Chapter 3

SELECTING A FILTER

Chapter 2 provides important background into filtering, specifically the filter marketplace and the important content categories that provide the primary mechanism for control with most commercial filters. Once you understand the basic workings of a commercial filter with their multiple categories of content, then you can make a decision about using one or not.

If you decide to use one, you need another level of understanding to ensure you select the best filter. Although filters have many similarities, their differences ultimately determine which filter will work best for your library and which ones will not work at all.

The filter selection committee

Once the library has decided to use a filter, it must define the right group of people to select the filter. Installing a filter affects not only the patrons using the Internet computers, but the floor staff who assist patrons, the technology staff who maintain the public access and staff computers, and the technology staff who will administer the filter.

Not only should representatives from each major department be involved in the filter selection process but so should the library director, technology manager, and a board and community representative.

Before selecting an Internet filter, clarify what is needed and what is possible. The filter selection committee must have some kind of filtering goal in mind. For example, the library's strategy might be to comply with CIPA by minimally filtering all computers one way and filtering the children's computers using a different filter entirely.

To realistically formulate a filtering strategy, the members of the filter selection committee should educate themselves about filtering, including how filters work and don't work, and what features they can expect to find in available filter products.

Only a diverse and well-informed committee can define the library's filtering goals and make decisions that will work for the entire library and the community.

Choosing where to install the filter

PC: personal computer

One of the most practical issues about choosing a filter is the design, size, and layout of the library's network. Client-based filters that must be installed and configured on each PC won't make sense for a large library or multibranch system because of the burden to configure and maintain each PC. The more computers to filter there are, the more important ease of management becomes and the more important the impact on the network could be.

Most filters suitable for libraries have been developed for school or business use and are server-based or network appliances. These network-level products have

the breadth of features necessary for managing the filters in a library setting and have more appropriate content categories. For this reason, this report focuses primarily on this class of filter, not client-based products. Situations exist, however, where installing a children's filter on a specific machine or in the children's department may make sense.

For example, the children's librarian would have complete control over how filtering is implemented if the one or two children's computers each had a client-based product designed specifically for younger children.

Although this installation would require more work on the librarian's part, it might be a better solution than relying on a system-wide filter that is completely controlled by people at the main library or the county or the consortium's headquarters.

Although choosing a single filter for the library is most cost-effective, some situations may exist where you can use alternative technology protection measures to supplement the main filter.

Remember, CIPA does not mandate that the library filter every computer with the same filter or to block content for all patrons in the same way. So, even if a minimal approach to blocking is in place for adults (which allows for CIPA compliance), the library might still choose to block content beyond the CIPA mandate to satisfy community demand. Such supplemental blocking can be done with a different filter from the one used to minimally block all PCs.

Using one filter for multiple locations

The ideal library filter for multiple locations is centrally installed and configured to enforce the library's Internet use policy (IUP) while providing for adequate administrative control throughout the library. This setup saves the library staff time as well as money for hardware and software.

The higher upstream (or near the point in the network where the Internet connection comes into the network) the filter is installed, the more cost-effective it becomes. But oftentimes the efficiency of installing the filter upstream comes at the cost of control for those downstream. To ensure that filters match the community's needs, local control cannot be sacrificed to the efficiencies of a filter completely controlled by network-level technology staff.

Selection tip
Install the filter as high upstream on the network as possible without sacrificing local control.

Sharing a filter with several branches or individual libraries can greatly reduce the time and cost associated with filtering. Under such a scenario, only one filter administrator would be required to handle the top-level management tasks such as generating reports needed by each library and keeping the filter functioning, patched, and updated.

A shared filter solution also saves in software and hardware costs because the appliance or server costs are shared. A filter serving multiple locations is generally installed upstream near the router and firewall.

Installing a single filtering "black box" on the library system's firewall or connecting to the router can represent a significant cost savings. But this solution only works if the filter can account for all the different filtering environments that exist behind that firewall or router.

The larger the system being filtered and the more individual buildings there are, the more important is selecting a product that allows the primary network administrator to delegate some filter administration functions.

Designated staff at each library must be able to define filter profiles that work in their environment without having to adjust their IUP to fit into another organization's policy, especially when that other organization is not a library.

Sharing a filter between organizations or agencies can save money, but it also requires coordination and trust that:

- Each entity's unique needs will be accommodated during setup.
- Requests for changes will be administered promptly.
- Each local entity has control over unblocking pages and turning off filters on individual PCs.

Selection tip
Using a shared filter for several locations or a filter installed outside the library requires coordination among the different organizations being filtered so the configuration of the filter accommodates everyone's needs.

Some libraries are connected through their county, city, or elementary school system. The filtering needs of a school system are different from a library's needs, so account for these differences when sharing a filter. The library also must be assured that the person and agency administering the filter will act on library requests promptly and will accommodate the unique requirements of a filtered library environment.

Specifically, staff at each library should have the ability to turn off filters or override blocked pages without help from a central filter administrator who is probably not on-site. Ideally, local staff should have a way to manage patron requests from the patron desktop or a Web-based interface.

Few products allow for this type of local control so select carefully. At least two products offer this feature: N2H2's filters include a delegated administration feature that allows the administrator to assign some degree of control to local administrators as does St. Bernard Software's iPrism product.

N2H2, www.n2h2.com/products/bess.php?device=features#da

St. Bernard Software, www.stbernard.com/products/iprism/products_iprism.asp

If the relationship between the organizations is not conducive to sharing a filter as described above, the library is better off installing its filter locally, on its own network segment.

Before selecting the product, identify the different filtering needs for each department of each library being filtered. it will not ultimately serve the needs of the library to install a filter that cannot be customized as needed locally. Finding the right balance of central and local control is key to identifying the right filter for the library.

Integrating filtering with other network services

Selection tip
Consider packaging the filter with other network services such as proxy or cache servers, patron-authentication programs, or public-access computer management systems.

Addressing other network needs may be possible when installing an Internet filter. For example, if your library has considered introducing patron authentication, session control, or print management systems for the public-access computers (PAC), you may be able to find a filtering solution packaged with a PAC management product.

If a library is considering introducing spam control, virus scanning, bandwidth management, application management, or firewall protection, a filter packaged with these types of security products, too, might be available

Comprise Technologies, www.comprisetechnologies.com

Cybraryn, www.cybraryn.com

3M, http://cms.3m.com/cms/US/en/2-115/cerlRFW/view.jhtml

Surfcontrol, www.surfcontrol.com

Comprise Technologies provides time, Internet, and print management software to libraries and schools. Its products can be used with many filter products to provide filtered Internet access with patron authentication and other services. Cybraryn provides PAC management tools to libraries including a filter product called FastTracker. 3M also packages PAC management tools with filtering using Surfcontrol as the underlying filter.

Many companies whose primary service is network management or network security provide Internet content filtering as part of a suite of services. For example, DynaComm i:filter is part of that company's i:series product line that includes i:scan for virus scanning and i:mail for e-mail monitoring. Smoothwall's primary product began as an affordable firewall product but its product list now includes traffic management and VPN products along with its content filter, Corporate Guardian.

Security products vendors that use a third-party filter include LogiSense, which developed its EngageIP product line—including NetManager and CacheManager. LogiSense uses the Cerberian content filter as a plug-in.

Barbedwire Technologies states that it has taken a "modular approach to an appliance based network security infrastructure" including intrusion detection and prevention, anti-virus, anti-spam, vulnerability assessment, wireless security, and application security. In addition, it provides a Web access control module, also based on Cerberian.

Even companies primarily focused on addressing Internet content filtering have introduced additional features into their products to address security issues associated with Internet access. For example, Websense not only provides a well-known content filtering product, but it also provides a bandwidth optimizer package and a client application manager designed to address spyware and malware (among other things) at the desktop. St. Bernard Software, maker of the iPrism content filtering network appliance, now sells an e-mail security product called ePrism.

Some filter companies that have been in business the longest also have a controversial history and have been the slowest adopters of new technology. More importantly, they have never seen libraries as one of their markets and have not responded to the library's unique filtering requirements.

This lack of response may be because library filtering requirements are not clear, or more likely, because library filtering requirements are more demanding. These products have generally filled the school filter niche and seem fairly content occupying only that space.

Some of these products have improved to some extent, but the biggest advances in filtering seem to be coming from different segments of the technology industry such as business and security. Many of the best filtering products available are now coming on the market. Don't discount newcomers simply because they are new.

DynaComm, www.dciseries.com/products/iseries

Smoothwall, www.smoothwall.net/products/family

VPN: Virtual private networking

LogiSense, www.logisense.com/products.html

Barbedwire Technologies, www.barbedwires.com/products/products.htm

Cerberian, www.cerberian.com

Websense, http://websense.com/products/about

St. Bernard Software, www.stbernard.com/products/products.asp

Open-source[12] options

At least two free, open-source filtering options are available that anyone looking into filtering should consider. These options are Squidgard, which runs on the Squid Web Proxy Cache, and Dan's Guardian. Squid (Squidgard's underlying proxy server) and Dan's Guardian form the basis for several commercially available products as well. LogiSense, N2H2 and iCognito all sell products based on Squid and Dan's Guardian forms the basis of Smoothwall's Internet filter.

Several lesser-known open-source filter products also are available or in the works, including Poesia, Dave's Naughty Stuff Blocker, Swiftsurf, and Middleman. ICRA is another organization to watch for free or inexpensive filters. It offers ICRAfilter, Optenet, and Xfilter.

Squidgard, www.squid-cache.org

Dan's Guardian, http://dansguardian.org/?page=smoothwall

Poesia, http://poesia.sourceforge.net

Dave's Naughty Stuff Blocker, http://sourceforge.net/projects/dns-block

Swiftsurf, http://swiftsurf.sourceforge.net/index-eng.html

Middle-Man, http://middleman.sourceforge.net

ICRA, www.icra.org/_en/icraplus/filters

Dan's Guardian, http://dansguardian.org

As of Dec. 23, 2003, a new subscription service was made available based on the black list used in Dan's Guardian. The new service, URLblacklist.com, contains a text-based list of categorized URLs that is maintained and checked by the company selling it. The subscription costs are under $480 per year depending on how often the library chooses to update its list. Using Squidgard or Dan's Guardian and this service offers libraries an affordable open-source option.

The primary advantage of open-source products is that they are infinitely customizable. None of the source code is hidden. Anything can be changed. Of course the library must have someone on staff who is familiar enough with Java, Perl, or PHP to take advantage of the customizability of open-source software.

But even if the library doesn't choose to make changes to the source code, open-source products offer advantages. The most tangible advantage is that open-source software is generally free or at least inexpensive.

Support for open-source software is provided through a network of programmers, developers, and other users of the product. Much open-source work is done on the Linux platform, which is itself an open-source operating system. Apache Web Server is a popular open-source Web server software. Half of all Web servers on the Internet run Apache.[13] (Other sources say this number is closer to 40%.)

oss4lib, www.oss4lib.org/about.php

Groups of librarians are already organized around the belief that open-source software makes good sense in libraries[14] such as members of oss4lib whose "mission is to cultivate the collaborative power of open-source software engineering to build better and free systems for use in libraries."

Resources for using Squidgard, http://mplcat1.meadvillelibrary.org/os/filtering

Kanguard, a filter based on open-source software, uses a library-specific block list. This filter is currently only available to Kansas public libraries, http://skyways.lib.ks.us/KSL/libtech/kanguard.

Resources for using Squidguard in libraries are available online or may be available from libraries successfully using open-source products—such as Kanguard—that just haven't posted technical information online.

Costs and personnel requirements

Numerous costs are associated with purchasing a filter including hardware, software, training, annual subscription, maintenance fees, and the costs and staff time associated with developing and incorporating new policies and procedures.

Depending on how many servers are needed to run the filter, whether the software is free, how many computers are being filtered, and what technical expertise is available on staff, the first-year costs could quickly equal any applicable E-rate discounts. The first-year costs will be dramatically higher than the subsequent years, but there are indeed significant costs to factor into your budget for each year a filter is used.

Because the costs will vary so much from installation to installation, providing any kind of useful generalization about what a filter will cost per user or per workstation is difficult. To facilitate comparing products, look at this filter product matrix where purchase and subscription costs for a 50-seat license have been reported by filtering vendors as follows:

Filter product matrix, http://libraryfiltering.org

Approximate cost for 50-seat license

Filter name	One-time fee	Annual subscription
i:filter	$1,136	$227 (first year free)
iPrism	$1,975 (hw included)	$1,165
EngageIP	$1,395 (hw included)	$520 ($900 w/reporting)
IF-2K	$1,222	No annual fee
Smartfilter	No one-time fee	$400
Bess	No one-time fee	$1,113
Minesweeper	$1,990	$1,050
Corporate Guardian	$406	$203
SurfControl	$1,500	$750 (first year free)
CyberSetting	$1,440	$645
FilterGate	$80	No annual fee
SurfPass	$2,350	No annual fee
CyberSitter	$497.50	No annual fee
Netsweeper	Hourly setup cost up to $500	$1,200
Squidgard	No one-time fee	$0 to $480/year[15]
Dan's Guardian	No one-time fee	$0 to 480/year

(*Source:* www.libraryfiltering.org)

Some products have many hidden costs such as installation and maintenance assistance, hardware and software upgrades, additional software purchases, and hardware purchases. The actual cost of ownership can't be easily derived solely from the costs paid to the filter company.

Skip Auld reported that his library (Chesterfield County Public Library) installed Websense in 2001. At that time, the cost of installing Websense on 161 public access computers cost $15,800, which included two servers. The annual license fee was $4,700.[16] Another library system filtering 500 PCs across 16 different buildings also uses two servers including SQL Server and spends $8,000 per year on the subscription alone.

The ALA E-rate Task Force has put together worksheets designed to help calculate filter-related costs. The purchase price and subscription (if applicable) is just the beginning. These spreadsheets help identify all the costs to consider when comparing prices.

ALA worksheets, www.ala.org/ala/washoff/WOissues/techninttele/erate/tools.htm

One worksheet compares costs among different vendor's products,[17] and another is designed to compare the costs of filtering against the value of E-Rate discounts to determine if a financial incentive exists to comply with CIPA.[18] Libraries should determine the total cost of ownership of the filter, not just the initial hardware and software costs associated with purchasing the filter.

Introducing new software always costs the library money in installation, support, and maintenance. Because the use of filters affects patrons and staff, everyone needs to be educated about what is being filtered and why, how to manage the filter, and what new procedures need to be followed. Introducing a

filter takes a significant amount of staff time. Don't overlook that time when designing budgets and calculating the total cost of ownership.

Hardware

Many server-based filters actually require more than one server to function fully. A database and reporting module is often separate from the actual filter itself. Sometimes, as in the case of Websense, SQL Server is required to support the reporting features. Depending on the requirements of the product, the platform, and the size of your network, these three modules may need to reside on separate servers.

Determine exactly what is required to run all the modules associated with the filter and reporting tools. A second server adds not only the cost of the hardware but also the operating system software and associated annual software maintenance costs.

Windows-based products are notorious for requiring separate servers for every function, which is one reason many system administrators prefer Unix- or Linux-based environments. Many filters support several different platforms so spend time comparing the pros and cons of running the chosen filter on the different platforms to see which makes the most sense for the environment.

Do an analysis of key networking devices such as the router, switch, hub, or firewall to ensure they are adequate for your new filtered environment. Replacing a router or switch can be costly in hardware costs as well as installation costs. Special technicians will probably be required to configure these network devices. Coordinate the configuration with the filter installation.

Installing the filter on a server already in use at the library may be possible. For example, many filters can be installed on an existing proxy server or firewall. In this case, you can probably save money in hardware and software, but be careful to account for the extra burden the filter will impose on the server.

Upgrading the RAM (and perhaps the storage capacity and the operating system) may be necessary. Don't assume that because the library has a Microsoft ISA Server (for example) and the filter's brochure states that it runs on this platform that the systems staff can simply plug the filter into it and be off and running in no time, and at no additional cost.

Vendors and technology staff should evaluate requirements specific to the library environment to determine whether using an existing server makes the most sense for the filter platform, and if so, what upgrades will be necessary or are recommended.

Network appliances

Network appliances are an attractive option for many reasons:

- Have no operating system costs
- Run on any network
- Use a Web-based management interface

They generally do not require the library to purchase an additional computer or upgrade an existing one. A network appliance also never requires the library to

be concerned about which operating system is running on the appliance or any costs associated with maintaining that operating system software.

Network appliances are designed to attach to the network in a black box fashion, meaning the filter and its underlying operating system are intertwined to hide the operating system. So even if the library runs all Windows servers, a network appliance (even one based on Linux) will not require that systems staff learn a new operating system. The network appliance doesn't require the systems staff to directly interact with anything but the filter's management tools.

Network appliances also are literally black boxes that connect to the network with no monitor or keyboard as found on a server with a management console. They are generally managed using a Web interface. A Web interface is attractive because the administrator can make changes to the system from anywhere on the network.

This distributed management function is important for libraries that don't have a dedicated network administrator sitting in the server room monitoring all network activity. If the network or filter administrator wears many hats, a good chance exists that that person won't be in the server room when a change needs to be made. Being able to do so from any library computer makes managing the filter that much easier and quicker.

Software and licensing

Know how many servers are required to run the filter and what operating systems will be used to run those servers. Unless the filter is a network appliance, an operating system will be required for each server.

The operating system running a library's servers is often referred to as the platform. Supported platform also sometimes describes the specific network device that the filter can be run on.

For example, Websense can be run on a Checkpoint, SonicWall, or PIX firewall or a Microsoft ISA or Squid Proxy server (among others)—all of which are network applications that themselves are installed on an underlying operating system. When choosing a server-based filter, an institution should stick to its chosen operating system platform so the systems staff doesn't need to be trained in several operating systems.

Network administrators who know Windows do not necessarily know Unix. Switching from one to the other takes time, training, and money. For example, even though you might be able to run your filter with fewer servers on a Unix platform, that platform may not be a good choice if your technology staff only knows Windows.

The cost of training the existing technology staff to support the second platform and the additional

Hardware and software costs associated with a single-server installation

- Operating system purchase (such as Windows Server 2003)
- Network application purchase (such as ISA Server)
- Filter purchase
- Server hardware purchase
- Annual software maintenance (another 15% to 20% of software purchase price)
- Annual hardware maintenance contract
- Annual filter subscription

skills required of future hires introduces several ongoing costs that might offset the savings of that second server.

In addition to the purchase of the underlying operating system for each server, you may have to purchase software such as proxy server software or caching server software or a firewall. This software has a purchase and maintenance cost too. Software licensing also is a significant cost.

Hardware and software costs associated with a two-server installation

Server one:

- Operating system purchase (such as Windows Server 2003)
- Network application purchase (such as ISA Server)
- Filter purchase
- Server hardware purchase
- Annual software maintenance (another 15% to 20% of software purchase price)
- Annual hardware maintenance contract
- Annual filter subscription

Server two:

- Operating system purchase (such as Windows Server 2003)
- Filter reporting module purchase
- Server hardware purchase
- Annual software maintenance (another 15% to 20% of software purchase price)
- Annual hardware maintenance contract

Hardware and software costs with open-source single-server installation

- Server hardware purchase
- Annual hardware maintenance contract

People often account for the cost of purchasing a new server—which involves a significant one-time hardware investment—without accounting for the ongoing costs associated with licensing software. A one-time cost purchasing cost almost always has an annual fee associated with it.

Most commercial software includes a software maintenance fee that is 15% to 20% of the original purchase cost. This fee covers all upgrades to the software and usually some level of support. Software maintenance fees are part of literally every type of software purchase including operating systems and network applications such as proxy server software.

The cost of hardware and the operating system is usually not included in the advertised price of a filter because the filter company does not sell the server and associated software that its filter will run on. Calculate these additional costs once the platform has been decided on or a network appliance has been selected. Using open-source software reduces several of the purchase costs and licensing fees associated with filtering software.

Many filters provide an optional reporting module. The reporting module can sometimes be installed on the same server as the filter or it may require its own server. As with all software, the reporting module and underlying operating system will have an initial purchase cost and an annual maintenance fee associated with it. Identify these costs up front.

Another set of software costs to keep in mind is at the desktop. If the library is running an outdated operating system such as Windows NT or Windows 98, the desktops may need to be upgraded to ensure they will function with the new filtered environment. Although desktop upgrades (also known as the client) are unlikely to be an issue with a network appliance or server-based filter, verify the client requirements before selecting a product because the cost in time and money of having to undertake such an effort is substantial.

Installation

In addition to the hardware and software costs associated with installing a new filter, additional expertise will likely be needed to help existing technology staff upgrade network components, install new servers, and configure the server software and filter components. If routers or switches need to be installed or reconfigured, this work will likely require outside assistance, too.

If possible, install and configure new software in a development environment rather than a production environment. A development environment is a network area where new components can be tested before going live. The production environment is the live network where library work is being performed. Given small technology staffs, tight technology budgets, and small working spaces, most libraries don't have this luxury.

More often, new software is installed on the servers in the middle of the night, and everyone hopes all the kinks are worked out before the library opens. Of course, this approach is becoming more difficult as libraries increase their virtual reach and provide some level of services 24 hours a day, even if the doors aren't open.

An outside consultant or the filtering vendor may be able to set up a temporary development environment to be used to train staff and to test different configurations before rolling out the filter throughout the library. Implementation and rollout is detailed in the next chapter.

Hosted solutions

Not all filters require hardware and software to be purchased. Some filters can be configured at the client to use the filter company's hardware and software over the Internet. Usually this hosting is done by configuring the client PCs to use the vendor's proxy server.

You might be able to use your Internet service provider's filter (if it offers this feature). This option might be viable for some libraries that have limited technical staff. The problem with this approach is that the library may lose too much control over how the filter is configured and operated.

Key product features

Libraries should not do without certain product features. The unique role of libraries and their relationship to patrons demands a level of transparency, privacy protection, and flexibility that other users of filter products do not require.

Many of these features are not included in the client-based products designed for home use. These features are generally only found in the business-oriented products or in some cases the products designed for school use. Several product features also can save library staff time.

Accuracy

All filters overblock. All filters underblock. No filter is 100% accurate because no one agrees on what being 100% accurate is. Still, librarians should evaluate whether a programmatic or inherent flaw exists in how sites are being categorized by a filter company before committing to using a filter.

Accuracy falls to an unacceptable level for library use because of the following:

- The practice of blocking all domains associated with a specific IP address
- The practice of blocking an entire domain because of how a single page was categorized
- Faulty programming
- The political or social bias of a filter company

Salon.com example, http://archive.salon.com/sex/feature/2001/06/18/zaftig

Some filters might place an entire domain into a category based on the analysis of a single page on that site. For example, here's an excerpt from a page on salon.com:

> Zaftig erotica
> "I was a thirty-something dyke watching this sweet ass slide its way down the hall." An excerpt from a collection of short stories.

IP: Internet Protocol

A good filter might categorize this page as erotica. But a filter that consistently overblocks could do one of at least two things wrong. It could, based on this one page, classify all of salon.com as erotica, which it is not. Or, even worse, it could classify everything at 206.14.209.40 (which is the IP address of salon.com) as erotica, which would not only classify all of salon.com as erotic but any other domains sharing that IP address as well.[19]

The filter should be sophisticated enough to distinguish between the different pages on the website rather than classifying the entire domain under one category.

Below is an example of how the different filters evaluated the above URL at the page level, the subdomain level (archive.salon.com), the domain level (salon.com), and the IP level:

URL	Smartfilter	Cerberian	Cyberpatrol	i:filter
archive.salon.com/sex/feature/2001/06/18/zaftig	mature	news/media	sexually explicit	not on list
archive.salon.com	entertainment	news/media	not on list	entertainment
salon.com	portal	news/media	sexually explicit	entertainment
206.14.209.40	not on list	pornography	not on list	not on list

This simple test shows Smartfilter categorizes pages to the page level. The above test doesn't make clear whether the other three products do so.

The fact that Cyberpatrol has classified salon.com as sexually explicit and that Cerberian has classified the IP address of salon.com as pornography also suggests potential problems with how categories are assigned to top level domains and to IP addresses.

Not all filter companies provide a way to test the categorization of a URL online but if such a tool is available, it is an excellent way to better understand content categories and the likelihood that unacceptable problems exist with the product's categorizing algorithms.

See the sites below for checking how URLs will be categorized by their product:

Smartfilter, www.securecomputing.com/cgi-bin/filter_whereV301.cgi

Cerberian, http://sitereview.cwfservice.net/sitereview.jsp?referrer=88

Cyberpatrol, www.cyberpatrol.com/support/#test

i:filter, www.futuresoft.com/ifilter/Categories/QueryCategory.asp

Ability to turn off keyword blocking

Keyword blocking was a technique used in the early filters that vehemently turned detractors against the use of filters. It resulted in massive overblocking, and the meaning of pages was altered by some keyword filters.

Keyword blocking is when a word is defined as forbidden and cannot be used in a URL or search box, or when websites containing the forbidden word are blocked. Such a simplistic technique for blocking content has no place in a library setting.

Although keyword blocking is more sophisticated than it was in the early years, it is still available and is sometimes a technique used to supplement URL filtering, which relies on a database of categorized URLs. Keyword blocking can help reduce the incidence of missed sites (but it still causes major overblocking problems).

When keyword blocking is offered as the only mechanism available for blocking content and cannot be turned off, the filter is not appropriate for library use (or possibly anywhere). When keyword blocking is optional, turn it off.

Selection tip
Make sure keyword blocking can be disabled.

Although some filters rely on some type of content analysis, which to some degree relies on evaluating keywords and phrases, this approach is generally not referred to as keyword blocking and should not be confused with it.

Multiple filter profiles

If the library is implementing filtering to do more than block all CIPA-mandated content, it should select a product that allows for multiple and flexible filter profiles. With CIPA, all patron and staff PCs must be filtered, but a library might wish to block beyond CIPA on certain PCs, such as the children's computers. Doing so requires a filter that allows multiple filter profiles.

For network-based products, linking to or importing usernames from the network directory, such as Active Directory or LDAP, is often possible. Defining a unique filter profile for each user in the library isn't usually necessary, but having filter profiles that can be applied to groups of users is required.

LDAP: Lightweight Directory Access Protocol

Importing users from the network directory can alleviate some of the setup associated with implementing a new filter, but it won't relieve the library of the job of setting up filter profiles for each group of users.

Libraries may wish to use separate filter profiles for young children, young adults, and adults. Depending on the filtering strategy being used and the capability of the filter, libraries also may have to create a separate *unfiltered* adult profile as well.

For example, some products do not offer a way to temporarily disable filtering for a specific user. In this case, the only way to do so is to log in the user using a different filter profile.

A staff profile might be set up with no categories selected for blocking to ensure that staff are not hampered by blocked sites and a handful of sites in the 'always bock' list to satisfy CIPA concerns.

When the filter being used hides the URLs in each category, staff must check erroneously blocked sites to make corrections to the filter. Don't filter staff computers to the same extent that patron computers are filtered when using a commercial filter's hidden content categories.

If the library is using a single block list of sites selected for blocking, then multiple profiles are not necessary. If all PCs, including staff PCs, are filtered using the same, transparent, library-customized list, then you can eliminate problems with erroneously blocked sites. Multiple profiles may not be needed.

Ability to customize default block page

Selection tip
Choose a product that provides for flexibility in designing the block page.

The ability to customize the block page is another important feature for libraries because the block page, or page that is displayed when a patron encounters a blocked site, serves as the primary interface between patron and filter. Some filters do not allow the administrator to customize the block page. Other filters, usually those designed for home use, display a generic browser error:

> **Not Found**
> The requested URL /homework/ was not found on this server.
>
> Additionally, a 404 Not Found error was encountered while trying to use an ErrorDocument to handle the request.

At least one product, CyberSitter 2003 sometimes redirects the users to a more suitable site instead of advising them that they have been blocked.

> Depending on how the site is blocked, CyberSitter will oftentimes redirect to a site that is child friendly or educational. We maintain a special server for this purpose as well as a database of family friendly websites. When a user attempts to access certain sites, they are taken to a random family friendly site instead.
>
> *—from CyberSitter help page of installed product*[20]

Products that allow for customization of the block page usually redirect the browser to an HTML page stored on the local network. As a standard Web page, it can be customized as needed by the administrator. Consider including the following information on customized block pages:

- Who to contact if there are problems with the filter
- The Internet use policy
- How filters are used to enforce library policies
- How to request review of a blocked page
- Where to find a computer that provides unfiltered access

Wording, such as the following, is helpful to patrons:

Sorry!
The content of this URL is currently blocked.
http://www.teenc.com/ is rated as:
nudity, pornography

If you require further assistance, please contact any library staff person.

Click Here to Override

In addition, different filters allow for some flexibility in configuring the block page with filter variables and special features, such as the ability to display the following:

- The URL being blocked
- The category causing the block
- A URL to the filter's website for requesting the filter company re-evaluate the site and its category
- A form for requesting that the local administrator immediately unblock the page
- A button that allows the patron to override the blocked page themselves (such as a warn page)
- A password-protected button for staff to use to override the blocked page

Password override

The default block page can sometimes be set up to include a password-protected override feature. Staff would need to know the password to help the patron. The ideal library filter allows floor staff to unblock a specific page for a specific user, on-the-fly, without involving the filter administrator.

A password-protected override also can be implemented by placing an icon for it in the system tray of the Windows desktop. Using a special keystroke combination or by right-clicking the icon, a staff person can enter the override password for the patron.

Password override features are implemented in numerous ways. Sometimes the password override simply overrides a single blocked page. As soon as the user visits another blocked site, the process must be repeated. At the other extreme are password overrides that effectively turn off all filtering.

Many of these types of blanket overrides remain in effect for a set period of time (10 minutes, 20

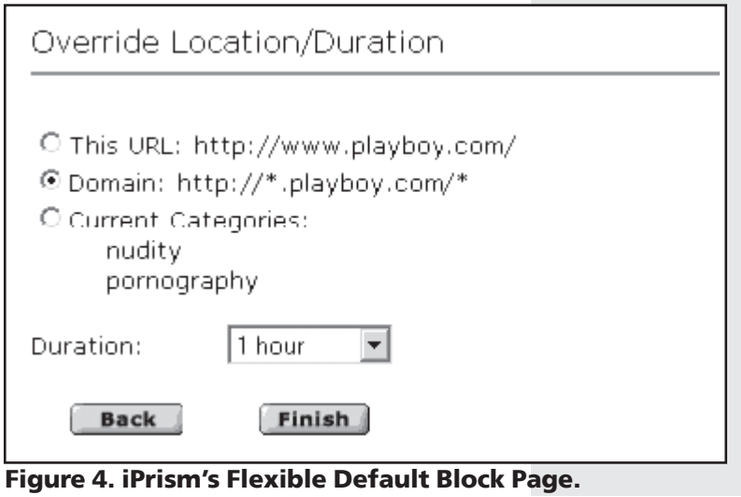

Figure 4. iPrism's Flexible Default Block Page.

minutes). The length of time the override stays in effect is generally set by the filter administrator. Still other products allow the override duration to be set when the password is entered.

Some products allow for a great deal of flexibility for handling overrides on-the-fly (see Figure 4). iPrism provides one of the best examples of maximum override flexibility. It allows the staff person to determine the duration of the override and whether the override shall apply only to the current page, the entire domain, or to all pages in the blocked categories.[21]

Ability to disable the filter for set amount of time

Closely related to the ability to unblock a single site is the ability to turn off the filter entirely for a given patron. Depending on how the password override option works, librarians may be able to override a single page or turn off the filter entirely, as needed, using the password override feature. Other filters will implement this feature a different way (or not at all).

Selection tip
Choose products with overide features that will allow you to enforce your Internet user policy.

One of the key outcomes of the Supreme Court opinions issued in the CIPA case was that the ease with which filters can be disabled was critical to its finding of constitutionality. The court stated that CIPA itself was constitutional, but if a library neglected to turn off filtering for an adult patron conducting bona fide research, constitutional problems could arise.

But not all filters are so easy to turn on and off. Although most filters allow some degree of overriding capability, entirely turning off filtering introduces a whole new set of problems including:

- Turning the filter back on
- Determining how long the filter remains disabled
- Informal monitoring of adults to ensure they are following the library's IUP and not viewing illegal content such as child pornography
- Lack of clarity about the library's responsibility to monitor the unfiltered computer to ensure a young person doesn't decide to use it while the adult is in the stacks or the restroom
- The fact that if the library is using the filter to control certain types of activities such as IRC (chat) , Instant Messenger or playing music or games, disabling the filter entirely disables these non-CIPA controls as well

In general, the library might be left with no option but to log out the patron from the current filtered profile and log them back in as a less-filtered patron—thus removing the blocks to constitutionally protected content but retaining other controls that prevent certain activities such as Internet chat or games.

Changing a patron's filter profile is one way to ensure that other controls handled by the filter remain in place while content filtering for CIPA is turned off. This approach, though, imposes a substantial burden on floor staff who have to be available to both log out the patron and then ensure the right filtered profile is in place for the next user of that machine.

ALA, www.ala.org/ala/washoff/woissves/civilliberties/washcipa/fccc.pa.htm

ALA has put together a useful Q&A section on the issue of disabling the filter and CIPA compliance.[22] At this point, the FCC has simply stated that the library must have filters installed on all computers and must be able to turn off the filter to allow adult patrons access to constitutionally protected speech.

Time controls

Many filters have the ability to apply different filtering profiles to a user or computer, based on the date and time. For example, if the public access computers are used largely by children during the after-school hours and by adults during the school day, setting the default filter to 'children' from 3 p.m. to 6 p.m. and set it to 'adult' at all other times might be convenient.

This flexibility can save the staff members from having to repeatedly unblock pages for the adult patrons using a public computer set with a children filter profile. Or it may prevent the need to repeatedly assist patrons who must log in and out to put the proper filter profile in place.

Ability to manage categories

Another important feature for libraries is the ability to recategorize websites that have been categorized by the filter company in a way that does not conform to the library's expectation or understanding of the Web page content and the filter's content categories. Although most products allow the administrator to add sites to an 'always allow' or an 'always block' list (Figure 5), this capability does not allow for the level of control over access that a library might desire.

Selection tip
Find a product that allows you to modify how Web pages are categorized.

'Always allow' list:
Customizable list of sites that are always exempt from filtering.

'Always block' list:
Customizable list of sites that are blocked for everyone.

For example, when devising a filtering strategy that encompasses children, youth, adults, and staff, libraries might want to allow access to sites for young adults that are not appropriate for the computers in the children's department. Undoubtedly the filtering company will have made mistakes that should be corrected by library staff. Having the ability to recategorize these sites will be important for libraries using multiple filter profiles.

Regardless of the reason, the ability to override the filter company's cataloging decision is key to ensuring that sites that should be blocked for children are not necessarily blocked for adults. When the only recourse for correcting a miscategorized site is to use the 'always allow' or 'always block' lists, then the library has to decide between overblocking adults and underblocking children.

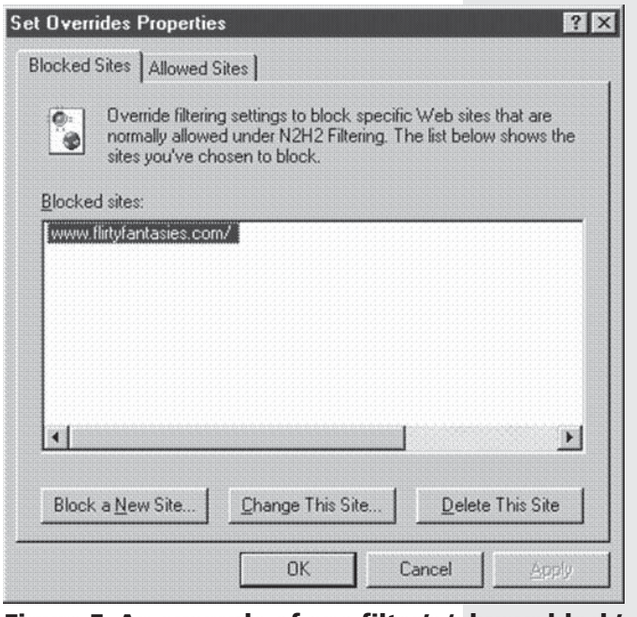

Figure 5. An example of one filter's 'always block' list showing the site listed always should be blocked.

The ability to recategorize or move a URL from one category to another is an important control for libraries to use to ensure their filter is as accurate as it can be.

In addition to moving a URL from one category to another, some filters also provide the ability to add new categories of content that can be used to build filter profiles. This feature is nice when a high level of granularity in the content categories is required or when the filter's simplistic categories are inadequate.

Example of recategorizing URL for accuracy

Sample library filtering policies:

Adult-blocked categories	Children-blocked categories
Extreme/obscene	Extreme/obscene
	Sex education

How it works:

1. URL categorized by filter as extreme/obscene: iwannaknow.com.
2. Using above profiles, site is blocked for children as well as adults.
3. Filter administrator recategorizes site as sex education.
4. Site becomes available to adults but continues to be blocked for children

To populate any new category created by the library, the sites that belong in that category will have to be located and entered into the new list—a fairly labor-intensive process.

Because filter companies do not generally reveal the URLs that fall within each content category, librarians cannot just move a URL from one category to another. Any site that the library wishes to override or place in its newly defined category must be identified via a Web search or by monitoring the filter's log files to focus on the URLs being accessed by patrons.

Ability to block images only (not text) within a content category

Selection tip
Consider products that allow you to block images instead of the entire Web page in certain categories.

For anyone comfortable with a strict interpretation of CIPA (blocking images not text), an important feature to watch for is the ability to block only images within a selected content category. Although many products allow the library to block by file type, these types of blocks (as with protocol style blocks) tend to apply to all filtering rather than being associated with specific categories of content.

More recently, products have sprung up that allow the administrator to select a content category (pornography, for example) and to block, within that category, certain file types such as .jpg, .gif, and other image files when they are loaded into the browser.

The result is that all text is loaded but the images associated with the pages are not. The effect is the same as setting your browser to not load images—but it is only done when the filter encounters a page in the categories selected by the filter administrator.

Filters that can block images but not text, by category: DynaComm i:flter (www.dciseries.com/products/ifilter) andNetweeper (www.netsweeper.com)

At least two products have this feature: DynaComm i:filter[23] and NetSweeper Intelligent Web Filter.[24]

Useful reports and monitoring tools

The extent to which predefined reports are made available varies considerably from product to product. Good reports can be significant timesavers and can help librarians monitor filtering closely. Oftentimes products with robust

reporting features tend to collect a large amount of information that also should be purged from the logs periodically to protect patron and staff privacy.

Depending on how the filter is implemented, the raw data may be contained in proxy server logs or in other logs created by the filter product itself.

The filter administrator should have the job of generating reports on a predefined schedule and then purging the logs that provided the source data contained in the reports. Retained reports should only contain summary data. No information that can be traced back to a specific person using the library computers should be retained longer than necessary.

The ability to generate useful reports and to purge data that is unsafe to retain in this age of the USA PATRIOT Act is an important consideration when selecting a full-featured filtering product. Each state has laws in place that cover issues of privacy. Apply these laws when creating retention schedules associated with logs and reports.

Examples of useful reports that might be provided by the filter's reporting module include:

- URLs requested that were blocked (by login or group, within a given category)
- Percentage of sites blocked, warned, allowed (by login or group, by category)
- Sites visited that use the most bandwidth
- Top-visited sites (allowed, warned, blocked)

Sometimes a library will choose to monitor activity by user or workstation. This monitoring would only be necessary when staff or administration suspects that the library computers are being used to engage in illegal activity. For such situations, different reports may be required such as:

- URLs requested by a given user
- URLs requested at a given workstation

Because of the way filters work, much private information is contained in the logs. The summary reports can provide useful information about the kind of information your patrons are most interested in locating. Review these reports and logs regularly to ensure the filter is configured appropriately for the library.

Logs can easily be misused, though, and should be purged frequently to protect the privacy of all computer users. A good filter provides useful summary reports and an easy way to purge logs.

Evaluations and reviews

Few sources exist for finding objective evaluations of Internet filters, especially for library use. Some studies of Internet filters focus on filter accuracy or the likelihood that they block constitutionally protected information in target content areas (health for example). But many of these studies accepted the default settings instead of configuring the filter to work better in a library setting.[25]

The best approach to evaluating a filter is to study the categories and features of several products and exclude the products that are not a fit because the categories or features aren't sufficient. For example, perhaps the content

categories aren't defined in a way that will help enforce the Internet use policy or the unblocking features are insufficient.

Next you would design and conduct a test of the product that matches the library's filtering goals. This test is the only way to ensure the filter is an acceptable match. Evaluating for accuracy makes the most sense when applied to a specific library with a specific filtering strategy.

Review the following evaluations of products when selecting a filter. These evaluations are not necessarily scientific studies but do provide some insight into issues associated with accuracy, effectiveness, intellectual freedom, civil liberties, and overall performance:

www.eff.org/Censorship/Censorware/net_block_report/net_block_report.pdf

http://kff.org/entmedia/3295-index.cfm

http://cyber.law.harvard.edu/people/edelman/mul-v-us

www.fepproject.org/policyreports/filteringreport.html

www.etestinglabs.com/clients/reports/usdoj/usdoj.pdf

http://galecia.com/included/docs/filters.pdf

- Online Policy Group and the Electronic Freedom Foundation, Internet Blocking in Public Schools: A Study on Internet Access in Educational Institutions, (San Francisco, CA: Online Policy Group, June 2003).
- Kaiser Family Foundation, See No Evil: How Internet Filters Affect the Search for Online Health Information. (Kaiser Family Foundation, December 2002).
- Edelman, Ben. Sites Blocked by Internet Filtering Programs: Expert Report for Multnomah County Public Library et al., vs. United States of America, et al. (Cambridge, MA: Ben Edelman, 2002).
- Heins, Marjorie, and Christina Cho. Internet Filters: A Public Policy Report, (New York: Free Expression Policy Project, fall 2001).
- Updated Web Content Software Filtering Comparison study, conducted by eTesting Labs on behalf of the Department of Justice. October 2001.
- Ayre, Lori. Internet Filtering Options Analysis: An Interim Report. Prepared for the Infopeople Project, May 2001.

Other types of reviews are available for client-based products especially those to be used in the home. But these reviews are generally strongly biased in favor of a certain social agenda such as www.filterreview.com, which is provided by the National Coalition for the Protection of Children & Families, an organization dedicated to "moving the people of God to embrace, live out and defend the biblical truth of sexuality."

Two notable websites provide evaluations of client-based products that might be more useful to libraries.

GetNetWise, www.getnetwise.org

GetNetWise provides an objective set of resources for parents seeking filtering or monitoring programs at http://kids.getnetwise.org/tools.

InternetFilterReview, www.internetfilterreview.com

- GetNetWise is a public service organization composed of Internet industry corporations and public interest organizations dedicated to "ensuring that Internet users have safe, constructive, and educational or entertaining online experiences."
- InternetFilterReview.com provides a more balanced approach to filtering in the home including reviews of several products.

Developing an RFP

The best way to compare the cost and impact of a filter is to distribute an RFP (request for proposal) to all filters vendors with a product you would like to consider for purchase. Rather than issuing an RFI (request for information), which essentially invites the filter company to send you their promotional material or a salesperson, an RFP requires the library to do some upfront work to define its needs and the network environment.

Any vendor responding to the RFP should be required to address each aspect of the RFP to be considered by the library. This requirement allows the library to fashion the RFP in a way that emphasizes the most important elements, ensures the proposals can be compared fairly, and makes rejecting proposals easy if the vendors don't meet the minimal qualifications.

Many resources are available for effectively writing an RFP, but writing an RFP for a filtering product is different from writing an RFP for other types of technology such as a self-checkout system.

Because of the tight integration of the filter with all library software systems and the ramifications for staff and patron use and ongoing support requirements, a thorough RFP is critical. It should include:

- A detailed description of your computing and network environment
- A statement about why the library is introducing filtering
- A list of functional requirements
- Instructions about how to respond to the RFP—in what format it should be delivered, how to find answers to questions, and the deadline for responses

Network description

In the description of the library's computer and network environment, describe all computer components in detail and require the responses to describe any additional hardware or software needed to be upgraded or purchased to deploy the respondent's filter product.

Put the burden on the filter company to recommend the best installation choice for your environment. Require that they recommend, or even purchase on your behalf, any additional servers or computers needed.

For the vendor to accurately respond, include a description of all servers including operating system (version and patch level), hardware (brand, model, processor, RAM, hard drive capacity, and configuration) and applications installed (version and patch level) on each server.

Also describe all installed routers and switches. Include model numbers and software installed (including version and firmware) and the number of computers to be filtered (including their hardware configuration, operating system, and all applications running on them).

Request that responses include not only a detailed description of all hardware and software that will need to be purchased (with cost quotes) but also ask for a list of any software you've listed in your inventory that cannot be run on the same server or PC as the filter software. Request that the responses describe how the filter must be installed on the server and on each client.

Goal statement

Give the vendors an opportunity to cater their responses to your needs by providing a statement about why the library is choosing to implement a filter at the current time. You have no need to be secretive about why you want to filter or how you hope to do it.

The filter vendor wants to sell you its product but if it knows its product is not well-suited to your needs, it won't want to sell it to you. A vendor's success depends on satisfied customers so give it the information it needs to satisfy your needs.

Functional requirements

The functional requirements section is the meat of any RFP. In this section, the library should describe the features that must be present in the filter or are highly desirable. This section can be designed in any number of ways but have a strategy for evaluating responses in place before writing this section.

For example, you could list the functional requirements and then require each respondent to respond to each requirement with a numerical response such as:

1—Product meets this requirement.

2—Product does not meet this requirement.

3—Third-party product can be added to address this requirement.

Devise a response system that allows those evaluating the responses to easily compare the responses and ideally to analyze them numerically.

Another approach is to identify the mandatory requirements from the highly desired requirements and then weight each one in the analysis phase. For example, perhaps the library has decided that having the content category causing the block on the default block page is not as important as having the blocked URL displayed on the block page.

When evaluating responses, the company whose product answered 'yes' to the ability to include the URL on the block page will receive more points than those that offer only the blocked category on the block page.

In addition to the functional requirements discussed in this report, look over the product features listed at www.libraryfiltering.org where more than 50 product features are described for several different Internet filters. This chart also provides information about how to contact several of the filter companies when sending out the RFP.

Information about the company

In addition to functional requirements of the filter, the RFP can be a vehicle for learning more about the filter company and terms of the sale. Ask for references from other libraries using the product. Ask the filter company to describe ownership and strategic partnerships for the company, and the primary customer base for the product.[26] The RFP is a good place to include these types of questions.

Other good questions to ask in the RFP are whether the company will provide a money-back guarantee or 30-day trial period. Why not ask for free technical support for the first year or complimentary installation assistance? Ask for what you want. The RFP is a competitive process. Give the vendors an opportunity to find a way to make the sale.

Response requirements

Every RFP should include a contact person for questions and any time frame associated with the RFP and responses. State in the RFP to whom the questions should be addressed and how questions will be answered. Ideally, the library will gather all questions from respondents and then answer all questions in one document, which they provide to all RFP respondents.

Remember the goal is to ensure all respondents have the same information about the environment. Calling a filter salesperson about your library and leaving out important details easily happens.

Using an RFP and addressing all vendors collectively until the initial evaluation of RFP responses is complete increases the likelihood that the library has covered all of the necessary bases with every prospective vendor.

Include information about how to submit questions, how the library will respond to the questions, the deadline for the final RFP response, and the format in which the response must be received. Request an electronic version as well as a paper version of all responses to allow for maximum flexibility in analyzing and evaluating responses.

Some vendors will want to provide documentation that can't be easily e-mailed so requesting a paper copy of the RFP response "with attachments" allows the vendor a more manageable way to include supplemental material without overwhelming anyone's inbox.

RFP and open source

Going through the process of developing an RFP is an important way to clarify the library's needs and a good way to compare commercial products. But don't neglect to compare those products and their associated costs with open-source alternatives, too.

The RFP process will likely exclude open-source options, so you may need to consult with an outside expert about the open-source options. Such a consultation is worthwhile if the resulting decision saves the libraries thousands of dollars in licensing costs each year.

Notes

[12] For a definition of open-source software and to learn more about how to locate open source programs, see http://opensource.org/docs/definition.php (Jan. 10, 2004).

[13] www.opensource.org/advocacy/faq.php (Jan. 10,2004).

[14] See Eric Lease Morgan's post about Open Source and Librarianship at www.oss4lib.org/listserv/msg00122.php (Jan. 10, 2004).

[15] Libraries do not need to pay any subscription costs if they wish to maintain their own 'block' list. A compatible subscription service is available from URLBlacklist.com Cost per year depends on how often a library wishes to update the list each week. Daily updates cost $480 per year.

[16] Auld, Hampton (Skip). Filters Work: Get Over It: A Virginia Library System Opts for Filtered Internet Access and Makes a Believer out of One Skeptic. *American Libraries,* February 2003, pp 39-42 (Jan. 10, 2004).

[17] www.ala.org/Content/NavigationMenu/Our_Association/Offices/ALA_Washington/Issues2/Civil_Liberties,_Intellectual_Freedom,_Privacy/CIPA1/compare.xls (Jan. 10, 2004).

[18] www.ala.org/Content/NavigationMenu/Our_Association/Offices/ALA_Washington/Issues2/Civil_Liberties,_Intellectual_Freedom,_Privacy/CIPA1/TCO.xls (Jan. 10, 2004).

[19] See "Websites Sharing IP Addresses: Prevalence and Significance" by Ben Edelman http://cyber.law.harvard.edu/people/edelman/ip-sharing for a thorough discussion of the significant overblocking problems that can occur when filters block IP addresses. Edelman reports that 87% of the active websites share their IP address with other domains. Therefore, when a URL is categorized as sexually explicit, for example, some filters automatically resolve the URL name to the IP address resulting in all other sites at that IP address being categorized as 'sexually explicit' whether they are or not (Jan. 10, 2004). [20] CyberSitter 2003 was installed to see how this feature works. And although the documentation and other users have verified that redirection occurs, the author was unable to replicate it using Windows XP. The author typed potentially objectionable words and phrases including 'nasty girls' and 'puberty'—both returned a generic 404 browser error. No redirection to an educational site occurred.

[21] iPrism 3.5 Administrator's Guide. www.stbernard.com/products/docs/ip35_adminguide/Chapter02.html#913765 (Jan. 20, 2004).

[22] Susman, Thomas M. Ropes & Gray LLP. December 2003. www.ala.org/Content/NavigationMenu/Our_Association/Offices/ALA_Washington/Issues2/Civil_Liberties,_Intellectual_Freedom,_Privacy/CIPA1/Q_and_A/Q_and_A.htm#dec03 (Jan. 20, 2004).

[23] For more information about this product, see the Library Filtering Product matrix at http://libraryfiltering.org/detail.php?pid=7&id=1 (Jan. 10, 2004).

[24] For more information about this product, see the Library Filtering Product matrix at http://libraryfiltering.org/detail.php?pid=7&id=17 (Jan. 10, 2004).

[25] The most recent example of an accuracy evaluation using the filter's default setting was done by the Colorado State Library and is available at www.aclin.org/cipa/downloads/Report_on_Free_Filters.pdf (Jan. 21, 2004).

[26] From the ALA Sample RFI Questions. www.ala.org/Content/NavigationMenu/Our_Association/Offices/ALA_Washington/Issues2./Civil_Liberties,_Intellectual_Freedom,_Privacy/CIPA1/RFI.pdf (Jan. 10,2004).