Global Information Assurance Certification Paper

Implementing an Internet Content Filtering and Reporting Program
Eric S Wilkens
SANS Security Essential GSEC Practical Assignment Version 1.3

Abstract:

This paper addresses the issues around implementing an Internet Content Filtering and Reporting program. In creating this document a holistic approach was used to point out the fact that this is not only an issue for Information Security but also an issue that should concern Human Resources, Management and Network Administrators.

The discussion includes the implementation of an actual Internet Content Filtering and Reporting program and what steps were necessary to implement the program. The paper challenges the user to take a look at different parts of the program including whether to filter Internet accesses, what an Internet filter is and reporting. While an actual program was used as a model to write this paper with it is only a template of what is possible in different organizations. While not the most glamorous job in a corporation the person responsible for Internet Content Filtering and Reporting fulfills a vital role to overall well being of each employee and the company.

Internet Statistics

As of January 2002, approximately 55 million American adults go online from work, up from 43 million in March 2000. Fifty-five percent of those with Internet access at work went online on a typical day in 2001, compared to 50% in 2000, and many were going online more frequently throughout the day than they had in 2001 (Pew Internet & American Life).

Workplace users had an average of 41 sessions during the month, while at-home users had 18 sessions (Nielsen/NetRatings).

82% of U.S. business executives surveyed by the consulting firm Dataquest (a division of the Gartner Group) believe Internet use should be monitored at their companies (Information Week).

28.83% of employees said their employer had caught them surfing non-work-related sites, although 54% of employers said that they have caught an employee surfing non-work-related sites at work (Vault.com).

53.2% of employees said it was ethical for employers to monitor Internet/e-mail usage (Vault.com).

Key Players

There are several levels and people in a company that play a significant part in and effective Internet Content Filtering and Reporting program. These include Management, Human Resources, Information Security, Network Administrators and the end users.

Management is responsible for supporting an acceptable use policy and being involved in corporate decisions pertaining to what the company accepts as appropriate usage. They should provide overall guidance, approve policy and processes, establish corporate culture, support both HR and IT efforts and stay involved.

Human Resources are an important player in the program because at its core the program is all about employees. Human Resources involvement in the development and implementation of the Internet Content Filtering and Reporting program helps to make sure it will be effective.

Network Administrators are responsible for providing the best possible infrastructure and applications. They are responsible for solving technical problems and for using the reports to help provide a better network environment based on business related Internet traffic.

Information Security is generally responsible for creating an acceptable use policy, instituting the Internet Content Filtering and Reporting process and monitoring the activity. They also generally have the responsible for the program in day-to-day operations.

End users will judge an Internet Content Filtering program based on their experiences and if they can find the information they need to do their job. The program will be a de-motivator to retaining good employees if the Internet filtering is seen as a hindrance to the performance of the employees. They also have the responsibility to know the policies that apply to Internet access.

To Filter or not to Filter: Two Viewpoints

One of the first questions a company needs to ask itself is do we need to filter what our employees see and do on the Internet? Secondly companies need to ask what can Internet content filtering software do for me and will it help me meet my objectives?

Internet content filtering products allow corporations to implement an appropriate use policy and an access policy in regards to using the Internet. You can allow employees access to certain sites, deny access to sites or even allow but warn employees about access certain types of sites. Product vendors claim that companies will see an immediate return on investment resulting from increased productivity and recaptured bandwidth, while reducing the company's legal liability.

There has been a large amount of debate about whether to filter Internet access or allow unabated access. The federal government protects a corporation's right to use Internet content filtering software. "Federal law authorizes and protects the good faith use of filtering software on all interactive computer services to restrict access to or availability of objectionable material and specifically preempts any contrary or state law. --47 U.S.C. '230(c)(2)" (Symantec).

There is debate in corporate America whether is worth the time and money to monitor an employee's access. What is your company's stance on employees using other technologies? Do you allow you employees to make brief local personal calls? Is it feasible then to not allow you employees to use the Internet for some personal use? The database manufacturer Oracle does not filter employee Internet access. According to Sally Hutchingson, Oracle's corporate public relations spokeswoman the company isn't

concerned that workers do personal tasks at work…. adding that certain behavioral guidelines must be followed (The Standard). Louis Maltby, president of the National work Rights Institute and former director of the ACLU's National taskforce on Civil Liberties in the Workplace is against Internet monitoring. "Under the Americans with disabilities Act, once a person has been hired, it is illegal to gather medical information without a compelling reason. An employer might end up violating the law because they're monitoring the Web sites employees are visiting" (The Standard).

Facts about Content filters

If your company decides to filter and monitor employee activity on the Internet you need to decide on a product to do your filtering. A very important item to be aware of up front when deciding to filter Internet content is that no filter can categorize every website on the Internet, it is simply growing to fast to keep up.

Manufacturers of Internet Content filters use two fundamental metrics when designing and implementing their products. These metrics are recall and precision. If a product has recall it identifies every site that should be blocked. The downside to having a high degree of recall is that the filters tend to incorrectly categorize web sites. If a product has precision it is able to distinguish between a pornography site and a medical site. The downside to being very precise is that the filter will not be able to categorize all the web sites. It is important to note that theoretically these two are inversely related. The concept of these metrics plays a role in the type of method vendor will use to categorize web sites for their filters (Kester).

There are two methods Internet content filtering vendors use to categorize sites. The first method is called dynamic filtering. In dynamic filtering the filter reads the web page and uses an algorithm to determine a page's classification in real time. These filters work by using keywords or artificial intelligence techniques. The downside to dynamic filtering is the additional system resources needed to analyze each page, longer wait times for a web page to be displayed and there tends to be a larger number of sites blocked that should not be blocked. The second method is a database method. In this method sites are classified and put into categories that reside in a database. As the user requests a site the filter looks up the URL and checks it against its database to find the category it is listed under. Once the category is found it is then checked against the filter list to see if the category open for viewing or is access should be denied. The downside to database filters is that they do not have the entire web classified. Web sites change their address each day and if your vendor only updates the list weekly, you may loose access to a good site or allow access to a bad site for a week's time.

Dynamic filters, effective at blocking porn sites and other obvious web content in real time are generally acceptable in homes or schools. Database filters are the preferred choice of corporations that demand less over blocking, more precision and a filter that mirrors its employees' Internet surfing patterns (Network World).

A problem with all Internet content filters is that they will produce both false positives and false negatives. A false positive occurs when the filter blocks access to a web site that should be allowed. A false negative is when a site that should be blocked is allowed through the filter. Just as large of a problem is vendor errors or a change in the

classification of a site. Take for example users are allowed access to all parts of Yahoo except the travel maps. For companies that have many people who travel in the field, this became an issue. Many people prefer to use Yahoo maps to Mapquest. Remarkably Mapquest was not classified the same as Yahoo maps. To resolve the issue employees were encouraged to use Mapquest and a vendor reclassification was requested the to classify Yahoo maps the same as Mapquest. The next week's update brought the reclassification and a happier employee base.

One of the nice things with the Secure Computing product is a special web page where you can check what category a site is located in for the various versions of the software. The site is located at http://www.securecomputing.com/cgi-bin/filter_where.cgi and will help you plan what categories you need to be concerned with and what classification is given to a web site. It is also a great tool to troubleshoot blocked sites.

Internet Content Filtering Products

Whatever product you use it should be easily customized, transparent to the employees and require minimum system requirements. There are several products on the market that perform Internet content filtering. Products that are in the database method include iPrism from Internet Products, i2100 Filtering Appliance from N2H2, WEBsweeper from MIMEsweeper which claims their URL Filter Category has 40 clearly defined categories that contain references to over 689 million Web pages, representing approximately 2.4 million domains, with provisions for at least 14 languages (MIMEsweeper). SmartFilter from Secure Computing which claims to be the first network-based solution in the industry allowing Internet management ranging from performance restrictions to complete blocking or URLs. Version 3.0.1 offers 30 categories and up to 10 user defined categories. It also offers support for groups including IP addresses and subnets, customizable coaching and blocking messages, e-mail notification and browser redirection (Secure Computing). Dynamic filtering products include IM Web Inspector by Elron Software. There are also products on the market that use a hybrid approach. SuperScout Web Filter by Surf Control and I-Gear from Symantec use the database method as a base and then use dynamic to categorize any sites not listed in the database.

What to block

Upper management is concerned with employee productivity, IT is concerned with network resource utilization, and Human Resources is concerned with eliminating a company's legal exposure.

It is important to spell out for your employees what they are allowed to access on the Internet. Do you have a time when users can openly surf the Internet? Is your content filtering software smart enough to know the difference? Do you have that ability?

Once you have installed your filtering product you will need to configure the categories that you will allow employees to access. When deciding which categories to block do not make ad hoc or impromptu decisions. You must understand what is included in each category and give upper management a true picture of what employees will and will not be able to do. Information Security, Legal Department, the network

group and your Firewall Administrator should review your proposed configuration. A lesson learned is to always reevaluate your decisions every six months.

Whenever you filter what employees can and cannot see you will inevitably block a web site that someone wants to see. A very important component of the program is having a method for handling complaints about web page access. One method of handling complaints is when an employee attempts to access a web site that is blocked the filter gives them an denied access page and lists the web site address and the category it falls under. There is then an e-mail link on the page to which employees can send a request for the web site review forms to be sent to them. Once completed the employee would submit the forms and the site would be evaluated and considered for access. Another possibility is to have the page take the user to an online form to complete and submit electronically. When reviewing the site it should be evaluated for relevance to the business, risk to the organization by allowing access to the site and appropriateness of the site. Before a web site can be evaluated access must be allowed to that site for the person doing the evaluation. SmartFilter has several ways to override the filter and allow the evaluation of the site. The filter override that works with the least amount of privilege is to allow the reviewers IP address to access the IP address of the web site in question. Another way is to allow the reviewer's IP address to access the category that the site resides in on the database. This allows only the reviewer to access the site. Once the evaluation is completed a recommendation is made to management to allow access to all employees, continue to block access, or ask the vendor for a reclassification in the next update. The final decision is reviewed by a member of the Senior IT management team and communicated to the employee who submitted the request.

Based on personal conversations with peers at other organizations the following general categories are block by a fair amount of companies due to the legal risk of allowing access:

- Criminal Skills
- Cults
- Dating and Introduction Services
- Extreme/Gross Content
- Gambling
- Hate Speech
- Illegal drugs
- Sex Related

Reporting tools:

Reporting tools offer you the ability to created comprehensive reports on what employees are doing on the Internet. They use the data the proxy server creates to give you an easy to understand and readable report. These reports should be able to give the administrator all visits, which were successfully accessed, which visits were denied and give the full URL. Management style reports with easy to understand graphs and the ability to do granular reporting on a specific user are all important features.

Some products that provide reporting include Super Scout by Surf Control and Cyfin

Reporter by Wavecrest Computing. After evaluating Surf Control and Cyfin Reporter, we chose to go with Cyfin Reporter for Sidewinder for the ease of integration with existing infrastructure.

Web Access Reports: Hits versus Visits:

There are two different types of web access reporting. Inbound traffic, used primarily for webmasters and corporations wanting to know how often their site is accessed and outbound traffic, users going to web sites on the Internet. Since this project is about what you should do if you want to filter and monitor employee access to the Internet, inbound activity is not of concern.

Outbound activity is primarily used to monitor the amount of activity employees engage in, and what type of sites they are accessing. One important thing to understand when using your reporting tool is the difference between hits and visits. Simply put hits are any browser related activity including banner ads, images and video. These are important items to review to see exactly what it is the employee is viewing, but it is not an accurate measurement of activity. Visits are deliberate clicks onto a web page or requesting a download. Therefore, visits are a more accurate representation of activity.

Determining the amount of time an employee spends on the Internet is not as easy as it might seem. Some reporting tools will try to tell you they can give you the amount of time an employee was using the Internet, beware of the snake oil. Some products attempt to use an On-Site time to tell you the amount of time an employee spends on the Internet. For example, an employee accesses a web site at 9:00am, gets a phone call at 9:02, attends an emergency meeting from 9:04 – 11:00, then returns to their desk and access a different site, how long was the user really on the Internet? Some products will count the number of hits as a measure of time spent on the Internet. This can also be misleading, if one user visits a web site that is a text only page this registers as one hit, but if a different user visits a complex web site, containing images, banners, and ads, it registers more hits. So which user spent more time on the Internet? By using this method, it is the user who visited the complex web site, which may not be correct. By looking at a detailed user report you can ascertain that approximately 75 percent of all hits are not visits. The best we can do estimate the time spend by taking the average time it takes to load a web site, 10 seconds, and multiple it by the number of visits. Although it may not be an accurate accounting of time on the Internet it is better to underreport activity time than to over report the activity time.

Implementing the Program:

Successful implementation of an Internet Content Filtering and Reporting Program requires three things, policies and standards, monitoring, and follow-up action. Polices and standards need to spell out what is considered proper use of the Internet. Effective polices and standards will be:

- Consistent with the company's culture
- Be well published and widely distributed
- Contain detailed, unambiguous, measurable and easily understood standards
- Be considerate of employee and corporate needs

- Contain language that allows for measurement of compliance
- Define allowable, non-allowable and any special provisions on Internet access.

Follow up action is required to gauge employee compliance or non-compliance with polices and standards. If usage is not in compliance management has several options:

- Modify the polices and standards
- Personnel actions
    - Training/Retraining
    - Counseling
    - Reprimand
    - Revocation of Internet access
    - Termination

The company's policy regarding Internet access should be provided to all employees in both print and electronic form. All employees should be required to sign an acknowledgement of the policy and that acknowledgement should then be part of the employee's permanent personnel folder. The company should also have a method for answering questions that may arise on what is considered acceptable use.

Before any action is taken is important that the reporting was accurate. Inaccurate reporting can have serious ramifications including unnecessary personnel issues and the loss of good employees or even lawsuits over wrongful termination.

A side benefit that reporting can provide is the ability to help with assessments of productivity, evaluating network bandwidth utilization, and for charge back programs where bandwidth is charged back to the business units.

There are three elements to a successful monitoring program. If a program is to be useful, polices and reporting need to work together and create a holistic picture. To make a program effective the reports must be accurate. Reliability is obtained when the metrics created by the reporting tool can be compared to policy standards.

Today SmartFilter by Secure Computing is utilized to do Internet content filtering. The program is set to automatically retrieve the updated control list weekly. While the download is automatic it is recommended to run the install manually. It is standard procedure to never let a system automatically update itself without administrator supervision.

A decision needs to be made whether to put the Internet content filtering software on a proxy server, or even on the firewall itself. The choice you make will depend on the network infrastructure and your network engineers. One possibility for an organization with a fairly small number of employees with Internet access is to install SmartFilter is directly on the firewall instead of a proxy server.

Getting the Data:

In order for your reporting to be useful the raw data must be obtainable. The raw data from your Internet Filter must contain the basic information your reporting needs, typically this is the URL accessed, a timestamp on the activity and some type of user ID.

In a real world example current firewall configuration means that there are two sets of access logs. The challenge became getting the logs from each firewall and creating one

log file in order to get a true picture of what employees were accessing on the Internet, no matter which firewall they accessed. One problem that arose was the need to get the access logs from the firewalls without needing to allocate additional personnel and transferring them securely.

The solution we implemented was a home grown set of applications. On the firewalls there are two scripts that deal with the handling of the log files. Both of the scripts are triggered automatically through crontab to run in the middle of the night to minimize any potential impact to production. The first script prepares the log and moves it to the ftp directory so it can be downloaded from the reporting workstation. The second script on the firewall removes the log from the system. There are two bat files that each have a text file associated with them to perform an ftp session out to the firewalls and retrieve the logs. The decision was made to go this way since it bad security to have a .netrc file on your firewall. After each log file is downloaded a program runs to move the files to a new directory where they are merged into one file and sorted. The Cyfin program then runs a site analysis report to give an overview of what happened in the previous days user activity. This report is automatically e-mailed to myself for review. After reviewing the report the determination is made whether or not there is any activity that would warrant further review. All files are triggered through scheduler complete the process before normal business hours. The process runs in the following order:

1. Logcopy.sh runs on both firewalls
2. CyfinLogs30.cmd runs
3. CyfinLogs33.cmd runs
4. Logdelet.sh runs on both firewalls
5. MoveAccessLogs.cmd runs and calls sort.exe
6. Cyfin runs a site analysis report
7. MergeAccessLogs.cmd runs
8. Rename.exe runs

A sample of the code necessary to make the process work is included in the appendix. The sort.exe and rename.exe files are written in the C programming language.

Analyzing and Using Reports

How do managers request reports on what their employees are doing. Is the employee notified? Is there a formal written report that goes along with each request to track the usage of the reporting ability?

A basic part of any information security program is reviewing the log files. If we do not review the logs then how will we know that there is a problem? The simple thing here is that is does not do you any good to monitor and generate reports if no one is looking at them.

One of the most important aspects is to make sure the reports are easy to read. It is easier for management to understand snapshots such as bar graphs and pie charts than it is to understand the 14,000 denied visits to self-help sites last month. What does this really mean is a question that you need to be able to answer through reporting that means

something. Most of the reporting tools will give you easy to read graphs and "management ready" reports.

Log File Retention

Retaining your log files is an issue that is based on the laws that govern your company, company's line(s) of business and any polices that you may have. The Office of Thrift Supervision requires that banking companies retain the records involved in any personnel issues for at least 300 days. They also require other records to be retained for 7 years. This has become our standard practice for the retention of Internet usage logs. In order to comply with this we currently zip our monthly logs with WinZip and every 2-3 months burn them onto CDs for long-term storage. These CDs are then stored in a locked container for retrieval at a later time if necessary. Depending on your procedures you may need to send a copy to offsite storage or secure them in a lockable container.

References:

Synder, Julene. "All work and No Play." 15 May 2000. URL: http://www.thestandard.com/article/0,1902,14759,00.html?body_page=2 (27 Feb 2002).

"The Case for Internet Content Filtering." 31 Mar 2000. URL: http://enterprisesecurity.symantec.com/article.cfm?articleid=14&PID=6526177 (26 Feb 2002).

"Pew Internet Project Part 1: The Internet and Work." 3 Mar 2002. URL: http://www.pewinternet.org/reports/reports.asp?Report=55&Section=ReportLevel1&Field=Level1ID&ID=243 (6 Mar 2002).

D'Antoni, Helen. "Behind the Number: Web Surfers Beware: Someone's Watching." 7 Feb 2000. URL: http://www.informationweek.com/bizint/biz772/72bzweb.htm (6 Mar 2002).

"Nielsen/NetRating Announces At-Work Internet Users Do Double-Time Online As Compared To At-Home Web Surfers." 22 Feb 2000. URL: http://www.nielsen-netratings.com/press_releases/pr_000222_work.htm (6 Mar 2002).

"Results of Vault Survey of Interent Use in the Workplace." URL: http://www.vault.com/surveys/internetuse2000/index2000.jsp (6 Mar 2002).

"WEBsweeper 4.1 Frequently Asked Questions." URL: http://www.mimesweeper.com/download/collateral/pdfs/faqs/wsw41_faq.pdf (5 Mar 2002).

Kester, Harold "Internet filters take two approaches." Network World February 18, 2002 Vol 19, No 7: 39

"Smart Filter 3.0.1 Frequently Asked Questions." URL: http://www.securecomputing.com/pdf/smartfilter301_faq_final.pdf (9 Mar 2002).

"Internet policy management from Secure Computing" Dec 1999. URL: http://www.securecomputing.com/pdf/smfilter_wp_internet.pdf (5 Mar 2002).

"The Case for a Policy-Based Web-Use Management Approach" URL: http://www.wavecrest.net/editorial/include/caseforPolicyBased.pdf (27 Feb 2002).

Tubell, W.J. "Implement Web-use Policies With Reliable Metrics." 2 May 2001. URL: http://www.wavecrest.net/editorial/include/reliableMetrics.pdf (27 Feb 2002).

"Employee Web-use Management: It's a People Issue" URL: http://www.wavecrest.net/editorial/include/peopleIssue.pdf (27 Feb 2002).

Appendix

Code samples

Logcopy.sh
```
cp /var/log/squid/access.log.0.gz /home/ftp/pub/access.log.0.gz
cd /home/ftp/pub
gunzip access.log.0.gz
mv access.log.0 access.log
```

Logdelet.sh
```
cd /home/ftp/pub
rm access.log
```

CyfinLogs30.cmd:
```
cd 130
ftp -s:cyfin.txt ipaddress >> sent.txt
```

CyfinLogs33.cmd
```
cd 130
ftp -s:cyfin.txt ipaddress >> sent.txt
```

Cyfin.txt
```
vailidftpuser
someone@yourcompany.com
cd pub
get access30.log
get access33.log
bye
```

MoveAccessLogs.cmd
```
cd audit files
cd 130
move "c:\audit files\130\access30.log" "c:\Audit Files\access30.log"
cd ..
cd 133
move "c:\audit files\133\access33.log" "c:\audit files\access33.log"
cd ..
sort.exe access30.log access33.log access.log
del access30.log
del access33.log
cd ..
```

Sort.exe

```
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#define MAXBUFFER 512

int getline(FILE * fd, char buff[], int nmax){
  char c;
  int n=0;
  while ((c=getc(fd))!='\n')
  {
    if(c==EOF)return EOF;
    if(n<nmax)
      buff[n++]=c;
  }
  buff[n]='\0';
  return n;
}
int stringMerge(char filename1[], char filename2[] , char filename3[]) {
  FILE *fd1, *fd2, *fd3;
  char buffer1[MAXBUFFER], buffer2[MAXBUFFER];
  int ln1, ln2, n;
  n=0;
  if ((fd1=fopen(filename1, "r"))==NULL)
  {
    perror("fopen");
    exit(1);
  }
  if ((fd2=fopen(filename2, "r"))==NULL)
  {
    perror("fopen");
          exit(1);
  }
  if ((fd3=fopen(filename3, "w"))==NULL)
  {
    perror("fopen");
    exit(1);
  }
  ln1 = getline(fd1,buffer1,MAXBUFFER-1);
  ln2 = getline(fd2,buffer2,MAXBUFFER-1);
  while ((ln1!=EOF) && (ln2!=EOF))
  {
          if (strcmp(buffer1,buffer2)<=0)
          {
```

```
    fprintf(fd3, "%s\n", buffer1);
    ln1 = getline(fd1,buffer1,MAXBUFFER-1);
   }
        if (strcmp(buffer1,buffer2)>=0)
        {
    fprintf(fd3, "%s\n", buffer2);
    ln2 = getline(fd2,buffer2,MAXBUFFER-1);
   }
        else
        {
         fprintf(fd3, "%s\n", buffer1);
    fprintf(fd3, "%s\n", buffer2);
          ln1 = getline(fd1,buffer1,MAXBUFFER-1);
          ln2 = getline(fd2,buffer1,MAXBUFFER-1);
        }
        n++;
 }
 while (ln1!=EOF)
 {
    fprintf(fd3, "%s\n", buffer1);
    ln1=getline(fd1,buffer1,MAXBUFFER-1);
    n++;
 }
 while (ln2!=EOF)
 {
    fprintf(fd3, "%s\n", buffer2);
    ln2=getline(fd2,buffer2,MAXBUFFER-1);
    n++;
 }
 fclose(fd1);
 fclose(fd2);
 fclose(fd3);
 return n;
}
int main(int argc, char *argv[]) {
 if(argc!=4){
  printf("Usage: %s sortedfile1 sortedfile2 mergefile\n", argv[0]);
  exit(0);
 }
 return 0;
}
```

MergeAccessLogs.cmd

```
copy "c:\audit files\January 2002\access.log" + "c:\Audit Files\access.log" "c:\audit
files\January 2002\access.log" >> merge.txt
```

Rename.exe

```
#include <dos.h>
#include <stdio.h>
void main()
{
 int i, x;
 struct date d;   //defined in date.h used to get the date
 char oldname[15] = {"access.log"}; // Oldname never changes.
  static char *newname[] = {
 "access0.log", "access1.log", "access2.log", "access3.log", "access4.log",
 "access5.log", "access6.log", "access7.log", "access8.log", "access9.log",
 "access10.log", "access11.log", "access12.log", "access13.log", "access14.log",
 "access15.log", "access16.log", "access17.log", "access18.log", "access19.log",
 "access20.log", "access21.log", "access22.log", "access23.log", "access24.log",
 "access25.log", "access26.log", "access27.log", "access28.log", "access29.log",
 "access30.log", "access31.log"};
  getdate(&d);
  i = (d.da_day - 1);
  if (i == 0)
 {
  x = 31;
  i = x;
  }
  rename(oldname, newname[i]);
}
```