How to Implement a Content Filtering System

Joshua S. Dean
December 30, 2003
GSEC Practical Assignment 1.4b
Option 1: Research on Topics in Information Security

Abstract

This paper is a guide for implementing a content filtering system. I cover the basics of an Internet Usage Policy (IUP), installing devices for content filtering, and finally enforcing the policy through disciplinary action. The installation portion will be specific to a particular product (WebWasher). I outline the capabilities and vulnerabilities in detail for anyone who may choose this product; however, I intend to remain objective and generic enough to offer ideas on how to approach other products. For background and understanding, I also discuss the purpose of content filtering and legal issues concerning monitoring employees.

My purpose is to show the importance of a filtering system. Through proper planning, such a system can provide another layer of protection for systems and mitigate legal exposure without creating new legal issues.

Background

Content filtering is a security method of restricting access to or blocking Internet material that may be inappropriate for viewing by a specific community or harmful to systems. In an enterprise, the community may be considered all employees or be broken down into groups (i.e. executives, security professionals, and everyone else). The inappropriate material may include pornography, chat, gambling, web mail, etc.

Aside from just being a "net nanny"[1] for employees, content filtering can also provide another layer of protection for systems. For example, the content filtering system could do antivirus scanning. It could also scan and block Java and ActiveX which attempts to perform malicious, or otherwise questionable, operations. Restrictions to certain file types may also be applied. This could keep clients from downloading harmful code in executables or other potentially dangerous file types.

Scanning HTTPS is yet another function of content filtering. The ability to scan and block HTTPS is extremely desirable today. This encrypted traffic has the potential of delivering harmful code to the client without being detected by firewalls or IDS (Intrusion Detection Systems). A content filtering system would mitigate much of this risk.

Why Bother?

It is true that content filtering systems do not provide an immediate or tangible return on investment. Some systems can be costly to install and support, and the additional overhead for network traffic actually slows down the rate of business. So, why even bother?

Content filtering does provide a high rate of return if you think about the amount of risk that can be reduced. The first risk that can be reduced is hostile work environments due to inappropriate browsing. Content filtering systems can block pornographic, prejudice, and other offensive sites. Inappropriate browsing

[1] NetNany. URL: www.netnanny.com

can create hostile work environments among employees, which may ultimately result in legal action, and waste your company's assets. The blocking of such material may also recover money lost through wasted employee productivity. However, I must also warn that content filtering systems and other technologies should never be considered as a replacement of management. These technologies should enable management and not take the place of it.

Content filtering can also reduce risks concerning records management. Do your employees use web mail or chat to conduct business? Using content filtering to block web mail and chat is common practice. By forcing employees to use only the tools you allow/provide, your company can have better control over records management. Also, web mail has been known to be a carrier of spam, worms, and viruses. Blocking and scanning of this content will reduce the risk of spreading malicious code to your trusted systems.

Other forms of malicious code can be blocked through content filtering. ActiveX Controls have been a major concern. ActiveX Controls can be written in just about any language. They are, therefore, limitless in the types of functions they can perform. For example, some controls may be used to access local file systems, create connections with other hosts, and transfer files. Some filtering systems can execute such code in what is frequently referred to as a "sand box". The nature of the code is determined before being delivered to the client. If the code is believed to be malicious, then it can be rejected before reaching trusted zones of your internal network. Overhead for these types of scans can be costly to system performance. Please keep this in mind when trying to size hardware and testing performance.

I have already mentioned that HTTPS can pose a security risk by surpassing conventional IDS or other sensors. When the information is encrypted, it is impossible to inspect through pattern matching and other scanning. Imagine then that a hacker launches an attack on one of your web servers while using HTTPS. The code comes within the tunnel past firewalls and IDS without detection. The server is compromised, and now you have no clue where it may have come from. The hacker has turned the control of confidentiality against you. By decrypting SSL and other encryption technologies, you could then scan for these types of attacks.

These risks and vulnerabilities could lead to events that cost large sums of money. This is why implementing and supporting a content filtering system should become a priority for your security infrastructure. Later I will discuss how content filtering can be implemented on lower end systems that you may even have lying around. Hopefully this will increase the return on investment for your content filtering system.

Privacy Issues

It is extremely important to become aware of some of the legal issues that may come into play when enforcing that policy. The most obvious issue is privacy. What are the rights of employees in the work place? Do employers have the right to monitor their employees?

The answer to the latter is, of course, yes. Does this mean that the employer can log every key stroke and trip to the bathroom? The Privacy Rights Clearinghouse[2] is an excellent source on privacy issues and provides a fact sheet for privacy in the workplace. They point out that though the employer reserves the right to monitor employees, they may have to abide by certain litigation, which I reiterate below.

Union contracts may only permit certain types of monitoring. Union contracts should be reviewed and possibly renegotiated before attempting to implement an IUP. Some monitoring may also fall into the category of search and seizure, which is protected by the Fourth Amendment[3]. Finally, some states may have their own regulations on employee monitoring. Such is the case with California.

The most important part of employee monitoring is communication. You must inform employees that they are being monitored, and provide detail describing the types of behavior considered to be inappropriate. It is also important that they know the possible ramifications for such behavior. This will be the basis of the IUP. In the case of a lawsuit, an IUP drafted in this manner may keep the judge from making a decision based on unreasonable search and seizure. This exposure can be limited by having the employees review and sign the IUP.

Writing the Internet Usage Policy

In writing an IUP, it is important to be clear and concise. It is a good idea to consult with legal and/or upper management when writing the IUP. Here is a sample to get started[4]:

Internet access to global electronic information resources is provided to assist employees in obtaining work-related data and technology. The following guidelines have been established to help ensure responsible and productive Internet usage. While Internet usage is intended for job-related activities, incidental and occasional brief personal use is permitted within reasonable limits.

All Internet data that is composed, transmitted, or received via our computer communications systems is considered to be part of the official records of COMPANY and, as such, is subject to disclosure to law enforcement or other third parties. Consequently, employees should always ensure that the business information contained in Internet e-mail messages and other transmissions is accurate, appropriate, ethical, and lawful.

The equipment, services, and technology provided to access the Internet remain at all times the property of COMPANY. As such, COMPANY reserves the right to monitor Internet traffic, and retrieve and read any data composed, sent, or received through our online connections and stored in our computer systems.

[2] Privacy Rights Clearinghouse. URL: http://www.privacyrights.org
[3] Find Law. : http://caselaw.lp.findlaw.com/data/constitution/amendment04
[4] adapted from Minnesota Rural Water Association. URL: http://www.mrwa.com/internetpolicy.htm

Data that is composed, transmitted, accessed, or received via the Internet must not contain content that could be considered discriminatory, offensive, obscene, threatening, harassing, intimidating, or disruptive to any employee or other person. Examples of unacceptable content may include, but are not limited to, sexual comments or images, racial slurs, gender-specific comments, or any other comments or images that could reasonable offend someone on the basis of race, age, sex, religious or political beliefs, national origin, disability, sexual orientation, or any other characteristic protected by law.

Abuse of the Internet access provided by COMPANY in violation of law or COMPANY policies will result in disciplinary action, up to and including termination of employment. Employees may also be held personally liable for any violations of this policy. The following behaviors are examples of previously stated or additional actions and activities that are prohibited and can result in disciplinary action:

- Sending or posting discriminatory, harassing, or threatening messages or images
- Using the organization's time and resources for personal gain
- Stealing, using, or disclosing someone else's code or password without authorization
- Sending or posting messages or material that could damage the organization's image or reputation
- Participating in the viewing or exchange of pornography or obscene materials
- Sending or posting messages that defame or slander other individuals
- Sending or posting chain letters, solicitations, or advertisements not related to business purposes or activities
- Using the Internet for political causes
- Using the Internet for gambling
- Engaging in any other illegal activities
-

I have read and understand the computer usage policy set forth by COMPANY. I agree to abide by the rules and regulations set forth by this policy.

Employee Signature: __________________________ Date: _________________

This IUP is well formed for the following reasons. The first paragraph makes a clear statement of proper Internet usage and behavior. The Internet may be used to access information for business use, and occasionally for personal use.

Notice the third paragraph. This is a clear statement of the limitations of monitoring e-mail and other Internet usage. The rest of the document outlines Internet usage that is inappropriate and the fact that such activity will result in disciplinary action. The examples of unacceptable behavior can be used as a check list when parameters for content filtering software have been set.

I have added the bottom section with the employee signature as a reminder that this document should be viewed and signed by all employees. Signing the IUP should be a part of employee orientation for new hires, and should be resigned whenever major changes are made. In addition, a current copy of the IUP should be readily available to all employees. Providing a link off of your Intranet homepage to the IUP is an excellent practice.

Installing the Content Filtering System

There are several solutions available when deciding on a content filtering system. The content filtering is usually software that may run on an appliance, such as a firewall or proxy, or on a separate server as a service. Some appliance solutions include BorderWare's MXtreme MX-200 Mail Firewall Appliance with 8e6 Technologies' R3000 for content filtering[5] or Network Appliance's NetCache with Secure Computing's SmartFilter[6]. If your current architecture already includes firewalls or proxies, then an on-the-box solution like these may be extremely cost effective and easy to setup. I suggest you do some research of your own and then consult your vendors for more information. Don't forget, however, that you still have the option of setting up a separate server with a content filtering service.

I have chosen to discuss installation of WebWasher Enterprise Edition[7] on a separate server. I feel that a guide like this is most useful since the on-the-box installation is more dependent of individual architectures. The software will run as a service on Windows 2000 Server. The guide assumes that the firewalls or proxy in place support ICAP 1.0[8], which is now outlined in RFC 3507[9].

Why WebWasher?

I have chosen WebWasher for a number of reasons. Most importantly, it met the criteria at the time I implemented a content filtering system. Here is a list of noteable functions:

- Sites are blocked at IP level (requests are treated the same whether made via URL, dotted-decimal IP, octal, binary, etc.)
- Blocking is decided on a hierarchy
- Ability to add sites to the categorization database without relying on WebWasher to make the change

[5] Spiwak, Mark. URL: http://www.crn.com/sections/security/security.asp?ArticleID=46440
[6] Network Appliance. URL: http://www.netapp.com/partners/application?origin=PSCCatalogSearch.jsp&event=bea.portal.framework.internal.refresh&pageid=PSCCompanyDetails&companyId=104
[7] WebWasher. URL: http://www.webwasher.com/enterprise/products/webwasher_products/webwasher_ee/index.html?lang=de_EN
[8] ICAP Forum. URL: http://www.i-cap.org/home.html
[9] RFC Editor. URL: http://www.rfc-editor.org

- Quick access to online categorization checker / URL submission
- Ad and pop-up blocking
- File type blocking through magic-bytes
- Able to enforce multiple policies based on an employee's credentials
- Customizable blocked messages

More important for the industry, is its ability to block Java and ActiveX, decrypt SSL traffic for scanning, and conduct antivirus scans.

How It Works

WebWasher handles requests in two modes: request and response. When a client attempts to access content on the Internet, the proxy or firewall handling the HTTP request sends an ICAP message to the WebWasher request mode service. The message includes the user's credentials and the URL of the request. WebWasher first places the user into a profile based on the user's credentials. The profile mapping may be done via LDAP, NTLM, NT domain, or even by IP. Access to the site is based off of the policy of the specific profile.

Now WebWasher goes through the hierarchy of the determined policy. The first check is the white list. This is a list that contains regular expressions. If the request matches any of the expressions, then it is allowed. Next is the black list. Exactly the same at the white list, only requests are blocked if a match occurs. It is important to keep these lists to a minimum. Every expression in the white list and black list is applied to every request (that doesn't match an earlier expression). This could be trouble considering regular expressions tend to be laboring in general.

The next step in the hierarchy is the extended list. This is where an administrator can add sites to the blocking database. As opposed to using a sequential ACL like most content filters, WebWasher allows sites to be added to its database processing. Sites can be added to existing categories or to any of five user-defined categories. If the site matches an allowed category, it is allowed, and if it matches a blocked category, it is blocked. If the site is in multiple categories, it only takes a single block for it to be blocked.

Finally, the request is checked through the DynaBLocator, which is the database WebWasher uses for its own site categorization. The same methodology as the extended list is used here to allow or block the site. If the site still hasn't met a match, then it is allowed. There is a URL Sweeper, which will block any sites that are uncategorized, but I would not recommend it.

Surprisingly, WebWasher is capable of making decisions in request mode extremely fast with little CPU usage. Response mode is a slightly different story, however. In response mode, the proxy or firewall sends each object via ICAP. The content of each object is scanned for certain criteria. Below is a list of possible scans. Note that the user's credentials are sent in each ICAP packet, so that the scans can be applied only to specific profiles.

- Classify the URL real-time by scanning the URL and meta tags for inappropriate words

- Scan for Java and ActiveX for illegal activity
- Scan magic-bytes to determine file types
- Scan files for known virus patterns
- Determine dimensions of graphics to block potential ads
- Kill pop-ups
- Decrypt SSL to scan for any of the above.

The amount of CPU and time per request goes up as more scans are applied. I have not conducted much testing with the Java, ActiveX, and antivirus scans. I do notice a decrease in system performance when enabling the response mode, which is required for these types of scans.

The real-time classifier has the potential of doing more harm than good. I have noticed that it is pretty touchy, and that many appropriate sites get blocked for very few additional blocks of unwanted content. It still doesn't block every inappropriate site. I have not tested the magic-bytes, but find it somewhat intriguing. Remember to weigh the performance degradation with the amount of risk that will be mitigated when deciding upon these scans. Response mode will definitely decrease response times and increase the potential of blocking more legitimate business activity.

Purchase

The WebWasher EE Software comes with Internet Access Management, Internet Content Filtering, and Reporting. Licenses can be purchased by number of users. Enterprise licenses (unlimited users) may also be purchased. Please note that the number of users allowed is licensed per box. For example, if you have 10,000 employees and purchase a license as such, then a single box running the filtering software would allow 10,000 users. If you install several servers for redundancy you'll notice that you'll more than likely never have all users connected to a single box. Ethically, you should purchase a license that fits the number of employees you have, but you may be able to get away with a smaller license.

There are two optional packages with the software that become available with the purchase of additional licenses. They are Virus Protection and Encrypted Content Scanning. Depending on your needs now, you may leave these packages out of your initial install. The functionality becomes available as soon as you enter valid licenses (no new downloads or re-installation required).

Hardware

WebWasher EE is extremely memory intensive. The system requirements suggest 512 MB for Windows, but I would recommend at least 1 GB. However, in my experience, the process is less intensive on CPU. Depending on number of users and other services on the server, a 400 Mhz system would suffice. Very little hard drisk space is required. Windows 2000 Server requires a 2 GB hard disk with at least 1 GB of empty space for installation[10], so 4 GB is definitely

[10] Microsoft. URL: http://www.microsoft.com/windows2000/server/evaluation/sysreqs/default.asp

enough. I also suggest a reasonably fast NIC. Every Internet request will be sent to this server, so I'd go with 100MB to 1GB, again depending upon your current network speeds and traffic.

When sizing the server, you should definitely conduct performance testing. Here I have given a baseline for system requirements. Sizes of systems will vary due to number of users, network traffic, and other functions of the server. Although, I would recommend dedicating a server to content filtering until an on-the-box solution can be implemented.

Security

Before installing WebWasher, it's important to incorporate security on the Windows 2000 Server. In addition to hardening the server in general, this includes creating a special account to run the WebWasher service and restricting access to the WebWasher directory.

WebWasher will run as a service, which means it will assume the privileges of the user running it. It's best to create a special user with only enough privileges to allow WebWasher to function properly. The user should have limited Internet access to WebWasher download sites, and only needs write privileges to the WebWasher EE directory.

Conversely, access to the WebWasher EE directory should be restricted except to the special user and the people that will support the content filtering system. I highly recommend that the directory be created on a separate drive or partition. Two gigabytes is plenty of space.

The server itself should also be hardened. Since this issue could possibly be a paper of itself, I'll simply direct you to this link:

http://www.systemexperts.com/win2k/HardenWin2K.html

Known Vulnerabilities

There are two noticeable vulnerabilities with WebWasher. The first is that the software configuration is done through a web interface. To accomplish this, the software installs a separate slimmed-down version of Apache, which also runs as a separate service. New software distributions are available every six months, so without patches, this service has the potential of being a soft spot in your network. Extra steps should be taken to monitor this threat. Already noted, by running this service as a low-profile user, the risk of a hacker launching a successful attack from the exploited service is minimized.

Secondly, the web interface is currently through HTTP. Access through the GUI is not encrypted. This means that none of the information should be considered private. Anyone with a sniffer will be able to get the username and password to access the GUI, along with any configurations. The GUI account then should be separate from logons to the server, and the password should be changed often. It is also wise to only allow supporting personnel to make HTTP requests on the designated GUI port (default is 9090).

Software Configuration

After installing the software, which shouldn't take more than ten minutes, it's time to get started with the configuration. Make sure that the WebWasher and Apache services are running as the special user, and now all configurations can be made using a web browser on supporting staff workstations. Access the GUI by making a similar request: http://*server*.*domain*:9090/conf. Notice that the request is made on port 9090. This is the default GUI port if you did not change it during installation. Supply the default username and password to begin.

The first step is to activate the license. The license is a simple text file that can be stored on the administrator's workstation. On the home page, select the "Browse" button under "Import License". Select the location of the license file, and hit "OK". Then select the "Activate License" button. Now all of the functionality of WebWasher that you purchased is ready for configuration/use. Note that this is the only step required if you purchase a different license.

Next, go to the "User Management" tab. This where you configure how Internet users will be grouped. Most importantly, in the "Administrative Rights" section you can change the GUI password. Make sure you change it from the default password immediately.

Now you will need to setup some things for networking. Go to the "Network Configuration" tab. Here you create the names of the request and response modes to be used by ICAP (I'll use wwreq and wwresp, respectively). Note that you will not need to enable response mode if you do not wish to do Java, ActiveX, or antivirus scanning. Also, you must setup the port for ICAP to use (the default is 1354). Under this tab, you can setup HTTP and FTP proxies. You can also configure how WebWasher can access the Internet to pull software and database updates. This is also where you must configure LDAP settings if you use it to profile users. WebWasher sends email messages for license expiration and virus alerts. Configure the SMTP server and admin email address to receive these messages.

Under the "User Management" tab you can create the separate profiles. Again, profiles can be determined via LDAP, NTLM, NT domain, and even by IP. The profile method can even be setup in a hierarchy (first check IP, then NT domain, then LDAP). For each method of authentication, you must enter the criteria for profile mapping.

Finally, under the "Policy" tab you turn on specific filters for each profile. Then configure each filter under the "Filter Maintenance" tab. For example, to enforce the IUP above, pornography, gambling, and criminal activity sites are not allowed. So, simply find these categories and select "Never" in their access pull-downs.

The only thing left to do to start filtering content is to configure the proxy or firewall to send ICAP messages to the content filtering server. Most proxies or firewalls will require ACLs to tell the appliance to send HTTP requests to the ICAP service, such as:

```
icap(req_farm) http
icap(resp_farm) http
```

Then configure the service farms as follows:

Enable	Name	Load Balance	By-Pass	Services
yes	req_farm	round_robin	yes	http://xx.xxx.xxx.x1/wwreq http://xx.xxx.xxx.x2/wwreq
yes	resp_farm	round_robin	yes	http://xx.xxx.xxx.x1/wwresp http://xx.xxx.xxx.x2/wwresp

Having "By-Pass" turned on keeps sites from becoming unavailable in the event the content filtering system goes down. For example, heightened network traffic at peak hours may cause ICAP to become sluggish. Without by-pass, the requests time out, and users receive an error message. The risk of exposure to illegal or otherwise harmful content during outages is less that the potential of disrupting legitimate business requests. Remember that users need to be able to conduct business. For this reason, it is my opinion that the services should fail open.

Enforcing the Policy

So, your content filtering system has been installed using WebWasher EE on Windows 2000 Servers. ICAP farms have been configured on proxies or firewalls that can handle ICAP 1.0. Internet traffic for your enterprise is now being filtered for pornography, criminal activity, and gambling in request mode. In response mode, content may be scanned for harmful viruses and Java and ActiveX. Now you must enforce the IUP and system you have set forth.

I suggest writing a few scripts that gain statistics on the number of blocked messages employees receive. This way you can find a good statistic representing the number of blocks employees receive under "acceptable Internet usage." Then, write a script that will generate a report detailing the top domains, time of day, and number of blocks employees receive per week. Generate a report as an aggregate of all employees for trending, and then one for each user over the threshold.

You should automate sending a message to abusers (users over the threshold). First send them a warning without copying management. Send them a copy of the report you generate, and include a copy of the IUP (or at least a link to the IUP on your Intranet). Be sure to allow them an opportunity to suggest sites that are needed for business. A private, automated message gives the user the opportunity to correct the action without involving other parties (such as management, HR, legal, etc.). On the second offense you may want to send another warning, and this time include management and/or HR.

However you chose to enforce the policy, make sure you are consistent and document every stage of monitoring and disciplinary action. Inconsistent action and poor documentation may increase legal exposure.

In addition to disciplinary action, you should also decide on your role in supporting blocked sites. I have a few scripts myself that parse through access logs to find URLs with bad words that may indicate future sites for blocking. I then submit these sites for review. Some providers offer this as a free online service. The success of your content filtering system now lies in the manner in which you support it. So, remember to stay on top of patches, updates, and new features.

Conclusion

I hope that I have provided enough of a guide to get you started with content filtering. Most systems can be installed and supported in the tens of thousands of dollars, which is a relatively low price for mitigating legal exposure and mitigating certain security risks.

The most important piece of the content filtering system is, of course, the IUP. The system itself should follow the guidelines set forth in that document. Such a document will ensure that your content filtering system mitigates legal exposure instead of creating it.

References

Cox, Phil. SystemExperts Corporation. Hardening Windows 2000 Guide. (2001). http://www.systemexperts.com/win2k/HardenWin2K.html

Find Law. *US Constitution: Fourth Ammendment*. (2001). http://caselaw.lp.findlaw.com/data/constitution/amendment04

ICAP Forum. (2003). http://www.i-cap.org/home.html

Microsoft. *Windows 2000 Server Requirements*. (2003). http://www.microsoft.com/windows2000/server/evaluation/sysreqs/default.asp

Minnesota Rural Water Association. *Sample Internet Usage Policy*. http://www.mrwa.com/internetpolicy.htm

Net Nanny. BioNet Systems, LLC. (2003). http://www.netnanny.com.

Network Appliance. *Partner Catalog: Secure Computing*. http://www.netapp.com/partners/application?origin=PSCCatalogSearch.jsp&event=bea.portal.framework.internal.refresh&pageid=PSCCompanyDetails&companyId=104

Privacy Rights Clearinghouse. (2003). http://www.privacyrights.org

RFC Editor. (2003). http://www.rfc-editor.org

Spiwak, Mark. New *Appliances Tackle Spam, Internet Filtering*. Dec 5, 2003. http://www.crn.com/sections/security/security.asp?ArticleID=46440

WebWasher. *WebWasher EE*. http://www.webwasher.com/enterprise/products/webwasher_products/webwasher_ee/index.html?lang=de_EN