Research Repository UCD

Title	Child Abuse Images and Cleanfeeds: Assessing Internet Blocking Systems
Authors(s)	McIntyre, T.J.
Publication date	2013
Publication information	McIntyre, T.J. "Child Abuse Images and Cleanfeeds: Assessing Internet Blocking Systems." Edward Elgar, 2013.
Publisher	Edward Elgar
Item record/more information	http://hdl.handle.net/10197/7594
Publisher's statement	Edward Elgar Publishing is the source and copyright holder of the work, and the article cannot be used for any other purpose elsewhere. The chapter is for private use only.

Downloaded 2023-04-07T10:52:29Z

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)

© Some rights reserved. For more information

Child Abuse images and Cleanfeeds: Assessing Internet Blocking Systems

TJ McIntyre¹

School of Law, University College Dublin

tjmcintyre@ucd.ie

To appear in: Ian Brown, ed., Research Handbook on Governance of the Internet (Cheltenham: Edward Elgar, forthcoming)

1. Introduction

One of the most important trends in internet governance in recent years has been the growth of internet blocking as a policy tool, to the point where it is increasingly becoming a global norm. This is most obvious in states such as China where blocking is used to suppress political speech; however, in the last decade blocking has also become more common in democracies, usually as part of attempts to limit the availability of child abuse images. Numerous governments have therefore settled on blocking as their “primary solution” towards preventing such images from being distributed (Villeneuve 2010).

Child abuse image blocking has, however, been extremely controversial within the academic, civil liberties and technical communities, and this debate has recently taken on a wider public dimension. At the time of writing, for example, public pressure has forced the German Federal Government to abandon legislation which would have introduced a police run system while the European Parliament has also rejected Commission proposals for mandatory blocking (Baker 2011; Zuvela 2011).

Why have these systems been so controversial? Two lines of criticism can be identified, which might be termed the practical and the principled. The practical argument claims that blocking is ineffective, with ill-defined goals and easily evaded by widely available circumvention technologies (see e.g. Callanan et al. 2009). The principled argument, on the other hand, is that blocking systems undermine the norms associated with freedom of expression in democratic societies (Brown 2008). This latter argument stems from the fact that blocking sits at the intersection of three different regulatory trends – the use of technological solutions (“code as law”), a focus on intermediaries and the use of self-regulation in preference to legislation – which individually and all the more so collectively create a risk of invisible and unaccountable “censorship by proxy” (Kreimer 2006; McIntyre & Scott 2008).

This chapter introduces and evaluates these claims by examining three prominent examples of child abuse image blocking – the United Kingdom Internet Watch Foundation (“IWF”) Child Abuse Image Content (“CAIC”) list, the European Union sponsored CIRCAMP system and United States hash value systems. It discusses the operation of each system and the extent to which the critics’ concerns are borne out. It concludes by considering the lessons which might be learned for proposals to extend blocking to other types of content.

2. Background and regulatory context

From the early days of the internet it was clear that the technology it embodied – in particular its possibilities for anonymity, decentralised distribution of content and regulatory arbitrage – threatened the ability of governments to control content such as child abuse images. Johnson

and Post (1996) famously expressed this “cyber-libertarian” view when they argued that “efforts to control the flow of electronic information across physical borders – to map local regulation and physical boundaries onto Cyberspace – are likely to prove futile”.

In response, however, “cyber-realists” argued that governments could adapt by shifting regulatory strategies. Three approaches in particular were identified and have since been widely adopted.

Regulation by code

The first, most associated with Lessig (1999), stressed the role of code (software) as a means of regulation. Lessig noted that while the first generation of the internet was structured in such a way as to provide for anonymous speech, decentralised distribution and the use of encryption, there was no guarantee that this structure would persist. Instead, he pointed out, the architecture of the internet could easily be remade to facilitate governmental control – and to do so in an automated manner which could be much more efficient than more traditional means of enforcement.

Intermediary-based regulation

The second, articulated by Boyle (1997) and Swire (1998), rejected the argument that the decentralised and international nature of the internet makes it difficult or impossible to control the conduct of users who may be anonymous or whose location might be uncertain. Instead, it was argued, regulators could simply resort to indirect enforcement, targeting intermediaries rather than end users. For example, Boyle presciently suggested that the state might target ISPs, pressuring or requiring them to “prevent copyright infringement through technical surveillance”.

This argument relied on the fact that the effect of internet disintermediation was oversold – while there has certainly been a great deal of disintermediation, there has also been the creation of entirely new intermediaries with greater technical and legal powers to control the actions of their users. For example, as compared with the post office an ISP or webmail provider has greater technical capability to screen communications, and may not be covered by older laws prohibiting this. Consequently, the ISP, search engine, hosting provider and others have become the new gatekeepers or “Internet points of control” and can be enlisted to stop the transmission of child abuse images (Zittrain 2003).

Self- and co-regulation

Closely related to the use of intermediaries, the third approach involved the promotion by governments of industry self- and co-regulatory schemes, which became so common in the internet context that they have been described as the presumptive starting points for regulation of information technology (Koops et al. 2006).

These schemes appeared to offer substantial benefits for states and industry alike. By harnessing industry expertise and responsiveness, they dealt with the objections that governments lacked the knowledge necessary to regulate the internet and that legislation could not keep up with the pace of change online. Self-regulation also offered governments the possibility of outsourcing enforcement and minimising the accompanying costs, while

industry was attracted by the promise of a flexible and light touch regulatory regime which might ward off more intrusive legislative intervention (Price & Verhulst 2005).

3. Development of child abuse image blocking

The three strategies mentioned above – a focus on intermediaries, regulation by code and the use of self- and co-regulation – neatly dovetail in the form of internet blocking which of its nature involves regulation by software and which generally (though not invariably) also involves ISPs and other intermediaries operating in a self- or co-regulatory context (McIntyre & Scott 2008).

Perhaps unsurprisingly, child abuse images have led the growth of blocking in democracies. Child abuse is a particularly abhorrent crime and as a result there has been a substantial degree of both domestic and international consensus as to the illegality of such images. Unlike many other types of content which governments seek to filter – such as adult pornography or file-sharing sites – the blocking of child abuse images has until recently generally provoked little public controversy (All Party Parliamentary Communications Group 2009, p.9).

There is also an important practical aspect which has favoured this type of blocking. As compared with other types of content, there are fewer websites or images which are potentially illegal. The IWF CAIC list, for example, currently contains about 500 URLs at any one time (Internet Watch Foundation 2011a). In addition, judgments about child abuse images are easier to make than judgments about other types of content. Whether something “glorifies terrorism” contrary to the UK Terrorism Act 2006 requires a difficult assessment of the context, including how it is likely to be understood by members of the public (Banisar 2008, p.21). By contrast, the evaluation of child abuse images does not generally present the same difficulty. As a result, the systems required to monitor, blacklist, and ultimately block child abuse images present fewer administrative and technological difficulties.

In relation to child abuse images, blocking by ISPs also appeared to solve the problem that states could not control material hosted beyond their national borders – enabling them to take action on a domestic basis against material hosted abroad without the international cooperation necessary to have it removed at source. Children’s advocacy groups therefore began to lobby for blocking as a form of situational crime prevention (See e.g. Carr & Hilton 2009).

These lobbying efforts have been remarkably successful, and during the last decade systems have been adopted in numerous jurisdictions including: the United Kingdom, Norway, Sweden, Denmark, Canada, Switzerland, Italy, Netherlands, Finland, New Zealand and most recently France (Villeneuve 2010; New Zealand Department of Internal Affairs 2010; La Quadrature du Net 2011).

In addition to these national systems, public and government pressure has led to many individual companies also adopting their own systems, with prominent examples including Google (search results), AOL (email attachments) and Facebook (uploaded images) (Office of the Attorney General 2010; Committee on Energy and Commerce 2006).

4. Case studies

These blocking systems all attempt to control the same basic subject matter. In almost every other way, however, they differ from each other. Consider, for example, one of the most basic issues: who decides what material is to be blocked? The United Kingdom has pioneered an industry-led approach where decisions are made by a private body (albeit one with extensive links to the state), most European jurisdictions have adopted a police-led approach where a designated unit within the police force is responsible, while within the United States at least one major ISP (AOL) has preferred to create a blocking list entirely in-house, concerned that it would be treated as a state actor if it relied on a government provided list (Tambini et al. 2008; Dedman & Sullivan 2008).

Other aspects also differ greatly. While some blocking systems are purely preventive, others have been used for police intelligence gathering and even prosecution purposes. The channels which are filtered also vary, with some systems focusing solely on the web while others extend also to email, search engines and filesharing. Similarly, the technologies used vary from the crude (DNS poisoning) to the more sophisticated (hybrid URL blocking, hash value matching). Some systems operate at a purely national level, while others have an international effect. Perhaps most importantly, only a tiny minority of blocking systems are underpinned by legislation, with the majority operating on a voluntary or self-regulatory basis (Callanan et al. 2009).

This diversity of approaches makes it difficult to generalise about the issues presented. For example, a system which blocks at the domain name level (blocking all access to example.com) will certainly raise concerns as to proportionality and fears that significant quantities of innocent material will be blocked; while more granular systems which block at the level of the individual file may require much greater scrutiny of the actions of users, thus raising fresh concerns as to user privacy and function creep.

The following section will tease out these issues by examining three of the most prominent schemes. These systems – the IWF CAIC list, the EU funded CIRCAMP network, and the United States hash value blocking systems – cover a variety of different technologies and stages at which blocking can be deployed. Figure 1 (adapted from Ofcom 2008) illustrates this point by depicting the internet content chain and showing the stages at which these systems operate. Although blocking is most commonly associated with controlling *access*, we will see from the US hash value systems that it can also be used as a means of controlling *availability* also, by scanning and blocking files at the point of uploading.

Figure 1 – Examples of blocking

4.1 IWF CAIC List (“Cleanfeed”)

Since 1996 the UK has seen the development of an industry-led response to child abuse images. A private body funded by the internet industry and the EU – the IWF – has acted as a hotline which works with police and government to receive public complaints and determine whether particular web pages contain potentially illegal material (including but not limited to child abuse images). If so, the IWF then forwards those complaints to the police and (where material is hosted in the UK) to the hosting provider in order to have that material removed (Walden 2010).

This approach has been remarkably successful at reducing the hosting of illegal material in the UK. It was, however, effective only in relation to domestic material. Where child abuse images were hosted abroad, takedown was dependent on the actions of local authorities and the material would remain available to UK users in the interim – or indefinitely where no local action was taken.

This limitation prompted British Telecom (“BT”) to develop a system which would block access to web pages hosted outside the UK. The technical system which they produced – dubbed “Cleanfeed” – represented a substantial step forward over the two main forms of web blocking then in use (IP address blocking and DNS poisoning). By using a two stage approach to blocking which combined redirection of traffic with the use of web proxies it filtered at the level of the full URL and appeared to minimise collateral damage. As compared with DNS poisoning, for example, it was capable of selectively blocking only <http://example.com/users/johndoe/lolita.jpg>, rather than all the material hosted at example.com (Clayton 2005). In addition, it should be noted that BT deliberately designed this system in such a way as to avoid logging data on users – effectively precluding its use for prosecution purposes and enabling them to present it as being solely for the protection of their customers (Hutty 2004).

Having developed this system, BT then persuaded the IWF to make its database of URLs available for blocking purposes. This was done in 2004, when the IWF first distributed its CAIC list to members. In mid-2004, therefore, BT began to trial the Cleanfeed system. Following the apparent success of this trial and the proof of concept it provided, there soon followed substantial pressure from politicians and children’s advocacy groups for other ISPs to follow BT’s example – including Home Office threats to introduce legislation compelling blocking unless ISPs “voluntarily” complied (Hargrave 2006).

This pressure convinced almost all UK ISPs to introduce filtering systems similar to BT’s Cleanfeed, and government plans for legislation were ultimately abandoned in 2009 following an Ofcom survey which established that 98.6% of home connections were subject to blocking systems. The UK government remains committed to 100% coverage, however, and has relied on consumer pressure as well as its own purchasing power as a means of encouraging compliance amongst the remaining smaller ISPs (Williams 2009; O’Neill 2010).

At the time of writing, therefore, there is near universal coverage of UK users by blocking systems which filter against the IWF CAIC list. There is also a spill over effect to ISPs in many other jurisdictions (such as Ireland) where the IWF list is used in the absence of a local blocking system (GSMA Mobile Alliance Against Child Sexual Abuse Content 2008). In addition, the IWF list is widely deployed in home, workplace and school filtering software and is also used by search engines (including both Bing and Google) on a worldwide basis to remove URLs from search results (Internet Watch Foundation 2011b). When considered in terms of numbers of users covered, therefore, the IWF list may well be the most widely used

blocking list ever. The UK model has also been influential elsewhere, and the name “Cleanfeed” has stuck as a generic term for UK blocking systems as well as related schemes in Canada and Australia (see e.g. Watt & Maurushat 2009).

It is striking, however, that this system has developed without any legislative basis, and has done so in a way which entrusts a private body with the role of determining whether content is “potentially illegal” with limited procedural safeguards and no judicial oversight. This became the subject of controversy in 2008, when the IWF added certain pages on Wikipedia to its URL list – before backing down and reversing its decision just five days later following a storm of public criticism (Davies 2009).

That episode focused public attention on the system and highlighted many issues raised by blocking. One of the first related to the blocked content itself. The pages blocked by the IWF did not match the public perception of child abuse images – instead, they contained a well known album cover from 1976 featuring a nude photograph of a prepubescent girl. While this image may well have been “potentially illegal” under English law the overwhelming public view was that it should not have been blocked – not least because the album itself remained for sale in UK record shops. This in turn focused public attention on the basis of the power of the IWF to make censorship decisions for the entire UK internet (Edwards 2009).

Substantial collateral damage also emerged. Despite the claimed superiority of two stage URL blocking systems, it soon became clear that many users found themselves unable to edit Wikipedia – even pages completely unrelated to the block – due to the use of proxy servers as part of the blocking system (Clayton 2008).

The Wikipedia incident also demonstrated a remarkable lack of transparency and procedural safeguards. There was no notice given to Wikipedia either before or after its pages were blacklisted, and most ISPs presented deceptive error messages to users who attempted to access the blocked pages – with the notable exception of Demon Internet which notified users of the blocking via the stop page illustrated in Figure 2.

Figure 2 – Demon Internet Block Page, <http://iwfwebfilter.thus.net/error/blocked.html> (accessed 16 May 2011)

In addition, as Wikipedia soon discovered, the IWF system does not provide for any judicial appeal against its decisions – while there is an internal review procedure, the only external input into that system comes from the police (Internet Watch Foundation 2010a).

Some of the issues raised by the Wikipedia incident have since been addressed by the IWF – in particular, new policies allow it to use greater discretion in relation to borderline cases where blocking is likely to be counterproductive, while greater emphasis is now placed on seeking the removal of material at source where possible (Internet Watch Foundation n.d.). There remains, however, substantial controversy as to the role of the IWF. The majority of commentators would appear to share the views of Edwards (2009), who argues that if a blocking system is to be implemented then it should be put on a statutory basis. As against that, however, there is a strong minority view which argues that the IWF – precisely because of its industry-led nature – has served as a buffer against further state regulation of the internet (see e.g. Walden 2010).

4.2 CIRCAMP

Within Europe, the single most common type of blocking is based on the EU funded CIRCAMP (COSPOL Internet Related Child Abuse Material Project) model. As with Cleanfeed, this also focuses on blocking at the ISP level – unlike that system, however, the CIRCAMP approach relies on police to designate what material is to be blocked (McIntyre 2010).

CIRCAMP has its origins in Norway which, in 2004, paralleled the UK by adopting a national child abuse material blocking system. Unlike Cleanfeed, however, the Norwegian system was police-led so that decisions as to which domains to block were made by the National Criminal Investigation Service. In addition, that system used DNS blocking only,

rather than the hybrid URL based blocking associated with most Cleanfeed implementations (Deibert & Rohozinski 2010).

The experience of the Norwegian police in operating their domestic blocking system later led to Norway becoming the primary driver of the CIRCAMP project. From 2006 onwards this project has helped national police forces to adopt Child Sexual Abuse Anti-Distribution Filters (CSAADF) which are closely modelled on the Norwegian system. Currently eight countries – Denmark, Finland, Italy, Malta, Norway, Sweden, Switzerland and New Zealand – are using CSAADF blocking systems. This is generally done on a voluntary basis by ISPs, without any legislative underpinning.

The CIRCAMP project has followed the Norwegian approach by promoting the use of DNS blocking over other forms of blocking. Interestingly – and unlike most other blocking systems – it embraces the resulting overblocking by claiming that it serves as a deterrent to domain owners:

The CSAADF focuses on blocking on domain level. We believe that this places the responsibility for the content of any domain or sub domain in the hands of the domain owner or administrator. If a domain owner places, accidental or willingly, child abuse material on his/her domain, and it is blocked by the police, the blocking will not be lifted until the material is removed. We believe that this will motivate content providers on the Internet to actively make an effort to avoid files with child sexual abuse on their systems/services. (CIRCAMP n.d.)

There is an exception, however, for certain hosting sites where CIRCAMP members will not block but will instead notify the owners seeking removal of the image:

In cases where a hosting company has been taken advantage of, like free photo hosting companies – CIRCAMP members will inform the owner/administrator of that domain that they are hosting child sexual abuse material. In most cases this will result in the removal of the files very quickly. Such services are not blocked as the implications for legal users and services would be substantial. (CIRCAMP n.d.)

The CIRCAMP project also provides for information sharing between national police forces and in particular the sharing of black lists – though the decision as to which material is to be blocked remains a decision for national police forces, applying national law. CIRCAMP has also worked with INTERPOL on developing a “worst of” list of domains containing images of particularly serious sexual abuse that would be illegal in almost all jurisdictions.

As compared with the early Cleanfeed systems, CIRCAMP makes some advances in relation to transparency and procedural safeguards. While the IWF would not (until recently) notify a domain owner that a site had been blocked, the CIRCAMP model requires notification in respect of image hosting sites and also in situations where a “legal domain or home page/company page of some sort” appeared to be compromised. In this case the site owner is contacted, told of the hacking or abuse and given the opportunity to stop the blocking by confirming that the child abuse material had been removed (CIRCAMP n.d.).

Similarly, while the IWF still does not require that users be notified about blocked pages the CIRCAMP system has from the outset emphasised the use of stop pages which contain “information about what kind of content the users browser tried to access, links to national legislation, contact information to complain about the blocking and to the police” (CIRCAMP n.d.). Figure 3 provides an example of a stop page from Malta.

**Cyber Crime Unit
Malta Police Force**

Il-browser ta' l-internet tieghek qiegħed jipprova jikkomunika ma' sit li qiegħed jiġi użat fit- tqassim ta' stampi ta' abbuż sesswali fuq tfal.

DAN HUWA KKUNSIDRAT BHALA REAT KRIMINALI SKOND IL-LIGIJIET TA' MALTA. (KAP. 9, ART. 208A).

L-iskop ta' twaqqif milli wieħed jaċċedi f'dawn is-siti huwa sabiex tiġi evitata l-kommissjoni ta' reat tat- tqassim ta' materjal li juri l-abbuż sesswali fuq tfal u sabiex jiġi evitat l-isfruttar ta' tfal li diġa ġew abbużati u fotografati. M'hiex qegħda tinzamm informazzjoni li tista twassal għal identifikazzjoni tiegħek meta tiġi muriġa din il-paġna.

Jekk għandek xi kummenti jew mistoqsijiet, inti mitlub/a tikkuntatja lis-Cyber Crime Unit ta' Pulizija ta' Malta fuq:
computer.crime@gov.mt

Dan huwa proġett kongunt bejn il- Pulizija u l- Provdituri ta' Servizzi Elettronici.

Contributors

Partners

Your internet browser is trying to connect to a website that is used in connection with the distribution of images depicting sexual abuse of children.

THIS IS A CRIMINAL OFFENCE IN TERMS OF MALTESE LEGISLATION (CHAP. 9, ART. 208A).

The purpose of blocking such pages is to prevent the criminal dissemination of documented sexual abuse of children and to prevent the further exploitation of children who have already been abused and photographed. No information that can be used to identify you will be stored when this page is displayed.

Should you have any comments or queries please contact the Malta Police Cyber Crime Unit by email on:
computer.crime@gov.mt

This is a joint project between the Police and Service Providers.

Figure 3 – CIRCAMP Stop Page, Malta, <http://www.mpfstopchildabuse.org/> (accessed 20 July 2011)

Also, as part of the CIRCAMP system EUROPOL now provides a web page for domain owners which enables them to seek a review of the blocking in each jurisdiction through a single request, rather than having to contact each jurisdiction individually (EUROPOL n.d.).

As with Cleanfeed, the system is not intended for prosecution purposes and CIRCAMP explicitly states that “the access blocking is purely preventive, no investigations against

persons are initiated as a result of an Internet user being blocked and the ‘stop page’ displayed”. However, the CIRCAMP model goes further and envisages that national police forces will also use blocking systems as an intelligence tool:

In most participating countries the ISPs grant the police access to web logs that are generated when the “stop page” is displayed. The IP-address of the Internet users has been removed from the logs, so they contain no identifying information. These logs are used for statistic purposes and will provide information about new sites that are unknown to the police. The statistics from these logs will also provide an overview of the Internet usage related to child sexual abusive material in addition to information about search words, type of operating system, browser, time of day that most Internet users are redirected to the “stop page” etc. (CIRCAMP n.d.).

The effect of this is made clear in a recent letter from Irish police to ISPs proposing the introduction of a CSAADF system. That letter acknowledges that users may have accessed a blocked site inadvertently, but goes on to request that in such cases the ISP should provide “details of other websites visited by the user” (Digital Rights Ireland 2011). This raises obvious privacy concerns, not least as it is often possible to identify users based on their internet history, and these are considered further at 5.6 below.

4.3 United States hash value blocking systems

The systems discussed above focus on blocking access to particular web addresses and between them reflect the majority of blocking systems in Europe.² There is a similar system in the US – since 2008 the quasi-public National Center for Missing and Exploited Children (NCMEC) has operated a “URL Project” which provides participating ISPs with a list of URLs it has found to contain “the worst of the worst” forms of child pornography.³ However that has not promoted blocking to the same extent as either the Cleanfeed or CIRCAMP models – while many ISPs subscribe to this list, the focus is on takedown of material hosted by those providers rather than blocking of material hosted elsewhere (Hakim 2008).⁴

Instead, a different form of blocking has been more prominent which focuses on the file itself rather than where it is located (see e.g. Anderson 2007). This approach relies on the use of hash values, which in effect serve as fingerprints to uniquely identify a particular file or photograph (for more detail see e.g. Salgado 2006). Where an internet intermediary has a database of hash values known to correspond to child pornography files then they can compare the hash values of files stored or transmitted by users and, if there is a match, they will be able to identify the file in question as constituting child pornography.⁵

AOL pioneered the use of this strategy through its Image Detection and Filtering Process (“IDFP”) which it has run since 2004. Figure 4 (adapted from Colcolough 2009) illustrates how it works.

Figure 4 – AOL's Image Detection and Filtering Process

As Figure 4 shows, the IDFP scans all emails sent by AOL members, generating hash values for any images being transmitted. Those hash values are then compared with an internal database containing the hash values of child pornography images previously dealt with by AOL. If there is a match AOL will block the email. At that stage, having knowledge of the child pornography, it is obliged by US mandatory reporting rules to notify the Cyber Tip Line at the NCMEC by sending a report containing the image, username, email address and zip-code of the user.⁶ The NCMEC will in turn notify the relevant law enforcement agency which can subpoena AOL for full details of the user.

This system has resulted in numerous convictions and has been influential in promoting other hash value blocking systems within the US. At the federal level, in 2008 Congress passed the PROTECT Our Children Act⁷ which specifically authorises the NCMEC to provide hash values to ISPs for the purpose of detecting and blocking child pornography (but doesn't require that ISPs either monitor or block users' communications). Similarly, at the state level the New York Attorney General's office has established its own hash value database, which is now being used by Facebook, MySpace, isoHunt and others to detect and block uploads of child pornography images (Office of the Attorney General 2010).

These systems are, however, controversial and AOL's IDFP in particular has been criticised for the way in which it scans private emails. Although Fourth Amendment challenges to the AOL system have been unsuccessful (as the courts have not accepted that AOL should be treated as a state actor) it has been argued that this type of mass surveillance is a worrying development – one which is easily capable of being extended to other material which might be suppressed by government (Soghoian 2010a, pp.12-14).

As against that, however, there is also an opposing view that the use of hash value blocking is minimally intrusive (similar to spam filtering) in that such automated monitoring reveals nothing about the contents of communications beyond a binary determination: that the file is,

or is not, known child pornography. Indeed, for this reason has been suggested that hash value scans should not be treated as searches for the purpose of the Fourth Amendment (see e.g. Salgado 2006; Morrison 2011).

It should also be noted that there is a division of opinion within the US as to how blocking systems should be implemented – in particular, whether there is a role for the state in distributing hash values of known child pornography images. For example, AOL has publicly stated its concern about using a government supplied list, fearing that by so doing it would be considered an agent of the government (Dedman & Sullivan 2008). Conversely, the New York example shows that Facebook and others are content to block against hash values supplied by New York law enforcement authorities.

Leaving aside this debate for the moment, however, it will be apparent that hash value blocking may have several advantages over either the Cleanfeed or CIRCAMP models. Systems such as those operated by the IWF or CIRCAMP members do not directly identify child pornography images, but instead point to locations. At best they can merely say that child pornography was found at a particular location at a particular time. Consequently, they require manual updating and review of each web address and will fail to detect the same image when moved to a new location. Each new location, therefore, will require fresh human intervention to block. Hash value blocking, on the other hand, does not rely on the image location and will correctly identify and block files even though they are being transmitted from a new location – and can also be applied in contexts (such as email or peer to peer) where DNS or URL based blocking will fail. While older forms of hash value matching (such as MD5 hashes) could be defeated by minor changes to files, newer “robust hashing” systems such as Microsoft’s PhotoDNA are capable of identifying and blocking photographs even if they have been edited, resized or cropped (Whittaker 2009). Hash value blocking may also minimise concerns about overblocking – depending on the precise system used, false positives should be minimal.⁸

5. Criticisms of blocking systems

Blocking systems have been questioned by many who fear that they may undermine freedom of expression online. The starting point for these critics is that internet blocking is, at its core, a form of restriction of freedom of expression and as such should comply with the democratic and constitutional norms associated with such restrictions. Instead, the argument runs, blocking may enable governments to sidestep these norms (Brown 2008). The following section considers these criticisms in light of the case studies above.

5.1 Transparency

A fundamental aspect of freedom of expression is that limitations of this right should be transparent and therefore subject to public oversight. Article 10 of the European Convention on Human Rights (“ECHR”), for example, states that any restrictions should be “prescribed by law” – which requires amongst other things that the legal basis for restrictions should be adequately accessible to the citizen.

However, blocking systems present significant challenges for transparency. Lessig has noted that regulation by code is inherently opaque, so in the case of internet blocking the user may not know that it is taking place, who is responsible or what material is being blocked. Consequently, he cautions that without “truth in blocking” these systems are likely to

undermine free speech (Lessig 1999). Some blocking systems (such as CIRCAMP) have responded to this concern by introducing “stop pages” which notify users when their access to a web page has been blocked. Unfortunately others (notably the IWF) do not require this, permitting the deliberate deception of users as to why content is unavailable, and hindering any attempts to remedy wrongful blocking.

The focus on intermediaries presents its own problems. Unlike traditional systems for controlling content (which generally target either the speaker or the reader) blocking can be deployed in a covert manner unbeknownst to anyone but the intermediary. In the same vein, controls which are drawn up by self-regulatory systems generally escape the publicity which would attach to legislation or judicial decisions. As a result, Deibert and Villeneuve (2004) have noted that blocking systems are generally murky in their operation:

as the practice of Internet content filtering and surveillance is largely new territory, the rules by which states implement such controls are poorly defined, not well known among the general public, and very rarely subject to open debate ... as it stands now such decisions are typically taken behind closed doors through administrative fiat.

These concerns are all the greater in the case of child abuse images where regulators will understandably seek to keep the list of blocked material secret. While secrecy may be necessary to avoid blacklists becoming an index for paedophiles, it also makes it difficult to monitor the operation of such systems and forces society to take a great deal on trust. Unfortunately, this trust may not always be warranted. Instead, where blacklists have come to public attention this has often revealed that these systems have been poorly operated.

A recent example came from a CIRCAMP system in 2010 when a police blacklist shared between Sweden and Denmark was leaked. Volunteers from the German anti-blocking group AK Zensur confirmed that the domains on the list were currently blocked in Denmark, and then visited each website to assess whether it was correctly listed. Out of a representative sample of 167 websites, they found that 92 sites had already had their hosting accounts terminated, 66 domains had expired and 6 sites did not contain any illegal content, leaving only 3 sites which in fact contained child abuse images. This appeared to demonstrate a failure on the part of the Danish authorities to keep the blacklist current and, more importantly, to ensure that legal content was not blocked – a failure which would not have come to light otherwise (AK Zensur 2010).

It also, significantly, illustrated a further challenge for transparency. The volunteers who visited each website were not named in the study – reflecting their fears that simply visiting the blocked sites might constitute an offence. Where the law presents such risks for researchers it makes it all the more difficult to exercise informal oversight by civil society – even though the formal oversight mechanisms might themselves be deficient.

5.2 Legitimacy and accountability

The IWF ... is supported by the Police and CPS and works in partnership with the Government to provide a 'hotline' for individuals or organisations to report potentially illegal content and then to assess and judge that material on behalf of UK law enforcement agencies.

– Crown Prosecution Service & Association of Chief Police Officers (2004)

I regret to inform you that the Home Office does not hold the information that you have requested regarding the relationship between the IWF and the Home Office. The IWF is a self regulatory, independent charity that has no formal links with the Home Office.

– Home Office, Response to Freedom of Information Act Request (2009)

Another common charge against blocking is that it lacks legitimacy and accountability. More precisely, the claim is that such systems – insofar as they can be adopted informally by private actors in response to government pressure – evade requirements that state measures which restrict freedom of expression should have a legislative basis, and avoid public law oversight mechanisms. As Marsden (2010) puts it “government favours more private censorship with loose – and therefore largely unenforceable – links to the government, but very strong policy and informal bonds”. This is not an inevitable feature of blocking systems, some of which do have a legislative basis. It is, however, extremely common.

A particularly good example is the Dutch system, adopted in 2007, which involved ISPs voluntarily blocking access to domains designated by the police, using DNS blocking. A study commissioned by the government found that this was unlawful and contrary to Article 10 ECHR in that it lacked any specific legal basis – ultimately forcing it to be abandoned (Stol et al. 2008; Stol et al. 2009). Remarkably, however, when this system was found to be illegal, the response of the Dutch government was not to provide a legal basis, but instead to try to further privatise blocking. The tactic adopted was to seek to persuade ISPs to develop a purely self-regulatory scheme – in which the sites to be blocked would be designated by a private body rather than by the police – thus avoiding the safeguards which would apply to a state run system (Bits of Freedom 2011).

The Dutch experience illustrates the shifting focus of these blocking systems: away from public bodies which are bound by constitutional constraints and towards private bodies such as ISPs which are insulated from judicial review. Lambers (2006) has described this approach as “tilting” where the “classical vertical state-citizen relationship on which... freedom of speech is founded, is short circuited since a second private party shifts between the state and the user: the ISP”. He graphically represents this “tilt” in Figure 5 below.

Figure 5 – Lambers’ model of “tilting”

Consequently, he argues, where non-legislative blocking is introduced the relationship between state and citizen becomes instead a relationship between ISP and user – one which is governed by private law only, deliberately displacing constitutional and public law rights.

This aspect of blocking has led critics such as Edwards (2009) to argue that if blocking systems are to be used then they should be reconstituted as public bodies – making them accountable to the ordinary mechanisms of public oversight and judicial review. As against that, however, there is a contrary view exemplified by Mueller (2010) which identifies the “saving grace of privatised governance” as the “ability of users and suppliers to vote with

their feet”, suggesting that if blocking is put on a statutory basis it is likely to become more rather than less pervasive. In practice, however, any significant customer response seems unlikely to happen for two reasons. First, the self-regulatory systems which we describe are often opaque in their nature, making it difficult for customers to understand what content is being restricted and by whom. Secondly, these systems are also often adopted on a universal or near universal basis, so that even where customers are aware of particular restrictions they may nevertheless have no realistic alternative. The UK example – where 98.6% of the population are covered by Cleanfeed type systems – offers an example of a situation where exit is not a realistic option for most users.

It is, therefore, difficult to argue with the recent report commissioned by the OSCE Representative on Freedom of the Media which rejects the use of “voluntary” or self-regulatory systems, concluding that:

There is concern that voluntary blocking mechanisms and agreements do not respect due process principles within the states in which they are used. In the absence of a legal basis for blocking access to websites, platforms and Internet content, the compatibility of such agreements and systems with OSCE commitments, Article 19 of the Universal Declaration and Article 10 of the European Convention on Human Rights is arguably problematic. Although the authorities’ good intentions to combat child pornography and other types of illegal content is legitimate, in the absence of a valid legal basis in domestic law for blocking access to websites, the authority or power given to certain organizations and institutions to block, administer, and maintain the blacklists remains problematic. Such a “voluntary interference” might be contradictory to the conclusions of the Final Document of the Moscow Meeting of the Conference on the Human Dimension of the CSCE and in breach of Article 19 and Article 10 of the European Convention on Human Rights unless the necessity for interference is convincingly established. (Akdeniz 2011, p.24)

5.3 Fair procedures

The complaint that internet blocking systems evade public law norms is particularly strong in relation to fair procedures – notably the right to be heard before a decision is made. This is not a facility which has been offered to site owners or users in most internet blocking schemes worldwide, despite the fact that blocking will operate as a prior restraint of speech – at best, the operators of internet filters generally provide (if at all) for review after the fact (Deibert & Villeneuve 2004). In response, it has been argued the norms of administrative decision making may not always be appropriate in the context of child abuse image blocking. For example, it has been claimed that to notify a site owner may jeopardise criminal enforcement (see e.g. Walden 2010).

Whether this reasoning would resist legal challenge will depend on the standards of each national system. In the United States, for example, the court in *Centre for Democracy and Technology v. Pappert*⁹ found that a legislative scheme whereby websites could be blocked by court order on an *ex parte* basis, with no notice or opportunity to be heard, did not meet the procedural requirements which the First Amendment required for a prior restraint to be imposed (see e.g. the discussion in Kleinschmidt 2010).

Of course, not all jurisdictions share the US suspicion of prior restraints. But at a minimum, notice after the fact and an independent appeal mechanism would appear to be necessary to provide adequate procedural safeguards. Most systems, however, do not provide for any notification of the site owner – even where users attempting to visit a site are presented with a block page (see e.g. Internet Watch Foundation 2010b). Similarly, none of the systems described here include any judicial oversight, and where appeal mechanisms are provided

they do not always provide for an independent review or even a right to make submissions. For example, in 2008 when the IWF blocked a number of pages on Wikipedia, the review which was carried out excluded any input from Wikipedia itself, causing their lawyer to comment that:

When we first protested the block, their 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. (Quoted in Davies 2009)

5.4 Overblocking

Internet blocking systems are often criticised as being disproportionate in their effect – that is, as being prone to causing collateral damage by blocking legal as well as illegal material. Both the IWF and CIRCAMP experiences bear this out – and it is striking that the CIRCAMP model deliberately adopts overblocking as a tactic to exert pressure on site owners.

The extent to which such overblocking takes place in any particular scheme will, of course, depend on a number of factors including the technological sophistication of the blocking system used and the diligence of those establishing and maintaining the blacklist. In general, however, the incentives faced by the ISPs and others who implement blocking systems favour overblocking. As Kreimer (2006) notes, the dominant motive of intermediaries is “to protect themselves from sanctions, rather than to protect the target from censorship”. This reflects empirical evidence showing that internet intermediaries make decisions in a manner which minimises their own financial, legal and reputational risk (see e.g. Ahlert et al. 2004). Consequently, there is likely to be a structural tendency towards overblocking in many blocking schemes.

5.5 Mission creep

Child pornography is great... Politicians do not understand file sharing, but they understand child pornography, and they want to filter that to score points with the public. Once we get them to filter child pornography, we can get them to extend the block to file sharing.

– Johan Schlüter, Chairman of the Danish Anti-Piracy Group (Quoted in Falkvinge 2011)

An important criticism of blocking systems is that they are prone to mission creep – that is, that once established for a particular purpose they may easily be extended to achieve a different goal. In relation to child abuse image blocking systems, this mission creep may take place in one of two ways.

The most commonly mentioned is that other material may be brought within their scope – for example, they may be extended to also block filesharing, suicide, pro-anorexia, etc. sites. Edwards (2009) points out that the UK government has considered extending child abuse image blocking to sites which “glorify terrorism” and argues that the IWF system enables this to be done in a way which is invisible to the public. Indeed, Mueller (2010) goes further by arguing that mission creep is a feature rather than a bug, noting that “emotional appeals to ‘the children’ have deliberately been exploited as the entering wedge for a broader reassertion of state control over internet content”.

It might be objected that mission creep is less likely in self-regulatory systems where ISPs have a financial incentive to minimise the scope of blocking. This argument is sometimes

made in the UK in defence of the IWF-led system – Ozimek (2009) for example typifies this view when he expresses a preference for its “slightly quaint, non-governmental route” as being “rather less threatening... than the more ‘efficient’ [state-run] models used elsewhere”.

There is undoubtedly some truth in this point, but it is significantly undermined by the fact that once a blocking infrastructure is in place it may be co-opted by others against the wishes of the ISP. Ironically, Cleanfeed itself illustrates this point. At the time of writing, the Motion Picture Association of America is suing BT, seeking an injunction requiring it to block access to a website (Newzbin) which is alleged to allow the illegal downloading of movies. According to a spokesman “BT was chosen because it’s the largest and already has the technology in place, through its Cleanfeed system, to block the site” (Williams 2011).

A potentially more difficult (though less often discussed) aspect of mission creep is that the objective of blocking may be expanded from crime prevention to also take on an intelligence, investigation or prosecution role – for example, by using a particular system to identify and prosecute users who seek to access or transmit child abuse images. This will be especially true in jurisdictions such as the United States where there is mandatory reporting of offences related to child pornography – in those cases, by operating a blocking system an ISP will come under an obligation to report those users whose actions have been flagged (see Morrison 2011).

As we have seen, some ISPs (notably AOL) have embraced this expansion of blocking to encompass a prosecution role, while others (such as BT) have sought to avoid this possibility by minimising the data which they log about their users. However, the US experience shows that any blocking system can easily be repurposed as a prosecution tool by introducing mandatory reporting by ISPs where they have knowledge of child pornography. In this case, voluntary blocking coupled with mandatory reporting can become, in effect, ongoing surveillance of the entire user base.

This is especially so with hash value systems as compared with other forms of blocking. Cleanfeed or CIRCAMP web blocking systems doesn't easily facilitate prosecution. These systems are intended to stop access to material hosted elsewhere – outside the control of the user – and the IWF and others have been at pains to stress that the main goal of such systems is to prevent “inadvertent exposure”. Consequently, if a user is prevented from accessing a site then there is little or no proof that they have committed or intended to commit a crime. Hash value blocking, on the other hand, can also be used for situations where a user attempts to make material available to others: for example, by scanning email attachments sent by a user (AOL) or images uploaded by a user to a shared group (Facebook). In these situations, if a blocking system detects a positive match then that in itself is evidence of the crime of possession on the part of the user and is likely to trigger any mandatory reporting requirement.

More generally, however, this type of mission creep presents significant risks for the criminal justice system. By introducing pervasive surveillance of all users – without any prior suspicion – even a low rate of false positives may result in the wrongful investigation, arrest and stigmatisation of many innocent users.

These risks can be seen by examining a previous large scale data-driven investigation of alleged child pornography offences. In 1999, a police investigation in the United States (“Operation Avalanche”) led to the seizure by the US Postal Service of credit card records

which appeared to implicate many tens of thousands of internet users in the purchase of child pornography. Of these, 7,272 records related to individuals in the United Kingdom. After these records were provided to the UK, in April 2002 the National Crime Investigation Service (NCIS) launched an investigation (“Operation Ore”) which ultimately resulted in the investigation of 4,283 individuals. As these cases proceeded, however, it became clear that many of those individuals had not paid for child pornography – instead, they had either been the victims of credit card fraud, or had paid for legal (adult) pornography sites which shared the same billing service (Campbell 2005). This, however, came too late for many of the individuals concerned, at least some of whom committed suicide as a result of the wrongful accusations against them while others lost their jobs as a result (Leppard 2005).

5.6 Privacy

Blocking systems pose a special challenge to legal norms relating to privacy, confidentiality of communications and data protection. These systems, of their nature, often involve the monitoring of internet traffic generally with a view to deciding which particular messages to block. Except in a few cases – for example, where blocking software is run at a purely local level under the control of the end-user – the operation of blocking can therefore involve third party pervasive surveillance of otherwise private communications (see e.g. Callanan et al. 2009, chap.6). There has, however, been relatively little examination of the issues this presents.

To the knowledge of this author, there have been no court cases which examine the operation of either the UK Cleanfeed system or the European CIRCAMP systems. In the United States there have been a number of defence challenges to prosecution evidence obtained as a result of the AOL IDFP system – in those cases, however, the challenges have invariably failed on the basis that the Fourth Amendment guarantee against “unreasonable searches and seizures” applies only against the state and not against an ISP acting in a private capacity. The most important case on point is *US v. Richardson*¹⁰ where the Fourth Circuit held that AOL was not acting as an agent of the government in scanning email, notwithstanding that it actively cooperated with law enforcement and was obliged by law to report any child pornography which it discovered to the NCMEC, based on a finding that there was “little evidence... to suggest that AOL intended to assist the Government” (see e.g. Morrison 2011).

In the US context, therefore, the voluntary nature of blocking may insulate it from judicial scrutiny.¹¹ It is probable, however, that a different result would be reached in a European context where both the European Convention on Human Rights and data protection guarantees recognise privacy rights which have horizontal effect so that they can be asserted against non-state actors. Indeed, a recent opinion of the European Data Protection Supervisor (“EDPS”) suggests that such systems may be in breach of the Data Protection Directive¹² and Article 8 ECHR where they are introduced without a statutory basis:

The EDPS underlines that monitoring the network and blocking sites would constitute a purpose unrelated to the commercial purpose of ISPs: this would raise issues with regard to lawful processing and compatible use of personal data under Article 6.1.b and Article 7 of the Data Protection Directive. The EDPS questions the criteria for blocking and stresses that a code of conduct or voluntary guidelines would not bring enough legal certainty in this respect. The EDPS also underlines the risks linked with possible blacklisting of individuals and their possibilities of redress before an independent authority. The EDPS has already stated at several occasions that “the monitoring of Internet user's behaviour and further collection of their IP addresses amounts to an interference with their rights to respect for their private life and their correspondence... This view is in line with the case law of the European Court of Human Rights”. Considering this interference, more appropriate safeguards are

needed to ensure that monitoring and/or blocking will only be done in a strictly targeted way and under judicial control, and that misuse of this mechanism is prevented by adequate security measures. (Hustinx 2010)

Despite these issues, however, privacy has often been overlooked in the literature on filtering. Bambauer (2009) for example has put forward a very useful four part metric for evaluating blocking systems which considers “openness, transparency, narrowness and accountability” – but leaves out of this metric any impact which particular systems may have on privacy of communications. Similarly Akdeniz’s recent analysis of European blocking measures focuses on freedom of expression, leaving privacy issues aside (Akdeniz 2010).

This tendency to neglect privacy may reflect a focus on systems such as Cleanfeed and CIRCAMP where material targeted is publicly available on the web, creating fewer privacy problems. Privacy issues are becoming more important, however, with the growth of hash value blocking systems such as AOL’s IDFP which – especially in conjunction with deep packet inspection – now make it feasible to target entirely private channels of communication such as email or instant messaging.¹³

It will be important, therefore, for future research to consider the privacy implications of these newer systems and whether indiscriminate and pervasive surveillance of this sort can ever be justified, however grave the material targeted. In particular, it would be desirable to assess individual measures with regard to their invasiveness and to reaffirm the principles of proportionality and necessity so that more invasive systems (such as the scanning of email) should only be used if it can be shown that less invasive systems (such as blocking of public web sites) would not achieve the desired goals.

5.7 Effectiveness

Are blocking systems effective? To answer this question we must first ask a preliminary question – effective in relation to what goals? This is a surprisingly difficult question to answer as few blocking systems set explicit objectives (see e.g. Stol et al. 2009). This (sometimes deliberate) vagueness reflects a tension between two competing factors – a political tendency to oversell what can be achieved and the technical realities which limit what can be done. However, we can take as our starting point the following summary from two prominent advocates of blocking:

- Blocking is a way of interfering with and disrupting the commercial trade of child abuse material
 - Blocking helps to prevent accidental access to this illegal and harmful content by helping the public
 - It helps to prevent deliberate access to child abuse material on the internet
 - It helps to reduce the customer base of illegal websites
 - It helps to prevent the re-victimization of those children who are or have been the victims of abuse.
- (Carr & Hilton 2011)

The distinction between deliberate and accidental access in this summary is significant – Carr and Hilton acknowledge that blocking can be circumvented, but go on to argue that it nevertheless has a role “in helping to prevent the casual, domestic consumer from stumbling across child abuse images by accident and in preventing those who might have a misguided sense of curiosity from gaining access”. In this they echo a rationale common to most such systems – i.e. that they can serve to protect the innocent or inquisitive user even if they are ineffective at stopping the deliberate criminal.¹⁴

It is easy to see why this paternalist rationale has become the dominant argument of advocates of blocking. Circumvention methods are no secret, and research such as that of Eneman (2010) has demonstrated that sex offenders – even those without any formal education or experience in working with computers – already find it easy to defeat blocking systems. In addition, public awareness of circumvention tools is on the rise. The use of blocking and geolocation as means of enforcing copyright has ensured that users are increasingly familiar with the use of proxy servers, alternative DNS providers and services such as TOR – whether to access sites such as ThePirateBay which are blocked by their ISP or to view services such as the BBC iPlayer which are not available in their country (see e.g. Svantesson 2008). Consequently, arguments based on stopping accidental and casual access take on greater importance as it becomes clear that blocking is at best only weakly effective at stopping deliberate viewing.¹⁵

To what extent, then, are blocking systems effective at preventing accidental or casual access to child abuse images? Here, unfortunately, we are hampered by a lack of data. In the first place, there does not appear to be any evidence that accidental exposure has been a significant problem. In their recent Dutch study Stol et al. (2009) point out that:

No interviewed expert, authority or other person involved was able to refer to a case in which a “decent” internetter was unexpectedly or incidentally confronted with child pornography on a website.

It may be that such systems are more effective at blocking casual viewing, but there is a lack of data in this regard also.¹⁶ Few blocking systems have made statistics available as to the extent of access attempts which are blocked, and where data has been made available it has generally been unreliable.

A well known example comes from the UK where BT has published statistics from its Cleanfeed system claiming (most recently) that it has blocked up to 45,000 hits per day. While these claims have been uncritically reported by the mainstream media as demonstrating the success of blocking, closer analysis has revealed substantial issues with those figures. Notably, by counting “hits” rather than “page visits” it overstates the issue, as an attempt to visit a single page will almost always generate multiple hits for the files which make up that page. In addition, sources familiar with the system have acknowledged that a substantial portion of that traffic is likely to be generated by malware or foreign users seeking to abuse open proxies within the UK, something which again undermines the claims that casual viewing is being prevented. Ironically, the steps which BT has taken in designing the system (for example, not logging the IP addresses which attempted to reach a blocked site) ensure that no conclusive analysis of the figures can be carried out. (Richardson 2004a; Richardson 2004b; Graham 2009).

Finally, it should be noted that there is a strong case that the use of blocking systems has been counterproductive, by distracting attention from international measures to achieve the removal of images at source. Villeneuve (2010), for example, has argued that “the introduction of filtering technology reduces the incentive for organisations with an already narrow conception of cooperation to further engage with relevant counterparts across international boundaries”. German anti-blocking group AK-Zensur illustrated this point in 2009, when using a leaked blocking list they succeeded in taking down 61 child pornography websites simply by contacting the hosting providers (Freude 2009). Research by Moore and Clayton (2008) has demonstrated that in relation to financial crimes it is possible to achieve

effective cross-border cooperation without any need to resort to national blocking systems, supporting the argument that child abuse images could similarly be dealt with.

6. Conclusion

It has often been claimed that the “success” of internet blocking for child abuse content should be followed by extending blocking to other forms of internet content. When examined more closely, however, it is apparent that the claims made for blocking must be heavily qualified and do not support the further extension of these systems.

As we have seen, child abuse images represent probably the best case scenario for blocking. There is near universal agreement on the illegality of such material and considerable public support for countermeasures. From a practical perspective, such images are relatively straightforward to identify and the comparatively small number of sites involved makes it technologically and administratively more convenient to introduce blocking systems. These advantages do not, however, apply to the majority of other content which states seek to control, making the experience of child abuse blocking marginally relevant at best.

More generally, however, this chapter has also identified significant problems with child abuse blocking systems themselves. All three systems examined show very significant shortcomings in relation to legitimacy, transparency and accountability, while claims for the effectiveness of blocking have also been undermined. In addition, two of the systems appear to prove the truth of concerns about privacy and function creep, insofar as they have moved beyond their original goals of simple crime prevention and towards an intelligence gathering and even prosecution function. There is, therefore, a very real risk that by promoting blocking the constitutional values associated with freedom of expression and privacy of communications may be sacrificed – and worse, may be sacrificed for systems which are ineffective for their stated goal.

References

- Ahlert, C., Marsden, Christopher & Yung, C., 2004. How “Liberty” Disappeared from Cyberspace: The Mystery Shopper Tests Internet Content Self-Regulation. Available at: <http://pcmlp.socleg.ox.ac.uk/text/liberty.pdf> [Accessed October 13, 2008].
- AK Zensur, 2010. Blacklists of Denmark and Sweden analysed. Available at: <http://ak-zensur.de/2010/09/29/analysis-blacklists.pdf>.
- Akdeniz, Y., 2011. *Freedom of Expression on the Internet: Study of legal provisions and practices related to freedom of expression, the free flow of information and media pluralism on the Internet in OSCE participating States*, Organisation for Security and Cooperation in Europe. Available at: <http://www.osce.org/fom/80723>.
- Akdeniz, Y., 2010. To block or not to block: European approaches to content regulation, and implications for freedom of expression. *Computer Law & Security Review*, 26(3), pp.260-272.
- All Party Parliamentary Communications Group, 2009. *Can we keep our hands off the net? Report of an Inquiry by the All Party Parliamentary Communications Group*, London. Available at: www.apcomms.org.uk/uploads/apComms_Final_Report.pdf.

- Anderson, N., 2007. Image hash database could filter child porn. *Ars Technica*. Available at: <http://arstechnica.com/tech-policy/news/2007/07/image-hash-database-could-filter-child-porn.ars> [Accessed September 12, 2009].
- Baker, J., 2011. European Parliament votes to remove offensive images at source. *Computerworld*. Available at: <http://www.computerworlduk.com/news/public-sector/3261164/european-parliament-votes-to-remove-offensive-images-at-source/> [Accessed April 9, 2011].
- Bambauer, D., 2009. Cybersieves. *Duke Law Journal*, 59(3), p.477.
- Banisar, D., 2008. *Speaking of Terror*, Strasbourg: Council of Europe. Available at: http://www.coe.int/t/dghl/standardsetting/media/Doc/SpeakingOfTerror_en.pdf [Accessed May 12, 2009].
- Bits of Freedom, 2011. Dutch providers abandon “ineffective” web blocking. Available at: <https://www.bof.nl/2011/03/07/dutch-providers-abandon-ineffective-web-blocking/> [Accessed March 9, 2011].
- Bourke, M. & Hernandez, A., 2009. The “Butner Study” Redux: A Report of the Incidence of Hands-on Child Victimization by Child Pornography Offenders. *Journal of Family Violence*, 24(3), pp.183-191.
- Boyle, J., 1997. Foucault in Cyberspace: Surveillance, Sovereignty and Hardwired Censors. *University of Cincinnati Law Review*, 177, p.186.
- Brown, I., 2008. Internet Filtering: Be Careful What You Ask for. In S. K. Schroeder & L. Hanson, eds. *Freedom and Prejudice: Approaches to Media and Culture*. Istanbul: Bahcesehir University Press. Available at: <http://ssrn.com/paper=1026597> [Accessed October 4, 2008].
- Callanan, C. et al., 2009. *Internet blocking: balancing cybercrime responses in democratic societies*, Dublin: Aconite Internet Solutions.
- Campbell, D., 2005. Operation Ore exposed. *PC Pro*. Available at: <http://www.pcpro.co.uk/features/74690/operation-ore-exposed> [Accessed May 6, 2011].
- Carr, J., 2004. *Child abuse, child pornography and the internet*, London: NCH. Available at: http://www.make-it-safe.net/esp/pdf/Child_pornography_internet_Carr2004.pdf [Accessed September 12, 2009].
- Carr, J. & Hilton, Z., 2009. Children’s Charities’ Coalition on Internet Safety Digital manifesto. Available at: <http://www.chis.org.uk/uploads/02b.pdf>.
- Carr, J. & Hilton, Z., 2011. Combating child abuse images on the internet - international perspectives. In J. Davidson & P. Gottschalk, eds. *Internet Child Abuse: Current Research and Policy*. Abingdon: Routledge.
- CIRCAMP, CIRCAMP overview. *CIRCAMP*. Available at: http://circamp.eu/index.php?option=com_content&view=article&id=11:circamp-overview&catid=1:project&Itemid=2 [Accessed March 27, 2010].

- Clayton, R., 2005. Failures in a Hybrid Content Blocking System. Available at: <http://www.cl.cam.ac.uk/~rnc1/cleanfeed.pdf> [Accessed April 17, 2009].
- Clayton, R., 2008. Technical aspects of the censoring of Wikipedia. *Light Blue Touchpaper*. Available at: <http://www.lightbluetouchpaper.org/2008/12/11/technical-aspects-of-the-censoring-of-wikipedia/> [Accessed March 28, 2009].
- Colcolough, D., 2009. Investigating and Prosecuting Computer Facilitated Crimes Against Children: An AOL Perspective. Available at: <http://www.childrensmn.org/web/mrcac/handouts/184933.pdf> [Accessed September 11, 2009].
- Committee on Energy and Commerce, 2006. *Making the Internet Safe for Kids: The Role of ISPs and Social Networking Sites.*, Washington, DC: US Government Printing Office. Available at: <http://ftp.resource.org/gpo.gov/hearings/109h/30530.txt> [Accessed April 5, 2011].
- Crown Prosecution Service & Association of Chief Police Officers, 2004. Memorandum of Understanding Between Crown Prosecution Service (CPS) and the Association of Chief Police Officers (ACPO) concerning Section 46 Sexual Offences Act 2003. Available at: http://www.iwf.org.uk/documents/20041015_mou_final_oct_2004.pdf [Accessed July 24, 2009].
- Davies, C., 2009. The hidden censors of the internet. *Wired*. Available at: <http://www.wired.co.uk/wired-magazine/archive/2009/05/features/the-hidden-censors-of-the-internet.aspx?page=all> [Accessed September 20, 2009].
- Dedman, B. & Sullivan, B., 2008. ISPs pressed to become child porn cops. *MSNBC*. Available at: <http://www.msnbc.msn.com/id/27198621/> [Accessed October 17, 2008].
- Deibert, R. & Rohozinski, R., 2010. Beyond Denial: Introducing Next-Generation Information Access Controls. In R. Deibert et al., eds. *Access Controlled: The Shaping of Power, Rights and Rule in Cyberspace*. Cambridge, MA: MIT Press.
- Deibert, R. & Villeneuve, N., 2004. Firewalls and power: An overview of global state censorship of the Internet. In *Human rights in the digital age*. London: GlassHouse.
- Digital Rights Ireland, 2011. Garda plans for web blocking referred to Data Protection Commissioner. *Digital Rights Ireland*. Available at: <http://www.digitalrights.ie/2011/03/29/garda-plans-for-web-blocking-referred-to-data-protection-commissioner/> [Accessed May 16, 2011].
- Edwards, L., 2009. Pornography, Censorship and the Internet. In L. Edwards & C. Waelde, eds. *Law and the Internet*. Oxford: Hart Publishing.
- Eneman, M., 2010. Internet service provider (ISP) filtering of child-abusive material: A critical reflection of its effectiveness. *Journal of Sexual Aggression: An international, interdisciplinary forum for research, theory and practice*, 16(2), p.223.

- EUROPOL, Funnel Web Introduction. *EUROPOL*. Available at: <http://www.europol.europa.eu/index.asp?page=FunnelIntro&language=> [Accessed March 21, 2010].
- Falkvinge, R., 2011. The Copyright Lobby Absolutely Loves Child Pornography. *TorrentFreak*. Available at: [http://torrentfreak.com/the-copyright-lobby-absolutely-loves-child-pornography-110709/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed:+Torrentfreak+\(Torrentfreak\)&utm_content=Google+Reader](http://torrentfreak.com/the-copyright-lobby-absolutely-loves-child-pornography-110709/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed:+Torrentfreak+(Torrentfreak)&utm_content=Google+Reader) [Accessed July 19, 2011].
- Figliola, P.M., 2010. *US Initiatives to Promote Global Internet Freedom: Issues, Policy, and Technology*, Congressional Research Service.
- Freude, A., 2009. Delete, don't block: It works! *UnPolitik.de*. Available at: <http://www.unpolitik.de/2009/05/28/delete-dont-block-it-works/> [Accessed July 16, 2010].
- Graham, I., 2009. Statistics Laundering: false and fantastic figures. *libertus.net*. Available at: <http://libertus.net/censor/resources/statistics-laundering.html> [Accessed March 1, 2010].
- GSMA Mobile Alliance Against Child Sexual Abuse Content, 2008. Implementation of filtering of child sexual abuse images in operator networks. Available at: www.gsmworld.com/documents/GSMA_Child_Tech_Doc.pdf.
- Hakim, D., 2008. Net Providers to Block Sites With Child Sex. *The New York Times*. Available at: http://www.nytimes.com/2008/06/10/nyregion/10internet.html?_r=2 [Accessed September 14, 2009].
- Hargrave, S., 2006. Surfing with a safety net. *The Guardian*. Available at: <http://www.guardian.co.uk/technology/2006/jun/29/guardianweeklytechnologysection> [Accessed May 1, 2009].
- Home Office, 2009. Response to Freedom of Information Request in relation to the relationship between the Internet Watch Foundation and the Home Office and network level blocking. Available at: <http://www.whatdotheyknow.com/request/5357/response/18479/attach/html/2/ResponseT11%209.doc.html> [Accessed February 27, 2009].
- Hustinx, P., 2010. Opinion of the European Data Protection Supervisor on the proposal for a Directive of the European Parliament and of the Council on combating the sexual abuse, sexual exploitation of children and child pornography, repealing Framework Decision 2004/68/JHA. Available at: http://www.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2010/10-05-10_Child_Abuse_EN.pdf.
- Hutty, M., 2004. Cleanfeed: the facts. *LINX Public Affairs*. Available at: <http://publicaffairs.linx.net/news/?p=154> [Accessed January 15, 2010].
- Internet Watch Foundation, 2011a. 2010 Annual Report. Available at: <http://www.iwf.org.uk/assets/media/annual->

reports/Internet%20Watch%20Foundation%20Annual%20Report%202010%20web.pdf.

Internet Watch Foundation, 2010a. Content Assessment Appeal Process. Available at: <http://www.iwf.org.uk/accountability/complaints/content-assessment-appeal-process> [Accessed February 15, 2011].

Internet Watch Foundation, 2010b. IWF Facilitation of the Blocking Initiative. *Internet Watch Foundation*. Available at: <http://www.iwf.org.uk/public/page.148.437.htm> [Accessed March 17, 2010].

Internet Watch Foundation, IWF URL List Policy and Procedures. Available at: <http://www.iwf.org.uk/services/blocking/iwf-url-list-policy-and-procedures> [Accessed February 15, 2011].

Internet Watch Foundation, 2011b. IWF URL List Recipients. *Internet Watch Foundation*. Available at: <http://www.iwf.org.uk/services/blocking/iwf-list-recipients> [Accessed May 18, 2011].

Johnson, D.R. & Post, D.G., 1996. Law and Borders - The Rise of Law in Cyberspace. *Stanford Law Review*, 48, p.1367.

Kleinschmidt, B., 2010. An International Comparison of ISP's Liabilities for Unlawful Third Party Content. *International Journal of Law and Information Technology*, 18(4), p.332.

Koops, B.-J. et al., 2006. Should Self-Regulation be the Starting Point? In B.-J. Koops et al., eds. *Starting Points for ICT Regulation: Deconstructing Prevalent Policy One-Liners*. The Hague: T.M.C. Asser Press.

Kreimer, S., 2006. Censorship by Proxy: The First Amendment, Internet Intermediaries, and the Problem of the Weakest Link. *University of Pennsylvania Law Review*, 155, p.11.

Lambers, R., 2006. Code and Speech. Speech Control Through Network Architecture. In E. Dommering & L. Asscher, eds. *Coding Regulation: Essays on the Normative Role of Information Technology*. Information Technology & Law. The Hague: T.M.C. Asser Press.

Leaseweb, 2009. LeaseWeb 1st Hosting Provider to Install Child Porn Filter. *Leaseweb blog*. Available at: <http://blog.leaseweb.com/2009/03/16/leaseweb-1st-hosting-provider-to-install-child-porn-filter/> [Accessed July 20, 2011].

Leppard, D., 2005. Child porn suspects set to be cleared in evidence "shambles." *The Sunday Times*. Available at: <http://www.timesonline.co.uk/tol/news/uk/article539974.ece> [Accessed May 6, 2011].

Lessig, L., 1999. *Code: And Other Laws of Cyberspace*, New York, N.Y: Basic Books.

Marsden, Chris, 2010. *Net Neutrality: Towards a Co-regulatory Solution*, London: Bloomsbury Academic.

- McIntyre, T.J., 2010. Blocking Child Pornography on the Internet: European Union Developments. *International Review of Law, Computers & Technology*, 24(3), pp.209-221.
- McIntyre, T.J. & Scott, C., 2008. Internet Filtering: Rhetoric, Legitimacy, Accountability and Responsibility. In R. Brownsword & K. Yeung, eds. *Regulating Technologies*. Oxford: Hart Publishing. Available at: <http://ssrn.com/abstract=1103030>.
- Metz, C., 2008. New York sends AOL “how-to-wiretap” slides. *The Register*. Available at: http://www.theregister.co.uk/2008/10/20/cuomo_pron_crusade_continues/ [Accessed June 30, 2009].
- Moore, T. & Clayton, R., 2008. The Impact of Incentives on Notice and Take-down. Available at: <http://weis2008.econinfosec.org/papers/MooreImpact.pdf>.
- Morrison, S.R., 2011. What the Cops Can’t Do, Internet Service Providers Can: Preserving Privacy in Email Contents. *SSRN eLibrary*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1729000 [Accessed February 23, 2011].
- Mueller, M., 2010. *Networks and States: The Global Politics of Internet Governance*, Cambridge, MA: MIT Press.
- New Zealand Department of Internal Affairs, 2010. Digital Child Exploitation Filtering System Code of Practice.
- O’Donnell, I. & Milner, C., 2007. *Child Pornography: Crime, Computers and Society*, Cullompton: Willan.
- O’Neill, S., 2010. Government ban on internet firms that do not block child sex sites. *The Times*. Available at: http://technology.timesonline.co.uk/tol/news/tech_and_web/the_web/article7055882.ece [Accessed March 12, 2010].
- Ofcom, 2008. Ofcom’s Response to the Byron Review. Available at: <http://www.ofcom.org.uk/research/telecoms/reports/byron/> [Accessed April 11, 2009].
- Office of the Attorney General, 2010. Attorney General Cuomo announces expansion of groundbreaking initiative to eliminate sharing of thousands of images of child pornography on social networking web sites. *New York State Attorney General*. Available at: http://www.ag.ny.gov/media_center/2010/june/june21a_10.html [Accessed March 30, 2011].
- Ohm, P., 2009. The Rise and Fall of Invasive ISP Surveillance. *University of Illinois Law Review*.
- Ozimek, J., 2009. A censorship model. *The Guardian*. Available at: <http://www.guardian.co.uk/commentisfree/libertycentral/2009/aug/02/internet-censor> [Accessed September 21, 2009].

- Price, M.E. & Verhulst, S., 2005. *Self-Regulation and the Internet*, The Hague: Kluwer Law International.
- La Quadrature du Net, 2011. French LOPPSI Bill Adopted: The Internet under Control? Available at: <http://www.laquadrature.net/en/french-loppsi-bill-adopted-the-internet-under-control> [Accessed February 15, 2011].
- Richardson, T., 2004a. BT on child porn stats. *The Register*. Available at: http://www.theregister.co.uk/2004/07/22/bt_ispa_cleanfeed/ [Accessed January 25, 2009].
- Richardson, T., 2004b. ISPA seeks analysis of BT's "Cleanfeed" stats. *The Register*. Available at: http://www.theregister.co.uk/2004/07/21/ispa_bt_cleanfeed/ [Accessed January 25, 2009].
- Richmond, R., 2011. Facebook's New Way to Combat Child Pornography. *New York Times*. Available at: <http://gadgetwise.blogs.nytimes.com/2011/05/19/facebook-to-combat-child-porn-using-microsofts-technology/> [Accessed July 20, 2011].
- Russell, D.E.H. & Purcell, N.J., 2005. Exposure to pornography as a cause of child sexual victimization. In N. E. Dowd, D. G. Singer, & R. F. Wilson, eds. *Handbook of Children, Culture, and Violence*. London: Sage, pp. 59–84.
- Salgado, R.P., 2006. Fourth Amendment Search and the Power of the Hash. *Harvard Law Review Forum*, 119, p.38.
- Soghoian, C., 2010a. An End to Privacy Theatre: Exposing and Discouraging Corporate Disclosure of User Data to the Government. *Minnesota Journal of Law, Science and Technology*.
- Soghoian, C., 2010b. Privacy And Law Enforcement: Caught In The Cloud: Privacy, Encryption, And Government Back Doors In The Web 2.0 Era. *J. on Telecomm. & High Tech. L.*, 8, pp.359–613.
- Stol, W. et al., 2008. Filtering Child Pornography on the Internet: An Investigation of National and International Techniques and Regulations. Available at: <http://www.wodc.nl/onderzoeksdatabase/internetfilters-tegen-kinderporno.aspx?cp=44&cs=6780>.
- Stol, W. et al., 2009. Governmental filtering of websites: The Dutch case. *Computer Law & Security Review*, 25, pp.251–262.
- Svantesson, D.J.B., 2008. How Does the Accuracy of Geo-Location Technologies Affect the Law. *Masaryk University Journal of Law & Technology*, 2, p.11.
- Swire, P.P., 1998. Of Elephants, Mice, and Privacy: International Choice of Law and the Internet. *The International Lawyer*, 32, p.991.
- Tambini, D., Leonardi, D. & Marsden, Chris, 2008. *Codifying Cyberspace: Communications Self-Regulation in the Age of Internet Convergence*, London: Routledge.

- Villeneuve, N., 2010. Barriers to Cooperation: An Analysis of the Origins of International Efforts to Protect Children Online. In *Access Controlled: The Shaping of Power, Rights and Rule in Cyberspace*. Cambridge, MA: MIT Press.
- Walden, I., 2010. Porn, Pipes and the State: Censoring Internet Content. *The Barrister*, (44), pp.16-17.
- Watt, R. & Maurushat, A., 2009. Clean Feed: Australia's Internet Filtering Proposal. *Internet Law Bulletin*, 12(2). Available at: <http://www.austlii.edu.au/au/journals/UNSWLRS/2009/7.html> [Accessed May 6, 2009].
- Whittaker, Z., 2009. Microsoft develops image DNA technology for fighting child porn. *ZDNet*. Available at: <http://blogs.zdnet.com/igeneration/?p=3655> [Accessed February 22, 2010].
- Williams, C., 2011. Hollywood studios ask High Court to block film website. *The Telegraph*. Available at: <http://www.telegraph.co.uk/technology/news/8597596/Hollywood-studios-ask-High-Court-to-block-film-website.html> [Accessed July 20, 2011].
- Williams, C., 2009. Home Office backs down on net censorship laws. *The Register*. Available at: http://www.theregister.co.uk/2009/10/16/home_office_iwf_legislation/ [Accessed October 16, 2009].
- Zittrain, J., 2003. Internet Points of Control. *Boston College Law Review*, 44, p.653.
- Zuvela, M., 2011. Deleting trumps blocking in fight against online child porn. *Deutsche Welle*. Available at: <http://www.dw-world.de/dw/article/0,,14968970,00.html> [Accessed April 7, 2011].

¹ Disclosure: the author is chairman of Digital Rights Ireland, which has been involved in lobbying against internet blocking measures. This chapter draws on material previously presented at BILETA, Glasgow Caledonian University 27-28 March 2008, and the 3rd International Conference on Legal, Security and Privacy Issues in IT, Prague, 3-5 September 2008.

² Although hash value systems are most commonly associated with the US, there are also some European initiatives in this area. In particular, the Dutch Ministry of Justice and the Dutch Hotline have cooperated with hosting provider Leaseweb and Swedish company Netclean to trial MD5 hash value blocking of images uploaded to certain sites (Leaseweb 2009).

³ While this chapter generally uses the term child abuse images, in this and other sections the term child pornography is used to reflect the terminology used by US law.

⁴ A web based blocking system was mandated by legislation in Pennsylvania in 2002 but was ultimately ruled unconstitutional in *Center for Democracy and Technology v. Pappert* 337 F.Supp.2d 606 (2004). This experience appears to have influenced later US developments, and may be responsible for government strategies which promote voluntary and self-regulatory blocking systems which may escape similar judicial review.

⁵ This is a deliberate oversimplification of the issues associated with hashing and in particular doesn't address the issue of possible hash value collisions where different files generate the same hash value, generating false positives.

⁶ For an example of such a report see *United States v. Brent Terry* 522 F.3d 645 (2008).

⁷ Public Law 110-401, 122 Stat. 4229-4253.

⁸ The likelihood of hash value collisions may, however, increase where robust hashing systems such as Microsoft's PhotoDNA are used. One Microsoft researcher has put the likelihood of false positives in PhotoDNA at one in 2 billion images (Richmond 2011).

⁹ 337 F.Supp.2d 606 (2004).

¹⁰ 607 F.3d 357 (2010).

¹¹ There is an argument that scanning and blocking of emails may violate either the Federal Electronic Communications Privacy Act or state surveillance laws, depending on whether either or both the sender and recipient consent to scanning (see e.g. Metz 2008; Ohm 2009). Such violations would not, however, result in the suppression of evidence, which explains why these arguments have not been made in cases such as *US v. Richardson*.

¹² Directive 95/46/EC.

¹³ A further application of hash values matching is in relation to private files which a user stores or backs up on a cloud computing service. With the move away from local storage and towards remote storage and backup this may result in all files stored by a user being scanned for contraband, irrespective of whether or not they are being sent to others. Although it is beyond the scope of this chapter, it is worth noting that in many jurisdictions there is lesser protection for remotely stored data than for data which is in the course of transmission, suggesting that hash value scanning of files stored remotely might be legally permissible even if blocking of those files in the course of communication would not be. On this point see Soghoian (2010b).

¹⁴ A variant of this argument is that blocking can prevent the accidental or casual viewer from developing a latent sexual interest in children, and can thereby prevent a progression to contact sexual offending (see e.g. Carr 2004). It should be noted that there is some debate as to whether viewing of child abuse images leads to “real world” offending. While some authors (e.g. Russell & Purcell 2005; Bourke & Hernandez 2009) suggest that it does, there appears to be no definitive study (compare the literature review in O’Donnell & Milner 2007).

¹⁵ In the United States in particular there is also a tension between different arms of government, with the State Department actively funding circumvention tools via its Global Internet Freedom strategy. Although intended for destinations such as China and Iran, such tools will undoubtedly also see a great deal of use domestically. See e.g. Figliola (2010)

¹⁶ This lack of data reflects the decentralised nature of most child abuse image blocking systems. Although the determination of what sites to block may be made by a central body, the implementation of that blocking is generally the responsibility of the individual ISP. As a result, there is no central repository of data or guarantee that any data is being logged. In addition, because individual ISPs may implement blocking in different ways any data which is logged may not be comparable with data from other sources.