



828 West Taft Avenue
Orange, CA 92865
714-282-6111
714-282-6117 Fax
www.8e6.com

Google and Yahoo SafeSearch: Are Your Students Really Being Protected from Harmful Web Content?

Web Filtering Technology Outsmarts Cybersmut

By The Forsite Group

Introduction

As the Internet is increasingly integrated into school curricula and classrooms, educators have become caught up in a constant struggle to prevent students from being exposed to pornographic images and content.

There are a wealth of solutions that promise to keep cyberporn off school networks: Client-side software filters implemented on individual PCs; server-based software that offers editable lists of keywords and regularly updated blacklists of blocked URLs; third-party integrated solutions that deploy proprietary software on various network devices; and industrial-strength filtering and monitoring appliances, which have proven particularly effective on enterprise networks.

Despite this array of technologies, however, stopping cyberporn from reaching students' desktops — whether by accident or by design — remains an elusive goal.

Although the problem is a complex one, it can be summarized fairly simply. Google and Yahoo, the two leading search engines, label their graphics in a way that prevents software and hardware filters from identifying them as cyberporn. Thus the filters, no matter how sensitive or powerful, don't realize that they should block these images. Instead, they pass them along, allowing pornographic thumbnails to be displayed on the Google and Yahoo search results pages.

Here's the kicker: Both Google and Yahoo have very powerful SafeSearch options that will completely prevent cyberporn — images and text — from being returned as search results. But there's a catch: *Any student at any desktop can disable these SafeSearch features with a few clicks — and there's no way for school IT managers to know they've done it.*

The situation appears to be an unsolvable dilemma: educators can't rely on Google or Yahoo SafeSearch because it's so easy to disable. They can't count on their filtering hardware and software to block pornographic images Google or Yahoo return. And they can't "outlaw" Google and Yahoo searches without affecting the educational value of their networks.

There is a solution: a new class of standalone hardware filter that makes it impossible for anyone but a system administrator to disable Google or Yahoo SafeSearch. This ability pays a double dividend: school IT professionals can ensure that SafeSearch remains the default setting. And they can use the filtering appliance to foil other attempts to reach age-inappropriate sites, while working with its advanced features to simplify management, deepen their defenses against cyberporn, and add a stable, scalable solution to their networks.

What's more, since this new solution is a dedicated appliance, there's no chance that protection will be traded off against performance — a common problem with software filters. Performance is just as much a concern in education as it is on the enterprise, as Charles Thompson, CIO of Orange County Public Schools (Orlando, FL) can attest. Thompson is responsible for 161 schools and 180,000 students. Currently, he's deployed more than 45,000 PCs. Slowdowns and bottlenecks (another common problem with software filters) would affect Internet access and application response times, undercutting the educational value of the network.

Dirty Tricks of the Trade

When it comes to cyberporn, educators have more to deal with than just technology issues. The online porn industry aggressively targets kids. For example, cyberpornographers buy up lapsed domain names for games, toys, TV shows, and movies that appeal to kids — and redirect them to porn sites. Similarly, they register URLs that are likely to be mistaken for general-interest sites. Probably the most famous instance of this is www.whitehouse.org and www.whitehouse.com, which for years linked to pornographic sites. (The correct URL for the White House is www.whitehouse.gov).

Another common practice is typosquatting, registering domain names that resemble actual sites: Yahooo for Yahoo, Gooogle for Google. A slip of the finger is all it takes to call up adult content.

These strategies, among others, have contributed to a statistic that, when it was published by The Third Way, a self-described moderate think tank, set off something of a firestorm: Children between 12 and 17 are the largest consumers of online pornography in the United States. (http://www.third-way.com/data/product/file/14/porn_standard.pdf)

That statistic really hits school IT managers where they live. They know that cyberporn has become a lightning rod for parents, community and religious leaders, and politicians. Its presence on a school network is the worst kind of publicity. They also don't want to expose their schools to possible legal action, although state and Federal laws regarding underage access to adult content are still being contested.

A few other statistics help reveal what IT managers are up against:

- According to the 2005 Pew Internet Project survey, 68 percent of all teenagers (about 16 million students) access the Internet at school. (http://www.pewinternet.org/PPF/r/163/report_display.asp)
- Family Safe Media estimates there are 4.2 million pornographic Web sites, which represents 12 percent of all sites worldwide. (http://www.familysafemedia.com/pornography_statistics.html)
- Only 3 percent of pornographic Web sites require age verification that goes beyond the honor system, according to *The Porn Standard*, a white paper published by The Third Way.

Anyone Under 18 Admitted

The “honor” system in the cyberporn industry typically comprises a front page to a Web site that asks visitors to click “Exit” if they are under 18. To discourage that choice many front pages offer graphic previews of what's available behind the “Enter” button, as well as invitations to “continue the virtual tour” or “download sample videos.” *The Porn Standard* indicates that 74 percent of adult sites offer free content.

Even members of the cyberporn industry realize that enticing underage viewers is fast becoming risky business. In a call for self-imposed age verification, Kathee Brewer, editor at large for Adult Video News Online notes, “Many attorneys who represent clients in the adult Internet industry now say that a simple choice between buttons or links that read ‘I'm 18 — let me in!’ and ‘No thanks’ is not sufficient to discourage underage access, because there's no penalty attached to lying.”

(http://www.avnonline.com/index.php?Primary_Navigation=Editorial&Action=View_Article&Content_ID=257108)

How a Bill Doesn't Become a Law

Other approaches to limiting access to porn sites have thus far failed. For example, the Federal government's 1997 Child Online Protection Act (COPA) has been hung up since 1998, while various courts attempt to determine if it violates the First Amendment.

Its most recent effort, the Internet Safety and Child Protection Act of 2005 (S.1507), looked as if it would sail through the Senate, until it ran afoul of another Constitutional issue. Besides strict age verification and a provision that prevents financial institutions from performing credit card transactions for unregulated sites, the bill also calls for a 25 percent tax on all adult Web sites. That tax would be used to fund a variety of projects, including R&D grants for “new filtering technologies” that will stop kids from accessing inappropriate content via “wireless and other emerging technologies.”

According to opponents, the tax is Constitutionally questionable. The Senate Finance Committee is currently reviewing the bill.

There's one other cyberporn-related issue that may ultimately affect school IT managers. In 2005 ICANN (Internet Corporation for Assigned Names and Numbers) proposed a new top-level domain (TLD) name for pornographic sites—dot-triple-x (.xxx). The suggestion, which would make it far easier to identify adult sites, found favor with no one. Some U.S. politicians argued that the new TLD should be mandatory rather than voluntary. Cyberporn purveyors claimed that .xxx was discriminatory. And some religious leaders decried the creation of a "virtual red light district" on the Internet. Dot-triple-x has been tabled, at least for now. (http://news.com.com/2010-1026_3-5176611.html)

Assessing Available Solutions

School IT managers obviously can't wait for the Federal government, ICANN, or the porn industry to solve a problem that's already reached the crisis point. What they can do, however, is to be certain they understand the strength and limitations of currently available solutions.

The place to start is with the SafeSearch options offered by Google and Yahoo — both because they're so powerful and because of the problems they create. To recap: With SafeSearch on, Google and Yahoo are very effective cyberporn filters for both image and ordinary Web searches. But any student can turn off SafeSearch. Once that's done, other filtering solutions can't block pornographic images and text because both search engines relabel their thumbnail images.

Google and Yahoo can be set for:

- Strict filtering (blocks explicit text and images)
- Moderate filtering (blocks explicit images only)
- No filtering

To see just how good a job Google does, consider a 9th grade class that's given an assignment to write a report on "Asian teens." The Internet-savvy students dutifully plug the search term into Google (set for no filtering) and get 30,700,000 results. The first 100 Web (content) results are all pornographic (100 is an arbitrary benchmark).

Google's Image search returns 31,200 thumbnails. Again, the first 100 are explicitly pornographic.

With moderate filtering enabled, Google's Web search returns the same or similar results. But the Image search delivers 677 results (rather than 30 million). The first 100, with one possible exception, are completely innocent.

Strict filtering returns 633 images. The questionable graphic noted above has vanished, along with 43 others moderate filtering let through. The Web search returns 9,340,000 results, rather than the 30+ million found when filtering was disabled.

What's particularly notable about Google's search algorithm is that even with strict filtering enabled, it's smart enough to discern that a link to "Sex, Drugs, and Rap Music" is actually a resource where "Asian teens get the facts about AIDS."

Shutting Down SafeSearch

Google SafeSearch is enabled from the Preferences page. It's disabled the same way. Unfortunately, both are a matter of a few clicks. There's no way to set SafeSearch as the permanent default.

The one difference for Yahoo is that it offers a password-protected SafeSearch Lock. The shortcoming is that anyone with a Yahoo password can unlock SafeSearch — even if he or she didn't set it.

Yahoo and Google both present a second problem in the way they identify the thumbnails returned when an Image search is performed. Although many of these images are culled from pornographic sites, when

they're displayed on Google and Yahoo they're no longer labeled with their original URLs (or HTML image source tags). Instead, they're assigned either a Yahoo or a Google temporary URL.

This plays havoc with Web filters, all of which use a database of URLs to identify and block content or images coming from a blacklisted site. These filters will block anything coming from www.porn.com. They have no reason to block images labeled www.yahoo.com or www.google.com.

This isn't a filter failure. These tools are doing exactly what they were designed to do, comparing URLs against their blacklists, stopping those that match and forwarding those that don't.

Thus, the double dilemma: Educators have no way of ensuring that Yahoo or Google SafeSearch remains enabled. That means they have to rely on some other filtering solution to block porn, with full knowledge that the filter will let images from Google and Yahoo slip by.

Ironically, Google has come under fire for the way it relabels thumbnails from the cyberporn industry. Perfect 10, an adult Web site, brought a copyright infringement suit against Google in 2004, claiming that the free thumbnails could adversely affect its attempts to sell downloadable images for cellphones. The suit was recently decided in Perfect 10's favor. Google has said it will probably appeal the decision. It also announced it will continue to host other thumbnails on its servers, so the February 2006 ruling isn't going to eliminate the image issue.

The Filth Filters

School IT managers basically have four types of tools that can help them block adult content: PC-based software filters, server-side software filters, third-party integrated solutions, and dedicated filtering hardware.

PC-based filters use password-protected blacklists and keywords, which can be edited by the administrator. In some cases, the purchase price includes a subscription service for updating the blacklist. Typically, these packages offer between 10 and 40 content categories (more categories mean finer-grained filtering).

Desktop packages have grown increasingly sophisticated. A few perform content analysis. One or two implement a rudimentary form of object analysis that attempts to identify porn by evaluating images. A graphic that contains a high percentage of flesh tones, for instance, would likely be screened out. These might be pornographic images; they also could be paintings by Rubens, Titian, or Picasso.

PC-based software is more suitable for home rather than school networks. Software must be installed and configured on every machine. With three or four PCs, that's a minor inconvenience; with 300 or 400, it's a major task. Schools that have thousands or tens of thousands of PCs would likely need a separate staff just to handle setup.

Another major drawback to client-side filtering is that it's resource-bound. These packages rely on the host's PC and memory to perform all tasks. As the number of filters, blacklisted URLs, and keywords increases, so does the load. Effective filtering can impede PC performance, sometimes dramatically.

Finally, there's evidence that PC-based filters are fairly easy to disable. According to a random phone survey of 1,209 respondents between the ages of 15 and 24 conducted by the Kaiser Family Foundation, 60 percent said they know how to get around filtering software or know someone who can show them how. (<http://www.kff.org/entmedia/loader.cfm?url=/commonspot/security/getfile.cfm&PageID=13719>)

Stopping Cyberporn with Servers

Server-side Internet filters also use keywords and blacklists, although the databases behind both are deeper than on their PC-based counterparts.

Increasing the number of keywords doesn't necessarily translate into more effective filtering. Keywords don't take context into consideration. If the word "breast" is blocked, for instance, content in which this word appears will be blocked: regardless if it's cyberporn, recipes, or research about breast cancer.

Further, keyword filtering only applies to content, not to images. So even if an HTML source tag for an image includes the word “breast,” a keyword filter isn’t going to screen it out. (This assumes, of course, that the image is returned by a search engine that doesn’t relabel results, as Google and Yahoo do.)

Blacklists of blocked sites filter both contents and images, either by IP address or URL. IP filtering is easier to set up, but it’s also coarser-grained. If several sites are virtually hosted at the same IP address, and the address is blacklisted, then all the sites will be filtered out regardless of content.

URL filtering can block specific Web pages on a site, so only the offending content is screened out. The critical factor with this approach is the database used to make decisions. With hundreds of new sites coming online daily, educators need to know how often the database is updated and reviewed for accuracy. Can they add and delete URLs as needed, and are their changes permanent — or will they be overwritten by the next update? How responsive is the vendor to requests to make changes to the database?

Equally important is the question: Do human beings maintain the database, reviewing and assessing its contents? Spiders and bots are a great way to identify potentially unsavory sites, but human intelligence is needed if the goal is to filter out porn while permitting access to as many Internet resources as possible.

The Impact of Server-based Filtering

Despite the name, server-based filters can be hosted on firewalls, routers, and (no surprise) servers. Like their PC-side counterparts, this software relies on its host’s horsepower to do its job. As a network grows or as the filter is asked to do more tasks, it needs more processing power — and there’s only one place it can get it.

Granted, servers, firewalls, and routers have more horsepower than desktop machines. But as the network scales up, or as the filter has to juggle more tasks, it still can degrade the performance of its host. If the filtering software is running on a firewall, that sort of cycle stealing can ultimately affect network security. If it’s running on a router or server, network speed and application performance may suffer.

In addition, since server-based software filters share memory with their hosts, there’s a limit to the number of URLs it can retain in its blacklist.

So why not deploy a dedicated server? It’s a matter of money, something that’s typically in short supply at educational institutions. If a second or third server is needed, it takes a bigger bite out of IT resources — both for hardware and for personnel to keep the servers humming along.

Integrated Filtering Solutions

A number of vendors sell an integrated approach to Web filtering and monitoring. In most cases, their proprietary filtering software is implemented on existing platforms, like firewalls, or requires new equipment, such as network access gateways, to be installed.

The former approach really doesn’t address server slowdowns since the vendor’s software, which is typically very sophisticated, still must contend for resources with its host. The latter approach deploys new devices so there’s no drag on servers and firewalls. One of the downsides to this scheme is that each new device increases network complexity and makes management more demanding. Increasing complexity, especially when it also adds to the management burden, presupposes that a school IT staff has the personnel and expertise to keep tabs on this equipment and make sure that failures can be recovered from quickly. Most school IT departments don’t.

Price in general is something to keep a close eye on with integrated solutions. They may offer far more capability than a school network really needs — or can afford.

The good news about integrated solutions is their databases tend to be very deep and content experts usually oversee them. School IT managers need to find out what options are available for bolstering the database with additional terms. Most companies allow customers to “suggest” URLs. Educators need to find out how often these suggestions are taken and how long before they show up in the database.

The final issue is architectural: all of these approaches use a pass-through architecture that requires the filter to “touch” every packet. Outbound requests are reviewed by the filter, which decides whether to block or allow them. If the request is OK’d, it’s passed back to the server’s host, which then forwards it until it reaches the Internet access point. This process is played out in reverse with incoming Web traffic.

The pass-through approach introduces yet another device into the connection path. If the filter is backed up, a bottleneck is inevitable. This can slow network performance, decreasing efficiency and increasing TCO (total cost of ownership). And even if Web traffic is flowing freely, server-based filters are a single point of failure.

Dedicated Filtering Hardware

Standalone Internet filters are the final option. Since these are dedicated devices, they’re equipped with their own CPUs and memory, so there’s no chance they’ll end up stealing cycles from a host. That also means performance remains stable as load increases, rather than degrading, which is one of the drawbacks of software-based filters.

A hardware filter, like its software counterpart, is only as good as its database. In order to make an informed decision, educators need to know how many entries the database contains, how many categories the content is grouped under, how often the database is updated, and to what degree human operators evaluate the URLs captured by spiders, bots, and other mechanisms.

School IT managers also need to ascertain how easily and efficiently they can alter the database, adding URLs to the blacklist and whitelisting others. Since this appliance is going to be used in an educational environment, they also should be able to define profiles for various user groups. Teachers, for instance, might have unlimited access (although their searches may be monitored). Junior high-school students would be assigned another level of permissions, and fourth graders yet another.

One key feature of a hardware-based solution is that it implements pass-by, rather than pass-through, technology. What this means is the filter sits outside the flow of traffic, monitoring Web requests and results and comparing them against its database. When it spots a match, it signals the Web server to kill the request and sends a block page to the offending PC. Thus, the pass-by filter can’t become a bottleneck or introduce a single point of failure.

Even with all these capabilities, though, an industrial-strength standalone filter can be outflanked as easily as a desktop package if a student turns off Google/Yahoo SafeSearch.

The 8e6 Answer

There is one filtering appliance that can lock down both Google and Yahoo SafeSearch so it remains in force, even if a student clicks the button that’s supposed to disable it: 8e6’s R3000 family of Internet Filters.

The R3000 enables school IT managers to set Google’s and Yahoo’s strict filtering, moderate filtering, and no filtering on any PC on their network—from a browser-based graphical user interface. They can alter settings just as easily, tightening the filters on a predefined group of machines when, for instance, 5th grade students are scheduled to follow the 12th grade advanced computer lab. To make things even simpler, all settings can be programmed to a specific time of day, so there’s no need for manual intervention.

The R3000 is a complete Internet filtering solution available for the education market. It lets school IT managers take full advantage of Google/Yahoo powerful SafeSearch capabilities, while giving them a state-of-the-art appliance loaded with advanced filtering capabilities.

Think of it this way: Once Google/Yahoo SafeSearch is locked down, the search engine uses its proprietary algorithms to filter thumbnails. Meanwhile the R3000 opens up powerful new filtering possibilities.

Since the R3000 employs noninvasive pass-by filtering technology, it keeps tabs on Internet traffic without having to interfere with every packet. It blocks or redirects only those that must be filtered out of the data stream. This highly efficient approach results in a robust, highly scalable solution that can support up to 30,000 users per appliance.

The R3000 also recognizes and filters a comprehensive range of Internet protocols, including URL and IP addresses, FTP, HTTP/HTTPS, and NNTP (newsgroup). It also can be programmed to trigger on a wide range of file types, including MP3, JPEG, and MPEG — so there's no chance that students will use school equipment to download pirated music or movies.

8e6 knows that an Internet filter stands or falls by its database, which is why it has combined artificial intelligence and human expertise to ensure comprehensive, up-to-date coverage. High-speed AI utilities collect Web sites on a 24/7 basis, so there's virtually no chance a URL that should be blocked gets by. To further refine its information, highly trained specialists individually verify and categorize each site. As a result, 8e6 has built a database of several million sites, organized into 90 overarching categories. In addition, IT and security managers can add or delete sites as needed.

When the R3000 is deployed with the standalone Enterprise Reporter, the only Internet reporting appliance, school IT managers can monitor exactly how their network is being used — or abused. For example, the Enterprise Reporter can generate detailed or summarized reports that match IP addresses with user names and activity. That makes it easy to spot the students who “accidentally” keep trying to access cyberporn.

Monitors also can be set to track sites visited in a particular category, number of pages viewed, amount of time spent on a specific site, and users who spend the most time on a specific site, to name only a few options. What's more, reports can be generated in a variety of formats, including bar and pie charts. When viewed online, all reported URLs are clickable for instant viewing, and data points can be drilled down to access more information.

8e6 also recognizes that IT resources need to be protected as much as networks and students. That's why it prides itself on delivering turnkey solutions that free educators to deal with the cyberporn threat without getting their hands dirty.



For more information on 8e6 Technologies and 8e6 appliance-based solutions for Internet Filtering and Web-use Reporting, visit www.8e6.com.