Global Information Assurance Certification Paper

Internet Content Filtering
Carol Woody
January 9, 2002

Background
Ease of access and removal of the physical limitations of paper have led to an expanded focus on the Internet as a content source. The challenge to all organizations establishing access to this source is the inability to limit use of these capabilities to the specific content areas appropriate for the organization. In addition to entertainment opportunities for news, music, movies, and shopping which can consume bandwidth along with employee time, there is potential access to controversial content (i.e. gambling and pornography) that carries legal restrictions in many geographic areas, as well as illegal content (i.e. obscene material and child pornography).

The business can be held liable for the illegal actions of employees using business facilities unless enforced company policies address inappropriate use. The Supreme Court in *Reno v. ACLU* declared that the local community standard is what must be applied to define the line between pornography and obscenity. This line is critical because pornography is protected by the Constitution under the First Amendment freedom of speech but not obscenity (Lane III). With the differences in regional perspectives, it is possible actions considered legal but controversial in one location will be illegal at another site. Without due care, the risk of exposure to illegal material is high. The online pornography industry is estimated to build to an annual level of $2 billion by the end of 2001. Its advertising dollars have heavily financed many of the available Internet search sites such as Yahoo! (Lane III). SexTracker, a Web service that monitors adult sites reports 26,000 active web locations with as many as 60 million unique visitors a day (Webb).

Organizations have been surprised by the amount of "surfing for fun" that employees choose to perform. Vault.com research established the following breakdown of surfing habits in a recent survey of 451 employees:

During an average workday, how much time do you spend surfing non-work-related sites?

9.6% Never
18.4% Up to 10 minutes
25.1% 10-30 minutes
22.4% 30 minutes to 1 hour
11.9% 1-2 hours
12.6% Over 2 hours

And what do they say they are doing?

Reading the news – 72%
Making travel arrangements – 45%
Making purchases – 40%

Job search – 37%
Visiting special interest sites – 37%
Checking stocks – 34%
Coordinating social events – 28%
Instant messaging friends – 26%
Downloading music – 13%
Playing computer games – 11%
Chat room surfing – 9%
Visiting pornographic sites – 4%

Employee monitoring raises issues of privacy invasion and harassment if not handled appropriately though consistently enforced policies. Management must carefully balance the needs of the organization with the rights of the employees to establish an effective outcome. Privacy Foundation estimates over a third of the online workforce is subject to some form of monitoring and suggests that the low cost of applying monitoring is the reason for the high volume. Telephone monitoring is less prevalent, but considered just as subject to inappropriate business use (Schulman).

Acceptable Use Policy

At a minimum, organizations must protect their actions through the implementation of acceptable use policies (AUP), which are enforced. The AUP (also called an Internet Use Policy or IUP) codifies guidelines for online communication with specific sanctions for inappropriate use. Without these in place, the organization has limited options of recourse against employees that make inappropriate use of employer provided facilities. In addition, the organization runs the risk of shared liability for inappropriate employee actions that are deemed illegal by local authorities. SurfControl, a content filtering software provider, has a sample AUP that provides a starting point available for downloading at **http://www.surfcontrol.com/resources/business/acceptable_use_policy/index.html**.

While much of the sexually explicit material available on the Internet may not be illegal, organizations that do not define appropriate use run the risk of employee lawsuits related to sexual harassment and workplace intimidation (Cohen). Management that chooses to monitor employees without an AUP in place runs the risk of employee lawsuits related to invasion of privacy. Congress has considered passage of the Notice of Electronic Monitoring Act (NEMA) that will require companies to tell new hires what information is gathered and stored, how it is used and who knows about it. Management will be required to share this information with all employees annually (Vault.com).

Content Filtering

Some companies chose to block sites rather than trust employees not to be tempted. This puts the company into a difficult monitoring role because filtering content slows down the network and no filtering software works effectively all the time. The performance impact may be less of an issue as organizations implement filtering to handle denial-of-service monitoring needs (Householder). An email article from InformationWeek poses the possibility that pornography addiction may become a disability subject to the Americans with Disabilities

Act protection (Soat). In such a case, filtering would be available as an appropriate deterrent to allow the employee to continue to be productive.

The application of content filtering requires two components: (1) a rating to be applied to each Internet address and (2) a filter module that uses the rating to determine whether to grant or block access to a site selection. Ratings can be assigned to the site through self-inclusion or via third party. Also ratings can be created within the filter module based on keywords that are considered objectionable. The filtering module can reside on the local machine, which is how a parent would monitor their children's access on the home computer; on a server at the organizational level using either a firewall or router location to initiate the screening; or at the Internet Service Provider level as is currently implemented by AOL.

The World Wide Web Consortium developed an open standard called PICS (Platform for Internet Content Selection) that provides a means for sites to incorporate a self-selected electronically readable rating within each web page. The PICS standard "establishes Internet conventions for label formats and distribution methods, while dictating neither a labeling vocabulary nor who should pay attention to which labels. It is analogous to specifying where on a package a label should appear, and in what font it should be printed, without specifying what it should say" (PICS). Rating options within the PICS standard are limited and do not offer context flexibility to separate nudity in pornography from nudity in art or objectionable violence in hate material from violent content in news reporting. Self-rating is subject to accidental or deliberate misinterpretation without penalty for errors.

There is much room for confusion in the implementation of the PICS standard. Use of stand-alone blocking software such as Net Nanny or CyberPatrol for the filtering module will allow the installer to select specific words to trigger a browser to block a site based on the site name or use of the selected word in a content page. Selected word blocking can also occur based on content in an ad appearing on the web page which may be unrelated to the actual content of the page and change with each individual request (Sobel).

The installer chooses how to inform the requestor of the blocking decision by blocking the full page or only blocking out the offending words. The latter approach can produce some very amusing responses. For example the poem "Owl and the Pussy Cat" and the nursery rhyme "pussycat, pussycat" suffer major revisions when the slang term "pussy" is blocked (TIFAP).

Implementation of a third-party rating system is done by categories created by the rating group (violence, profanity, sex, nudity, etc.) and screening is based on blocking some or all of the categories as defined at the filtering module. A web site is assigned a category by the rating group which is stored within a proprietary database maintained by the rating group and queried by a site based filtering module to make each access decision. Human researchers who check each web site and assign a category build the database. Sites are periodically rechecked for validity and new sites are added constantly. The vendor to initially categorize sites sometimes uses automation, but human review to confirm the designation is still required.

Vendors that automatically classify sites continue to appear; the cost savings is huge since it takes many resource hours to manually evaluate web pages and update current solutions, but none of the automated options have emerged as market winners so far.

Filtering Providers

Many of the older monitoring and blocking options were developed in response to parental concerns of exposing their child to unacceptable content. Independent rating groups such as United Federation of Child Safe Web Sites offer a ICCS Certified Child Safe rating for sites that submit their content to the group for review (Sobel). A quick search for "Internet filtering" using MSN Search brings up the link for "Internet Filtering Products for Business" (Group1Internet A.M.). This site provides links to a vast array of products organized for varying business sizes.

The 2001 leader in the market place for selling monitoring software is by reported sales dollars is Websense (www.websense.com) claiming 256 of the Fortune 500 as customers. This product can be incorporated in a firewall, router, proxy server, application server, or load balancer to identify and optionally block out content from the Internet. Options are available for restrictions based on time of day, employee role, rating category, or keyword. Websense has constructed the following groupings for web pages, which is accompanied with a strong disclaimer about liability for a web page assignment: abortion advocacy, advocacy groups, adult material, business & economy, drugs (based on US laws), education, entertainment, gambling, games, government, health, illegal/questionable, information technology, internet communication, job search, militancy/extremist, news & media, productivity management, bandwidth management, racism/hate, religion, shopping, society & lifestyle, special events, sports, tasteless, travel, vehicles, violence, and weapons.

The Websense master database provides category assignments for 600 million web pages with updates for an additional 25,000 weekly. Each filtering point requires installation of an appropriate filtering module and access to a copy of the database, which must be maintained frequently to stay current. Category assignments are made through a combination of automated and manual procedures. The vendor claims to have covered "the most frequently accessed sites on the Web." These assignments are built manually and the vendor offers a procedure for web page owners to dispute the category decision.

Many firewall and router packages can accept rules to evaluate packets for content and screen them from availability eliminating the need for purchase and maintenance of additional software. Appropriate implementation of screening becomes the responsibility of already overworked network support staff. Inappropriate filtering is easily included and the rules are hidden from the general user lending credibility to a censorship attack. The reported processing impact on the network has not been acceptable resulting in sluggish download times in experiments on high volume environments (TIFAP).

Un-rated Sites
Steve Lawrence and Lee Giles of NEC Institute researchers identified 320 million " pages as the 1999 size of the Web (Lawrence & Giles). As of November 2001 another source reported over 16 million hosts within 95 top-level domains and an unknown number of web pages (RIPE NCC). Estimates of growth range from a top rate of 1000% in 1994 down to low rate of 100% (Coffman). The volume of information available on the Internet and the global expansion pace makes it impossible for any third party source to classify every site leaving vast amounts of content without a rating.

Early options based on ratings required a decision to accept or reject un-rated pages unconditionally. More recent implementations allow for key word screening to remove unacceptable language.

None of the filters currently available address images. At best, they rely on the text file name of the image to provide sufficient description to evaluate the contents. There is much research underway seeking to identify appropriate ways to handle images. Junichi Tatemura has constructed the Web Graphics Navigator (http://graphics.media.iis.u-tokyo.ac.jp/webgnavi.cgi?uid=UYVlsudgdQdXEPHo&lang=en) as an analysis tool to experiment with filtering-by-example capabilities, but none of this is ready for extensive use.

Filtering Reviews
Research has shown that all existing implementations of content filtering exhibit many inaccuracies. The Electronic Privacy Information Center (EPIC) reported the conclusion that available filtering was faulty based on the wide difference between the number of matching documents returned by each filtered search versus unfiltered searches using single word search requests (Sobel). A recent test conducted by Consumer Reports comparing a selection of popular filtering options reported a failure rate of 14 – 23% in blocking against objectionable sites (Magid).

Many of the category designations are subjective. The U.S. Anti-Defamation League (ADL) has issued software called HateFilter to block anti-Semitic, racist, and extremist Web sites. Cyber Patrol was attacked by censorship foes for blocking all homosexual references in 1998 and had to rewrite their software filtering module to include previously blocked sites. The Commission on Online Child Protection issues a warning on their web page that the "technology raises First Amendment concerns because of its potential to be over-inclusive in blocking content. Concerns are increased because the extent of blocking is often unclear and not disclosed" (COPA).

The Internet Filtering Assessment Project (TIFAP) ran from April to September 1997. This volunteer project involved forty librarians testing seventeen products in an effort to understand how the filtering processes worked within a library setting. While their approach was not scientific, the results provide a broad perspective on the use of filtering in an information intensive environment. Their results show that keyword blocking does not work without appropriate adjustments for the needs of each site. Keyword blocking impacted

access to nursery rhymes and simple words such as "button." Category selection was frequently too broad to provide effective filtering of offensive materials; filters perform badly blocking needed content in 35% of the searches. The conclusion is that individual evaluation is still needed a great deal to effectively differentiate appropriate content from inappropriate materials.

Current Status

Organizations must consider the need for clearly defining to employees why Internet access is available and the policy applicable for filtering to counter attacks of unreasonable censorship. The filtering process and results must match this limitation structure or employers risk restricting workers from information needed to perform their job.

Employee monitoring of Internet use is awkward at best. An acceptable use policy is critical to defining what is acceptable and providing a means of justification for monitoring or filtering. The use of software to filter content must be carefully applied to avoid censorship of critically needed materials, and current vendor offerings may be too limiting for your organization.

Resources must be applied to establish and maintain the filtering process. Bandwidth must be applied to review each requested packet and screen out offensive material. If screening is based on vendor-supplied categories, updates will be required on a daily or weekly basis to keep pace with Internet growth.

Reference List

Coffman, K.G., Odlyzko, A. (2001). *The Size and Growth Rate of the Internet.* http://www.firstmonday.dk/issues/issue3_10/coffman/.

Cohen, Alan. No Web For You!. *FSB, October 2000.*

COPA, Commission on Online Child Protection, http://www.copacommission.org/report/filteringblocking.shtml

Group1Internet A.M. http://www.group1iam.com/cgi-bin/iam/index.html

Householder, A., Manion, A., Pesante, L., Weaver, G.M., (2001) *Managing the Threat of Denial-of-Service Attacks,* Carnegie Mellon University.

Lane III, F. S. (2000). *Obscene Profits*. New York: Routledge.

Lawrence, S., & Giles, L. (1999). *How Big is the Web?* http://www.neci.nj.nec.com/homepages/lawrence/websize.html

Magid, L. (2001). *Rating the Internet filtering programs.* http://www.larrysworld.com/articles/sjm_filtering.htm

Privacy Foundation, http://www.privacyfoundation.org/workplace/technology/extent.asp

PICS Standard, http://www.w3.org/PICS/iacwcv2.htm

RIPE NCC. http://www.ripe.net/statistics/hostcount/2001/11/index.html

Schulman, A. (2001). *The Extent of Systematic Monitoring of Employee E-mail and Internet Use*. http://www.privacyfoundation.org/workplace/technology/extent.asp

Soat, J. (2002). *John Soat: Confidential,* InformationWeek BetweenTheLines, email distribution January 6, 2002.

Sobel, D. L. (Ed.). (1999). *Filters & Freedom* (First ed.). Washington, DC: Electronic Privacy Information Center.

TIFAP. Learning from TIFAP, http://www.bluehighways.com/tifap/learn.htm

United States Congress (1998). *Internet Filtering Systems: Report of the Committee on Commerce, Science, & Transportation*. Washington, DC: United States Congress US GPO.

Vault.com. http://www.vault.com/surveys/internetuse2000/results2000.jsp;jsessionid=0urqaiykr2d10a3296?results=2&image=employee. http://www.vault.com/nr/newsmain.jsp?nr_page=3&ch_id=420&article_id=19332&cat_id=1422

Webb, G. (2001). Sex and the Internet. *Yahoo!, May 2001,* 88-95,136-137.

Websense, www.netpart.com/products/about/wse/index.cfm.

Tatemura, J., Santini, S., Jain, R., Social and Content-based Information Filtering for a Web Graphics Recommender System. Proceedings of the 10th International Conference on Image Analysis and Processing (ICIAP). http://www.computer.org/proceedings/iciap/0040/00400842abs.htm

Last Updated: September 25th, 2017

Upcoming Training

Community SANS Sacramento SEC401	Sacramento, CA	Oct 02, 2017 - Oct 07, 2017	Community SANS
SANS DFIR Prague Summit & Training 2017	Prague, Czech Republic	Oct 02, 2017 - Oct 08, 2017	Live Event
SANS October Singapore 2017	Singapore, Singapore	Oct 09, 2017 - Oct 28, 2017	Live Event
SANS Phoenix-Mesa 2017	Mesa, AZ	Oct 09, 2017 - Oct 14, 2017	Live Event
SANS Tysons Corner Fall 2017	McLean, VA	Oct 14, 2017 - Oct 21, 2017	Live Event
CCB Private SEC401 Oct 17	Brussels, Belgium	Oct 16, 2017 - Oct 21, 2017	
SANS Tokyo Autumn 2017	Tokyo, Japan	Oct 16, 2017 - Oct 28, 2017	Live Event
SANS vLive - SEC401: Security Essentials Bootcamp Style	SEC401 - 201710,	Oct 23, 2017 - Nov 29, 2017	vLive
SANS San Diego 2017	San Diego, CA	Oct 30, 2017 - Nov 04, 2017	Live Event
SANS Seattle 2017	Seattle, WA	Oct 30, 2017 - Nov 04, 2017	Live Event
San Diego Fall 2017 - SEC401: Security Essentials Bootcamp Style	San Diego, CA	Oct 30, 2017 - Nov 04, 2017	vLive
SANS Gulf Region 2017	Dubai, United Arab Emirates	Nov 04, 2017 - Nov 16, 2017	Live Event
Community SANS Vancouver SEC401^	Vancouver, BC	Nov 06, 2017 - Nov 11, 2017	Community SANS
Community SANS Colorado Springs SEC401~	Colorado Springs, CO	Nov 06, 2017 - Nov 11, 2017	Community SANS
SANS Miami 2017	Miami, FL	Nov 06, 2017 - Nov 11, 2017	Live Event
SANS Paris November 2017	Paris, France	Nov 13, 2017 - Nov 18, 2017	Live Event
SANS Sydney 2017	Sydney, Australia	Nov 13, 2017 - Nov 25, 2017	Live Event
SANS San Francisco Winter 2017	San Francisco, CA	Nov 27, 2017 - Dec 02, 2017	Live Event
Community SANS St. Louis SEC401	St Louis, MO	Nov 27, 2017 - Dec 02, 2017	Community SANS
Community SANS Portland SEC401	Portland, OR	Nov 27, 2017 - Dec 02, 2017	Community SANS
SANS London November 2017	London, United Kingdom	Nov 27, 2017 - Dec 02, 2017	Live Event
SANS Khobar 2017	Khobar, Saudi Arabia	Dec 02, 2017 - Dec 07, 2017	Live Event
SANS Austin Winter 2017	Austin, TX	Dec 04, 2017 - Dec 09, 2017	Live Event
Community SANS Ottawa SEC401	Ottawa, ON	Dec 04, 2017 - Dec 09, 2017	Community SANS
SANS Munich December 2017	Munich, Germany	Dec 04, 2017 - Dec 09, 2017	Live Event
SANS Bangalore 2017	Bangalore, India	Dec 11, 2017 - Dec 16, 2017	Live Event
SANS vLive - SEC401: Security Essentials Bootcamp Style	SEC401 - 201712,	Dec 11, 2017 - Jan 24, 2018	vLive
SANS Cyber Defense Initiative 2017	Washington, DC	Dec 12, 2017 - Dec 19, 2017	Live Event
SANS Cyber Defense Initiative 2017 - SEC401: Security Essentials Bootcamp Style	Washington, DC	Dec 14, 2017 - Dec 19, 2017	vLive
SANS Security East 2018	New Orleans, LA	Jan 08, 2018 - Jan 13, 2018	Live Event
SANS Amsterdam January 2018	Amsterdam, Netherlands	Jan 15, 2018 - Jan 20, 2018	Live Event