defines the manner in which they and other bodies must evaluate and resolve [free speech] claims — a [free speech] "due process," if you will' (Monaghan 1970). In particular, courts have required that stringent procedural safeguards be in place in the context of *prior restraints on speech* – restraints on speech that are imposed prior to a judicial determination of the speech's illegality.[iii] Nationwide Internet filtering systems, like that implemented in the U.K. and Australia, constitute prior restraints on speech. Prior restraints on speech – in contrast to subsequent punishment – are restrictions that are imposed prior to publication and/or before a judicial determination that the speech in question is illegal. In setting forth procedural requirements for prior restraints, courts have developed 'a comprehensive system of procedural safeguards designed to obviate the dangers of a censorship system' (Monaghan 1970).

The prohibition against prior restraints has been a foundational principle of freedom of expression in the Anglo-American tradition. Indeed, in William Blackstone's Commentaries,

'Freedom of the Press' is defined simply as the right to be free from prior restraints. In his Commentaries, Blackstone explained that 'the liberty of the press is indeed essential to the nature of a free state; but this consists in laying no previous restraints upon publications, and not in freedom from censure for criminal matter when published.' Indeed, following Blackstone, some have argued that the sole purpose of the First Amendment was to foreclose in the United States any system of prior restraint such as was embodied in the English censorship system (for a discussion of this historical process see, especially, Emerson [1955]).

In order to restrict harmful speech, governments generally have the option of imposing prior restraints, such as those imposed via the types of filtering systems discussed above, or forms of subsequent punishment, such as by criminally prosecuting content providers who make available harmful speech, for example, under obscenity or child pornography statutes. Regulations that proceed via subsequent punishment generally provide vastly greater procedural protections for speech and are likely to be implemented in ways that are far more speech-protective than regulations embodying the method of prior restraints. Pre-eminent First Amendment theorist Emerson (1955) explains that systems of prior restraint are likely to operate in a manner that is much more hostile toward free speech than systems of subsequent punishment.

First, Emerson explains, systems of prior restraint are prone to adverse decisions. Such systems are designed to make it easier, and more likely, that the censor will rule adversely to free expression. As Emerson observes, '[an] official thinks longer and harder before deciding to undertake the serious task of subsequent punishment. . . . Under a system of prior restraint, he can reach the result by a simple stroke of the pen. Thus, [in the case of subsequent punishment], the burden of initial action falls upon the government; in the other, on the citizen….[Accordingly], a decision to suppress in advance is usually more readily reached, on the same facts, than a decision to punish after the event' (1955, 657).

Second, under a system of prior restraint, the issue whether to suppress expression is determined by an administrative procedure, instead of via criminal procedure. Accordingly, 'the procedural protections built around the criminal prosecution – many of which are constitutional guarantees – are not applicable to prior restraint. The presumption of innocence, the heavier burden of proof borne by the government, the stricter rules of evidence, the stronger objection to vagueness, the immeasurably tighter and more technical procedure – all these are not on the side of free expression when its fate is decided [in the context of prior restraints]' (Emerson 1955, 657).

Third, within a system of prior restraints, the decision to restrict speech rests with a single administrative entity instead of with a judge and/or jury. Both judge and jury function to provide important safeguards against abuses of power and are designed to secure independent and objective judgments. Such safeguards are not necessarily present within an administrative system implementing prior restraints, such as the Internet Watch Foundation.

Fourth, systems of prior restraint like that implemented by the Internet Watch Foundation are more likely to operate out of the public view and in such a manner that they are hidden from public scrutiny, appraisal, and accountability. In contrast, subsequent punishments take place in a context that assures greater public scrutiny and accountability. As Emerson explains,

> [In systems of prior restraint,] decisions are less likely to be made in the glare of publicity that accompanies a subsequent punishment. The policies and actions of the licensing official do not as often come to public notice; the reasons for his action are less likely to be known or publicly debated; material for study and criticism are less readily available; and the whole apparatus of public scrutiny fails to play the role it normally does under a system of subsequent punishment. . . . . [T]he preservation of civil liberties must rest upon an informed and actve public opinion. Any device that draws a cloak over restrictions on free expression seriously undermines the democratic process." (1955, 658)

Finally, and perhaps most importantly, the institutional framework in which systems of prior restraint operate are such that they favor suppression of expression. As Emerson explains, 'The function of the censor is to censor…. He is often acutely responsive to interests which demand suppression…and not so well attuned to the forces that support free expression. . . . The long history of prior restraints reveals over and over again that the personal and institutional forces inherent in the system [of prior restraints] nearly always end in . . . unnecessary and extreme suppression' (1955, 659).

These considerations make clear that the historical opposition to systems of prior restraint in the tradition of Anglo-American jurisprudence does not arise as a result of arbitrary historical accident, but follows directly from the importance of according meaningful protections for free expression.

Nationwide filtering systems impose "prior restraints" -- restraints on speech prior to and apart from a judicial determination of the speech's illegality. Instead of imposing punishment on such speech after it has been published and adjudicated illegal by a court, these systems regulate and ban the speech at issue before a court has made the determination that such speech is illegal. Such prior restraints on speech impose grave dangers to freedom of expression and impose much greater harms on freedoms than subsequent punishments of speech. Even though the restraints on speech imposed under a nationwide filtering system like those operationalized in the U.K. and Australia are not necessarily imposed before the speech is made available to the relevant public in the first instance, such restraints still embody the dangers of prior restraints, because they are imposed prior to a judicial determination that the blocked content is actually illegal. Such "midstream" prior restraints entail similar types of harms as those imposed ex ante. Prior restraints can be imposed by governments ex ante, via pre-publication licensing schemes, as occur in the context of motion picture censorship boards, or nationwide, centralized filtering schemes in countries such as China. Alternatively, *midstream* prior restraints include those restraints on speech that are imposed after the content's initial circulation but before a judicial determination that the content is illegal. Because midstream prior restraints are imposed prior to a judicial determination of the content's illegality, they are constitutionally suspect. Nationwide

filtering systems that work from evolving blacklists of websites maintained in response to tips from Internet users, such as that implemented by the Internet Watch Foundation in the U.K., embody midstream prior restraints that are as constitutionally suspect as ex ante prior restraints.

The United States Supreme Court struck down a censorship regime involving a midstream prior restraint in the case of *Bantam Books v. Sullivan* (1963). In that case, the Rhode Island Commission to Encourage Morality in Youth was responsible for investigating and recommending prosecution of booksellers for the distribution of printed works that were obscene or indecent. The Commission reviewed books and magazines that were already in circulation, and issued notices to distributors of cases in which a book or magazine had been distributed that the Commission found objectionable. In reviewing the constitutionality of this scheme, the U.S. Supreme Court held that, even though the restrictions on publication were imposed by a non-government agency and after initial circulation and distribution, the Commission's actions nonetheless effectuated an unconstitutional prior restraint. The Court explained that 'the separation of legitimate from illegitimate speech calls for . . . . sensitive tools' and reiterated its insistence that regulations of speech 'scrupulously embody the most rigorous procedural safeguards.' The Court observed that, under the scheme at issue, 'the publisher or distributor is not even entitled to *notice and hearing* before his publications are listed by the Commission as objectionable' and that there was 'no provision whatever for *judicial superintendence* before notices issue or even for *judicial review* of the Commission's determinations of objectionableness.' The Court concluded that, in the context of this system of midstream prior restraint, the 'procedures of the Commission are radically deficient' and unconstitutional (*Bantam Books v. Sullivan* 1963, 66).

In *Bantam Books* and other prior restraint cases, the U.S. Supreme Court has articulated a series of procedural safeguards that must be in place for any system of (ex ante or midstream) prior restraint to be constitutional, which provides a helpful starting point for other countries in establishing nationwide filtering systems. Translated into the context of nationwide filtering or blocking of Internet speech, these procedural safeguards require, first, that the filtering scheme *operate in an open and transparent manner*, such that affected Internet users and content providers are provided with *notice* that the content was filtered and the reason for such filtering; second, that any filtering be imposed subject to *clear and precise definitions of the speech to be regulated*; and third, that the filtering system provide Internet users with the *opportunity to appeal any such blocking decisions* to a *judicial body in an expeditious manner*. These procedures do not themselves dictate what categories of speech are deemed harmful or dangerous and as such do not interfere with the prerogative of each country to make such substantive determinations. Rather, they impose meaningful, process-based safeguards on the implementation of restrictions on whatever categories of speech are deemed harmful or dangerous by each country.

A. Ability to Meaningfully Challenge Decisions to Filter/Block Content

Both the International Covenant on Civil and Political Rights, as construed by La Rue, and prior restraint jurisprudence require that any filtering decision be subject to meaningful challenge before an impartial decision-maker. In interpreting the ICCPR, La Rue explains that:

> Any legislation restricting the right of freedom of expression must be applied . . . with adequate safeguards against abuse, including the possibility of challenge and remedy against its abusive application. (2011, 8)

Similarly, United States courts have consistently emphasized the importance of the availability of *prompt judicial review* of censorship determinations in the prior restraint context. As the United States Supreme Court explained, 'because only a judicial determination in an adversary proceeding ensures the necessary sensitivity to freedom of expression, only a procedure requiring a judicial determination suffices to impose a valid final [prior] restraint' (*United States v. Pryba* 1974, 405). In order for a nationwide filtering system to effectuate a constitutionally valid prior restraint, such a system must provide for an opportunity to secure prompt judicial review of a censorship decision (for more on the Court's interpretation of the expeditious requirement see *United States v. Thirty-Seven Photographs* [1971, 372-74]; *Kingsley Books v. Brown* [1957]; *Interstate Circuit, Inc. v. City of Dallas* [1968, 690]; *Bantam Books* [1963, 70]; *Redner v. Dean* [1994, 1501-02]; *East Brook Books, Inc. v. City of Memphis* [1995, 225]).

Attempts within the U.S. to filter Internet speech at the ISP level have been held unconstitutional because they have failed to provide for judicial review in an adversary proceeding of the decision to censor. In the *Center for Democracy and Technology v. Pappert* (2004), for example, Pennsylvania sought to combat online child pornography by enacting the Internet Child Pornography Act, which required ISPs serving Pennsylvanians to block access to websites believed to be associated with child pornography. The Act permitted the Pennsylvania Attorney General or Pennsylvania district attorneys to seek an ex parte court order requiring an ISP to remove or disable access to items accessible through the ISP's service, upon a showing of probable cause that the item constituted child pornography. The Act did not require a final judicial determination in an adversary proceeding that the material to be blocked constituted child pornography before it was placed on the blacklist. In consultation with the affected ISPs, the Attorney General's office decided to implement the Act by proceeding without even securing ex parte court orders and instead by providing 'Informal Notices of Child Pornography' to ISPs that certain websites allegedly contained child pornography. The Informal Notice directed the ISP to remove or disable Pennsylvania citizens' access to the suspected material within five days of receipt of Notice.

The statute was challenged as an unconstitutional prior restraint lacking the requisite procedural safeguards. In defense of the statute, the attorney general argued that only material that its office had probable cause to believe constituted child pornography was identified for removal. The court found that this probable cause showing did not save the statute (nor did the fact that the attorney general only issued 'Informal Notices' not court orders, and that the process was therefore 'voluntary' not coercive[iv]). First, the court explained that in order to comply with the Supreme Court's exacting procedural requirements for prior restraints, to be constitutional, a prior restraint must be imposed by a judicial determination in an adversary proceeding. The

attorney general's determination that there was probable cause that the material was illegal was insufficient. Further, even an ex parte judicial determination that the material was illegal would not suffice to impose a constitutional prior restraint because it did not result from an adversarial proceeding. As the United States Supreme Court explained (*Freedman v. Maryland* 1965, 58) 'only a judicial determination *in an adversary proceeding* ensures the necessary sensitivity to freedom of expression.' Ex parte judicial determinations that are made in the absence of notice and an opportunity to be heard on the part of the adversely-affected speaker are constitutionally deficient, and ex parte *nonjudicial* determinations are constitutionally deficient by an even greater measure.

Under many nationwide filtering systems, provisions do exist for some sort of appeal of the censorship decision. However, such provisions generally do not provide for *judicial* determination (in an adversary proceeding or otherwise) and instead merely provide for a second look by the administrative body that made the censorship determination in the first place. In the U.K., for example, the IWF website indicates that 'any party with a legitimate association with the [blacklisted] content . . . who believes they are being prevented from accessing legal content may appeal against the accuracy of an assessment' (Internet Watch Foundation 2012a). The appeal procedure provided by the IWF, however, does not contemplate judicial review. Rather, the procedure for appeal involves a second look by the IWF itself (as occurred in the Virgin Killer Wikipedia page incident discussed above), and following that, a review by a police agency, whose assessment is final (Internet Watch Foundation 2012b). Such provisions for appeal – because they do not provide for a judicial determination of the affected parties' rights – fail to accord the requisite protections for freedom of expression.

In Australia, the current process for appealing an ACMA decision to blacklist a website (as specified on ACMA's website) contemplates only an appeal from the agency's Classification Board to the agency's Classification Review Board (Australian Communications and Media Authority 2012b). Furthermore, the separation of powers in the Australian Constitution largely insulates ACMA's decisions from judicial review. The High Court (and the rest of the judicial branch) are barred from acting as appellate bodies for the administrative decisions of executive agencies on the merits.

To comport with the procedural requirements for protecting speech articulated in the ICCPR and in prior restraint jurisprudence, nationwide filtering systems like those in the U.K. and Australia should be modified to provide adversely-affected Internet users with a meaningful opportunity to appeal adverse decisions to an impartial judicial body before which the users' interests are represented.

B. Meaningful Notice to Affected Internet Users

The International Covenant on Civil and Political Rights as construed by La Rue, as well as U.S. prior restraint jurisprudence, require that individuals affected by nationwide filtering systems must at a minimum be made aware of such a decision to filter so that they can effectively challenge a decision, as discussed above. The right to meaningfully challenge and

appeal a filtering decision presupposes that affected individuals have *notice* of any such censorship. As La Rue explains, transparency is essential in any such system:

> Any limitation on the right to freedom of expression must . . . be provided by law, which is clear and accessible to everyone (principles of predictability and *transparency*). (2011, 8)

Filtering systems in which the affected parties are not made aware that content has been filtered fail this threshold requirement. For example, as discussed above, ISPs implementing the U.K.'s naionwide system provide no notice to the censored content providers that their websites have been blocked. Nor do the IWF or the implementing ISPs provide meaningful notice to Internet end users when the content they have requested has been blocked. Instead, Internet end users are merely provided with a "File Not Found" or "Forbidden" error message. Indeed, in the incident in which the Virgin Killer Wikipedia page was blocked, it took tech-savvy Internet users employing substantial investigative skills to establish that their access was blocked.

Countries implementing nationwide filtering systems to restrict their citizens' access to content that they deem harmful should at the very least operate these systems in an open and transparent manner, in which the restrictions on speech are provided by law and are clear and accessible to everyone, to adhere to the principles of predictability and transparency articulated in the ICCPR and in prior restraint jurisprudence. These systems should operate in a manner such that (1) Internet users are made aware of the operation of such filtering systems generally, and (2) affected users – both content providers and end users -- are specifically informed of instances in which the filters operate to block access to a particular website. Only then can affected content providers and end users have the meaningful notice necessary to challenge the decision to censor.

C. Categories of Prohibited Speech Should Be Clearly and Precisely Defined

A third procedural requirement for nationwide filtering systems is that the censor's discretion be meaningfully constrained by clearly defined and precise guidelines to help ensure that the censors adhere to narrow and precise definitions of what content is proscribable. While countries may reasonably differ in their determinations of what categories of speech are illegal content – pornography, hate speech, Holocaust denial, etc. – it is important that, within each country, the definitions of illegal speech – and especially definitions of any illegal speech subject to prior restraint -- be precisely defined so as to constrain the initial censor's discretion. Providing precise definitions of proscribable content helps to ensure that the speech restriction is the least restrictive means required to achieve the purported aim,' consistent with La Rue's interpretation of the ICCPR's procedural protections for speech. Similarly, the U.S. Supreme Court has strictly scrutinized the discretion of censors in systems of prior restraint and has rejected as unconstitutional any restrictions on speech that do not embody the least restrictive means of achieving the government's goals or that confer unbounded discretion to determine whether or not speech is protected. For example, in *Shuttlesworth v. Birmingham* (1969, 149-50) the Court evaluated the constitutionality of a parade permitting system that vested the City

Commission with the broad discretion to deny parade permits in cases where 'in [the Commission's] judgment the public welfare, peace, safety, health, decency, good order, morals or convenience require that [the parade permit] be refused' (149-150). In ruling on a challenge to the statute, the Court held that, because the permitting scheme constituted a prior restraint on expression that conferred 'virtually unbridled and absolute power' on the Commission, it failed to comport with the fundamental due process requirement that any law subjecting the exercise of First Amendment freedoms to prior restraint of a license must embody 'narrow, objective, and definite standards to guide the licensing authority' (150-51).

Requiring that the criteria by which the censoring authority makes the decision to censor be set forth with precision helps to cabin administrative discretion and also helps to limit 'mission creep' within the censoring body. Without a precise and detailed specification of the criteria for censorship, the censor can exercise unbridled discretion to restrict speech.

While the Internet Watch Foundation's website provides some information about what type of content it will censor, this entity seems to be susceptible to the problem of mission creep. Although the IWF's initial mission was solely to restrict access to images of child sexual abuse, the target of its censorship was subsequently expanded to include 'criminally obscene' adult content, as well as 'incitement to racial hatred content.' Recently, IWF's responsibility for incitement to racial hatred content was committed to another entity, True Vision, which is under the auspices of the Association of Chief Police Officers. In Australia, the categories of Internet content that will be subject to filtering by the ACMA are also in a state of flux, as the Australian government is undertaking a review of its Refused Classification categories. The fluidity and uncertainty regarding what categories of speech will be censored by the IWF and the ACMA are highly problematic and fail to accord the requisite procedural protections for Internet content.

Conclusion

In summary, state-mandated Internet filtering systems – the likes of which are now being imposed by over forty countries worldwide – embody prior restraints on speech, which violate principles of due process, absent the inclusion of fundamental procedural protections for speech. These free speech procedural protections mandate that such filtering systems be implemented (1) such that affected Internet users are provided with the *opportunity to appeal any such filtering decisions*, to an *impartial judicial body and in an expeditious manner;* (2) *in an open and transparent manner*, such that affected Internet users and content providers are provided with information that the content was blocked and the reason for such blocking; and (3) such that any restraints on speech are imposed subject to *clear and precise definitions of the speech to be regulated*. Only such "sensitive tools" for distinguishing between protected and unprotected speech can adequately protect individuals' free speech rights on the Internet.

[i] Demon returns the following error message for those attempting to access images on the IWF blacklist: "We have blocked this page because, according to the Internet Watch Foundation (IWF), it contains indecent images of children or pointers to them; you could be breaking UK law if you viewed the page."

[ii] The ICCPR provides further, in Article 20, that any propaganda for war or advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence, is prohibited by law. Further support for freedom of expression comes from the European Convention for the Protection of Human Rights, which has been signed by 47 nations, is considered binding, and is enforced by the European Court of Human Rights (ECHR). This Convention, however, like the Universal Declaration and the ICCPR, allows signatories to carve out exceptions and limitations to the protections granted to speech.

[iii] This strong presumption against the legality of prior restraints is also shared by Latin American countries.

[iv] On this point, the court explained that the informal and technically noncoercive nature of the attorney general's removal requests did not insulate them from constitutional scrutiny. The court

explained that removal requests issued by law enforcement officials were not interpreted by the recipient ISPs as being voluntary, even if technically they did not have the force of law.

Reference List

Australian Communications and Media Authority. 2012a. "What is prohibited online content?" Australian Communications and Mediat Authority. Accessed July 19. http://www.acma.gov.au/WEB/STANDARD/pc=PC_90169#prohib

Australian Communications and Media Authority. 2012b. "Rights of review and appeal." Australian Communications and Media Authority. Accessed July 19. http://www.acma.gov.au/WEB/STANDARD/1001/pc=PC_410100

Bambauer, Derek. 2009-10. "Filtering in Oz: Australia's Foray into Internet Censorship." *University of Pennsylvania Journal of International Law* 31(2): 493-531.

Bantam Books v. Sullivan, 372 U.S. 58 (1963).

Center for Democracy and Technology v. Pappert, 337 F. Supp. 2d 606 (E.D. Pa 2004).

Clinton, Hillary Rodham. 2010. "Remarks on Internet Freedom." United States Department of State. http://www.state.gov/secretary/rm/2010/01/135519.htm

Clinton, Hillary Rodham. 2011. "Internet Rights and Wrongs: Choices and Challenges in a Networked World." United States Department of State. http://www.state.gov/secretary/rm/2011/02/156619.htm

Deibert, Ronald J., John G. Palfrey, Rahal Rohozinski, and Jonathan Zittrain, eds. 2010. *Access Controlled: The Shaping of Power, Rights, and Rule in Cyberspace*. Cambridge, MA: MIT Press.

East Brooks Books, Inc. v. City of Memphis, 48 F.3d 220 (6th Cir. 1995).

Department of Broadband, Communications, and the Digital Economy. 2012. "Internet service provider (ISP) filtering." Accessed July 19. http://www.dbcde.gov.au/online_safety_and_security/cybersafety_plan/internet_service_provider_isp_filtering

Edwards, Lilian. 2006. "From child porn to China, in one Cleanfeed." *SCRIPT-ed* 3(3): 174-75. doi: 10.2966/scrip.030306.174

Emerson, Thomas I. 1955. "The Doctrine of Prior Restraint." Faculty Scholarship Series. http://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?article=3761&context=fss_papers

Farrior, Stephanie. 1996. "Molding the Matrix: The Historical and Theoretical Foundations of International Law Concerning Hate Speech." *Berkeley Journal of International Law 14* (1): 3-98.

Freedman v. Maryland, 380 U.S. 51 (1965).

Heverly, Robert A. 2011. "Breaking the Internet: International Efforts to Play the Middle Against the Ends: A Way Forward." *Georgetown Journal of International Law* 42(4): 1083-1122.

Internet Watch Foundation. 2012a. "What is the criterion for a URL to be added to the list?" Internet Watch Foundation. Accessed June 20. http://www.iwf.org.uk/services/blocking/blocking-faqs#WhatisthecriterionforaURLtobeaddedtothelist

Internet Watch Foundation. 2012b. "IWF Content Assessment Appeal Process." Internet Watch Foundation. Accessed June 20. http://www.iwf.org.uk/accountability/complaints/content-assessment-appeal-process

Interstate Circuit, Inc. v. City of Dallas, 390 U.S. 676 (1968).

Kingsley Books, Inc. v. Brown, 354 U.S. 436 (1957).

Krotozynski, Ronald J., Jr. 2006. *The First Amendment in Cross-Cultural Perspective: A Comparative Legal Analysis of the Freedom of Speech*. New York: NYU Press.

La Rue, Frank. 2011. *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. United Nations Human Rights Council. http://www2.ohchr.org/english/bodies/hrcouncil/docs/17session/A.HRC.17.27_en.pdf

MacBean, Nic. 2009. "Internet filter blacklist leaked on web." ABC News. http://www.abc.net.au/news/2009-03-19/internet-filter-blacklist-leaked-on-web/1623890

Malinski v. New York, 324 U.S. 401 (1945).

Monaghan, Henry. 1970. "First Amendment Due Process." *Harvard Law Review* 83(3): 518-51.

Murdoch, Steven J., and Ross Anderson. 2008. "Tools and Technologies of Internet Filtering." In *Access Denied: The Practice and Policy of Global Internet Filtering*, edited by Ronald J. Deibert, John G. Palfrey, Rahal Rohozinski, and Jonathan Zittrain. Cambridge, MA: MIT Press.

Redner v. Dean, 29 F.3d 1495 (11th Cir. 1994).

Reporters Without Borders. 2012. *Enemies of the Internet Report 2012*. Reporters Without Borders. http://issuu.com/rsf_webmaster/docs/rapport-internet2012_ang?mode=window&backgroundColor=%23222222

Sedler, Robert A. 2006. "Freedom of Speech: The United States versus The Rest of the World." *Michigan State Law Review* 377 (2): 377-84.

Shuttlesworth v. City of Birmingham, 394 U.S. 147 (1969).

Travalione, Karina. 2009. "Internet Censorship in Australia – A 'clean-feed'?" Mannkal Essay Competition. http://www.mannkal.org/downloads/scholars/internet-censorship-in-australia.pdf

UN General Assembly. 1966. "International Covenant on Civil and Political Rights." *United Nations, Treaty Series* 999: 171-346. http://www.unhcr.org/refworld/docid/3ae6b3aa0.html

United States v. Pryba, 502 F.2d 391 (D.C. Cir. 1974).

United States v. Thirty-Seven Photographs, 402 U.S. 363 (1971).

Wei, Weixiao. 2011. "Online Child Sexual Abuse Content: The Development of a Comprehensive, Transferable International Internet Notice and Takedown System." Internet Watch Foundation. http://www.iwf.org.uk/assets/media/resources/IWF%20Research%20Report_%20Development%20of%20an%20international%20internet%20notice%20and%20takedown%20system.pdf