Harmful Content on the Internet and in Video Games

1. My name is Dr Richard Clayton and I am a researcher in the Security Group of the Computer Laboratory at the University of Cambridge. Over the past few years I have published several academic papers on various schemes for blocking Internet content.[1]

ISP Blocking of content

2. The Government (by which I mean senior civil servants and ministers) appear to be under the impression that it is now possible to require ISPs to block undesirable Internet content and that this will be effective.[2] Since your inquiry may be tempted to endorse blocking by ISPs, it is important to explain that this impression is – in almost all practical respects – entirely mistaken.

3. Blocking can only practically be done at individual ISPs – there is no "Internet backbone" where it can be done for everyone at once. Hence for a blocking scheme to work at a national level in the UK, every single ISP (and there are nearly 70 major suppliers, and quite a number more of smaller ones), needs to deploy suitable equipment. Since all of these ISPs have different network designs, they would each need to design their own particular scheme for content blocking. However, despite all this complexity, it is possible to distinguish four basic schemes by which blocking can be done:

4. The first is to block particular IP addresses – particular machines on the Internet. Schemes that do this are reasonably low cost if only a few machines (a few thousand) are to be blocked, but can become very expensive indeed if tens of thousands of machines are to be blocked (leaving aside the question as to where such an extensive list might come from). The key problem with this blocking method is that many machines are shared between multiple content providers. Thus a machine in Romania might be hosting not only some undesirable content such as DraculaBitesYou.com, but also the Transylvanian Tourist Board. Both sites will have the same IP address and so both would be blocked – which is clearly undesirable for the Romanian tourist industry. Similarly, IP address blocking could only block the whole of FaceBook, or geocities.com, rather than individual parts of these enormous sites.

5. The second scheme is to block particular domains by arranging that an invalid response is given when the domain name is "resolved" to give the IP address where it is currently hosted. This permits the blocking of DraculaBitesYou.com (which would fail to be resolved properly) whereas the Tourist Board could still be available. This scheme is very low cost, and "scales" well – viz: it is still low cost when tens of thousands of domain names are to be

---

[1] Richard Clayton: *Failures in a Hybrid Content Blocking System*. In George Danezis and David Martin, editors, Privacy Enhancing Technologies, Fifth International Workshop, PET 2005, Cavtat, Croatia, May 30–June 1 2005, volume 3856 of LNCS, pages 78–92, Springer.

Richard Clayton: *Anonymity and Traceability in Cyberspace*. Technical Report UCAM-CL-TR-653, University of Cambridge Computer Laboratory, November 2005.

Richard Clayton, Stephen J. Murdoch and Robert N.M. Watson: *Ignoring the Great Firewall of China*. In George Danezis and Philippe Golle, editors, Privacy Enhancing Technologies, Sixth International Workshop, PET 2006, Cambridge, UK, 28–30 June 2006, volume 4258 of LNCS, pages 20–35, Springer.

[2] "*Recently, it has become technically feasible for ISPs to block home users' access to websites irrespective of where in the world they are hosted"* Vernon Croaker, Hansard, 15 May 2006, Column 715W.

blocked. However, it is still ineffective if it desired to block only one part of a large website such as FaceBook or GeoCities.

6. The third scheme involves the use of "proxy" machines. Here the ISP arranges that instead of connections being made directly to the remote machine, the connection is forced to be made to an ISP machine (the proxy) – which pretends to be the remote machine. This proxy relays all of the traffic to and from the remote machine, but can apply filtering rules as material passes by. This scheme can be made very precise, individual images from a page can be suppressed, and there are none of the potential "overblocking" issues that occur with the previous two schemes. However, it is an extremely expensive scheme for an ISP to implement, because all of the traffic must be handled by the proxy (which must therefore be a very capable and hence very costly device). Also, in practice – to avoid "single points of failure" – ISPs will need to purchase multiple proxies, and this only adds to the cost.

7. The fourth scheme is that of "deep packet inspection" whereby traffic is examined by devices at the ISP as they pass the packets of data to and from the customers. Unlike the previous scheme, there is no proxy masquerading as the remote machine – instead, if a "bad" connection is being made, further packets will be discarded (or perhaps some extra "reset" packets will be introduced to persuade the machines to close the connection). Deep packet inspection is expensive, albeit not as expensive as using proxies. As a scheme it is famous because it is used for censoring the Internet by the Chinese Government – as a key part of their "Great Firewall of China" – and it is also currently being used by Comcast, the American ISP, to block some types of peer-to-peer file-sharing traffic.[3]

8. It is also possible to create hybrid schemes, of which the most famous is BT's Anti-Child-Abuse Initiative, commonly known as CleanFeed. In the CleanFeed system traffic to particular IP addresses is specially treated. However, instead of simply being blocked (as in the first scheme above), it is instead passed through a proxy machine which then applies the necessary filtering. Hence it is an amalgam of schemes 1 and 3 and is a distinct improvement on both of them. Essentially it combines all the exactness of the proxy scheme, but it need only handle a small proportion of the traffic, and so the proxy machines can be smaller and cheaper: the traffic to both DraculaBitesYou.com and the Transylvanian Tourist Board is redirected to the proxy, and the proxy then permits unfettered access to the tourist information whilst blocking inappropriate pages from the Dracula site.

9. Unfortunately, although CleanFeed is a good engineering design, it suffers from a significant problem in that it is possible to reverse engineer the list of sites that are being blocked. If this list is of paedophile sites (as is currently the case in the BT deployment) then this permits people who are interested in such material to obtain information about places to visit which they might not otherwise have known about. This runs counter to the public policy aims of the system.

10. CleanFeed's problems aside, the real problem with all of the blocking schemes is that they are all pretty trivial to evade. Two generic ways of avoiding blocking are encryption and proxy services:

11. If traffic is encrypted this defeats any system that relies on looking at the packets to determine what they contain, either by deep packet inspection or within a proxy. Since a lot of traffic is encrypted anyway (when people are doing online shopping, or using a webmail

[3] Peter Svensson, Associated Press, *Comcast Blocks Some Internet Traffic*, 19 Oct 2007. http://www.breitbart.com/article.php?id=D8SCEBLG0 For much more detail about Comcast's activity see: http://www.eff.org/wp/detecting-packet-injection

system) it isn't possible to just decide to block encrypted material. It is worth noting in passing that encryption is becoming very widespread in the peer-to-peer world of file sharing. This is because of the deployment of peer-to-peer blocking systems and "traffic shaping" (systems that slow down peer-to-peer traffic). The use of encryption allows file sharing to continue at full speed – hence its growing popularity. It must be expected that if content blocking disrupted what people wished to do on the Internet then a similar evolutionary process would occur and encryption would rapidly become very popular so that access to the content could continue.

12. Proxy services work just like the ISP proxies already described, except by being located in another jurisdiction, they are not bound by UK rules on what to block, so they will not filter anything but will allow unfettered access to content. Since the connections to these proxies are encrypted, and proxies have legitimate uses for people who need anonymity – so that blocking them would be unreasonable – they provide a simple and easy-to-use way of evading blocking schemes. Their only downside (apart from making connections slightly slower) is that the providers of proxies usually charge for usage – although there are free systems available.

13. There are more complex anonymity systems available such as Tor,[4] originally developed by the US Navy to permit them to surf the web without revealing their identity, and which currently has several hundred thousand users. Someone who uses Tor will also evade local content blocking schemes – and if the "exit node" is in an appropriate jurisdiction, will evade blocking altogether.

14. Besides these generic evasion schemes, other more specific methods of evasion are possible: for example, if ISP A runs CleanFeed then it will be correctly resolving domain names; so the customers of ISP B – where blocking is done by arranging for incorrect name resolution – can use ISP A's resolvers and thereby evade B's scheme!

15. Furthermore, all of the discussion so far has assumed that it is possible to construct a list of what is to be blocked in the first place, and that's considerably more difficult than it seems:

16. A taste of the difficulties can be seen by examining the criminals who set up fake bank websites (phishing) or who illegally sell pharmaceuticals online. It is far from uncommon for them to purchase thousands of domain names (with a complaisant registrar, they can avoid paying for them, provided they recycle the domains within 5 days)[5] and then arrange for all of the domains to point at a single website – they then send out their spam to advertise their wares, and anyone who wants to block access must learn the thousands of domain names, and arrange to block them all.

17. The other scheme that is now widely used by the most successful phishing gangs is called "fast-flux". Here it is arranged that a domain resolves to a particular IP address for just a few minutes, thereafter it resolves to another machine for a few more minutes, and so on. The machines are not owned by the criminals, but are consumer machines onto which they have planted some malicious software (such as might be done to recruit them into a "botnet"). This software will either serve up the content directly, or will cause the machines to act as a proxy and relay traffic to the criminals' actual website.

18. Thus it can be seen that it is relatively straightforward to make the blocking systems' problem become that of blocking thousands of domains, which resolve to thousands of

[4] http://www.torproject.org

[5] Joel Hruska, *Cybersquatting daisy-chain leads to Dell trademark infringement lawsuit*, ars technica, 30 Nov 2007. http://arstechnica.com/news.ars/post/20071130-cybersquatting-daisy-chain-leads-to-dell-trademark-infringement-lawsuit.html

addresses, with this all changing dynamically every few minutes. Quite clearly, this is a difficult problem and would make creating and distributing blocking lists very difficult indeed. The banks (who try to remove phishing websites) do manage to tackle their similar problem, with websites removed within a day or so – but they have the law on their side: the phishing website is illegal in every jurisdiction. If what the blocking system is preventing access to is legal content where it is hosted, then it is far less likely to be removed. An example of this sort of difficulty occurs even with indecent images of children – often thought to be a clear example of material that is illegal everywhere. However, if the images have been computer-generated, then the United States Supreme Court held that these were lawful[6] (no child was harmed in their production), whereas they are illegal to make or possess in the United Kingdom (where we assume that they incite viewers into moving on to further activity that does involve real children).

19. Nonetheless, despite all of the issues I have described above, and despite informed criticism from the ISP industry, the UK Government continues to believe that it is possible to block Internet content. The Home Office has instructed UK consumer ISPs to block all websites on the Internet Watch Foundation list (these are sites that the IWF is unable to get removed from the Internet, occasionally because the content is lawful where they are hosted, but mainly – in my view – because of the ineffective manner in which they contact the hosting sites, passing reports via multiple law enforcement agencies which seldom leads to timely action). This blocking is, in my view, for the reasons set out above, rather a waste of time and money, and I urge you not to consider extending this system or endorsing it in any way.

End-user blocking systems

20. What does make some sense, and I believe you should be seriously considering the complex issues that surround it, is the voluntary use of blocking software on end-user machines. This software can have direct access to the requests being made by the user, and the data that is returned. It should therefore, in principle, be unaffected by the use of encryption or of proxy systems.

21. Furthermore, because it is specific to a particular machine, or even to a particular person using the machine, it can be customised to block precisely what is appropriate. It could, for example, block different sites for an 8-year old boy, for his 14-year old sister, whilst permitting an adult to view what they wished. ISPs are of course unaware of who is using each of the machines within a particular house, and what their age is, and so ISP-level blocking can only be "on" or "off" for everyone.

22. However, there are a number of generic problems with end-user blocking software:

23. Some brands are better than others – but there is currently no way for a potential purchaser to distinguish. The industry has been developing a BSI "kite-mark" for several years, but the launch continues to be delayed and so consumers cannot tell if a product meets a minimum standard. Previous surveys by Which? have indicated that many do not – a 2005 report by Intertek[7] summarises the situation across Europe.

24. Some filtering software is trivially easy to circumvent or to switch off altogether, and of course this is one of the issues as to whether or not it is "fit for purpose" and should be given a kite-mark! Although considerable skill may be needed to first discover how to do this, it is the sort of information that will rapidly circulate in chat-rooms and playgrounds, so "obscurity"

[6] Ashcroft v. Free Speech Coalition (00-795) 535 U.S. 234 (2002) 198 F.3d 1083, affirmed
[7] http://www.anec.org/attachments/ANEC-R&T-2006-ICT-002%20(1).pdf

cannot be seen as an alternative to robust protection mechanisms. Famously, the system endorsed by the Australian Government in summer 2007 was circumvented within 30 minutes by a 16 year old child[8] – and the history of this type of software is littered with similar incidents. Nevertheless, in a family situation, it may well be that even though a filter could be turned off, a reasonably well-behaved child would not turn it off as a matter of course: so it would still provide some level of protection and some reassurance to parents.

25. Most of the filtering software on the market is fairly simple-minded when considering what should be blocked. Typical designs contain "blacklists" of sites that should always be blocked, "whitelists" of sites that should never be blocked, and then some heuristic rules that attempt to deduce whether a site that has never been heard of should or should not be blocked. The lists should be relatively trouble-free (though one current system is reported to block the CBeebies site when it is accessed as www.bbc.net.uk/cbeebies rather than the normal www.bbc.co.uk address), but the heuristics are generally extremely fallible, and are unable to distinguish discussions (and images) of breast cancer from breast enhancement, or SuperBowl XXX sites from "XXX" porn sites – hence the use of the sitelists to fix up the shortcomings.[9]

26. The anecdotal evidence (anecdotal because the blocking software firms do not publish detailed information about their products) is that what is blocked is informed by the concerns of the middle-class, white, Calvinist, Connecticut males who commissioned the software – so there is wholesale blocking of sites discussing homosexuality, the occult or gambling. This may well be appropriate for many younger children, but many parents (especially those not from a "WASP" tradition) may not be so worried about these issues; and in particular blocking of informational sites on controversial topics may make little sense for teenagers trying to do research for their schoolwork.

27. In this context, it is very instructive to look at survey results from the schools blocking system in the Republic of Ireland (where the schools use a specialist ISP for Internet access, which implements a centralised blocking system). Here some 85% of primary school teachers believed that the blocking system was "just right", whereas in the secondary schools, this fell to 52% – with 40% of survey respondents asking for a way of overriding the system on a case by case basis.[10] The lesson here is that blocking systems have significant limitations for older children where subtle judgments must be made as to whether particular content should be accessible. In particular, these judgments depend as much on the context of the access as the nature of the material – distinctions that no automated filtering system can hope to address.

28. It should be noted that although considerable lip service is paid to the desirability of end-user blocking software, in practice it is relatively rarely used. The reason for this appears to be that when it comes down to it, parents are not prepared to pay ongoing subscriptions for software that contains up-to-date lists of websites. Without an income stream the software providers are not prepared to keep their lists current. Trying to break this vicious circle with Government money would raise questions of state subsidy and competition law – so unless some sort of charity comes forward to put serious money behind the shipping of free software to all, it seems that this software will remain highly praised – and rather more seldom used.

---

[8] Nick Higginbottom and Ben Packham, *Student cracks Government's $84m porn filter*, news.com.au, 26 Aug 2007. http://www.news.com.au/story/0,23599,22304224-2,00.html

[9] The ACLU report "*Censorship in a box*" contains a number of other examples of unexpected blocking decisions. http://www.aclu.org/privacy/speech/14915pub20020916.html

[10] Ronan Byrne: *Content Filtering on Ireland's Schools Network: Delivering a Safer Online Environment for Irish Schools?* TERENA 2007, Copenhagen, Denmark, 21–24 May 2007. http://tnc2007.terena.org/programme/presentations/show.php?pres_id=39

Website labelling

29. Finally, I wish to briefly comment upon another Bad Idea, which continues to have remarkable traction in the UK Government and within the EU Commission. This is the notion of website self-labelling.

30. The idea is a simple one – there is much to be said for end-user blocking systems, but they have difficulties in rating sites they have not previously encountered. So shouldn't the sites rate themselves and the blocking systems can use those ratings?

31. This idea first surfaced in the mid-1990s with an RSACI scheme that was almost identical to their existing videogame rating system. It later evolved into an ICRA scheme that had slightly more sensible categories. Meanwhile, since there were a handful of labelling schemes, the W3C organisation developed PICS as a way of permitting multiple labels on a single webpage.

32. Labelling was never very popular; although tens of thousands of websites were labelled, this was only ever tiny fractions of a percent of the total. The difficulty was that it was just too hard to label sites correctly, because once you move away from sites consisting of sales brochures, extremely complex judgments arise. For example, for a Guy Fawkes themed site in 1996, I was involved in extensive discussions[11] as to what rating should be applied to a webpage that discussed 1605 torture techniques and explained what "hung drawn and quartered" actually meant. In practice, any website that carries anything other than completely innocuous material will face these types of decisions every day – and will rapidly conclude that spending their staff time on these complex decisions makes very little economic sense.

33. This abrogation of responsibility for self-labelling can be seen on the UK Department of Health website – which, in common with many other UK Government sites is ICRA labelled. However, chapter 19 of the *Inquiry into Child Abuse in North Wales* is viewable at http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/Browsable/DH_4927518 and is labelled[12] "no potentially offensive language" despite the presence of a four letter word for sexual intercourse within some reported speech – which is precisely what the labelling is supposed to warn about.

34. The bottom line on labelling is that although it sounds a neat idea, more than a decade of experience shows that it is completely impractical; take-up has always been miniscule; when used to pay lip service to political correctness it is often inaccurate; and the burdens on webspace creators are so significant that even its advocates find it just too much work. It is high time that politicians stopped endorsing self-labelling schemes as an apparently easy out when considering how the web should be rated. I trust that this inquiry will have more sense!

Finally

35. I believe that the above discussion, though rather long, clearly sets out the technical issues relating to content blocking systems. If I can be of further assistance to the inquiry on this, or other technical matters, I would be pleased to help.

---

[11] For a longer version of this section, along with hyperlinks to relevant material, see: http://www.lightbluetouchpaper.org/2007/09/17/web-content-labelling/

[12] http://www.icra.org/cgi-bin/rdf-tester/labelTester.cgi?lang=en&url=http%3A%2F%2Fwww.dh.gov.uk%2Fen%2FPublicationsandstatistics%2FPublications%2FPublicationsPolicyAndGuidance%2FBrowsable%2FDH_4927518