Software Evaluation
Internet Watch Foundation
Child Abuse Image URL Service

Prepared for InternetNZ
January 2006

by Technology Research and Consultancy Services
Principal Consultant: Mark Harris

TRACS
TECHNOLOGY RESEARCH AND CONSULTANCY SERVICES

File:TRACS-INZ001

Executive Summary

We tested the Internet Watch Foundation's Child Abuse Images database, in conjunction with the Department of Internal Affairs Censorship Compliance Unit, to determine it's fitness for purpose as an Internet feed filter for the New Zealand environment.

Our initial belief was that we would be testing a complete software solution. This was not the case and we tested only a filter list of URLs. End to end solutions have been developed in the UK based on this list.

Tests were conducted on the integrity of the list/process and false positives (Test 1), proof of concept (Test 2), coverage (Test 3) and change (Test 4). Additionally, some ad hoc testing for localisation was undertaken.

In my opinion, the list has internal integrity.

The 'proof of concept' test was regarded as partially successful, as sites were blocked in the software that was used but not consistently.

Coverage, from a New Zealand perspective, was disappointing but not unexpected.

Updating the list was simple and the integrity of the list was not affected.

Overall, it's not a silver bullet.

The major problem with any list compiled in a foreign jurisdiction is that it won't necessarily apply the same legislative criteria. New Zealand's legislation covers more areas than the UK legislation, and things such as bestiality and necrophilia won't be blocked by this list.

It is a first step in providing protection to NZ Internet Users from images of this nature, and it is the recommendation of this report that InternetNZ endorse it as such. We also recommend that InternetNZ continues to have discussions with the major players in this field and keep a watching brief over other related initiatives.

Background

1. The Public Policy Committee of InternetNZ (previously, the Legal and Regulatory Committee) became involved in discussions regarding the effectiveness of a child abuse image filter maintained by the UK Internet Watch Foundation and the need for testing it in the New Zealand environment.

2. The testing of the filter is a co-operative endeavour involving InternetNZ, Internet Safety Group, Department of Internal Affairs and the Office of Film and Literature Classification.

3. The tests set out to establish two things:

   - What proportion of sites containing child pornography, legally classified as objectionable, are blocked by the IWF filter
   - What proportion of sites blocked by the filter do not contain objectionable material

4. The Internet Watch Foundation (IWF) was formed in 1996 following an agreement between the UK government, police and the internet service provider industry that a partnership approach was needed to tackle the distribution of child abuse images (often referred to as child pornography) online.

5. Essentially, the IWF operate the only authorised 'hotline' in the UK for the public to report their inadvertent exposure to illegal content on the Internet. They provide a ‘notice and take down’ service to ISPs in the UK so they can remove potentially illegal content from their servers and work closely with law enforcement agencies at home and abroad to help them trace offenders.

6. The IWF works in partnership with UK Government Departments such as the Home Office and the Department of Trade and Industry to contribute to initiatives and programmes developed to combat online abuse. This dialogue goes beyond the UK and Europe, to ensure greater awareness of global issues and responsibilities.

7. They are funded by the EU and the UK internet industry including Internet Service Providers (ISPs), Mobile Network Operators & manufacturers, Content Service Providers (CSPs), telecommunications & software companies and credit card bodies.

8. The project to test the IWF filter grew out of discussions between the Internet Safety Group and British Telecom (BT), who have implemented a solution in the UK using the IWF filter. This solution is known internally as Cleanfeed and used by BT to safeguard their own customers. They welcome ISPs to join the initiative and are willing to share their technology. This should not be confused with CleanFeed (www.cleanfeed.co.uk), a transparent proxy service run by Legend Communications for their own broadband customers. As far as can be determined, Legend is not making use of the IWF list.

9. InternetNZ became involved because of our connections and support of the Internet Safety Group and we offered to supply neutral technical capability and machines to be used in testing. Because of the nature of the material being tested, it was vital to include the Department of Internal Affairs Censorship Compliance Unit (CCU) and the Office of Film and Literature Classification (OFLC). Steve O'Brien of the CCU made space available within the Unit of InternetNZ to set up the machine used and provided the ADSL connection. The officers of the Unit also advised on whether individual images were in breach of the law when the tester was unsure.

10. Testing was undertaken by Mark Harris, a Councillor of InternetNZ over a period of 5 days in December 2005.

What was tested.

11. Some confusion existed when we began as to what was being tested. InternetNZ understood that a software package of some nature was involved and supplied 2 PCs running Windows XP with an ADSL router and a 56K modem, prepared to do performance tests. Firefox was used as the browser.

12. On our behalf, Steve O'Brien made contact with the IWF and obtained permission and access to the IWF Child Abuse Images URL Service. The IWF supplied passworded access to their Service and update facility at no charge. Once the package was downloaded and examined, and the IWF website researched in depth, we realized that the IWF simply supply a URL list and it is up to the IWF member making use of the list to incorporate it into a filtering mechanism.

13. So, there was no software to test for performance, and the 56K machine became irrelevant. Later discussions with Liz Butterfield at the ISG indicated that, because ISG's discussions with BT had sparked the initiative, and BT referred to their “Cleanfeed” implementation, we had become confused about what was actually available for testing. Note that InternetNZ minutes on the project refer to the “Cleanfeed trial”.

14. I rapidly revised our simple test plan. There were 3 key things we wanted to find out:

    - did the list block sites for reasons other than objectionable material (e.g. political sites);
    - what proportion of objectionable material was blocked; and
    - what legitimate material was blocked (i.e. false positives)

15. The IWF list is compiled by hand. URLs are reported by the public via the Hotline and these are checked by hand before being added to the database. URLs are re-checked at regular intervals and removed from the database if they are no longer existent or have removed the material.

16. Copies of the list are downloaded by subscribing members, which may be ISPs or large companies wishing to filter staff access.

17. The list was very detailed, sometimes down to a specific image. It contained 857 individual URLs, many of which pointed to locations on the same server. Some of these locations were to specific user directories on public image repositories (such as Photobucket). Some were to pay-to-view commercial sites. Some were to message boards that, in themselves, were not objectionable, but contained links to objectionable material, especially links to file-sharing sites.

18. Because each entry to the list is manually verified, it is unlikely to be a false positive. The value of the list is directly related to the integrity of the provider. The IWF is very aware of this and go to some lengths to preserve that integrity. Although they work closely with the Home Office and Police, they stress they are a non-governmental organisation (see http://www.iwf.org.uk/public/page.148.htm). In order to verify the assumption that the list contained no false positives, I decided to manually verify each URL. This would also serve to verify that blocks were only placed on sites for the stated reasons and not because of political influence. This became **Test 1**.

19. The list is of no use without some form of filtering software. Ideally, this will be at the firewall/proxy server level, rather than the user machine, as this is more effective and less likely to be compromised. I decided to download some client level programs to see if they would accept the list and make use of it, as a proof of concept. This became **Test 2**.

20. **Test 3** stayed as originally planned – perform Google searches based on standard

keywords and check the list to ascertain whether casually obtained sites would be blocked.

21. **Test 4** would be to update the list using the function provided by the IWF and compare the two lists, reviewing and verifying any changes.

Results

22. For obvious reasons, I will not be specifying URLs in this report, and some technical detail may be under-reported as it is not my intention to provide opportunities for individuals to breach New Zealand law. I am prepared to answer questions, of course, and hopefully satisfy any inquirer that any concerns they may have have been considered.

Test 1: Analysis and verification of the IWF list.

23. The list contained 857 individual URLs. Each URL was regarded as a separate entity, although many referred to the same domain name, in the same way that many bloggers use blogger.com as their base domain. A significant number referred to different pages belonging to the same website – these were also regarded as separate, as they would have been the results of separate complaints. To get entered on the IWF list, a specific complaint or report must be made for each URL through the IWF Hotline (see appendix 1 for process)

24. If a site proved to be objectionable under New Zealand legislation, a PDF print of the site was taken for evidential purposes. Advice was occasionally sought from CCU personnel as to the likelihood of illegality, although the CCU stress that formal decisions require the OFLC to review the material. As we were not seeking classification of the images but merely verification of the list, I decided not to send material to the OFLC in order to save time and resources. Discussions were held with the Chief Censor and OFLC staff regarding the nature of the material and the service and they were comfortable with this approach.

25. A number of URLs (around 40%) returned a message indicating that material had been taken down in response to public complaints or a simple 404 message. Similarly, a number of URLs returned “server not found” messages, indicating that some action to remove the site had been taken by an upstream provider.

26. A number of URLs returned pages in Foreign languages or character sets. Unless an illegal image could be seen on the page, it was assumed that the page had been withdrawn.

27. Many of the URLs lead into 'link farm' networks, which are lists of sites purporting to be the material that the inquirer is searching for, but links go to similar lists of links ad nauseam. This is a standard ploy in the amateur (and professional) porn industry and serves 2 purposes – inflating a site's Google PageRank(tm) rating by having multiple links leading to the site, and having large numbers of page impressions to market to advertisers and collect fees. Even so, many of these sites contain thumbnails of images that fall into the objectionable category, and the sites themselves would fall afoul of Section 1(A): “ the publication promotes or supports, or tends to promote or support... [t]he exploitation of children, or young persons, or both, for sexual purposes”.

28. The statistical breakdown of failures ("failure" being "no child abuse images found") was:

403	404	NoServer	Adult	Foreign	No result	Hosting	LinkFarm	SubForm	Other
3.36%	8.05%	6.04%	34.90%	8.72%	6.71%	25.50%	0.67%	2.01%	4.03%

29. The codes are as follows:

- **"403"** is an HTTP error code indicating that access to a URL is not permitted due to a lack of authorisation. This may be an indication that a set list of permitted users has been put in place to block access, or that a hosting provider has withdrawn access privileges to all users.
- **"404"** is an HTTP error code that indicates a specified URL cannot be found.
- **"No Server"** indicates that a domain name could not be found. This was taken as indicating that domain names had been cancelled by providers.
- **"Adult"** means that attempting to view the URL resulted in a redirect to an adult porn site with a different URL. The interesting thing was that it was almost always the same adult site that the redirect went to, indicating either that there is a favourite adult porn site in the industry, or that a lot of these separate URLs are more interconnected at the business level than would superficially appear. A prevalence of adult URLs in this list would cause concern about the integrity of the list but, as they came from redirects, I believe that this is not the case and indicates hasty takedowns to avoid prosecution.
- **"Foreign"** indicates that I was unable to read the content of the page, but context indicates no objectionable material. In some cases, links were available and I followed them in order to gain context. Some foreign pages, led to objectionable material, but most indicated that the host had removed or blocked access to the material.
- **"No Result"** indicates that nothing was received by the browser, not even an error message.
- **"Hosting"** indicates a message from the hosting company that the site has been removed or an account has been terminated.
- **"Link Farm"** indicates a page of links which don't present objectionable material (see para 27). It could be argued that these pages are objectionable in themselves, under the "support and promote" provisions of the NZ legislation.
- **"SubForm"** indicates a programattic submission form to be included in a site. If no objectionable images were included, such a form was deemed 'not objectionable'.
- **"Other"** - there were a small number of other reasons for deeming a URL "not objectionable" under NZ legislation, including "young girl in bath, not erotic". I stress that this was my opinion, and would need to be verified by the OFLC if the list was formally used in New Zealand.

30. After 3 days of viewing 300 sites, "porn fatigue" set in and the remaining URLs were sampled at random. While allowing that some foreign-language sites may have been incorrectly included in the list, as it was not possible to effectively translate those sites, no sites were found that had been listed for anything other than child abuse images. (See Appendix 2)

Test 2: Incorporating the IWF list in Filtering Software

31. I downloaded 5 trial products to attempt a proof of concept for making use of the filter list. The products were:

- Anti-Porn.exe http://www.tueagles.com/anti-porn/
- BrowseControl.zip http://www.panvasoft.com/eng/4259/
- IPfree.exe http://www.softforyou.com/ip-index.html
- ParentalFilterSetup.exe http://www.ecommsec.com/pf/
- Puresight_pc.exe http://www.puresight.com/products/isp.shtml

32. Results were patchy. Most of these products only allow you to enter individual URLs rather than a large list. I did enter a number of URLs into each program by hand, and achieved some blocking, although user agent caching overrode the blocking on two occasions. I was very aware of the time limitations of this project, and did not persevere with Test 2. I am confident that the list can be used easily in server-level filtering mechanisms.

Test 3: Performing Google searches with keywords

33. A Google search on a common keyword produced 385,000 English pages. Analysing the first 100 results against the IWF list gave poor results, in percentage terms, as only 2 of the sites were on the list. As the aim was to determine how much casual access the list would block, this had to be counted a failure for the list.

34. Discussions with CCU personnel indicated that this may be due to the sheer volume of material on-line purporting to be child abuse material. Much of it is not, or would not fall under the provisions of New Zealand legislation, and the Google search contained a number of URLs that lead to sites that appeared to have no connection with such material. The IWF list, on the other hand, only included URLs that had been verified as containing objectionable material. It was hardly a surprise, then, that the percentage of total material available on Google would contain much that was not listed in the IWF database.

Test 4: Test the Update Function and Verify Changes.

35. On day 4 of the test (2 December), I updated the list, having kept a copy of the previous list. There were still 857 URLs in the list, but comparison showed that 84 from the original list had been dropped to be replaced by 84 new URLs. The 84 that had been dropped were URLs that had no objectionable material on them when I checked, and random assessment of the 84 new URLs indicated that the distribution between objectionable material and missing content was similar to the results for Test 1. The update process was seamless and would be simple to automate.

Additional Testing

36. CCU personnel expressed concern about the UK-centric nature of the database and noted that much spam received in NZ points users to localised versions of the content (i.e. using Asia/Pacific domain names and IP addresses). CCU supplied me with copies of the emails and the URLs were compared to the database. No matches were found.

Issues and comments

37. The IWF list is based upon UK law. NZ law is more restrictive, especially with regard to the "promote and support" provisions. While the URLs in the database may have material that would be classed as objectionable, material that doesn't meet the UK criteria might still be objectionable in New Zealand. The same problem would exist in using a database from any other jurisdiction.

38. The IWF list is based on Internet user reports. This means that someone has to view the material and object to it before any action is taken. Unfortunately, an alternative approach does not exist as yet, as automated tools lack the judgement required.

39. The IWF list does not block entire servers based on one URL proving to be objectionable. The positive aspect of this is that blocking URLs on shared servers will not result in inappropriately blocking material on that server that doesn't meet censorship criteria. The

negative side is that other objectionable material on the server can still be accessed.

40. The IWF CAI Service only covers images of child abuse. While the IWF Remit includes criminally obscene content and criminally racist content, it can only work with this material on sites hosted in the UK. Thus, images featuring acts of extreme sexual activity such as bestiality, necrophilia, rape or torture are not included in the database, although these would likely be objectionable under New Zealand law..

41. URLs in the database are monitored and polled randomly to see whether they are still live. If not live, they are removed from the database. If they are still live, and the content has changed, they are referred to analysts for reclassification. Thus, the database doesn't get stale and full of dead URLs

42. The IWF has a formal complaints and appeals process for content providers who believe they have been unfairly added to the database.

43. The IWF database does not list URLs for UK hosted material as, under the Sexual Offences Act 2003, the IWF is empowered to issue a formal takedown notification, with which UK ISPs are required to comply.

44. Security around use of this list is paramount, as it amounts to a simple shopping list if in the wrong hands. Users of the list in Europe are aware of this, and the IWF has standard security processes it requires subscribers to implement, under their contract.

45. ECPAT, in Sweden, manage a similar list with similar processes, and are part of an implementation initiative. I attended a presentation at DIA from their New Zealand representative regarding the service. Similar initiatives are also being or have been implemented in other European countries and the United States.

Conclusions

46. While it was initially thought that under test was a software solution for filtering illegal content, testing was achievable only on a component of such a system – the database of URLs.

47. In order to determine its usefulness, tests were conducted on the integrity of the list/process and false positives (Test 1), proof of concept (Test 2), coverage (Test 3) and change (Test 4). Additionally, some ad hoc testing for localisation was undertaken.

48. In my opinion, the list has internal integrity. Half of the live URLs checked had material that is objectionable under New Zealand law. Of the half that did not, 35% had redirects to adult porn (the same adult sites recurred regularly, indicating a hasty redirect), 49% were no longer available and the remainder were foreign and difficult to interpret.

49. The 'proof of concept' test was regarded as partially successful, as sites were blocked in the software that was used but not consistently. As the list was not meant for client level tools, this was not too concerning, and problems may have been due to hasty mis-configuration of the tools in question. Implementing the list at a routing level would be more effective.

50. Coverage, from a New Zealand perspective, was disappointing. Given the large number of sites of a similar nature easily findable through Google et al, and especially localized spam messages, a list of 857 URLs was never going to have a huge impact. Our initial understanding was that there would be 1500 URLs on the list, but even that would not be hugely effective. Sadly, it's a drop in the bucket, which doesn't rule it out completely.

51. Updating the list was simple (and about to get more so, as the IWF are moving to an encrypted email update process, rather than relying on users to pull the file down) and the integrity of the list was not affected.

52. Overall, it's not a silver bullet. It won't touch the serious paedophile networks, which operate in chat rooms and file sharing services. It may have an impact on the casual surfer, but even this would be minimal at this stage. The main benefit of this sort of operation is the capability for users to report material to “the authorities”, and it is not clear that New Zealanders would be able to do that effectively with regard to the IWF.

53. More importantly, it does not prevent knowledgeable users from finding ways around it to get to material they know exists.

54. The major problem with any list compiled in a foreign jurisdiction is that it won't necessarily apply the same legislative criteria. New Zealand's legislation covers more areas than the UK legislation, and thinks such as bestiality and necrophilia won't be blocked by this list.

55. That's not to say it's a complete failure – it does do what it purports to do, but it is only a small and relatively expensive part of any solution, if the £5000 price tag discussed is the final individual cost to subscribers.

56. DIA are very interested in the concept of encouraging ISPs to put filtering technology in place. It would be beyond the resources of most ISPs to pay this sot of money for something that does an incomplete job. There was some discussion that a central body might subscribe to the list and redistribute it within New Zealand, suitably enhanced for the NZ environment. As mentioned above, ECPAT presented on a network they are part of in Sweden. A combination of the two lists might be a good starting place.

57. It can be fairly questioned whether this is part of InternetNZ's remit, and whether we should be involved in anything more than evaluating systems for 'uncapturability'. During Council's planning sessions, there seemed to be a consensus that we want the Internet to be seen as a safer, more trustworthy place ton increase usage. Encouraging and participating in discussions and initiatives such as this is a useful way to achieve that goal.

58. The main question from our members to be asked of this testing was “Does the list block material other than child pornography?”, thereby chilling freedom of speech. The answer is “No, it does not”. That doesn't mean it could not, at some time in the future, be used in that way, and some ongoing effort would need to occur to ensure thatdidn't happen. However, the risk is low, given the constitutional setup of the IWF, and the list of current members/subscribers who would resist that course of action (see http://www.iwf.org.uk/funding/page.64.htm).

59. This issue will obviously raise some discussion among InternetNZ members, regardless of the position taken by Council, from the “won't someone think of the children?” brigade to “sharp end of the wedge” people. This is a good thing. The issues around what is acceptable online and what is not need to be constantly revisited and reviewed.

60. There is a question as to whether a partial solution like this is worth endorsing or even considering, and it's a valid one. However, instead of worrying about the 'last mile' in terms of what is not covered, I believe we need to consider the 'first mile' – the first steps in removing objectionable material from users' experience of the Internet, in order to make it a more trusted means of communication. This list is a baby step at best, but may be a good means of learning how to do it better.

Recommendations

61. I recommend that:

- InternetNZ endorse the integrity of the IWF Child Abuse Images URL Service as a useful tool in preventing inadvertent viewing of child abuse images, with the caveat that it is not a complete solution in itself;

- InternetNZ continue in dialogue with the entities in this area, namely, DIA (notably CCU and OFLC), the Internet Safety Group, ECPAT, NZ Police and others, to ensure that the Internet User viewpoint is considered in any programme regarding filtering of Internet material.

Appendices

Appendix 1

IWF process flow – as determined from their website and confidential documents supplied during evaluation period

Appendix 2

Obfuscated testing results