Expert Report of Benjamin Edelman

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

I. Qualifications 1
II. Purpose and Summary 2
III. Documentation of Specific Misclassifications in Popular Blocking Programs 4
Selection and Configuration of Blocking Programs 4
Sites Tested 5
Testing Methodology 6
Archiving of Blocked Sites 6
A Guide to Appendices A, B, and C 8
Results 10
IV. The Design and Operation of Internet Blocking Software 14
Installation Method and Location 14
Category Lists and Criteria for Blocking 15
Overview of Web Site Classifications 16
The Use of Automated Systems for Site Classifications 17
The Use and Omission of Human Review 18
The Scope of Classifications 19
The Unavailability of Site Categorization Lists for Review by Customers 20
Failure to Review Sites Again Once They Are Classified 21
Customizing Site Categorization Lists 21
V. Problems with Blocking Software 23
Overblocking 23
Human Error and Judgment Calls 23
Technical Error 25
Underblocking 27
Inability to Block Images Only 28
Other Flaws 29
VI. Compliance of Blocking Software with the Special Provisions of CIPA 32
Creating Exceptions for Research Purposes 32
VII. The Flaws of Blocking Programs are Fundamental 35

I. Qualifications

My name is Benjamin Edelman. I am a systems administrator and multimedia specialist at the Berkman Center for Internet and Society at Harvard Law School, and I am a senior at Harvard College concentrating in Economics. My Berkman Center work includes substantive research as well as systems administration. In the course of this work, I have become generally familiar with systems that attempt to restrict access of Internet users to particular web sites or other Internet content.

During the past ten years, I have worked as a computer purchasing consultant, a network designer and systems integrator, a custom software designer, a database specialist, and a designer of database-driven web sites. My experience includes six years as an Internet web server administrator, including operation of a server ordinarily receiving more than 20,000 hits per day. In addition, I have seven years of experience with the TCP/IP protocol on which the Internet is based, including six years administering TCP/IP-based networks.

I have previously provided expert testimony in Pittsburgh federal court in National Football League, et al., v. TVRADIONOW Corporation.

I have attached a current copy of my resume.

II. Purpose and Summary

I have been retained by Ann Beeson and Chris Hansen, who are counsel for the plaintiffs in this case, to provide expert testimony in this case. I received $[redacted] per hour for my work. Specifically, I was asked to identify and describe the capabilities and flaws of widely-used Internet blocking systems. I was also asked to design and implement systems to identify particular sites that are blocked by four specific Internet filtering programs but which do not fit within the programs' self-defined categories for blocking; I also verified that certain other sites were blocked, including a number of web sites operated by plaintiffs in this case. In addition, I prepared CD-R's with archives of the content on the blocked sites as well as other information about the sites.

My testimony is based on my expertise as a systems administrator; my general knowledge of blocking programs; my review of the manuals and other technical documentation associated with blocking programs; my testing of Surfcontrol Cyber Patrol, N2H2 Internet Filtering, Secure Computing SmartFilter, and Websense Enterprise; and my attendance and review of transcripts and documents from the depositions in this case of representatives from these Internet blocking software companies.

I have concluded that installation in libraries of Internet blocking programs configured to block particular categories of Internet content will inevitably block Internet content that does not meet the programs' self-defined category definitions. I have further concluded that these programs 1) are incapable of blocking only images on the Internet that meet certain definitions; 2) are incapable of blocking all content on the Internet that meets certain definitions; 3) are incapable of blocking Internet content that meets certain definitions and is communicated through certain non-Web-based protocols; and 4) do not provide a practical or efficient system for configuring blocking programs to allow access only to a specific blocked site during a limited time period for a particular user for a specific purpose.

My general understanding of the Children's Internet Protection Act (CIPA) is that it requires libraries and public schools who participate in certain federal programs to install automated systems to prevent Internet access to images that are obscene or harmful to minors, or that depict child pornography. I further understand CIPA to allow the temporary disabling of blocking software for bona fide research purposes for some users.

III. Documentation of Specific Misclassifications in Popular Blocking Programs

Selection and Configuration of Blocking Programs

I first designed and implemented research in order to test and document specific instances of overblocking by four leading blocking programs recommended to me by counsel for the plaintiffs. The specific programs tested were Surfcontrol Cyber Patrol 6.0.1.47, N2H2 Internet Filtering 2.0, Secure Computing SmartFilter 3.0.0.01, and Websense Enterprise 4.3.0.[1] I installed each program on a fresh Windows 2000 Server with Microsoft Proxy Server 2.0.[2]

At the instruction of counsel for the plaintiffs, I configured the programs to block the following categories:[3]

Cyber Patrol: Adult/Sexually Explicit

N2H2: Adults Only, Nudity, Pornography, and Sex, with "exceptions" engaged in the categories of Education, For Kids, History, Medical, Moderated, and Text/Spoken Only

SmartFilter: Sex, Nudity, Mature, and Extreme

Websense: Adult Content, Nudity, Sex

I otherwise left blocking programs in their default configurations.[4]

[1] These version numbers reflect the software programs used in final testing. In some instances, initial testing used earlier versions of software from these companies. When new versions were released during the course of testing, I upgraded to the latest versions as available.

[2] The four programs tested are all server-based programs. In section "Installation Method and Location," I discuss the reasons why server-based programs are preferable to client-based programs for many libraries. Although each Windows 2000 Server installation was newly-created for the purpose of this testing, I of course applied all current service packs and security fixes to each system.

[3] These categories reflect the configurations of the latest version of blocking software, as used in final testing. When testing prior versions of the software, in limited instances I engaged different categories as available in those versions.

Sites Tested

I began my formal evaluation of these programs by preparing a database of sites to be tested for possible blocking.[5] This database included substantial portions of site listings from the popular Yahoo directory, as well as numerous additional sites indexed by the search engine Google. Counsel for the plaintiffs provided guidance, both general and specific, as to sites to be included and excluded from my testing.

Much of my testing focused on sites from the Yahoo directory because Yahoo directory categorizations are, in my experience, relatively accurate. This may be the case because the entirety of Yahoo's list is prominently posted for public review on Yahoo's web site. It may also be the case because Yahoo explicitly seeks assistance — from site creators as well as interested users — in continually updating its categorizations.[6] Thus, when Yahoo places a site in a particular category (for example, Government) but blocking programs classify it differently (say, as Sex), there is reason to investigate the blocking, as this blocking may well be in error. I separately obtained Google Directory categorizations for most sites, as available; these independent categorizations provided further information to inform inferences that certain sites were wrongly blocked.

After forming a list of Yahoo sites for which Yahoo's classification of each site differed significantly from the classification of one or more blocking programs, I asked Google for sites that it considered to be "related." While Google's definition of "related"

[4] Of course, configurations varied from defaults in that I had to specify administrator passwords as well as credentials to retrieve updated site categorization lists. In some instances, I also customized "access denied" pages to guarantee reliability of testing scripts and to simplify their implementation.

[5] In some instances, sites were added to this database after its initial creation. For example, after I determined that certain sites from Yahoo's listing were blocked, I also tested sites that Google reported were similar or related. Although testing proceeded in a series of rounds, each site ultimately reported in the Appendices to my testimony was verified to be blocked by at least one filtering program, as configured, on at least one instance, and for most sites in several instances. The Appendices report the specific dates of testing of each site.

[6] "How to Suggest Your Site." <http://docs.yahoo.com/info/suggest/>

is not publicly documented, it is likely that related sites include at least some of similar keywords, links to each other, and links to and from overlapping sites. I included these Google “ related” sites in my subsequent testing and analysis.

Testing Methodology

After downloading and installing the most recent versions of site categorization lists for the four blocking programs, I began testing to determine which sites in my site database were blocked by any of the products. I conducted site testing via an automated system that I developed for this purpose, designed to efficiently determine whether particular URLs were blocked by particular blocking programs. I began informal testing in June 2000; the Appendices attached to my testimony reflect the results of testing conducted between February and October 2001. Note that the Appendices report, among other details, the specific dates on which particular sites were blocked by particular blocking programs. Most sites listed on the Appendices were tested and found to be blocked on multiple occasions over a period of several months.

Archiving of Blocked Sites

In addition to verifying that all sites to be submitted into evidence were indeed blocked, counsel for plaintiffs asked me to attempt to preserve each site as it stood on the date of my final testing of each site.

Accordingly, after verifying that each URL was still blocked, I archived the publicly-accessible web content available at each URL. I purchased a commercial web archiving program known as Teleport Exec and made by Tennyson Maxwell Information Systems. I used this program to retrieve and archive the specific web pages that had been blocked by one or more blocking systems. I also retrieved any additional files necessary to display those pages, typically including images, Shockwave and Flash files, and most other forms of web-accessible content. In addition, for a subset of blocked web pages specifically designated by counsel for the plaintiffs, I separately configured Teleport to retrieve all directly linked web pages (and all files used therein) within the same URL hierarchy (i.e. on the same web server, and within the same directory if the original URL referred to a file located in a directory). This subset of pages is available in Appendix A, while other pages are archived in Appendix B.

For most sites, the archival procedure proceeded as anticipated, yielding a set of files that accurately reflect the web pages as they stood on a date of testing.[7] Archiving closely followed the verification that sites were blocked: More than 89% of the subset of blocked sites were archived within 6 hours, 94% within 30 hours, and 100% within 48 hours. More than 77% of the other blocked sites were archived within 6 hours, 95% within 10 hours, 99% within 24 hours, and 100% within 48 hours.

For a variety of technical reasons, some sites on CD-R Appendices A and B were unable to be archived in their entirety. Depending on the type of problem and on the specific browser software used, these problems may cause display of blank pages, of error messages within the web browser, or of popup (dialog box) error messages.[8] For this reason, Appendices A and B also provide links to the current versions of sites as available on the web. I also printed and retained hard copies of the all sites on CD-R Appendix A.

[7] In general, most sites were archived immediately after the mid-September date in which they were verified to be blocked. Some sites were archived immediately after the early-October date in which they were verified to be blocked. Each site was only archived immediately after a test verified that it was blocked by at least one filtering program, configured as previously described.

[8] In some instances in which I noticed these problems and was able to isolate them, I manually corrected the archives; in other cases, I manually archived the affected sites via screen snapshots or Internet Explorer's File-Save As command.

A Guide to Appendices A, B, and C

I have archived each web site blocked by one or more blocking programs on two CD-Rs, Appendices A and B. These CD-Rs are readable in any standard personal computer with an installed web browser. Full viewing of certain pages may require the installation of the Macromedia Flash player.

The printed Appendix C includes additional characteristics about each web site found to be blocked, as detailed below. The CD-Rs also report this information on their respective index pages.

The first row of each site listing gives the page's title, as retrieved from Google. This data element ordinarily reflects the HTML TITLE (of the specific page found to be blocked), as written by the page's creator, when the page was last indexed by Google.[9] The first row also includes a unique numeric identifier for each site.

The second row notes the URL of the specific page found to be blocked.[10] In the CD-R version of the archive, this URL is a clickable link to archives as stored on the CD. In the CD-R version of the archive, the subsequent text "Current Web Version" allows direct access to the current contents of the corresponding page, as retrieved from the web.

[9] For selected sites, when the Google title was blank or uninformative, I manually obtained the proper title of the page by selecting either the page's current HTML TITLE, as retrieved within 72 hours of testing, or (when the HTML TITLE was blank or uninformative) a series of representative words in the page's body text.

[10] The testimony of representatives from SmartFilter, N2H2, and Websense indicates that blocking companies ordinarily block on the level of an entire site, including its root, all subdirectories, and all subsidiary pages; internal documentation from N2H2 concurs. Accordingly, there is a strong likelihood that other pages within the same domain or directory may be blocked also. However, it is possible that only a single page on a site might be blocked. For example, if my tests determined that http://www.aclu.org was blocked by a particular program, it might nonetheless be the case that http://www.aclu.org/projects remained accessible. While complete determinations here were beyond the scope of my testing, I did verify that Websense consistently blocked all plaintiff sites at the level of the entire server, while N2H2 blocked some plaintiff sites at the level of the server and others on a directory- or page-level basis. For additional discussion of this subject, see the section entitled "The Scope of Classifications Across Pages."

The third row lists the programs that blocked the specific page tested as well as the date or dates of testing. This row also includes the category or categories in which each program classified the site.[11]

The fourth row lists Yahoo categories that include this specific URL (if any). When a site is in multiple categories, this section may expand to include multiple lines.

The fifth row lists any Google categories that include this specific URL.

The sixth row gives the site's description, as retrieved from Google when the page was last indexed. This data element ordinarily reflects the page's META DESCRIPTION tag, as written by the page's creator and included in the page's HTML header, but in some instances may be edited by Google's staff.

The CD-R of Appendix A also contains a listing of plaintiff web sites along with information about when each site was blocked, by which blocking programs, and (when available) for membership in which categories of these programs. For these tests, I configured Cyber Patrol, N2H2, SmartFilter, and Websense to block sites in all of their categories.[12]

The CD-R of Appendix B also contains a listing of web sites that are included in the random sample of expert Joe Janes. The population of sites used by expert Joe Janes

[11] Cyber Patrol was configured to block only the Adult/Sexually Explicit category. SmartFilter and Websense report categories of site blocking on their "page denied" error pages. For N2H2, I retested each blocked site with N2H2 configured to block only one category of sites at a time; in this way, I learned which sites were blocked due to classification in which N2H2 categories.

[12] I did not verify that plaintiff site <http://members.aol.com/parker4congress> and <http://www.afraidtoask.com> were blocked until October 2001 testing. However, I have reviewed documentation from Bennett Haselton of Peacefire.org that reports that the former site was blocked by N2H2 as of November 7, 2000 and March 16, 2001, and that the latter was blocked by Cyber Patrol and N2H2 as of March 16, 2001.

represents the combination of all sites listed in Appendices A and B (save for the plaintiff sites, unless otherwise listed in Appendix A or B).[13]

Results

Appendices A, B, and C detail the complete results of my testing, including a full listing of specific sites blocked.

In total, my research yielded 6777 distinct web page URLs that were blocked by at least one of the filtering programs tested, as configured. Of these pages, council for the plaintiffs designated 395 sites for which I was also instructed to retrieve and archive internally linked pages; these sites are listed and archived on the CD-R of Appendix A. The other 6382 sites are in Appendix B. Appendix C lists all 6777 sites in printed format.

These sites fall into all of Yahoo's categories, as detailed in the table below. Note that many sites were classified in multiple Yahoo categories.

Yahoo Category	Sites in Appendix A	Sites in Appendix B	Total
Arts	63	730	784
Business and Economy	86	2766	2853
Computers and Internet	7	173	181
Education	21	34	57
Entertainment	36	1325	1361
Government	20	44	64
Health	27	60	87
News and Media	17	244	261
Recreation	33	741	775
Reference	4	16	20
Regional	84	1394	1478
Science	12	85	97

[13] However, due to errors in data processing, Appendices A and B mistakenly include three sites (<http://www.agi-usa.com>, <http://desires.com/1.6/Sex/Museum/museum1.html>, and <http://desires.com/1.7/Word/Bookrevs/Docs/mv.html>), which were not included in the population of sites used by Janes; Appendix B also omits the site <http://home8.inet.tele.dk/aaaa/index_m.htm> which was included in Janes' population.

Yahoo Category	Sites in Appendix A	Sites in Appendix B	Total
Social Science	8	55	64
Society and Culture	80	1461	1543

These sites fell into all Google's categories except Adult, as detailed in the table below.

Google Category	Sites in Appendix A	Sites in Appendix B	Total
Arts	26	426	452
Business	5	79	84
Computers	15	113	128
Games	1	51	52
Health	8	38	46
Home	3	11	14
Kids and Teens	1	14	15
News	2	12	14
Recreation	10	255	266
Reference	6	3	9
Regional	47	416	463
Science	6	20	26
Shopping	12	241	253
Society	47	356	404
Sports	4	59	64
World	5	89	94

Sites were blocked by each of the blocking programs, as detailed in the table below.

Blocking program	Sites in Appendix A	Sites in Appendix B	Total
N2H2	193	4768	4961
SmartFilter	105	1485	1590
Cyber Patrol	128	1854	1982
Websense	123	2065	2188

Many sites were blocked by a combination of multiple blocking programs, as detailed in the tables below. In the first table, a " yes" in one of the first four columns indicates that the corresponding row counts sites that were blocked by that program (as configured). The table's fifteen rows reflect all combinations of blocking in which at least one of the four blocking programs refused access to the specified site. For example, the second row of the table indicates that 398 sites were blocked by all four of the programs. The second table lists the number of sites blocked by one, two, three, or four of the blocking programs tested.

	N2H2	Cyber Patrol	SmartFilter	Websense	Number of Sites Blocked
Sites blocked by:	Yes	Yes	Yes	Yes	398
	Yes	Yes	Yes	No	144
	Yes	Yes	No	Yes	144
	Yes	Yes	No	No	221
	Yes	No	Yes	Yes	286
	Yes	No	Yes	No	384
	Yes	No	No	Yes	542
	Yes	No	No	No	2842
	No	Yes	Yes	Yes	46
	No	Yes	Yes	No	64
	No	Yes	No	Yes	0
	No	Yes	No	No	1750
	No	No	Yes	Yes	76
	No	No	Yes	No	192
	No	No	No	Yes	606

Number of Blocking Programs Blocking Each Site	Number of Sites Blocked
1	5390
2	1287
3	620
4	398

Sites were blocked due to inclusion in a variety of blocking program categories, as detailed in the tables below. Note that many sites were classified in multiple categories by a single blocking program and that many sites were classified by multiple blocking programs.

N2H2 Category Combinations	Number of Sites Blocked
Adults Only	455
Adults Only, Nudity	81
Adults Only, Nudity, Pornography	21
Adults Only, Nudity, Pornography, Sex	41
Adults Only, Nudity, Sex	7
Adults Only, Pornography	750
Adults Only, Pornography, Sex	12
Adults Only, Sex	115
Nudity	1066
Nudity, Pornography	47
Nudity, Pornography, Sex	1
Nudity, Sex	71
Pornography	1138
Pornography, Sex	13
Sex	920

SmartFilter Category Combinations	Number of Sites Blocked
Extreme	92
Extreme, Mature	4
Mature	411
Nudity	141
Nudity, Extreme	3
Nudity, Mature	98
Sex	601
Sex, Extreme	14
Sex, Mature	10
Sex, Nudity, Extreme, Mature	4
Sex, Nudity, Mature	91

Websense Category	Number of Sites Blocked
Adult Content	924
Nudity	191
Sex	799

My testing began with only a small sample of sites on the Web. Thus, it is virtually certain that my research details only a small portion of sites that are wrongly blocked by the programs I have tested.

IV. The Design and Operation of Internet Blocking Software

By design, Internet blocking software seeks to limit access to specified portions of the Internet. Most blocking software emphasizes restrictions on access to the Internet's World Wide Web, and discussion throughout this document is generally limited to blocking of the web. See also infra Section V, Other Flaws.

Installation Method and Location

In general, blocking software prevents web page access by monitoring user requests and by interceding between user and connection to the Internet. Blocking programs designed for use on a single computer are able to monitor and interrupt page requests because they reconfigure a computer's means of Internet access to send all such requests through the installed blocking program. Programs designed for use on an entire network anticipate installation on a centralized network device (such as a " proxy server") or in some other context between a network and its single connection to the Internet.[14] In this special position, Internet blocking programs may also provide a number of additional features beyond refusal of access to particular sites. For example, such programs may log

[14] Network-based blocking programs are ordinarily considered preferable when computer terminals share a single high-speed connection to the Internet, while PC-based filtering programs are ordinarily considered preferable when each computer has its own modem connection to the Internet. Furthermore, network-based filtering programs ordinarily bring about efficiencies from centralized management, configuration, and updates, while PC-based systems are easier to bypass using instructions and programs available on the web. For these reasons, most blocking program customers that provide Internet access on multiple terminals to multiple users are likely to prefer network-based filtering programs.

web page accesses and may notify administrators or supervisors when particular web pages or types of web pages are accessed.

Category Lists and Criteria for Blocking

All blocking programs classify web sites into a variety of categories created and defined by their producers. For example, Websense classifies web sites into some 75 categories including “Abortion Advocacy — Pro-Life,” “Abortion Advocacy — Pro-Choice,” and “Advocacy” through “Vehicles,” “Violence,” and “Weapons.”[15] Different blocking programs use different categorization schemes, and producers often change these schemes over time. Customers ordinarily configure blocking programs to prevent access to sites classified to be within specified categories. For examples, a customer can configure Websense to block access only to sites categorized as “Abortion Advocacy — Pro-Choice,” while in principle allowing access to the entirety of the rest of the Internet.

While each producer determines its own category lists, in general these category lists include one or more categories related to nudity and sexually explicit or “adult” content. For example, among the 30 categories offered by SmartFilter are Extreme/Obscene/Violence, Mature, Nudity, and Sex.

However, none of the four programs I tested has categories that map specifically to CIPA’s definitions of material that is obscene, is harmful to minors, or depicts child pornography, and I am unaware of any other blocking program that uses CIPA’s specific categories and definitions. In depositions, staff from N2H2, SmartFilter, and Surfcontrol explicitly agreed that their companies cannot speak to the compliance of their systems

[15] “About Websense Enterprise: Master Database."
<http://www.websense.com/products/about/database/version4.cfm>

with CIPA's requirements and that they cannot guarantee that their products block only images targeted by CIPA.[16]

Of the blocking programs I reviewed, all employ category definitions that apply both to text and images. None of them categorizes and blocks only images.[17]

Overview of Web Site Classifications

The practical effect of a blocking company's decision to place a web site into a particular category is to block access to that site by all users who access the Internet through a system configured to prevent access to sites in that category. Because blocking programs depend heavily on the accuracy of their categorizations of web sites, it is important to understand the procedure for the creation of these lists.

Based on my review of publicly-available documentation, of confidential documents produced by blocking companies, and of deposition transcripts of blocking company employees, it is my understanding that site categorization lists are formed by blocking program companies in roughly three steps as discussed below: 1) developing a list of web sites for possible categorization; 2) using automated systems to examine each page or site and to recommend possible inclusion in one or more blocking categories; and 3) in many (but not all) instances, using human reviewers to make the ultimate decision about whether and how to categorize each page or site.

The review and classification of web sites is difficult because the Internet is both large and changing, For example, leading search engine Google reports that it has

[16] Dussome Dep. (N2H2), at 12-3, 119 (Confidential). Gallagher Dep. (SmartFilter), at 8-9, 69 (Confidential). Blakeman Dep. (Cyber Patrol), at 8.

[17] Dussome Dep. (N2H2), at 118-9 (Confidential). Blakeman Dep. (Cyber Patrol), at 73-4. Gallagher Dep. (SmartFilter), at 8-9, 50-1 (Confidential).

indexed 1.6 billion web pages,[18] while studies indicate that even the best search engines reflect only a fraction of content available on the Internet.[19] Furthermore, many of these billions of web pages change frequently.[20] These two factors — the size of the Internet and its rate of change — are problematic for providers of blocking software which seek to efficiently and cost-effectively classify a large number of diverse sites into a finite and fixed set of categories.

The Use of Automated Systems for Site Classifications

All producers of blocking software use automated site review and classification systems to produce the list of sites for possible categorization as well as to focus that list. These automated classification systems are necessary because the list of sites for possible categorization is ordinarily quite large, making human review of all such sites impractical.[21]

Blocking companies report that they use a variety of automated methods to create site lists for possible classification. These include reviewing lists of newly-registered domain names, following links from a variety of online directories (e.g., web sites that provide links to self-labeled adult content), requesting pages related to certain keywords from ordinary search engines, buying or licensing lists from third parties, and reviewing log files and other submissions from customer installations of their blocking software.[22] In principle these methods operate on a continuous basis, though representatives of

[18] <http://www.google.com>, September 30, 2001.
[19] " Accessibility and Distribution of Information on the Web." <http://wwwmetrics.com/>
[20] " Rate of Change and other Metrics: A Live Study of the World Wide Web." <http://www.usenix.org/publications/library/proceedings/usits97/douglis_rate.html>
[21] Dussome Dep. (N2H2), at 48-57 (Confidential). Gallagher Dep. (SmartFilter), at 22-25 (Confidential). Blakeman (Cyber Patrol), at 19-23.
[22] Dussome Dep. (N2H2), at 26-8 (Confidential). " Surfcontrol' s Accurate and Relevant Filtering," Appendix B. Surfcontrol [re search engine keywords]. Gallagher Dep. (SmartFilter), at 22-3 (Confidential).

blocking programs report that in some instances certain methods are only used intermittently.[23]

Leading blocking programs use keyword analysis to identify web sites for possible classification. Keyword analysis may be direct — the designer of the blocking program searches for web sites that contain specific words or phrases. Alternatively, keyword analysis may be indirect via a so-called “artificial intelligence” engine — which produces and runs such rules on its own after a “training” period.[24]

The documentation I have reviewed reflects that automated systems used by leading blocking programs to make classification recommendations do not include image recognition technology.[25] Indeed, image recognition technology is thought by many to be unreliable.[26] Rather, the automated systems rely exclusively on keyword analysis.

The Use and Omission of Human Review

Most blocking companies advertise that a staff person reviews every site before its addition to a categorization list. For this purpose, producers of blocking programs employ staff who compare web content with category definitions; Surfcontrol reports that it has about 40 such staff (some of whom work only part time), while N2H2 has twelve and SmartFilter has eight.[27] In addition, the repetitive nature of site review work prevents these staff from working entirely on this task; instead, they also fulfill a number of other

[23] Gallagher Dep. (SmartFilter), at 25 (Confidential).

[24] During training, an artificial intelligence engine would review numerous web pages that human review had already found to meet a certain category definition. A properly-designed artificial intelligence engine analyzes patterns in these documents and subsequently classifies new documents on the basis of their similarity to these patterns. “Surfcontrol’s Accurate and Relevant Filtering,” Appendix B. Blakeman Dep. (Cyber Patrol), at 26.

[25] Dussome Dep. (N2H2), at 118 (Confidential). Blakeman Dep. (Cyber Patrol), at 39-40.

[26] Blakeman Dep. (Cyber Patrol), at 149.

[27] Dussome Dep. (N2H2), at 68-9 (Confidential). Gallagher Dep. (SmartFilter), at 45 (Confidential). Blakeman Dep. (Cyber Patrol), at 56. “Surfcontrol’s Accurate and Relevant Filtering.” Dussome Dep. (N2H2), at 68 (Confidential).

customer support and administrative tasks, further reducing human review capacity.[28] A N2H2 representative says that their human review staff is expected to review about 50 sites per hour.[29]

In at least some cases, at least some blocking companies admit that sites are categorized without human review. In particular, N2H2 apparently adds certain sites to its Pornography category, perhaps among others, on the basis solely of recommendation by an automated system, without subsequent human review.[30] A SmartFilter representative reports that SmartFilter is considering a similar approach.[31]

The Scope of Classifications

Representatives of N2H2, SmartFilter, and Surfcontrol testified in their depositions that in the majority of instances, human reviewers classify all web content on a site based on their review of only the front page of that site and, in some instances, a sampling of other pages. They subsequently place the entire web site within a particular category even without having reviewed much of the content on that site.[32] In this context, " web site" can refer to either an entire server or, in the case of servers known to host content from many users (such as http://www.geocities.com), a single directory on that server (such as http://www.geocities.com/bedelman).

Theoretically, blocking lists could categorize (and thus allow blocking of) only a particular page or pages on a web site, a particular directory, a particular server, or a range of servers (with one or several IP addresses, or with a common domain name).

[28] Blakeman Dep. (Cyber Patrol), at 57.
[29] Dussome Dep. (N2H2), at 71 (Confidential).
[30] Dussome Dep. (N2H2), at 51 (Confidential). SmartFilter email (August 23, Phyllis Houseman) (Confidential).
[31] Gallagher Dep. (SmartFilter), at 158-9 (Confidential).

However, the more specific and limited approaches of file or directory-level blocking are used infrequently due to the size of the web, as discussed below, as well as due to the number of staff assigned human review and their time limitations.[33]

When sites are categorized via an automated system, categorization scope is even less specific and may be determined for a batch of sites simultaneously.

The Unavailability of Site Categorization Lists for Review by Customers

Blocking companies periodically make available updated versions of their site categorization lists — sometimes as often as once per day — for automated download by the blocking programs installed at customer locations. Such downloads take place in an encrypted format precisely intended to prevent human review of the list. In particular, blocking list formats are designed to be impossible to read either with generally-available tools or with custom tools created for this purpose. Having invested significant resources in the creation of these lists, blocking companies seek to avoid public dissemination of the contents of the list. In recent years, blocking companies have been successful in keeping the contents of their lists secret.

While case-by-case testing is possible, this approach is only partially informative. This method ordinarily only determines whether a page is blocked, but fails to report which of several criteria caused the block, fails to include the status of this site at various points in the past, and fails to describe the scope of the block across pages. It is also time-consuming, unreliable, and difficult to automate.

[32] Dussome Dep. (N2H2), at 62-4 (Confidential). Gallagher Dep. (SmartFilter), at 50 (Confidential). Blakeman Dep. (Cyber Patrol) at 61.

[33] Dussome Dep. (N2H2), at 61-2 (Confidential). Gallagher Dep. (SmartFilter), at 49 (Confidential). Blakeman Dep. (Cyber Patrol), at 62-3.

Failure to Review Sites Again Once They Are Classified

Once a site is added to a blocking company's classification of web sites, it is generally not reviewed again. Blocking company N2H2 reports that it ordinarily reviews already-categorized sites only in response to customer feedback.[34] SmartFilter notes that in one instance they re-evaluated sites in selected categories, but this was seemingly an ad hoc procedure that does not ordinarily reoccur on a scheduled basis[35]; a different SmartFilter representative reports that sites are reevaluated on an occasional basis, not on a regular schedule, and that some sites may never be reevaluated.[36] Surfcontrol reports that site reevaluation is conducted only as staff time allows and that other staff obligations likely make this a low priority.[37]

Furthermore, at least some blocking companies do not ordinarily conduct quality control studies to verify the accuracy of decisions made by staff. A representative of Surfcontrol reported that such evaluations are performed only for new employees but not as a matter of course for existing employees.[38] An N2H2 representative said that N2H2 had only recently begun to conduct random audits of site classifications.[39]

Customizing Site Categorization Lists

A blocking program customer generally has two options when she wishes to unblock a site classified by the program (or to block a site that is not classified by the program): She can create a customized list, or she can send a request to blocking program staff and ask them to review the site's classification. In practice the first option proves difficult due to the nature of the task as well as the design of popular blocking

[34] Dussome Dep. (N2H2), at 74 (Confidential).
[35] SmartFilter email (August 23, Phyllis Houseman) (Confidential)
[36] Gallagher Dep. (SmartFilter), at 33 (Confidential).
[37] Blakeman Dep. (Cyber Patrol), at 79.
[38] Blakeman Dep. (Cyber Patrol), at 67.

programs. For one, it is nontrivial to create an internal system for testing, review, and classification of sites, as well as to deploy customized site lists to blocking servers. Furthermore, it is difficult for customers to know what sites are currently being blocked (and thus what customizations are appropriate) because, as already discussed, the underlying blocked site list is not available for review. The difficulty, cost, and complexity of these tasks is likely to reduce the interest of most companies or institutions in doing so.

In addition, while blocking programs may in principle permit the import and export of customized blocking lists, they do not explicitly encourage customers to share lists. For example, there is no standard means of transferring customized lists between programs, nor of annotating each customization with an explanation for the rationale for the change. Thus, there is little economy of scale in performing customizations.

Instead, via a variety of customer feedback mechanisms, blocking companies encourage administrators to submit changes for approval by blocking companies' own staff, thereby discouraging the development of extensive end-user customizations of blocking lists. This decision may be optimal for strategic reasons (as it increases the relative value-add of the blocking company as against a world in which individual administrators extensively customize their blocking software installations) or to reduce customer support costs.[40]

[39] Dussome Dep. (N2H2), at 93-94 (Confidential).
[40] Extensive administrator site customizations add additional complexity and unpredictability to the end-user support responsibility of providers of blocking programs. N2H2 internal documentation.

V. Problems with Blocking Software

Overblocking

Blocking programs prevent access to a substantial number of sites that do not contain content that fits within the blocking company's stated category definitions. Such overblocking may result from a number of factors discussed below, among them various types of error by blocking company employees as well as shortcomings in the design of blocking programs. Representatives of blocking companies agree that overblocking affects their products.[41]

Human Error and Judgment Calls

Overblocking may occur due to omission of human review or due to errors in the human review process. As previously discussed, blocking companies admit that in some circumstances their programs classify sites without individual human review of each such site. While N2H2 and SmartFilter representatives assert that these circumstances are limited, my testing suggests that human review may be omitted in a number of instances beyond the contexts that company representatives admit. For example, no human reviewer is likely to find "Red Hot Mama Event Productions" (a web site offering the services of a California event planner) to meet definitions of Sex (SmartFilter) or Adult/Sexually Explicit (Cyber Patrol). Nonetheless, <http://www.redhotmama.com> was classified in these categories by the respective programs on multiple occasions, as documented in the Appendices of my testimony. Similarly, the site at <http://www.the-strippers.com> is in fact a wooden furniture varnish removal service, yet it is classified as Adult/Sexually Explicit by Cyber Patrol and as Sex by Websense. My research has produced numerous other examples of this sort, and others on the web discuss similar problems at length.[42] In each instance, page titles and domain names include terms that likely satisfy certain rules of keyword-based searching systems — suggesting the use of such systems, since no human is likely to misclassify these small, straightforward, and unambiguous sites.

Even when humans do review a site before classifying it, overblocking may occur because, for large sites in particular, reviewing each page of a site is time-consuming and impractical. As a result, web page reviewers are ordinarily instructed to start their review with the front page of a web site but are told that they need not review every page on a site in order to classify it.[43] However, the most obvious pages on a web site do not always properly indicate the contents of all other pages on the site, and some pages on a site may differ substantially from the majority of others. In particular, even sites widely known to contain some sexually-explicit pictures may also contain extensive non-explicit text content, as in a news section.

Overblocking may also occur because blocking company employees misclassify a site or portion of a site due to ambiguity in classification rules or error in rule application to particular pages. Many sites fail to fit clearly into the category definitions established by blocking companies; while company staff may write additional documentation explaining the intent of their classification systems, categorization accuracy remains uncertain, especially as rules increase in number, length, and complexity. Furthermore,

[41] Dussome Dep. (N2H2), at 163-4 (Confidential). N2H2 Seven Week Sample of Review Mailbox Requests — Action Taken. (Confidential). Gallagher Dep. (SmartFilter) at 10. Blakeman Dep. (Cyber Patrol) at 100.
[42] For example, " Blocked Site of the Day." <http://www.peacefire.org/BSOTD>
[43] Dussome Dep. (N2H2), at 62-4 (Confidential). Gallagher Dep. (SmartFilter), at 50 (Confidential).

many sites contain content that might be found to fit a number of categories; in such instances it can be especially difficult to properly categorize a site.

Finally, blocking company staff may cause unintended web page blocks as a result of design flaws or subtleties in the blocking database. In particular, certain kinds of site categorizations and exemptions may interact unpredictably with the unusual configurations of certain web servers (especially web servers hosting public content for multiple users, as at an ISP or a free hosting provider such as geocities.com). In such circumstances, the effective scope of categorization may differ from the staff person's intent, preventing access to more content than the staff person found to meet a program's category definitions.[44]

Technical Error

Even when blocking companies know for certain which content they intend to classify and into which categories, certain technical challenges may make it difficult for them to successfully do so. For example, a single web server may host hundreds of domain names using only a single IP address. While the blocking company's efficiency and database size benefit from blocking that single IP address rather than the many associated domain names, doing so risks unintentionally blocking other content that does not fit category definitions.

In addition, blocking servers do not always work precisely as intended; like all software, they may have a variety of " bugs." In some instances, flaws in blocking programs result in user inability to access sites for which the blocking company

[44] SmartFilter Control List Training Guide, " Example Two." [redacted]-

specifically seeks to permit access.[45] Other flaws may mistakenly prevent users from accessing certain web-based services such as email; in particular, users with certain account names cannot access web-based email accounts via the N2H2 blocking system due to a flaw in that program's URL parameter blocking methods.[46] Errors of this sort may block large amounts of content as well as web-based email and possibly other web-based applications.

In addition, a variety of services on the Internet provide proxy servers, translation servers, and other methods by which a user might retrieve Internet content via a third party rather than directly from the content provider. The use of such devices may stem from an interest in privacy, since proxy servers can prevent web server operators from gathering a variety of facts about a web user. Proxy servers may also provide other helpful services, such as translation of web content into other languages, addition of links to sources of related content elsewhere on the web, removal of unwanted or potentially-hazardous software code otherwise present in some web pages, or removal of advertisements. However, such servers also provide a possible means of circumventing the restrictions of popular blocking programs. Thus, it has been documented that blocking programs seek to prevent access to these proxy servers even when such blocking is not requested by the administrator of a blocking program and even when such sites are not within the specific descriptions of categories requested for blocking.[47] My testing found multiple examples of blocking of these sites, including translation service tranexp.com and privacy service idzap.com.

[45] [redacted]

[46] [redacted]

Underblocking

Blocking software also fails to block all content meeting specified category definitions. To some extent, this shortcoming reflects the difficulty of reviewing and categorizing the entirety of the Internet; even search engines, which ordinarily have vastly greater information-processing resources and which need not attempt human review of each site, typically fail to review most of the Internet.[48] Automated systems address a portion of the challenge posed by the Internet's size and rate of change, but they do not solve the problem completely because of performance concerns. When specifically asked why Surfcontrol's automated systems do not review more of the web, a Surfcontrol representative explained that even automated systems offer limited performance due to delays in page retrieval as well as delays in application of classification rules.[49] [redacted] Thus, the size of the Internet prevents even automated site classification systems from reviewing all pages. Of course, even if automated systems were faster, their accuracy would remain problematic, as previously discussed.

A portion of underblocking may result from the failure of blocking companies to accurately categorize non-English sites. Some blocking companies hire site reviewers able to read multiple languages, but their work is ordinarily limited to sites in specific target markets,[50] and other blocking companies do not include this requirement in their job descriptions.[51] Furthermore, staff of N2H2 and Surfcontrol were unaware of the

[47] " BESS's Secret LOOPHOLE." <http://sethf.com/anticensorware/bess/loophole.php>

[48] Generally known. See, for example, <http://www.siliconvalley.com/docs/news/svfront/030266.htm> .

[49] Blakeman Dep. (Cyber Patrol), at 114 (Confidential).

[50] [redacted]

[51] Dussome Dep. (N2H2), at 68-69 (Confidential).

extent of their respective companies' multilingual automated categorization efforts, while a SmartFilter representative said that focus to date had been on English URLs.[52]

Underblocking reflects a fundamental tradeoff faced by blocking companies as they seek to balance overblocking against underblocking. In particular, when judgment calls determine the classification of a site, or when technical or practical constraints prevent perfect classifications, any action is likely to contribute to either overblocking or underblocking.

Representatives of blocking companies consistently agree that underblocking affects their products' accuracy.[53]

Inability to Block Images Only

I understand the text of CIPA to require only the prevention of access to certain images defined by the law; it does not require the prevention of access to text which meets the same definition. In particular, even if a site contains one or more images that must be blocked because they meet the criteria of CIPA, users may still read any text on the same page as well as view any images that do not meet CIPA's criteria.

The blocking programs I have reviewed do not do this. Rather, when a blocking program receives a request for a page in a category to which it has been configured to deny access, the program prevents access to the entirety of the page. In particular, I know of no program designed to block solely the images which fit certain criteria on a site while leaving the rest of a page intact.

[52] Dussome Dep. (N2H2), at 106-107 (Confidential). Gallagher Dep. (SmartFilter), at 115-6 (Confidential). Blakeman Dep. (Cyber Patrol), at 51-3.

[53] Dussome Dep. (N2H2), at 13, 164 (Confidential). Gallagher Dep. (SmartFilter), at 93 (Confidential). Blakeman Dep. (Cyber Patrol), at 14.

Staff of multiple blocking companies agree that their products do not specifically include this feature.[54] To the best of my knowledge, no generally-used blocking program includes this feature. In addition, as discussed previously, none of the leading blocking programs provide blocking categories that classify images only, and their automated systems are incapable of making recommendations for classification based on image analysis.

While it may be possible to configure some proxy servers or blocking programs to refuse access to all graphic files (not just those that meet certain category definitions), this configuration would cause many web sites to become unusable due to their dependence on graphics for navigation, spacing, and other purposes.

I know of no blocking program that can differentially allow access to graphic files on the basis of a site's classification (of only graphic files) in that blocking program's site database.

Other Flaws

Additional overblocking and underblocking may both result when blocking companies fail to classify web content behind an access control system and when blocking companies mistakenly guess the possible classification of such content without reviewing the specific contents of that site. Blocking company representatives report that their staff members do not complete registration forms or pay registration fees to web sites that require these procedures for access to full site contents.[55] While in some cases the contents of a site may be clear from its registration page, this need not be the case in

[54] Dussome Dep. (N2H2), at 65, 119, 165-8 (Confidential). Gallagher Dep. (SmartFilter), at 69, 144 (Confidential). Blakeman Dep. (Cyber Patrol), at 81-2.

[55] Gallagher Dep. (SmartFilter), at 42, 63, 76 (Confidential). Blakeman Dep. (Cyber Patrol), at 117-8 (Confidential).

general. Indeed, an access control system might restrict access to diverse content; in such cases, having decided not to complete site registration procedures, blocking companies must either speculate as to site contents or fail to categorize a site altogether, in either case possibly causing overblocking or underblocking, respectively.

Blocking software also suffers from underblocking in its inability to block Internet content that is delivered by means other than via the World Wide Web but that nonetheless meets CIPA's definitions.

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. When email is retrieved using a web-based email program (such as Hotmail), blocking programs have no way to differentiate between content that meets their criteria versus content that does not.

Like email, streaming video playback relies on programs other than a web browser, thereby circumventing blocking software that solely restricts web page access. In some instances, videos may be retrieved via the HTTP protocol, giving blocking programs an opportunity to differentially allow access depending on the specific content request and depending on the design of the blocking program. However, so-called "streaming video" is delivered via alternative protocols such as RTSP, PNM, and MMS in preparation for playback in a specialized application (such as the RealNetworks RealPlayer or Microsoft Windows Media Player) outside the web browser. While some of the programs I have reviewed can block all such streaming video access, none can block only access to certain sources of streaming video content[56]. Furthermore, while blocking companies may seek to limit access to streaming video by limiting access to web pages that link to streaming video, users may nonetheless directly access the desired content by manually entering specific streaming video locations into a player's File-Open Location (or similar) dialog box, thereby circumventing blocking at the level of web pages. Thus, available blocking software is unable to prevent access to certain streaming content while allowing access to all other streaming material.

Certain forms of interactive content may yield pictures that fit CIPA's definitions. For example, a real-time chat room may provide a means to transfer files, including graphics that meet CIPA's definitions. However, while blocking programs may be able to prevent all access to certain web-based chat sites, they are ordinarily unable to block access to only certain rooms within such sites and they are unable to block access to only certain messages; they are almost certainly unable to block image transmissions.[57] Furthermore, blocking programs are especially ill-equipped to block chat rooms located other than on the web; I know of no blocking program that classifies certain types of IRC or instant-messaging communications. Depending on the configuration of a non-real-time discussion boards, these areas may be similarly difficult for blocking programs to classify; here too, it is impossible for blocking programs to prevent access only to those discussions with images meeting their definitions while allowing access to all other content.

[56] Gallagher Dep. (SmartFilter), at 65 (Confidential).

[57] Dussome Dep. (N2H2), at 39-40 (Confidential). Gallagher Dep. (SmartFilter), at 63-5 (Confidential). Blakeman Dep. (Cyber Patrol), at 40-1, 88-9.

VI. Compliance of Blocking Software with the Special Provisions of CIPA

Creating Exceptions for Research Purposes

It is my understanding that CIPA provides that administrators may grant access to a blocked site for a limited period of time by a single user for "bona fide research purposes." While blocking programs purport to have capabilities that would facilitate this process, numerous practical problems make it unlikely that, in practice, such a system could be successfully implemented using existing network-based blocking software.

One set of implementation difficulties results from the skills of librarians and the likely need for extensive special training to administer blocking software. While blocking servers ordinarily offer a number of configuration options, their flexibility comes at the cost of complexity. Indeed, the manual for the SurfControl Cyber Patrol program includes some 35 pages of documentation of rule creation and management systems,[58] and this complexity is representative of the blocking programs I reviewed. In particular, the administration systems of leading blocking programs offer customization along multiple axes in a variety of combinations; while these many options may allow flexibility in system configuration, they also make simple changes less intuitive to novice administrators. In Cyber Patrol, for example, granting a user the ability to access a single otherwise-blocked site would require creating a "URL Exception Object" and associating that object with a particular user or workstation — a procedure involving multiple dialog boxes and multiple opportunities for error.

The lack of granularity in blocking server administration permissions adds additional risk to the granting of research exceptions. In particular, in order to let specific librarians add temporary exceptions to blocking restrictions, a systems administrator must grant each one the ability to administer the blocking server. Since blocking servers ordinarily offer only a single level of administrative access — complete control over the entire blocking system — this delegation could have a variety of serious side effects. For example, an erroneous URL Exception Object could mistakenly allow all access to all pages, rather than to only the single site intended, or it could just as easily deny all accesses. Application of a rule to the wrong users or systems, potentially even to all users or systems, could similarly have dramatic unanticipated consequences that would cripple library Internet access or prevent CIPA compliance. It is no doubt for these reasons that the documentation of Cyber Patrol specifically advises against creating blocking rules on an ad hoc basis.[59]

Concerns about granularity of exceptions may also pose practical difficulties in allowing access for bona fide research. For example, when the blocked page is located in a directory of a web server, librarians must decide whether to allow access to pages only in that directory or to the entirety of the server. The former may prove inadequate for the research at issue, while the latter may allow access to content that must be blocked for CIPA compliance. The decision is made even more complex because many content providers distribute content from multiple servers (for a variety of reasons including reliability, ease of administration, security, and performance); in these cases, exceptions may have to allow access not just to multiple directories but to multiple servers. The ease of linking and embedding on the web makes this factor especially serious; a page on one server might link to other pages on dozens of different servers, or for that matter draw

[58] SurfControl Cyber Patrol Manual, pages 30-64.
[59] SurfControl Cyber Patrol Manual, page 56.

images or other supporting content from dozens of different servers. In these cases, it will surely be especially tempting for a librarian to temporarily remove all blocking, granting the researcher unfettered access to the entire Internet. However, this may not comply with the requirements of CIPA. Furthermore, specific decisions in such circumstances will require additional training as librarians are forced to learn about the necessary scope of exceptions on a variety of types of web pages as well as about techniques for diagnosing the specific cause of blocking of particular web pages.

The need under CIPA to limit, document, and terminate temporary research exemptions further complicates the use of blocking programs. In particular, standard blocking programs provide no way to cause a blocking rule to expire automatically when a user leaves an Internet terminal or after a specified period of time; thus, librarians would be required to periodically remove all temporary research exceptions, a time-consuming and potentially error-prone task. Furthermore, standard blocking programs provide no method for logging librarians' changes to blocking configurations, nor for recording and preserving the librarians' specific rationale for such exceptions; thus, compliance with CIPA might require maintenance of separate records of exceptions granted and why they were granted.

In addition, software licensing and design problems may hinder the deployment of blocking administration systems for distributed granting of research exemptions. Widespread administration of blocking servers would require the installation of specialized administration software on every computer to be used for adjusting system configurations by a person authorized to grant exceptions, at sizable expense both due to the cost of technical staff (whose assistance might be required to properly install the software) and due to possible additional licensing fees payable to blocking companies for installation of multiple copies of the administration software. The need to make configuration changes via specialized administration software also prevents administrators from granting exceptions from the computer terminal of an affected patron; instead, an administrator would have to make such changes from administration computers for use by staff, slowing the process of granting an exception. Finally, in the context of centralized blocking servers serving large libraries or multiple branch libraries, there might be conflicts among multiple administration tools used simultaneously. Technical documentation from Cyber Patrol speaks specifically to the problem of simultaneous remote administration by multiple distinct administrators, noting that this can cause corruption in the system's database from which recovery may be difficult or impossible.[60] In depositions, staff of blocking companies also agree that their designs did not contemplate multiple simultaneous administration operations, [61] and that such use may cause unexpected behavior.

Many of the concerns described above apply also to PC-based blocking systems, while certain of these problems would be solved by the use of PC-based systems. However, PC-based systems entail other serious shortcomings that make them unsuitable in the context of large libraries and libraries with high-speed Internet connections.

VII. The Flaws of Blocking Programs are Fundamental

After reviewing numerous blocking programs as well as depositions of their representatives, it is my conclusion that blocking programs are fundamentally unable to

[60] "Readme," Cyber Patrol v6.0.1 for Microsoft Proxy Server.

block all Internet content that meets specific category definitions while simultaneously allowing access to all other content.

Overblocking and underblocking of web content are both inevitable as a result of the size of the web and the rapid rate of modifications of web pages.[62] Automated systems remain too crude to properly classify many web pages, but human review is slow and costly, preventing timely and continuous review and re-review of all web pages. In the case of sites with extensive or frequently-updated content, including news sites and sites that host submissions from the general public, it is effectively impossible for blocking programs to compare each individual page with category definitions. But even certain sites that change less frequently are likely to remain effectively impossible to properly categorize; for example, blocking companies' site classification methods are ineffective in finding and classifying small or obscure sites that are neither listed in search engines nor linked from other web pages.

These web content blocking flaws are fundamental and cannot be addressed in the short run. The size of the web is unlikely to decrease, and neither will its rate of change slow in the foreseeable future; thus, it will remain impossible to conduct human review of every site to be blocked. Furthermore, automated systems for content classification remain immature, especially when attempting to classify solely images but not surrounding text. The dual requirements of reducing underblocking and reducing overblocking remain fundamentally in conflict, preventing any increase in efficiency by sacrificing one to improve the other.

[61] Gallagher Dep. (SmartFilter), at 84-5 (Confidential).

[62] " Rate of Change and other Metrics: A Live Study of the World Wide Web." <http://www.usenix.org/publications/library/proceedings/usits97/douglis_rate.html>

In addition, blocking programs systematically fail to differentially allow access to certain Internet other than that distributed over the Web. For example, blocking programs are unable to differentially restrict access to images that meet their category definitions but are transmitted via email, streaming media, or a variety of interactive systems. This problem is fundamental in that there is no way for blocking programs to, for example, prevent access to content (that fits category definitions) that is received via web-based email without also blocking access to all other web-based email (that does not fit category definitions). This problem is also fundamental in that there can be no guarantee that blocking programs will be able to identify and (if desired) block content distributed via methods yet to be developed or widely deployed.

In short, then, currently available blocking programs are fundamentally unable to satisfy the requirements of their own specifications and category definitions. They cannot prevent access to all Internet content meeting certain category definitions while simultaneously allowing access to all other content, and they will remain unable to do so for the foreseeable future.