South African Journal
of Information Management
[DoE accredited]

Peer Reviewed Article Vol.4(4) December 2002

Controlling Internet abuse through effective content filtering: a higher education implementation

C.W. Rensleigh
Department of Information Studies
Rand Afrikaans University
cr@rau.ac.za

Contents

1. Introduction
2. Computer systems and Internet services abuse
3. Arguments for Internet content filtering
4. Defining Internet content filtering
5. Implementing Internet content filtering in a higher education library environment
6. Internet services abuse
7. Possible future developments
8. Conclusion
9. References

1 Introduction

The Internet is a rich and educational resource for information, ideas and entertainment. No other medium has provided society with so much information so easily. At the same time, the Internet has raised concerns about privacy and access to information.

The tremendous benefits associated with the Internet have in recent years propelled information resources in the higher education library environment away from the traditional paper-based material to on-line content (Willson 2000:196). Today's students, using higher education Internet infrastructures, have the world at their fingertips. For many academic courses, access to the Internet is a prerequisite and compulsory as a medium to communicate and have access to the required information resources.

Advances in technologies are making information resources available in these ever increasing 'wall-less' environments (terms such as virtual universities or virtual libraries are being used) that in turn offer unique opportunities to share and distribute knowledge and information on a world-wide scale (Furnell 1997:61). On the other hand, managing student access to these on-line information resources are becoming increasingly challenging (Powell 1997).

In this article is a discussion on the need and necessity of Internet content filtering mechanisms and in particular the implementation of such a content filtering solution in a higher education library environment to better manage and increase the use of student Internet workstations.

top

2 Computer systems and Internet services abuse

2.1 Computer abuse

Computer abuse can be defined as 'any intentional act associated in any way with computers where a victim suffers or could have suffered a loss, and a perpetrator made or could have made a gain' (Parker 1998:333, as quoted by Lee 2002:61). It includes all crimes against hardware, software, data and computer services (Straub 1990:46). As networked systems have grown and matured, so too has the nature of abuse within the networked environment. In the earlier days of computing, abuse was largely restricted to fraud and theft related activities, which simply represented the extension of traditional crimes into the electronic environment. However, as time progressed, new and more advanced forms of abuse have emerged (e.g. computer viruses, invasion of privacy and hacking) (Furnell 1997:62).

2. 2 Internet abuse

A 1999 study by the American Management Association found that more than 50% of all Internet activity taking place within companies is not business-related, causing billions of dollars a year in lost production (Greengard 2000:22). In a 2000 study by the Saratoga Institute, it was found that nearly two-thirds of USA firms have disciplined employees for Internet abuse, and slightly less than a third have actually terminated employment (Greengard 2000:22).

Internet abuse in the workplace includes, but is not limited to, accessing sites that are not work related, e-mail abuse, on-line chatting, gaming, investing, shopping, and downloading programs of personal interests, such as MP3s and movies. It also includes using the Internet too often at work, which is commonly referred to as 'cyberslacking' (Siau 2002:75). Ultimately, these abuses refer to employees being on-line at work and not doing work-related tasks.

For the purpose of this article, Internet abuse is classified into the 10 categories defined in Table 1, as listed by Siau (2002:76).

Table 1 Different categories of Internet abuse

General e-mail abuse Includes spamming, harassments, chain letters, solicitations, spoofing, propagations of viruses and worms, and defamatory statements.	**Hacking** Hacking of Web sites, ranging from denial of service attacks to illegally accessing organizational databases.
Unauthorised usage and access Sharing of passwords and access into networks without permission.	**Copyright infringements or plagiarism** Using illegal or pirated software. Copying of Web sites and copyrighted logos.
Transmission of confidential data Using the Internet to display or transmit confidential documents.	**Pornography** Accessing, displaying, distributing and surfing of sexually explicit sites from

	work.
Leisure use of the Internet Surfing the Internet, which include shopping, sending e-cards and personal e-mail, gambling on-line, chatting, auctioning, stock trading and other personal activities.	**Download or upload that is not work related** Propagation of software that ties up organizational bandwidth. (Programs like Gnutella, Napster and Kazaa allow the transmission of movies, music and graphical material.)
Newsgroup postings Posting of messages on various topics that are not work related.	**Moonlighting** Using office resources such as networks and computers to conduct personal business.

top

3 Arguments for Internet content filtering

The major reasons for the implementation of Internet control software appear to be concerns about 'access to potentially objectionable material', commonly understood as sexually explicit material (Willson 2000:198). Controlling access to the Internet by means of filtering software has become a growth industry in the USA and elsewhere. Its use has increased as a mandatory response to the current plagues of society, namely pornography, violence, hate and, in general, anything seen to be unpleasant or threatening (Rosenberg 2000).

The current preferred method of choice to limit access is to filter content either by blocking access to specific Web sites (URL filtering) or by using a large set of keywords to prevent access to sites that contain one or more of these words. Some argue that the excessive Internet usage is a recent phenomenon with addictive elements for the search for stimulation through interactive services and an escape from real-life difficulties (Armstrong 2000:538).

As computer technology and the Internet are rapidly dispersed, e-generation employees are encountering more unethically attractive situations than ever when using a computer (Lee 2002:57). Internet abuse has a direct negative impact on the following:

3.1 Bandwidth loss

Surfing that is not work related congests digital pipelines. Heavy graphics, video clips, video and audio streaming and in particular peer-to-peer file sharing are notorious for clogging bandwidth.

3.2 Productivity loss

The very same business tools that contribute to increased productivity of the workforce can be used as a mechanism for 'cyberslacking'. Internet use that is not work related is costing industries billions.

Using third party filtering and blocking software is a significant means of controlling and limiting these losses. The key argument for imposing controls is to prevent access to potentially objectionable material and to curb activities that are not work related (Willson 2000:199).

top

4 Defining Internet content filtering

The Computer Professionals for Social Responsibility (CPSR), an organization concerned with a variety of social issues associated with the use of computers, defines content filtering as 'one or more pieces of software that work together to prevent users from viewing material found on the Internet' (Hocheiser 1998). This process consists of two components:

4.1 Rating

Value judgments are used to categorize Web sites based on their content. These ratings could use simple allowed and disallowed distinctions like those found in programs like CyberSitter or NetNanny, or they can have many values, as seen in ratings systems based on Platform for Internet Content Selection (PICS).

4.2 Filtering

With each request for information, the filtering software examines the resource that the user has requested. If the resource is on the 'not allowed' list, or if it does not have the proper PICS rating, the filtering software tells the user that access has been denied and the browser does not display the contents of the Web site.

Another definition, by the American Library Association (ALA), describes blocking or filtering software as a mechanism used to restrict access to Internet content (American Library Association 2000):

- Based on an internal database of the product
- Through a database maintained external to the product itself
- To certain ratings assigned to those sites by a third party
- By scanning text, based on a keyword or phrase or text string
- By scanning pixels, based on colour or tone
- Based on the source of the information

Internet content filtering software tends to take two primary forms, namely:

4.3 Text screening

One form screens documents before allowing access. If the screening detects forbidden words, such as 'pornography' or 'sex', access is denied. Text screening evaluates content based on the presence or absence of forbidden terms. This technique is needed when recognizing changes in context: for example, text-based filters may block the downloading of files containing information on breast cancer because the document contains the word 'breast'.

One way to alleviate the severity of the automatic blocking rule is to purge the offending terms, but this can result in the changing of the context, for instance removing the word 'homosexual' from a sentences such as 'Traditionalists oppose homosexual marriage' will result in 'Traditionalists oppose marriage', which is completely out of context (Balkin, Noveck, and Roosevelt 1999).

Text-based filtering software is very exact (include or exclude) but is being enhanced to include more sophistication to try and capture the contextual meaning by considering factors such as the repetition and proximity of forbidden terms. But it seems that context evaluation is too complex for a mere mechanical evaluation.

4.4 Evaluation

Third-party evaluation

This approach relies on third-party organizations to evaluate content by inspection and then

generating lists of acceptable (white-listed) and unacceptable (black-listed) Web sites. The filtering software can then either restrict access to the unacceptable sites or allow access to only the acceptable sites. These lists are continuously updated. Given the dynamic growth of the Internet, the content filtering software is required to regularly download the updated banned Web site list from the Internet. In addition to the third-party lists, system administrators can add newly discovered unacceptable sites to the banned list.

Self-evaluation
Another approach is to encourage or even require every Web site to rate their content along several dimensions, including violence, language, sexual explicitness and nudity. The Web site rating system is similar to systems used to rate music on CDs, movies and television shows.

A system administrator can create a profile that characterizes a set of numerical ratings for acceptable sites. When a Web site's ratings exceed the profile on any dimension, the content filtering software will block access to that Web site. For this system to work, it is necessary that Web sites rate themselves; a default condition must be that non-rated sites are automatically blocked (Rosenberg 2000).

top

5 Implementing Internet content filtering in a higher education library environment

With the shrinking budgets of higher education libraries and greater demands on existing campus information technology infrastructure, it becomes increasingly important to professionally manage these resources and to maximize the use for the intended purposes. Many higher education libraries have student Internet access from within the library. Predominantly, these student Internet workstations are dedicated to access to academic information resources such as electronic databases and online public access catalogues (OPACs), as apposed to general or open access student computer laboratories where students can use the workstations for diverse purposes like typing assignments, e-mail, chats, SMS, etc.

Unfortunately, the temptation will always be there to iniquitously use and abuse these facilities. The abuse ranges from Internet surfing that is not academic related to maliciously crippling (hacking) the system and services (Siau 2002:76). In most cases, an acceptable Internet use policy (AIUP) alone, if not strictly enforced, is not enough to deter system abuse. Most academic libraries impose some form of control on student access to the Internet and some have implemented controls on employee access as well.

Finding solutions for these predicaments should be of high priority, as Lee (2002:57) points out that the frequency of computer abuse and the losses associated with such abuse are expected to grow due to highly sophisticated and educated abusers armed with knowledge about the latest information technology.

top

6 Internet service abuse

The Rand Afrikaans University (RAU) campus is experiencing a growth in its student population. This growth is increasingly placing pressure on the existing student computing facilities. Currently there are a number of computer laboratories where students can do anything from assignments to private surfing and e-mail. The problem is that these facilities are so in demand that there tend to be queues of students waiting to make use of them.

It did not take long for students to realize that the workstations in the library could be used for non-academic purposes. The popularity of the workstations had to be controlled, as those wanting to search for information on academic information resources had to queue. A solution had to be found.

6.1 Regulating Internet access

One solution is regulating Internet access. Content filtering and blocking software is an acceptable way of addressing the concern of library users accessing illegal and inappropriate and non-academic material (Rosenberg 2000; Willson 2000:199). One of the easiest ways to regulate content is to set up and use Microsoft Internet Explorer's content advisor. For more advanced content management, there is a variety of commercial software available. The Microsoft Explorer's approach was used to develop a rating system where the sensitivity for four categories (language, nudity, sex and violence) can be set. Additionally, all sites can be blocked and only approved sites can be accessed. The set-up of the content advisor has to be done on each workstation accessing the Internet.

The amount of administration on each workstation each time an additional URL was added warranted an investigation into a centralized administration system. The investigation led to the next step where an inline proxy and content filtering system was implemented to centrally manage the Internet content that can be viewed on all of the student workstations (Rensleigh 2002).

The student workstations in the library were configured with different Internet configurations to perform dedicated functions. The content filtering software made the necessary different configurations. The different student Internet access configurations are:

Exclude all URLs that are off campus: Student workstations do not have open Internet access, but only access to services situated on the RAU campus (intranet), for example OPAC. Access to off-campus Web sites are blocked.

Exclude all URLs except the URLs specified: Student workstations have access to only certain Internet sites, for example only the electronic databases that RAU subscribe to. In other words, these student workstations are only able to use the five or ten Web sites specified. Access to other Web sites are not available they are blocked.

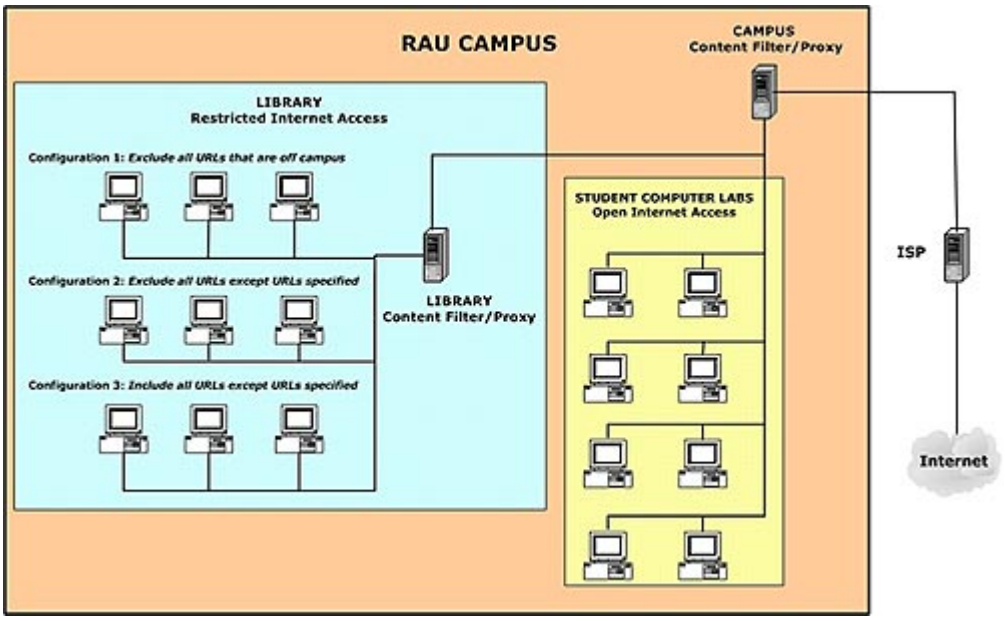
Include all URLs except the URLs specified: Student workstations have Internet access with access to the whole Web but can only accept certain content. In other words, these workstations can access all the Internet Web sites except the ones that were specified by the administrator. Typical sites that were added by the administrator include: e-mail, chats, SMS, etc.

On top of the different configurations, which are defined by the system administrator, filtering of content takes place, based on certain words appearing in the Web sites (text screening), such as pornography and racism. This facility is part of the system and when enabled is updated on a regular basis. The administrator can add additional lists of words that the system must filter out. In addition, the system can also point out which sites are being visited most, how often and for how long.

The library content filtering system implemented is independent of the campus proxy and content server and is used in a proxy chain configuration and does not interfere with the campus proxy and content filtering. It is an addition on top of the more general system. One of the main benefits of this system configuration is that it is administered from within the Library Information Technology (IT) department.

The content filtering proxy system ensures that the student workstations are used efficiently and effectively for their intended purpose: access to academic information resources. In Figure 1 is a layout of the configurations.

Figure 1 Layout of the RAU Library Internet configuration



top

7 Possible future developments

Now that the Internet abuse on the library student workstations has been brought under control, a logical next step would be the filtering of the library employee Internet activities. Currently all staff have open access to the Internet at any time and any place. Certain staff members do not need to have access to the Internet other than e-mail. The current Internet content filtering system can be extended and configured to block access to the Internet by default and only open access for certain periods, such as during lunch time, coffee breaks and after and before working hours.

As for most organizations, content filtering and the controlling of access to the Internet will yield increased productivity and merit careful consideration.

top

8 Conclusion

Today's business environments are increasingly dependent on the Internet. Surveys have indicated that employees tend to abuse their Internet privileges, thus having a negative impact on productivity and Internet bandwidth. The high price of South African bandwidth necessitates that the access to and the content viewed from the Internet be controlled and properly managed. This is applicable not just to businesses, but also to the higher education sector.

The academic library environment is in a process of change. It is inevitable that the information technology infrastructure plays an increasingly important role in delivering information services to clients. Effective management of information and communication technology (ICT) resources maximizes use and increases competitiveness.

top

9 References

American Library Association. 2000. Statement on library use of filtering software. [Online]. Available WWW: http://www.ala.org/alaorg/oif/filt_stm.htm. (Accessed 9 November 2002).

Armstrong, L., Phillips, J. and Saling, L.L. 2000. Potential determinants of heavier Internet usage. International Journal of Human-Computer Studies 53(4):537–550.

Balkin, J.M., Noveck, B.S. and Roosevelt, K. 1999. Filtering the Internet: a best practice model. [Online]. Available WWW: http://www.copacommission.org/papers/yale-isp.pdf. (Accessed 9 November 2002).

Furnell, S.M. and Warren, M.J. 1997. Computer abuse: vandalizing society. Internet Research: Electronic Networking Applications and Policy 7(1):61–66.

Greengard, S. 2000. The high cost of cyberslacking. Workforce 2000 79(12):22–24.

Hochheiser, H. 1998. Filtering FAQ, Computer Professionals for Social Responsibility (CPSR), [Online]. Available WWW: http://www.cpsr.org/filters/faq.html. (Accessed 9 November 2002).

Lee, J. and Lee, Y. 2002. A holistic model of computer abuse within organizations. Information Management and Computer Security 10(2):57–63.

Parker, D.B. 1998. Fighting computer abuse – a new framework for protecting information. New York: John Wiley and Sons.

Powell, A. and Gillet, M. 1997. Controlling access in the electronic library. In Ariadne. [Online]. Available WWW: http://www.ariadne.ac.uk/issue7/access-control/. (Accessed 9 November 2002).

Rensleigh, C. and Swanepoel, B. 2002. Managing student Internet workstations as part of an information infrastructure: an academic library perspective. The 2002 Conference on Information Technology in Tertiary Education (CITTE): Durban.

Rosenberg, R.S. 2000. Controlling access to the Internet: the role of filtering. [Online]. Available: http://www.copacommission.org/papers/rosenberg.pdf. (Accessed 9 November 2002).

Siau, K., Nah, F.N. and Teng, L. 2002. Acceptable Internet use policy. Communication of the ACM 45(1):75–79.

Straub, D.W. and Nance, W.D. 1990. Discovering and disciplining computer abuse in organizations: a field study. MIS Quarterly 14(1):45–62.

Willson, J. and Oulton, T. 2000. Controlling access to the Internet in UK public libraries. OCLC Systems and Services 16(4):194–201.

Disclaimer

Articles published in SAJIM are the opinions of the authors and do not necessarily reflect the opinion of the Editor, Board, Publisher, Webmaster or the Rand Afrikaans University. The user hereby waives any claim he/she/they may have or acquire against the publisher, its suppliers, licensees and sub licensees and indemnifies all said persons from any claims, lawsuits, proceedings, costs, special, incidental, consequential or indirect damages, including damages for loss of profits, loss of business or downtime arising out of or relating to the user's use of the Website.

top

ISSN 1560-683X

Published by InterWord Communications for the Centre for Research in Web-based Applications, Rand Afrikaans University