

Expert Rebuttal Report of Benjamin Edelman

Multnomah County Public Library et al.,
vs.
United States of America, et al.

I. Purpose and Summary	1
II. Inability of Blocking Programs to Restrict Access Only to Non-Web Internet Content that Meets Specific Category Definitions.....	1
Images Transferred via Email	2
Images and Videos Transferred via Streaming Video	3
III. Blocking of Search Engines and Other Web Services	6
IV. Flaws in Methodology of eTesting.....	7
V. Flaws in Methodology of Finnell	12

I. Purpose and Summary

I have been asked by counsel for the plaintiffs to review and respond to certain aspects of expert reports submitted by Finnell and eTesting on behalf of the government. In some cases my responses elaborate on research reported in my first expert report, submitted October 15, 2001.

II. Inability of Blocking Programs to Restrict Access Only to Non-Web Internet Content that Meets Specific Category Definitions

The eTesting report implies that blocking programs are capable of blocking content that is delivered from a source other than via the World Wide Web but that nonetheless meets CIPA’s definitions.¹ As also described in my first report, that claim is significantly overstated because the blocking programs are at best able to block all content transmitted via protocols other than HTTP, regardless of whether the content

meets specific category definitions. In particular, they are not able to block access to specific content transmitted via a non-HTTP protocol while simultaneously allowing access to other non-HTTP content.²

The following table summarizes my findings regarding my ability to view sexually-explicit non-web Internet content even when selected categories of Internet sites were blocked (as described in my first expert report).

	Cyber Patrol	N2H2	SmartFilter	Websense
Allowed viewing of sexually-explicit images via Hotmail or Yahoo Mail	Yes	Yes	Yes	Yes
Allowed viewing of sexually-explicit streaming video	Yes	Yes	Yes	Yes

Images Transferred via Email

Blocking programs ordinarily fail to restrict content delivered by email even when such content meets their category definitions. When email is retrieved via a standalone email program (such as Microsoft Outlook or Eudora), blocking programs ordinarily have no opportunity to review the content prior to its receipt by a user. Furthermore, when email is retrieved using a web-based email program (providing email access via an interface within a web browser, as via Hotmail), blocking programs have no way to differentiate between content that meets their criteria versus content that does not. I personally verified that even when Cyber Patrol, N2H2, and Websense were configured as described in my first report, I was able to retrieve via email to my Hotmail account images that were retrieved from URLs classified as sexually-explicit.³

¹ Report of eTesting Labs, at 4, 12.

² Expert Report of Benjamin Edelman, at 30-1.

³ I tested a sexually-explicit image from a site blocked by all four filtering programs, configured as described in my initial report, that was also classified by Finnell's expert report as correctly blocked. After

Smartfilter addressed this potential underblocking by blocking all image content delivered through Hotmail, regardless of whether the content was consistent with Smartfilter's category definitions. In my testing, absolutely no Hotmail image content was available; empty space replaced both Hotmail's navigation buttons (essential to the navigation of Hotmail's user interface) and all email images received (including those that did not meet Smartfilter's category definitions). However, even this method, which resulted in serious overblocking, was ineffective in preventing access to the specified sexually-explicit content. When I instead attempted to receive the same testing image using a web-based Yahoo Mail account, still accessed from a computer restricted by Smartfilter, I was able to view the image without difficulty.

Thus, none of the blocking programs were able to differentially allow access to sexually-explicit image content received via web-based email. Three programs allowed access to all email image content, including sexually-explicit image content, while one allowed access to no image content. However, no program differentially prevented access only to sexually-explicit images, and I know of no blocking program that can do so or that purports to be able to do so.

Images and Videos Transferred via Streaming Video

Blocking programs are also unable to prevent access to only specific content delivered via streaming video. Streaming video playback is ordinarily delivered via specialized protocols such as RTSP, PNM, and MMS and is ordinarily played in a specialized application (such as the RealNetworks RealPlayer or Microsoft Windows

I used Internet Explorer to save this file to my local disk, I emailed it to my Hotmail and Yahoo Mail accounts from an ordinary email account.

Media Player) outside the web browser.⁴ As a result, ordinary blocking software, which solely or primarily filters only the HTTP protocol, is unable to differentially allow access to certain streaming videos but not to others. While some of the programs I have reviewed suggest methods of configuring network routers or other equipment to block all streaming video access,⁵ [REDACTED]

Furthermore, no blocking program can block access only to certain portions of specific video files (such as the specific sexually-explicit scenes of ordinary movies that might be posted on the web) without also blocking the entirety of the files.

In some instances, blocking companies may seek to limit access to particular sources of streaming video by limiting access to the web pages that link to streaming video meeting blocking programs' category definitions. However, this method is ineffective in preventing access to streaming video content because users may nonetheless directly access the desired content, without passing through the web pages that link to the streaming video content. In particular, in leading streaming video players, it is possible to manually enter specific streaming video locations into a player's File-Open Location (RealPlayer) or Open URL (Windows Media Player) dialog box. In this way, users may readily circumvent streaming video blocking that operates at the level of

⁴ In some instances, videos may be retrieved via the same HTTP protocol used by ordinary web content; in these cases, blocking programs can differentially block access only to videos that contain material that meets certain category definitions. However, the most popular and technically robust methods of delivering streaming video do not use the HTTP protocol and therefore cannot be filtered in this way.

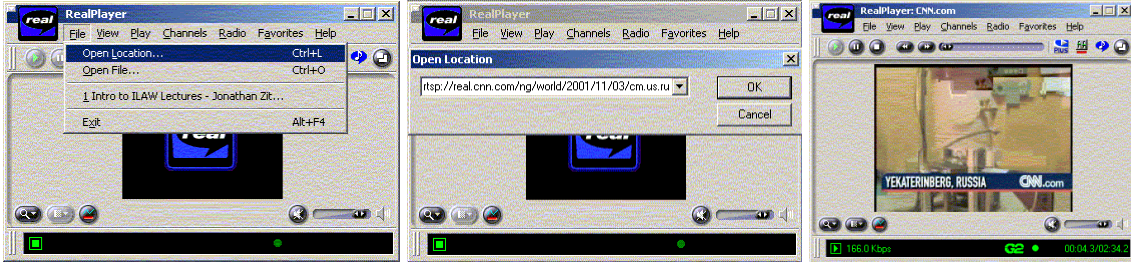
⁵ "Packet Filtering for Other Popular Internet Services."
<http://www.n2h2.com/support/router/other_services.php>

web pages that link to streaming video clips.⁷ With the failure of this attempt at differentially allowing access to certain streaming content but not others, blocking programs are left with only all-or-nothing access control; thus, available blocking software is unable to prevent access to certain streaming content while allowing access to all other streaming material.

To verify this behavior, I used search engines to find web sites that provide free sexually-explicit streaming videos in RealVideo and WindowsMedia formats. For each resulting site, I used an ordinary computer (with access not restricted by blocking software) to determine the address of a specific streaming video clip from each site. On computers with access restricted by blocking programs, I then used RealPlayer and Windows Media Player to view the selected content by entering the corresponding addresses directly into the Open Location and Open URL dialog boxes of the RealPlayer and Windows Media Player, respectively. In every instance, I was then able to view the selected video content even though, in most cases, the underlying web site (which had provided the link to the video content) was inaccessible from the computers with filtered Internet access.

The screen shots below depict the use of the File-Open Location feature in version 8 of the RealPlayer. By entering the RTSP://... URL directly into the RealPlayer, as shown below, it was possible to view streaming video from CNN without first visiting the CNN web site. A similar method was used to view sexually-explicit content even from a computer with access restricted by blocking software. Windows Media Player offers a similar interface for performing the same task.

⁷ “Information regarding monitoring or blocking streaming audio and video.”
<http://www.surfcontrol.com/support/knowledge_base/showrecord.html?id=321>



Thus, in my testing, no blocking program was able to differentially block access to all sexually-explicit streaming video content. Some programs suggested ways to configure network hardware to prevent access to any streaming video content, but no program differentially prevented access only to sexually-explicit streaming video while allowing access to all other streaming video. I know of no blocking program that can do so or that purports to be able to do so.

III. Blocking of Search Engines and Other Web Services

The Finnell report notes that search engines “posed a challenge” to accurate classification. Finnell finds one specific instance of overblocking of search engine results pages, but the problem is larger than he suggests.⁸ Indeed, in a variety of instances, blocking software may prevent access to certain search result listings at Internet search engines. (I mean here to refer not just to the pages throughout the web that are linked by search engines, but to pages actually on the search engines’ own web sites.) For example, when N2H2 is configured to block the category “Search terms,” it did not allow searches via any of Yahoo, Google, or Lycos on keywords including “sex education,” “sexual health,” “sexual reproduction,” and “sexually transmitted disease.” N2H2’s decision to block search terms seems to result from a simple “string contains” test;

⁸ Expert Report of Cory Finnell, at 6.

assumptions that lead to systematic underreporting of the proportion of sites wrongly blocked. In particular, they assume that sites wrongly blocked are sites that “by the nature of their content might be confused by content filtering software.” However, as my first expert report shows and the Finnell report confirms,¹¹ many sites are wrongly blocked despite no obvious inclusion of such content. For example, it is not obvious that Southern Alberta Fly Fishing Outfitters (<<http://www.albertaflyfish.com/>>) contains any content likely to confuse any human or automated review process, yet that site was classified by N2H2 as pornography and by Wensense as sex in a total of five distinct tests over a period of three months. Similarly, the Action for Police Accountability site (<<http://www.bayswan.org/APA.html>>) is not obviously likely to confuse any rating system, yet it was classified as by N2H2 as sex, by Smartfilter as sex, by Cyberpatrol as Adult/Sexually-explicit, and by Websense as Adult Content in a total of fourteen tests over four months. Appendices A and B of my expert report are filled with numerous additional examples of such sites – sites that are seemingly blocked randomly, and that clearly lack any of the “confusing” content that eTesting took to be the primary cause of overblocking. Many of the wrongly blocked sites on Finnell’s Attachments B, C, and D also lack any content that is likely to be confused with sexually-explicit materials; for example, the Disney Music Page (<<http://www.dismusic.com>>) is blocked by Cyber Patrol according to Finnell’s testing, even as it lacks any content obviously related to sex, instead offering only music heard in Disney movies and at Disney theme parks.

It is my sense that overblocking results from two general causes: “confusing” misclassified sites (like those described by eTesting) and “randomly” misclassified sites

¹⁰ “BESS's Secret LOOPHOLE.” <<http://www.sethf.com/anticensorware/bess/loophole.php>>

¹¹ Expert Report of Cory Finnell, at 6-7.

(like those just described).¹² eTesting’s study provides an estimate of the frequency of the former problem, but fails to take into account the latter problem, which can only add to the proportion of sites misclassified by blocking programs. As a result, eTesting’s analysis systematically underestimates the true proportion of sites wrongly blocked by the blocking programs tested. The thousands of wrongly blocked sites documented in my first report confirm this.

eTesting’s conclusions about underblocking may also fail to properly quantify the true rate of underblocking of web sites meeting blocking companies’ category definitions. eTesting reports that they selected distinct URLs “randomly ... using a variety of searching techniques,” but this description of their methodology raises significant doubts as to the representativeness of their sample relative to the full set of web sites meeting category definitions. The word “random” suggests that all URLs meeting category definitions were equally likely to be selected into the eTesting sample, but this is sure to be false since search engines index only a small portion of the web.¹³ As a result, it is likely that the eTesting sample overrepresented well-known sites and underrepresented more obscure sites that may also meet category definitions. This effect is even more pronounced if eTesting selected sites from the top pages of results (those results first returned by search engines in response to their selected search terms), rather than from the entirety of results pages.¹⁴ [REDACTED]

¹² My initial expert report discusses certain reasons why these “randomly” misclassified pages may have been misclassified. In particular, these pages may have been misclassified in the course of overbroad classification of a larger realm of content, such as an entire domain name, an entire IP address, or an entire block of IP addresses.

¹³ “Accessibility and Distribution of Information on the Web.” <<http://wwwmetrics.com/>>

[REDACTED] Thus, since the eTesting sample oversamples well-known sites, it produces an overestimate of the rate of correct blocking of sites meeting blocking programs' category definitions.

To quantify this effect, I retrieved from Google the 797 distinct URLs Google found to be associated with the search criteria "free adult sex."¹⁶ The plaintiffs' counsel asked me to remove two sites which clearly did not meet the category definitions for sexually-explicit content of the four programs; I discarded these two sites, and proceeded with the remaining 795. In particular, I attempted to access each of these URLs on computers with Internet access restricted by Cyber Patrol, N2H2, Smartfilter, and Websense, each configured as described in my first report. I found that the blocking programs blocked a higher proportion of the highest-rated sites from Google's list than they did of the lower-rated sites on Google's list. In this way, I confirmed that if eTesting in fact sampled from the first several hundred sites on Google's list, their analysis did indeed overstate the rate of blocking of content meeting category definitions.

The following table details my specific results. In this table I have arranged Google's results into groups of 100, and I have calculated the proportion of sites blocked within each group.¹⁷

Rank Range	Proportion of sites blocked
1-100	0.8625

¹⁴ eTesting's report says that they went beyond the first fifty listings when forming their sample of URLs for CBR estimation. However, it is likely that they did not go far enough; my testing suggests that this is the case, as discussed below.

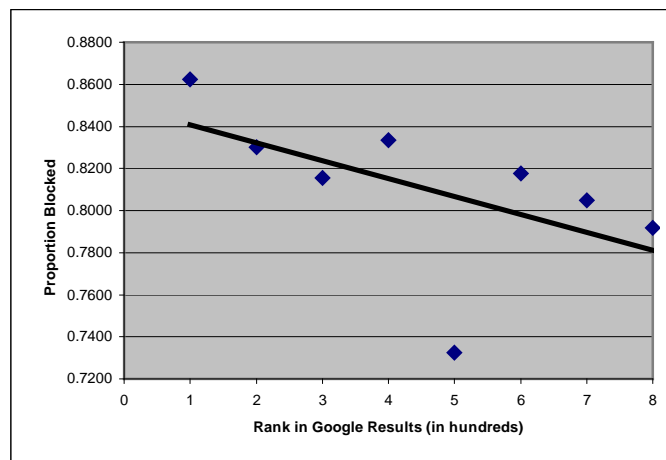
[REDACTED]

¹⁶ Google initially reported that more than one million URLs matched this search expression. However, after I retrieved the eighth page of results (in 100-URL batches), Google reported that there were no more results available.

¹⁷ Each of the 100 sites in each grouping could be blocked by as many as four blocking programs. Thus, the first seven rows of the right column reflect averages over 400 observations (4 tests for each of 100 URLs). The eighth row reflects an average over 384 observations (4 tests for each of the 96 URLs).

101-200	0.8300
201-300	0.8157
301-400	0.8333
401-500	0.7325
501-600	0.8175
601-700	0.8050
701-795	0.7917
Overall	0.8113

I also prepared a scatter plot that shows the decline in proportion of sites blocked as a function of rank in Google’s search results. The plot includes a linear trendline, as prepared automatically by Microsoft Excel; I understand this trendline to reflect the results of an ordinary least squares regression.



Computing results separately for each of the four blocking programs, I found that in the sample tested, the blocking programs blocked the following proportions of the 795 web sites:

Blocking Program	Proportion of sites blocked
Cyber Patrol	0.8163
N2H2	0.8943
Smartfilter	0.7346
Websense	0.8000

In contrast, when I restricted analysis to the top fifty sites from Google’s search results, blocking proportion results were as follows:

Blocking Program	Proportion of sites blocked
Cyber Patrol	0.9400
N2H2	0.9600
Smartfilter	0.8200
Websense	0.9200

On the basis of this analysis, I cannot agree with eTesting's estimates of the Correct Blocking Ratio. Instead, my tests suggest that Websense, N2H2, and Smartfilter likely correctly block a significantly smaller portion of sites than eTesting estimates.¹⁸ (Of course, blocking programs are even less accurate than my sample above suggests; as previously discussed, blocking programs are likely to be relatively more successful at preventing access to the well-known sites listed on search engines than to more obscure sites.) I conclude that notwithstanding eTesting's description of a methodology designed to sample more than the most popular sites returned, their error likely results from oversampling the most popular sites. This conclusion is strengthened by the notable similarity between eTesting's results and my results when analysis is restricted to the top fifty results from Google.

V. Flaws in Methodology of Finnell

The Finnell study also uses methods that prevent or weaken inferences that apply their findings to a broader set of Internet sites. Finnell biases his estimates of the rate of overblocking of N2H2 and Websense via flaws in his method of analyzing updates in current versions of the programs' site lists. When Finnell checks for updates to blocking programs' site lists, his report reflects that he tests only those sites that he had initially found to be wrongly blocked; in each case, he finds that a portion of these sites have since been reclassified and are thus no longer blocked. As a result, Finnell concludes that the proportion of wrongly blocked sites has decreased since the data was collected, and

he reports these new lower error rates. However, Finnell's procedure fails to consider the possibility that sites that were previously correctly not blocked have since been misclassified and would therefore be blocked were library patrons to attempt to access them at this time. As a result, Finnell's adjustment procedure systematically underreports the amount of overblocking taking place; his analysis includes corrections to the list that reduce errors, but it does not include updates to the list that in fact cause additional erroneous denial of access.

It seems likely that this error causes a substantial bias in Finnell's adjusted estimates. By Finnell's analysis, his adjustment procedure causes the most likely estimate of Websense's error rate to decrease from 8.14% to 6.69%. Similarly, his estimate of N2H2's error rate decreases from 8.14% to 6.92% as a result of this adjustment procedure. Assuming each library used up-to-date site lists, this adjustment procedure should reflect only changes to site lists that were made by blocking companies in the period between data collection and Finnell's analysis. In the case of Websense, this period is less than two weeks – from October 1-3 to not later than October 15. In the case of N2H2, the period is approximately six weeks, from August 2-15 through not later than October 15.

Increases in accuracy of two to three percent in such a small amount of time are not consistent with my observations of trends in error rates over the past year. While I have not sought to quantify changes in accuracy rates over time, my testing (taking place over many months) and my background knowledge both suggest that, if blocking programs are becoming more accurate over time, their rate of improvement is considerably slower than Finnell's data suggests. [REDACTED]

¹⁸ Report of eTesting Labs, at 4

[REDACTED]

[REDACTED] Thus, it is especially important to take into account that, had libraries used newer versions of site lists, they would have wrongly denied access to numerous additional sites even as they allowed access to certain of the sites wrongly blocked in the logs Finnell examined. Finnell's adjusted estimated error rates fail to include this consideration and are, for that reason, likely to be in error.