

IWF, Wikipedia and the “Wayback Machine”

Dr Richard Clayton

`richard.clayton@cl.cam.ac.uk`



UNIVERSITY OF
CAMBRIDGE

Computer Laboratory

UKNOF13, Sheffield
28th May 2009

Outline

- The IWF & UK blocking of child sexual abuse images
- The blocking of Wikipedia (Dec 2008)
- The blocking of the Internet Archive (Oct 2008 – Jan 2009)
- What is on the IWF URL block-list today ?
- How fast are child sexual abuse images removed ?

The IWF

- Internet Watch Foundation
- Set up in 1996 to address issue of child pornography on Usenet
 - phrases “child pornography” or “kiddy porn” seen to trivialise issue
 - politically correct term became “child abuse images” (CAI)
 - or rather more recently “child sexual abuse images”
- IWF operates a consumer “hot-line” for reports
- UK institution, but cooperates via INHOPE with other hotlines
- Funded by industry and also by EU (seen as leading light)
- Now mainly concerned with websites
- Has a database of sites not yet removed (for efficiency)
- Database now underpins various blocking systems

Most (all?) UK filtering is proxy based

- Comparison of URLs in proxy means no “overblocking”
- Proxying all web traffic very expensive (and other downsides)
- So select only traffic that might need filtering
 1. DNS poisoning
 - resolve dubious domains to address of web proxy
 - low cost, and highly scalable – widely used in UK
 - assumes customers using the local DNS server!
 2. custom iBGP
 - resolve dubious domains and route their /32 to web proxy
 - mechanism used by BT’s “cleanfeed” system
 3. exotica (DPI, WCCPv2 etc)
 - scaling issues, so used mainly by smaller ISPs

Whitehall comprehension?

- Blocking considered “impossible” until BT deployed CleanFeed
- Parliament told: *“Recently, it has become technically feasible for ISPs to block home users’ access to websites irrespective of where in the world they are hosted”*
- In my view, doubtful that they actually understood the cost, fragility or ease of evasion of these blocking systems, let alone the reverse engineering of the blocking lists.
- Ministers want all (consumer?) broadband suppliers to filter
 - original target date of end of 2007 else “review our options”
- ISPA claimed 80% (more recently 95%) of consumers covered by systems that block illegal child images
 - methodology for count unclear (& not all ISPs filter all customers)

Wikipedia

- Member of public reports Virgin Killer album cover to IWF
- IWF conclude it is an indecent image, and add URLs to blacklist
- List rolled out midday Friday December 5th 2008
- Large numbers of UK accesses to Wikipedia now proxied
 - this breaks Wikipedia security model!
- Mechanism rapidly identified, as is particular image
 - propriety of keeping image debated in May 2008
- Many instances of image located (some on Amazon US)
- On Monday 8th IWF considers Wikipedia “appeal” & rejects it
- On Tuesday 9th IWF board decide to remove URL from list
- Wikipedia blocked elsewhere for some time thereafter!

What was blocked?

- #1: Main page was blocked
 - http://en.wikipedia.org/.../virgin_killer
 - blocked entire text about The Scorpions album, not just the image
- #2: Image description page was blocked
 - http://en.wikipedia.org/.../Image:Virgin_Killer.jpg
 - this is also a text page (despite the URL!)
- Did not block ../Virgin_Killer (there are four duplicate URLs!)
- Some blocking systems were case sensitive, some were not
- Caused considerable confusion as to what blocking was in place
 - general lesson about this event and the archive.org event; most consumer reports were almost entirely inaccurate!
- Evidence that some ISPs did not block until Monday
 - possibly just slow, possibly because a high-traffic website

What about the proxies?

- Wikipedia security model is that wicked page alterations (spam, or the losing side in “edit wars”) means edit privilege revoked
- But identity for anonymous editors is tied to IP address
- 95% of UK were now on less than a dozen IP addresses
- So anonymous editing rapidly impossible from UK
- Cannot create new signed-in identities (because IP is wicked)
- Wikipedia have a fix for this, which is to rely on the IP address in the **x-Forwarded-From** header from trusted caches...
 - ... but (a) many ISPs weren’t generating XFF headers
 - ... and (b) it took time to add the caches to the trusted list

Blocking of the “Wayback Machine”

- The Internet Archive automatically archives websites
- Some archived material is child sexual abuse images ☹️
- When they are found, the site is regularly added to IWF list
- Demon users reported problems with links pages
 - links are to `iwfwebfilter.thus.net` (which doesn't serve content)
 - from early October 2008 onwards; cause never pinned down
- On Jan 14th 2009 one such report makes it into The Register
- Comments include a report from Romania that they also see corrupted links pages pointing at Demon cache
- Finger points at problem at Internet Archive...
... fault identified and fixed by mid-evening UK time

What was the failure mechanism?

- Wayback machine holds generic versions of pages & sites
- Does dynamic replacement of “www.example.com” text

`http://web.archive.org/web/20010217021148/http://www.example.com`

- Uses a header passed from a front-end cache for this
- Unfortunately exactly this header was being sent by Demon
- Hence pages incorrectly constructed – and served to all-comers
 - NB: an attacker could have spoofed entire summary pages!
- Fix is for archive.org to remove clashing incoming headers
 - hence not Demon’s (or the IWF’s) fault at all!
- Note: actual URL that was blocked never externally identified

What is the IWF currently blocking ?

- My 2005 paper reverse-engineered “Cleanfeed” list
- Latest idea (NB: does not access the sites, since that’s illegal!)

```
for $hostname in (list of all valid hostnames)
  if (resolve(hostname) == cache-IP-address)
    print “hostname is blocked”
```

- List of hostnames comes from ISC “passive DNS” dataset
 - systems collecting anonymised copies of DNS responses
- c 120 million hostnames – 40 million are DNSBLs etc
- Further clean-up gives c 70 million hosts to check
- Takes about 2 days (and 22Gbytes) over home ADSL
- NB: does not identify URLs, merely hostnames

Current results (this is ongoing research)

- IWF list currently holds about 450 URLs (says a mole)
- 40% not identified by the methodology (too obscure?)
- 35% clearly (from hostname) intentionally wicked
- Remaining 25% are legitimate “free” hosting sites (etc)

`100free.com, 2st.jp, 3dn.ru, 4shared.com, 50webs.com, adultdreamhost.com, adultshare.com, awardspace.biz, awardspace.info, bbs.zgsm.com, beam.to, boulay.be, byethost3.com, clan.su, club.telepolis.com, depositfiles.com, dump.ru, filehoster.ru, freeforum.tw, funkyimg.com, gayhomes.net, gratisweb.com, grou.ps, hotshare.net, i037.radikal.ru, image5.poco.cn, imagecross.com, imagevenue.com, imgsrc.ru, indexjunkie.com, ipicture.ru, letitbit.net, mail.su, megaupload.com, multipics.net, my1.ru, nakido.com, oo.lv, opendirviewer.com, pic.ipicture.ru, pic2us.com, picsbuddy.us, pornhome.com, pornspaces.com, pridesites.com, rapidshare.com, sapo.pt, sendspace.com, surge8.com, uploading.com, uppic.net, zshare.net`

IWF removal process

- Bank phishing websites removed in 4 hours (when known about), 2 days (fast-flux systems), 10 days (not known about)
- Part time volunteers remove scam websites in 1-7 days
- Child Sexual Abuse Image sites: average lifetime ~ 4 **weeks**
- Only thing removed slower is fake pharmacy websites
 - and they are not tackled by any group we can locate
- We were amazed to uncover this, and consider it a scandal
- Main reason appears to be lack of prompt contact with hosters
 - IWF “not authorised” to contact foreign hosting providers
 - INHOPE rules mean local hotline must act, not the IWF
 - IWF not going after domain names, only the hosting
 - IWF (& INHOPE) confused as to whether aim is to remove content or to catch the criminals

IWF, Wikipedia and the “Wayback Machine”

<http://www.lightbluetouchpaper.org>



UNIVERSITY OF
CAMBRIDGE

Computer Laboratory