Chapter e66

Content Filtering

Pete Nicoletti, CISSP, CISA, CCSK
Director of Security Solutions and Compliance at Virtustream, Inc.

1. DEFINING THE PROBLEM

No one can deny that the Internet is one of the greatest inventions of the late 20th century. It has touched many aspects of our everyday lives including, how we communicate, how we conduct business, or simply how we entertain ourselves. While there are many legitimate uses for the Internet like, social networking, entertainment, and work, there are also some very serious risks, some of which include: spam, viruses, worms, Trojans, keystroke loggers, identity theft, and Internet scams. While these risks have the potential to impact us all, businesses today have some very unique risks when it comes to leveraging the power of the Internet:

- Loss of productivity
- Sensitive data loss through intentional and unintentional means
- Application performance issues caused by bandwidth intensive applications
- Infection and destruction of corporate information and computing resources due to increased exposure to Web-based threats such as viruses, worms, Trojans, spyware, etc.
- Legal liability when employees access/download inappropriate and offensive material such as pornography, racism, violence, etc.

Surfing the Web is the most common of all Internet activities. It offers global access to all types of information, banking, buying and selling goods and services from the comfort of our computer, and online bill paying, and it's so entertaining in many different ways. For business, access to Web apps is mission critical, and other sites are productivity tools. Accessing the Internet from the office is constantly presenting new challenges to manage. Some of the negative impacts of doing the wrong thing and going to the "wrong places" include:

- Lost productivity due to nonbusiness-related Internet use
- Higher costs as additional bandwidth is purchased to support legitimate and illegitimate business applications
- Network congestion; valuable bandwidth is being used for nonbusiness purposes, and legitimate business applications suffer
- Loss or exposure of confidential information through chat sites, nonapproved email systems, IM, peer-to-peer file sharing, etc.
- Infection and destruction of corporate information and computing resources due to increased exposure to Web-based threats (viruses, worms, Trojans, spyware, etc.) as employees surf nonbusiness-related Web sites.
- Legal liability when employees access/download inappropriate and offensive material (pornography, racism, etc.)
- Copyright infringement caused by employees downloading and/or distributing copyrighted material such as music, movies, etc.
- Negative publicity due to exposure of critical company information, legal action, and the like

Casual non-business relatad Web surfing has caused many businesses countless hours of lost productivity and occasionally hostile work environments have been created by employees who view and download offensive content. RIAA takedown notices and copyright infringement threats, fines, and lawsuits are increasing as employees use file-sharing programs to share their favorite music and movies. Government regulations and legal requirements are getting teeth with fines and penalties as company executives are being held accountable for their employees' actions. Corporate executives and IT professionals alike are now becoming more concerned about what their employees are viewing and downloading from the Internet. However, we are starting to see a paradigm shift with regards to how companies are choosing to leverage the power of the Internet, most importantly around brand recognition, advertising which often involve social media. There is a growing need to give employees access to the Internet as a business aide, thus ensuring the

safety of company assets is becoming increasingly difficult.

Government regulations on Internet access and information security are being enforced by many countries and individual states: the Children's Internet Protection Act (CIPA) for schools and libraries, Japan's Internet Association's SafetyOnline2 to promote Internet filtering, HIPAA, Sarbanes-Oxley, Gramm-Leach-Bliley, and "duty of care" legal obligation legislation are just a few. Many independent reports and government agencies (such as the FBI) are now reporting that employees are the single highest risk and are the most common cause of network abuse, data loss, and legal action. Because employers can be ultimately held responsible for their employees' actions, many businesses are now working aggressively with their Human Resources departments to define acceptable Internet usage. Reports of Internet abuse include:

- Global consulting firm IDC reports that 30–40% of Internet access is being used for nonbusiness purposes.
- The American Management Association reports that 27% of Fortune 500 companies have been involved in sexual harassment lawsuits over their employees' inappropriate use of email and Internet.
- The Center of Internet Studies have reported that more than 60% of companies have disciplined employees over Internet and email use, with more than 30% terminating employees.

Numerous stories of employee dismissal, sexual harassment, and discipline with regard to Internet use can be found on the Internet[1]:

- The Recording Industry Association of America (RIAA) and the Motion Picture Association of America (MPAA) have relentlessly pursued legal action against schools, corporations, and individuals over the illegal downloading of music and movies from the Internet. The RIAA recently won a case against an Arizona company for $1 million.
- An oil and gas company recently paid $2.2 million to settle a lawsuit for tolerating a hostile work environment created by the downloading and sharing of Internet pornography.

To address these issues, companies are choosing to create Internet usage polices and develop education programs to train employees on how to safely use the Internet for browsing and communication in a manner that protects all parties from legal action and financial losses. To become compliant with many new policies and compliance regulations, corporations have chosen to start monitoring and controlling Web access. As companies are depending on the Internet for their businesses to flourish and cyber criminals are evolving and becoming increasingly more sophisticated, Web2.0 content filtering is becoming an essential part of every Internet and network security strategy.

2. WHY CONTENT FILTERING IS IMPORTANT

This section examines the motivating factors that each of the entities below consider when purchasing a content filtering solution:

- Schools and Libraries
- Commercial businesses
- Financial organizations
- Healthcare organizations
- Local, state, and federal government
- Parents

Each of the preceding faces different risks. But, they are all trying to solve one or more of the following challenges:

- Maintain compliance
- Protect company and client sensitive data (DLP: Data Leakage Prevention)
- Maximize employee productivity
- Avoid costly legal liabilities due to sexual harassment and hostile work environment lawsuits
- Preserve network and Internet resources
- Enforce company acceptable use policies (also known as Internet access policies)
- Control access to customer records and private data
- Monitor communications going into and out of a company
- Protect children

Let's look at the specific risks driving these entities to review their motivation to filter content.

Schools

In 2000, the United States Congress enacted the Children's Internet Protection Act (CIPA) to address concerns about children accessing harmful or obscene content over the Internet. This act imposes requirements on schools and libraries that receive discounts or funding for Internet access through the government's E-rate program. Schools and libraries may not receive E-rate assistance unless they can certify that they have an Internet safety

1. Stories of workers being dismissed for porn surfing: "IT manager fired for lunchtime Web surfing," www.theregister.co.uk/1999/06/16/it_manager_fired_for_lunchtime; "Xerox fires 40 in porn site clampdown," www.theregister.co.uk/2000/07/15/xerox_fires_40_in_porn; "41 District workers have been fired/suspended for visiting pornographic Web sites," www.wtopnews.com/?sid=1331641&nid=25.

policy that includes technology protection measures, which include blocking/filtering access to: (a) obscene content; (b) child pornography; or (c) anything harmful to minors, for all computers that are accessible by minors. For additional information on CIPA, please review http://www.fcc.gov/guides/childrens-internet-protection-act[2]

Why Content Filtering is Important

There are many reasons companies consider implementing content filtering, which can range from improving employee productivity, to blocking web based threats, and even prevent data leaks. In commercial businesses the motivations will vary based on the industry, any compliance or regulation that might be required, and the organization's culture. Many companies prefer to allow employees to access recreational content for a limited amount of time each day, this helps to strike a balance of maintaining employee productivity and employee morale.

Financial Organizations

Financial organizations have unique privacy and security concerns due to the fact that they need to protect their customer's personally identifiable information such as credit card numbers, Social Security numbers, and other financial related information which means they have no room for error. Data leakage is a very serious concern for financial institutions because an event can lead to financial losses, loss of reputation, competitive information, customers, and legal action. This risk tends to be the primary driver for financial institutions to consider content filtering and reporting on employees' Internet activity.

Healthcare Organizations

In today's fast-paced technological world, healthcare organizations must keep their costs and risks down. When employees inadvertently disclose sensitive information through intentional or unintentional means, it puts health organizations in financial and legal jeopardy. To help Healthcare organizations protect confidential patient information, The Department of Health and Human Services created the Health Insurance Portability and Accountability Act of 1996 (HIPAA). This act sets national standards for the protection of certain health information and outlines how medical records should be securely shared for legitimate purposes. HIPAA also requires that an audit log be kept for medical record access, which is one of the many reasons healthcare organizations implement content filtering.

Internet Service Providers

ISPs have unique motivations with regards to content filtering. One of the primary reasons is, ISPs have to comply and produce logs in response to requests from law enforcement, including assisting the government with the USA PATRIOT Act and CIPA, and must have technology in place to monitor the activities of their customers.

U.S. Government

In the United States, the threat from websites that host malicious software presents a significant risk which can be easily managed through content filtering. In addition, the military is attempting to reduce the number of Internet connection points from thousands to hundreds, to reduce the risks we've described. National secrets, military plans, and information about soldiers and citizens cannot be exposed to our enemies by inadvertent surfing to the wrong places and downloading Trojans, keystroke loggers, or other types of malware. Additionally, logging is another key requirement that the U.S. Government leverages content filtering for; it provides classification of visited sites, which makes identifying trends and reading reports easier and more meaningful.

Other Governments

Other countries only allow their government employees access to whitelisted sites, to control access to news and other information that censors determine to be inappropriate. In Russia the ruling party routinely shuts off access to political rivals' Web sites. China has deployed and maintains a new and virtual "Great Firewall of China" (also known as the Golden Shield Project). Ministry of Public Security Authorities determine sites that they believe represent an ideological threat to the Chinese Communist Party and then prevent their citizenry surfing, blogging, and emailing to blocked sites with sophisticated content-filtering methods, including IP address blocking and even DNS cache poisoning. Russia, Tibet, North Korea, Australia, China, Iran, Cuba, Thailand, Saudi Arabia, and many other repressive governments use similar technology. The ironic part of these massive content-filtering efforts is that many U.S.-based technology companies have been involved in their construction, sometimes as a contingency for doing business in China or other censoring countries.[3]

2. Children's Internet Protection Act http://www.fcc.gov/guides/childrens-internet-protection-act

3. U.S. companies' involvement in the "Golden Shield" Chinese content-filtering project, www.forbes.com/forbes/2006/0227/090.html, www.businessweek.com/magazine/content/06_08/b3972061.htm.

Libraries

The use of Internet filters or content-control software varies widely in public libraries in the United States, since Internet use policies are established by local library boards. As mentioned above, many libraries adopted Internet filters after Congress conditioned the receipt of universal service discounts with the use of Internet filters through the Children's Internet Protection Act (CIPA). Other libraries do not install content control software, believing that acceptable use policies and educational efforts address the issue of children accessing age-inappropriate content while preserving adult users' right to freely access information. Some libraries use Internet filters on computers used by children only. Some libraries that employ content-control software allow the software to be deactivated on a case-by-case basis, based on submitting an application to a librarian; libraries that are subject to CIPA are required to have a policy that allows adults to request that the filter be disabled without having to explain the reason for their request. Libraries have other legal challenges as well. In 1998, a U.S. federal district court in Virginia ruled that the imposition of mandatory filtering in a public library violates the First Amendment of the U.S. Bill of Rights.[4]

Parents

There are many ways parents can protect their children from inappropriate material on the Internet; and, web content filtering should be one of them. There are dozens of products for parents to consider when it comes to content filtering, as many browsers and consumer-grade endpoint security packages include a content filter, all with varying degrees of flexibility. Parents should research which solution will work best for their family based on children's age and whether or not the children are technology savvy. Additional features such as, a password based bypass, reporting, and parental alerting when an attempt to access blocked content is made, should also be considered before making a decision.

3. CONTENT CATEGORIZATION TECHNOLOGIES

There are many technologies that can be used to categorize content. Most commercial products use a number of techniques together to optimize their capability, and below is a list of some of the most popular techniques.

4. "Library content filtering is unconstitutional," Mainstream Loudon v. Board of Trustees of the Loudon County Library, 24 F. Supp. 2d 552 (E. D. Va. 1998).

Keyword Lists

The keyword lists method allows the creation of a blacklist dictionary that contains keywords or phrases. URLs and Web content are compared against the blacklist to block unauthorized Web sites. This technology relies on a manual update process, with vendors providing blacklists as starting points, requiring customers to manually update/tune the lists by adding or excluding keywords. Since updates are usually performed manually, filtering accuracy may be impacted due to the rate at which content changes and the fact that the Internet has progressed from a publishing model, to a community model with interactive content. This progression is making it difficult to rely on keyword lists as a primary tool. For example, if a user goes to the BBC sports site to lookup stats from the World Cup Soccer Tournament, their browser will make over 15 connections to servers in 6 domains possibly spanning multiple countries and languages. Simple keyword lists won't provide protection if one of those domains is compromised with malware or contains material that would be considered offensive While this feature is widely available in many commercial and freeware products today, it has largely been replaced with more advanced technologies such as Bayesian and reputation analysis.

URL Lists

URL lists contain full and/or partial URLs, which are compared to the URL in an HTTP get request. This technology is different from keyword lists because keyword lists look at content both within the HTML page and the URL, whereas URL lists just look at the address in the get request. However, URL lists tend to be a little more precise when looking to block a specific web site. One similarity between URL and keyword lists is that both are maintained through manual updates, which is why neither is used as primary technologies when categorizing content. The drawback to technologies that require manual updating is depending on the frequency of the updates, the lists may fall out of compliance with the corporate policy. While there are many uses for this specific technology, most organizations use URL lists to locally recategorize content that has been rated using another technology.

Content Categorization as a Service

The goal of organizing websites into categories based on the content in a web site is to make administration and updating of web filters easier and more accurate. By leveraging static categories, system administrators can simply choose which categories to block and which to

allow. This type of service is typically offered as part of a fee based subscription but there are some free services that are available as well. The key to picking which service to use should be based on cost, technology used to rate content, granularity of categories, and frequency of updates. Today, most web-filtering companies utilize multiple technologies to accurately rate content, some of which are proprietary. Below is a list of the most popular techniques:

- Keyword and pattern matching
- Multilanguage pattern matching
- Bayesian Analysis
- Anti-Virus and Anti-Malware tools
- Traffic patterns
- Content Based Image Filtering
- Domain registration and reputation (Whois)
- Counter intelligence (subscribing to proxy avoidance, and malware hosting lists)

Bayesian Filters

Particular word combinations and phrases have certain probabilities of occurring together on Web sites. For instance, the word "breast" can have a number of contexts depending on other words around it. The first could be cooking, as in Chicken Breast, second could be health as in Breast Cancer Awareness, and finally it could be porn. Bayesian Analysis, as it is used in web site content categorization, looks for patterns such as the words: breast, cooking, baking, seasoning, degrees, recipe. In this example the website would be classified as food related. While Bayesian analysis works well for this purpose there is a drawback which is it needs to be trained to recognize these patterns. To train the filter, the user or external "grader" must manually indicate whether a new web site is a porn site, cooking site or healthcare related site. Eventually, the filter will learn enough from the training process to make decisions on it's own and can continue to learn new word patterns without too much human tuning. Typically most fee based subscription services, such as the ones we will explore later in this chapter, have gone through the process of training their filters, so you won't have to.

Content Labeling

Content labeling is considered another form of content-control. The Internet Content Rating Association (ICRA), now part of the Family Online Safety Institute, developed a content rating system to self-regulate online content providers. Using an online questionnaire, a Webmaster describes the nature of their Web content. A small file is generated that contains a condensed, computer-readable digest of this description that can then be used by content-filtering software to block or allow that site.

ICRA labels are deployed in a couple of formats. These include the World Wide Web Consortium's Resource Description Framework (RDF) as well as Platform for Internet Content Selection (PICS) labels used by Microsoft's Internet Explorer Content Advisor.

ICRA labels are an example of self-policing and self-labeling. Similarly, in 2006 the Association of Sites Advocating Child Protection (ASACP) initiated the Restricted to Adults (RTA) self-labeling initiative. The RTA label, unlike ICRA labels, does not require a Webmaster to fill out a questionnaire or sign up to use. Like ICRA, the RTA label is free. Both labels are recognized by a wide variety of content-control software.

Content-Based Image Filtering (CBIF)

The latest in content-filtering technology is CBIF. All the text-based content-filtering methods use knowledge of a site and text matching to rate it's content. This technique makes it impossible to filter visual and audio media. Content-based image filtering may resolve this issue, as shown in Figure e66.1 below. The method consists of examining the image itself for flesh tone patterns, detecting objectionable material and then blocking the offending site.

Step 1: Skin Tone Filter

First the images are filtered for skin tones. The color of human skin is created by a combination of blood (red) and melanin (yellow, brown). These combinations restrict the range of hues that skin can possess (except for people from the planet Rigel 7). In addition, skin has very little texture. These facts allow us to ignore regions with high-amplitude variations and design a skin tone filter to separate images before they are analyzed.

Tip: Determine if the image contains large areas of skin color pixels.

Step 2: Analyze

Since we have already filtered the images for skin tones, any images that have very little skin tones will be accepted into the repository. The remaining images are then automatically segmented and their visual signatures are computed.

Tip: Automatically segment and compute a visual signature for the image.

Step 3: Compare

The visual signatures of the potentially objectionable images are then compared to a predetermined reference

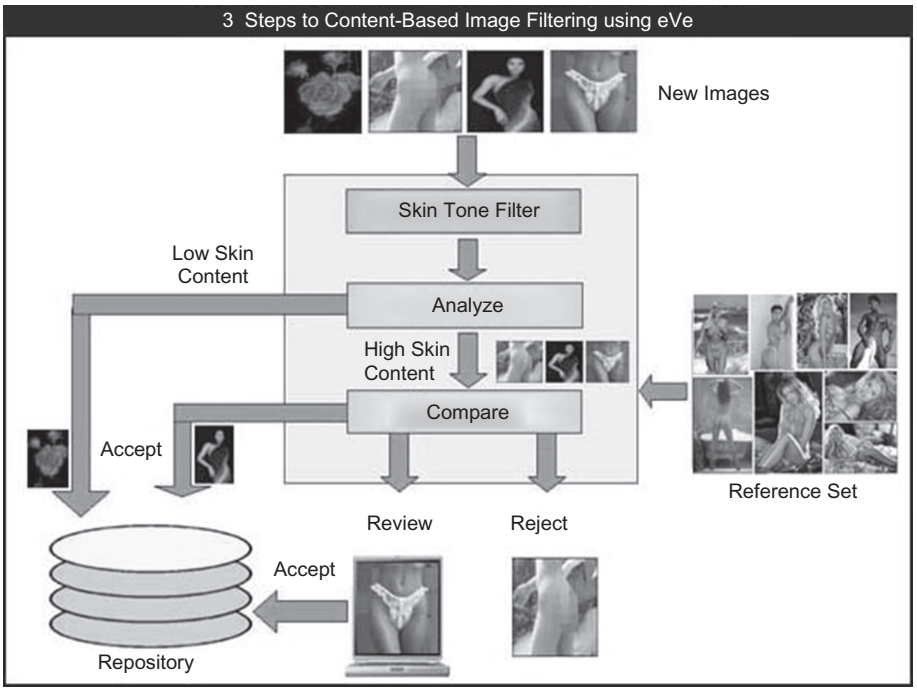

FIGURE e66.1 Content-based image filtering.

data set. If the new image matches any of the images in the reference set with over 70% similarity, the image is rejected. If the similarity falls in the range of 40−70%, that image is set aside for manual intervention. An operator can look at these images and decide to accept or reject them. Images that fall below 40% are accepted and added to the repository. These threshold values are arbitrary and are completely adjustable.[5]

Tip: Match the new image against a reference set of objectionable images and object regions.

4. PERIMETER HARDWARE AND SOFTWARE SOLUTIONS

There are several different technologies that help facilitate Web monitoring, logging, and filtering of HTTP, FTP sites, and other Web-related traffic. The methods available for monitoring and controlling Internet access range from manual to fully automated systems designed to scan, inspect, rate, and control Web activity in real-time.

Solutions can range from software that runs on Intel-based servers to purpose built appliances, like firewalls and proxies, that offer an administrator many different integration options. Two deployment options are inline and out-of-band, with each having their own benefits and risks, so it's very important to consider your deployment while deciding which solution to use.

Inline deployments are typically seen as the easiest from an installation perspective because you don't have to re-route traffic to the appliance, it's pretty much a drop in deployment. This design does tend to have some risks depending on the type of solution you decide to deploy. If the device is an application proxy, some applications might not function correctly which means you will have to spend time tweaking and bypassing non-proxy friendly traffic, which in a large network can be painful for users and time consuming for administrators. Please examine Table e66.1 for the risks and benefits of each deployment type.

Proxy Server vs. Firewall

There tends to be a lot of debate around which is device provides better security, a proxy server or a stateful network firewall. In early versions of both of these devices they served very different roles in network and Internet security. Early stateful firewalls operated at a much lower level in the OSI model than proxy servers do. The early firewall would allow and deny traffic based on OSI layers 3 and 4, and was great at performing this function. Proxy servers, sometimes called application proxies, are an intermediary between a client and a server operating at OSI layer 7, the application layer. This means that if a user Bob is sitting behind a proxy server that is connected to the Internet, and opens the browser on his workstation

5. Envision Search Technologies (permission to use), www.evisionglobal.com/index.html.

TABLE E66.1 Risks and Benefits of each Deployment Type.

	Benefits	Risk
Inline Deployment	• Visibility in to all Internet traffic • Easy to deploy • Easy to manage and troubleshoot	• Application compatibility issues • Single point of failure • Potential bottleneck • Limited scaling options for high availability • Requires outage to install
Out-of-band Deployment	• Forward only interesting traffic • Greater scalability and high availability options • Doesn't require network outage to install	• Often requires changes to network and/or client configurations - More difficult to deploy • Lack of visibility in to all Internet traffic • Difficult to troubleshoot

and requests a web page, his browser will ask the proxy server for that web page. The proxy server will then go out and get the page Bob requested from the Internet, maybe scan it for threats, make sure it's allowed, and send it over to Bob. The power of a proxy server is it can make decisions based on application layer data and even block granular functions within an application. As an example let's look at HTTP, an application proxy can determine if the traffic going over TCP port 80 is HTTP and if it's not, it can choose to block it. Going a little deeper, if someone attempted to tunnel Peer-to-Peer traffic over TCP port 80, they would not be able to get past an application proxy, but a firewall wouldn't be able to distinguish between these two applications because the firewall policy would likely allow TCP port 80 regardless of which application was using it. Another argument for using an application proxy is if you wanted to protect an application, such as a banking application, that was being published to the Internet, which is called a reverse proxy. A forward proxy requests resources from the Internet on behalf of users inside a LAN, and a reverse proxy services requests from the Internet and makes requests to a server or application sitting behind it; usually on a protected LAN or DMZ not allowed to connect directly to the Internet. Since a proxy server is an intermediary between a client and a server, it prevents a bank's customers, and more importantly hackers, from directly communicating with the server hosting the banking application. The administrators of the reverse proxy can choose to only allow certain HTTP methods, like the method GET, for most of the application's pages and only allow the method POST for pages or scripts that require use of that method. This allows a security administrator to mitigate some of the risk and gain greater control of Internet facing applications. There are many other benefits of using a proxy server that are outside the scope of this book and you should be familiar with them before deciding whether or not to use one. For all of the proxy's benefits, it's does have a number of limitations, such as it doesn't support a wide range of applications or protocols and they tend to require a lot of time and expertise to install not to mention ongoing administration. If they are not configured correctly and maintained regularly, the protection they provide can be greatly reduced and/or introduce issues in to the application that wouldn't otherwise be there.

The landscape changed when next-generation application firewalls became available to the market. Now, you could have a single device that had application awareness in addition to being a stateful firewall. The biggest difference between an application firewall and a proxy is an application firewall takes a client's original request and modifies the header, performs some inspections and if allowed, sends the original request to its destination. Whereas a proxy server will receive the request from the client, and initiate its own request to the destination to fulfill the client's request. Application firewalls quickly became a favorite of network administrators who were looking to protect their users while using the Internet and to protect applications that were published to the Internet. Since application firewalls have visibility all the way up to the application layer they can perform many of the functions that an application proxy does, like application validation, and content filtering.

Next-Generation firewalls (NGFW) are also known as UTM Appliances or Unified Threat Management Appliances. There are some people that consider these terminologies to be completely different; the fact is that UTM appliances are the same as NGFW as they both have many of the same features including application layer (OSI layer 7) firewall. While marketing departments are likely to blame for this confusion, you should consider them to be the same since they both describe the same type of technology and you should differentiate based on individual features that the various vendors offer. UTM appliances offer a wide range of services such as NGFW,

AV scanning, IPS/IDS, Web Filtering, and Application Control just to name a few.

Internet Gateway-Based Products/Unified Threat Appliances

One of the fastest-growing markets over the past several years has been the UTM appliance space, as shown in Figure e66.2. Let's review who the leaders are in both the UTM and Proxy markets. While there are many tools you can use to compare the products from the market leaders, like Gartner's Magic Quadrant, there is no replacement for a good old-fashioned proof of concept or "Bake-off" to determine which product will work best in your environment!

Fortinet

Fortinet's unique approach to protecting networks against the latest vulnerabilities involves several key security components. By combining many key security functions into one hardware platform called the FortiGate Multifunction Security Firewall, which benefits from their custom Application Specific Integrated Circuit (ASIC) accelerated security processor, Fortinet has developed the world's first Dynamic Threat Prevention System. With this platform, Fortinet is able to quickly identify threats and proactively block them at the network level before they reach the endpoints to cause damage. Customers can create custom protection policies by turning on any of the security functions—Stateful Firewall, IPSec and SSL VPN, Antivirus, IDS & IPS, Web Content Filtering, Anti-Spam, and Bandwidth Shaping—in any combination and apply it to their traffic. Fortinet's Web Content Filtering technology allows customers to take a wide variety of actions to inspect, rate, and control Web traffic. Fortinet uses their own cloud-based service called FortiGuard for threat research, which provides web filtering, anti-virus, IPS, application control, etc., for all of their security platforms including their FortiGate Firewall.[6]

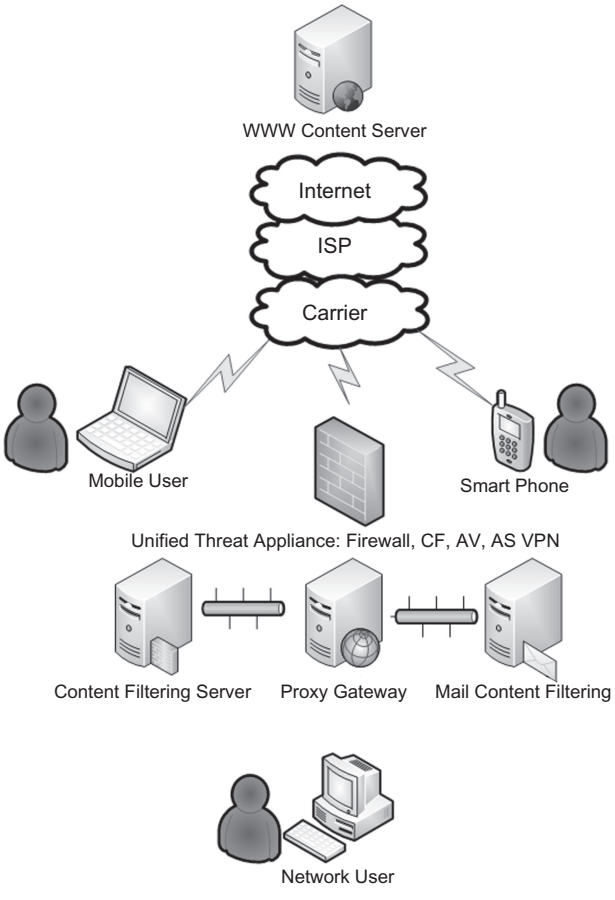

FIGURE e66.2 The UTM appliance space.

Websense

Websense started out as just a web filtering solution that would plug in to a customer's existing firewall infrastructure. They have always been known for great flexibility in deployment and perhaps the most granular category list, and for a long time were considered the leaders in reporting, for which they received numerous awards. Today, Websense is branching out to provide solutions in other growing areas of Internet security such as email protection, DLP, and mobile security, in addition to offering their own hardware appliances. While they may offer some similar features to UTM, they are not considered a UTM company since they don't offer firewall, which is a signature feature of UTM vendors.

Blue Coat

Blue Coat is widely considered the industry leader in Application Proxies. They started as a company that sold caching appliances to ISPs and large companies and built their security business around that product. Today they offer a cloud-based web filtering technology that works with their proxy platform and has recently introduced a total cloud-based service, which acts like a "proxy in the cloud" that users can access anywhere in the world from virtually any device. The benefit of the total cloud-based solution is it uses a subscription model, which doesn't require the customer to make a significant up-front investment in hardware. This new service allows smaller companies, that couldn't afford high-end content filtering, to leverage the same protection as larger companies. Blue Coat also offers solutions outside of their core web filtering and proxy products, which include anti-virus scanning, WAN optimization, traffic shaping and DLP. Their

6. Fortinet Multi-Threat Security Solution, www.fortinet.com/doc/whitepaper/Webfilter_applicationNote.pdf.

web filtering solution has a real-time feedback component where users' unrated URLs are automatically submitted to the cloud for real-time analysis, often receiving a real-time response. They have also connected their anti-virus appliance to their cloud, which provides real-time visibility in to sites that contain malware and they share this data amongst their customer base. There are many other companies including Secure Computing, Aladdin Knowledge Systems, Finjan, Marshal, FaceTime Communications, Webroot Software, Clearswift, CP Secure, IronPort Systems, ISS/IBM Proventia, Trend Micro, McAfee, MessageLabs, Barracuda Networks, ContentKeeper Technologies, Computer Associates, Cymphonix, Pearl Software, and St. Bernard, that all compete in the content filtering space each with their own unique features.

PC Based

PC software such as Norton Internet Security includes parental controls. Operating Systems such as Mac OS X and Microsoft's Windows Vista operating system also include content-control software. Other PC-based content-filtering software products are CyberPatrol, Cybersitter, EnoLogic NetFilter, iProtectYou Pro Web Filter, Net Nanny, Norton Internet Security, Safe Eyes Platinum, SentryPC, Bess, Crayon Crawler, Cyber Snoop, Covenant Eyes, K9 Web Protection, Naomi, Scieno Sitter, Sentry Parental Controls, Websense, Windows Live Family Safety, Windows Vista Parental Control, WinGate, X3Watch, and PlanetView. For the Apple users there are many options and the list is growing, some popular titles include Covenant Eyes, DansGuardian, Intego, and Mac OS X Parental Controls.

Remote corporate PCs and now the ever-increasing sophistication and capabilities of smart phones and tablets make them challenges for content-filtering deployments (see Figure e66.3). There are multiple ways to deal with these challenges one of which is to load a client on the device that provides local or cloud-based filtering, and VPNs with split tunneling disabled to force all traffic to go out the corporate protected Internet connection.

ISP-Based Solutions

Many ISPs offer parental control browser-based options, among them Charter Communications, EarthLink, Yahoo!, and AOL (see Figure e66.4). Cleanfeed is offered by British Telecom in the U.K. and is an ISP administered content-filtering system that targets child sexual abuse content using offensive image lists from the Internet Watch Foundation. Many ISPs in the US are starting to offer families a cloud-based content filtering solution for an additional monthly fee.

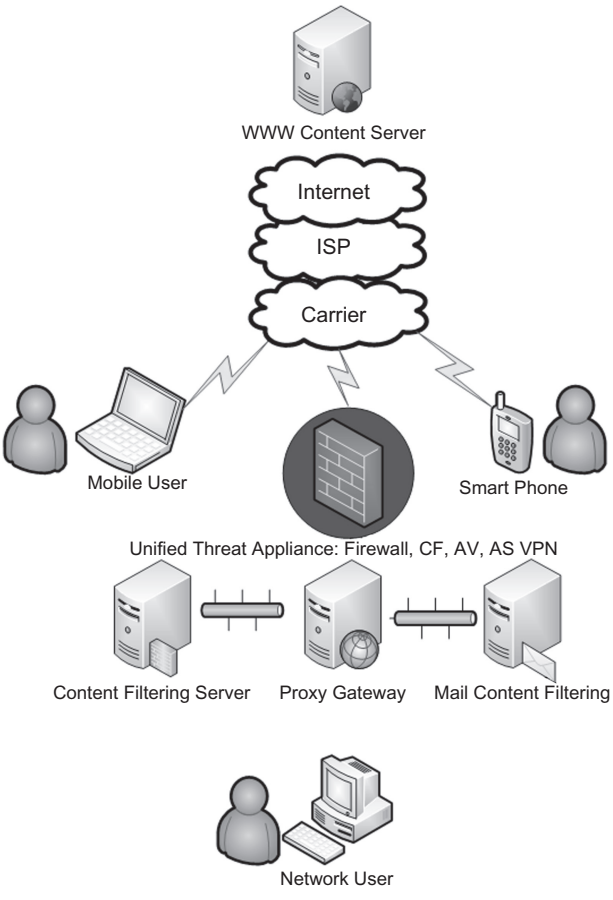

FIGURE e66.3 Content-filtering deployments.

Internet Based: Search Engine Safe Search

Safe Search is a feature that today is supported by most major search engines like Yahoo! and Google. Safe Search filters pornography from search engine results and can be toggled on or off in the browser. Most major content filtering companies have a feature to force it on regardless of the browser setting. In 2008 the Google search engine adapted a software program to faster track child pornography accessible through its site. The software is based on a pattern recognition engine. Content distribution networks that are attached to the Internet, such as Akamai, Limelight, Panther Express, EdgeCast, CDNetworks, Level 3, and Internap, manage the content in their networks and will not distribute offensive images (see Figure e66.5). In addition, some argue that using content-control software may violate Articles 13 and 17 of the Convention on the Rights of the Child.[7]

7. Convention on the Rights of the Child, www.unhchr.ch/html/menu3/b/k2crc.htm.

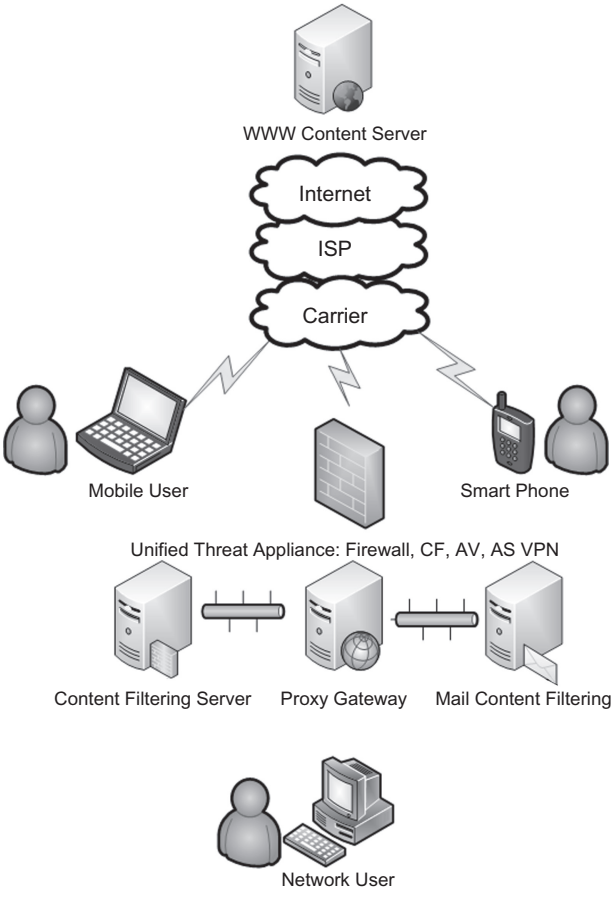

FIGURE e66.4 Many ISPs offer parental control browser-based options.

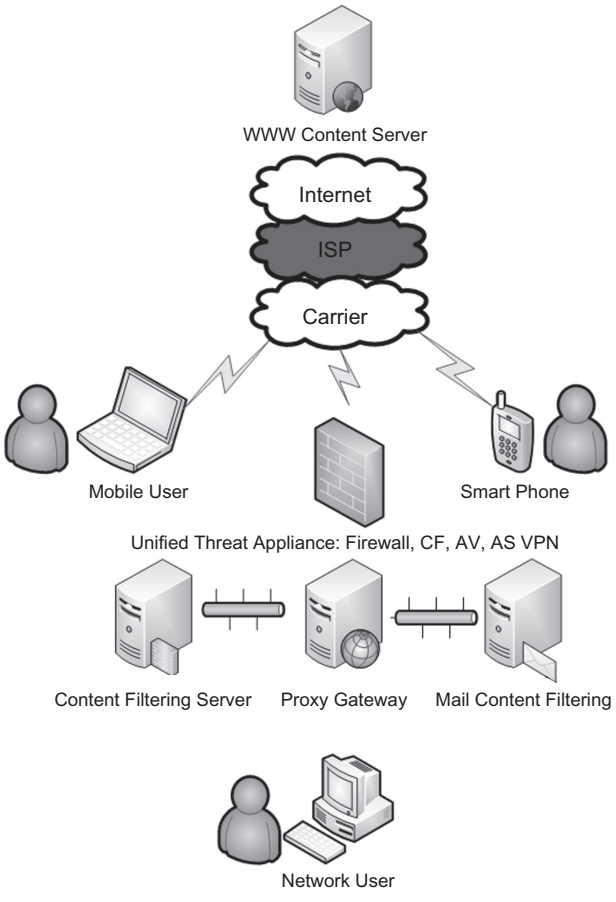

FIGURE e66.5 Content distribution networks that are attached to the Internet, such as Akamai, Limelight, Panther Express, EdgeCast, CDNetworks, Level 3, and Internap, manage the content in their networks and will not distribute offensive images.

5. CATEGORIES

It's tricky to determine the number of active web sites that are on the Internet today and due to its seemingly unstoppable growth, it will likely always remain difficult to determine an exact number. Most content-filtering companies have only rated a fraction of the entire Internet and do so using many different techniques that range from crawlers that attempt access every IP address to real-time services that examine URLs as users click on them. Social Media sites have traditionally been difficult to classify due to the broad range of services they offer. Many companies have chosen to apply multiple categories per URL that way they can distinguish between Social Networking: Games and Social Networking: Shopping and just Social Networking. It's important to consider how these sites are classified when choosing a solution because your corporate policy may be to allow social networking sites but deny games and shopping from both social networking sites and non-social networking sites (see checklist: "An Agenda For Action For Implementing Content-Control Filtering Software"). Table 66.2 below is the category list from Fortinet's Fortiguard service. Figure e66.6 shows the flow of categorization updates to the end user.

6. LEGAL ISSUES

There are many legal issues to consider in content filtering. And once you think you have a handle on your particular organizational requirements and have ensured that they are legal, a court will make a ruling that changes the game. A number of Internet technology issues and related challenges have not yet been fully addressed by legislatures or courts and are subject to a wide range of interpretation. For example, virtual child pornography, pornographic images and text delivered by SMS messages, sexual age-play in virtual game worlds, the soft porn Manga genre of Lolicon and Rorikon are all challenges to current laws and issues that will need to be addressed as our society comes to grips with the Internet

TABLE 66.2 79 Categories are Organized into the following 6 Main Groups.

General Interest - Business	Armed Forces, Business, Finance and Banking, General Organizations, Government and Legal Organizations, Information and Computer Security, Information Technology, Search Engines and Portals, Secure Websites, Web Hosting, Web-based Applications
General Interest - Personal	Advertising, Arts and Culture, Brokerage and Trading, Child Education, Content Servers, Digital Postcards, Domain Parking, Dynamic Content, Education, Entertainment, Folklore, Games, Global Religion, Health and Wellness, Instant Messaging, Job Search, Medicine, Meaningless Content, News and Media, Newsgroups and Message Boards, Personal Privacy, Personal Vehicles, Personal Websites and Blogs, Political Organizations, Real Estate, Reference, Restaurants and Dining, Shopping and Auction, Social Networking, Society and Lifestyles, Sports, Travel, Web Chat, Web-based Email
Bandwidth Consuming	File Sharing and Storage, Freeware and Software Downloads, Internet Radio and TV, Internet Telephony, Peer-to-Peer File Sharing, Streaming Media and Downloads
Controversial	Abortion, Adult Materials, Advocacy Groups, Alcohol, Alternative Beliefs, Dating, Extremist Groups, Gambling, Lingerie and Swimsuit, Marijuana, Nudity and Risque, Pornography, Sex Education, Sport Hunting and War Games, Tobacco, Weapons
Potentially Liable	Child Abuse, Discrimination, Drug Abuse, Hacking, Illegal or Unethical, Plagiarism, Proxy Avoidance, Violence
Security Risk	Malicious Websites, Phishing, Spam URLs

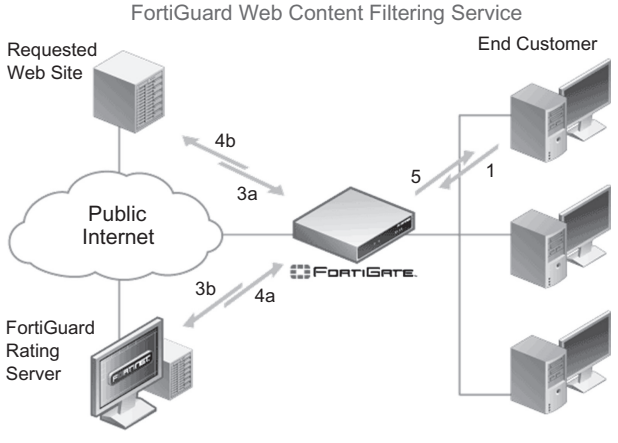

FIGURE e66.6 The flow of categorization updates to the end user.

and what is "out there." The following discussion centers on the most relevant laws in the content-filtering space.

Federal Law: ECPA

The Electronic Communications Privacy Act (ECPA)[8] allows companies to monitor employees' communications when one of three provisions are met: one of the parties has given consent, there is a legitimate business reason, or the company needs to protect itself.

8. U.S. Code: Wire and Electronic Communications Interception and Interception of Oral Communications, www.law.cornell.edu/uscode/18/usc_sup_01_18_10_I_20_119.html.

If your company has no content access policy in place, an employee could argue that he or she had a reasonable expectation of privacy. However, if the company has implemented a written policy whereby employees are informed about the possibility of Web site monitoring and warned that they should not have an expectation of privacy, the company is protected from this type of privacy claim.

CIPA: The Children's Internet Protection Act

CIPA provisions have both the "carrot and the stick." The U.S. government will pay you for equipment to access the Internet, but you have to play by its rules to get the money! Having been rebuffed by the courts in its previous efforts to protect children by regulating speech on the Internet, Congress took a new approach with the Children's Internet Protection Act (CIPA). See, for example *Reno v. ACLU*, 521 U.S. 844 (1997) (overturning the Communications Decency Act of 1996 on First Amendment grounds). With CIPA, Congress sought to condition federal funding for schools and libraries on the installation of filtering software on Internet-ready computers to block objectionable content.

CIPA is a federal law enacted by Congress in December 2000 to address concerns about access to offensive content over the Internet on school and library computers. CIPA imposes certain types of requirements on any school or library that receives funding for Internet access or internal connections from the E-rate program, which makes certain communications technology more affordable for eligible schools and libraries. In early

An Agenda for Action for Implementing Content-Control Filtering Software

Social networking sites and non-social networking sites are organized into between 10 and 90 various categories. Typical subjects of content-control software include (Check All Tasks Completed):

_____**1.** Illegal content with reference to the legal domain being served by that company.

_____**2.** Promote, enable, or discuss system cracking, software piracy, criminal skills, or other potentially illegal acts.

_____**3.** Sexually explicit content, such as pornography, erotica, nudity, and erotic discussions of sexual topics such as sexuality or sex. Promote, enable, or discuss promiscuity, lesbian, gay, bisexual, transsexual, sexual activity outside of marriage, or other lifestyles seen to be immoral or alternative.

_____**4.** Contain violence or other forms of graphic or "extreme" content.

_____**5.** Promote, enable, or discuss bigotry or hate speech.

_____**6.** Promote, enable, or discuss gambling, recreational drug use, alcohol, or other activities frequently considered to be vice.

_____**7.** Are unlikely to be related to a student's studies, an employee's job function, or other tasks for which the computer in question may be intended, especially if they are likely to involve heavy bandwidth consumption.

_____**8.** Are contrary to the interests of the authority in question, such as Web sites promoting organized labor or criticizing a particular company or industry.

_____**9.** Promote or discuss politics, religion, health or other topics.

_____**10.** Prevent people who are hypochondriacs from viewing Web sites related to health concerns.

_____**11.** Include social networking opportunities that might expose children to predators.

_____**12.** Potentially liable: drug abuse, folklore, hacking, illegal or unethical, marijuana, occult, phishing, plagiarism, proxy avoidance, racism and hate, violence, Web translation.

_____**13.** Controversial: abortion, adult materials, advocacy groups/organizations, alcohol, extremist groups, gambling, lingerie and swimwear, nudity, pornography, sex education, sport hunting and war games, tasteless, tobacco, weapons.

_____**14.** Potentially nonproductive: Advertising, brokerage and trading, digital postcards, freeware, downloads, games, instant messaging, newsgroups and message boards, Web chat, Web-based email.

_____**15.** Potentially bandwidth consuming: Internet radio and TV, Internet telephony, multimedia download, peer-to-peer file sharing, personal storage.

_____**16.** Potential security risks: Malware, spyware.

_____**17.** General interest: Arts and entertainment, child education, culture, education, finance and banking, general organizations, health and wellness, homosexuality, job search, medicine, news and media, personal relationships, personal vehicles, personal Web sites, political organizations, real estate, reference, religion, restaurants and dining, search engines, shopping and auction, society and lifestyles, sports, travel.

_____**18.** Business oriented: Armed forces, business, government and legal organizations, information technology, information/computer security

_____**19.** Others: Content servers, dynamic content, miscellaneous, secure Web sites, Web hosting.

2001, the FCC issued rules implementing CIPA. CIPA made amendments to three federal funding programs: (1) the Elementary and Secondary Education Act of 1965, which provides aid to elementary and secondary schools; (2) the Library Services Technology Act, which provides grants to states for support of libraries; and (3) the E-Rate Program, under the Communications Act of 1934, which provides Internet and telecommunications subsidies to schools and libraries. The following are what CIPA requires[9]:

- Schools and libraries subject to CIPA may not receive the discounts offered by the E-Rate Program unless they certify that they have an Internet safety policy and technology protection measures in place. An Internet safety policy must include technology protection measures to block or filter Internet access to pictures that (a) are obscene, (b) are child pornography, or (c) are harmful to minors (for computers that are accessed by minors).
- Schools subject to CIPA are required to adopt and enforce a policy to monitor online activities of minors.
- Schools and libraries subject to CIPA are required to adopt and implement a policy addressing: (a) access by minors to inappropriate matter on the Internet; (b) the safety and security of minors when using electronic mail, chat rooms, and other forms of direct electronic communications; (c) unauthorized access, including so-called "hacking," and other unlawful activities by minors online; (d) unauthorized

9. What CIPA Requires, www.fcc.gov/cgb/consumerfacts/cipa.html.

disclosure, use, and dissemination of personal information regarding minors; and (e) restricting minors' access to materials harmful to them. Schools and libraries are required to certify that they have their safety policies and technology in place before receiving E-Rate funding, as follows:

- CIPA does not affect E-Rate funding for schools and libraries receiving discounts only for telecommunications, such as telephone service.
- An authorized person may disable the blocking or filtering measure during any use by an adult to enable access for bona fide research or other lawful purposes.
- CIPA does not require the tracking of Internet use by minors or adults.

"Harmful to minors" is defined under the *Act as: Any picture, image, graphic image file, or other visual depiction that (i) taken as a whole and with respect to minors, appeals to a prurient interest in nudity, sex, or excretion; (ii) depicts, describes, or represents, in a patently offensive way with respect to what is suitable for minors, an actual or simulated sexual act or sexual contact, actual or simulated normal or perverted sexual acts, or a lewd exhibition of the genitals; and (iii) taken as a whole, lacks serious literary, artistic, political, or scientific value as to minors.*

Court Rulings: CIPA from Internet Law Treatise

On June 23, 2003, the U.S. Supreme Court reversed a District Court's holding in *United States v. American Library Ass'n*, 539 U.S. 194 (2003).[10] It held that the use of Internet filtering software does not violate library patrons' First Amendment rights. Therefore, CIPA is constitutional and a valid exercise of Congress's spending power.

The Court held, in a plurality opinion, that libraries' filtering of Internet material should be subject to a rational basis review, not strict scrutiny. It explained that, because collective decisions regarding printed material have generally only been subject to a rational basis review, decisions regarding which Web sites to block should likewise be subject to the same test. It reasoned that libraries are no less entitled to make content-based judgments about their collections when they collect material from the Internet than when they collect material from any other source.

Further, it reasoned that heightened judicial scrutiny is also inappropriate because "Internet access in public libraries is neither a 'traditional' nor a 'designated' public forum" *(Id.* at 2304). Therefore, although filtering software may overblock constitutionally-protected speech and a less restrictive alternative may exist, because the government is not required to use the least restrictive means under a rational basis review, CIPA is nonetheless constitutional.

Moreover, the Court held that Congress did not exceed its spending power by enacting CIPA because, when the government uses public funds to establish a program, it is entitled to define its limits. By denying federal funding, the government is not penalizing libraries that refuse to filter the Internet, or denying their rights to provide their patrons with unfiltered Internet access. Rather, it "simply reflects Congress' decision not to subsidize their doing so"(*Id.* at 2308).[11]

The Trump Card of Content Filtering: The "National Security Letter"

The FBI, CIA, or DoD can issue an administrative subpoena to ISPs for Web site access logs, records, and connection logs for various individuals. Along with a gag order, this letter comes with no judicial oversight and does not require probable cause. In 2001, Section 505 of the PATRIOT Act powers were expanded for the use of the NSL. There are many contentious issues with these laws, and the Electronic Frontier Foundation and the American Civil Liberties Union (ACLU) are battling our government to prevent their expansion and open interpretation.[12]

State of Texas: An Example of an Enhanced Content-Filtering Law

Texas state law requires all Texas ISPs to link to blocking and filtering software sites. In 1997, during the 75th Regular Session of the Texas Legislature, House Bill 1300 was passed. HB 1300 requires ISPs to make a link available on their first Web page that leads to Internet "censorware" software, also known as "automatic" blocking and screening software. The two most important portions of the law are shown here:

Sec. 35.102. SOFTWARE OR SERVICES THAT RESTRICT ACCESS TO CERTAIN MATERIAL ON INTERNET.

(a) A person who provides an interactive computer service to another person for a fee shall provide free of charge to each subscriber of the service in this state a link leading to fully functional shareware, freeware, or demonstration versions of software or to a service that, for at least one operating system,

10. CIPA and E-Rate Ruling, www.cdt.org/speech/cipa/030623decision.pdf.

11. IETF Fights CIPA, http://ilt.eff.org/index.php/Speech:_CIPA.

12. ACLU Sues Over Internet Privacy, Challenges ISPs Being Forced to Secretly Turn Over Customer Data, www.cbsnews.com/stories/2004/04/29/terror/main614638.shtml.

enables the subscriber to automatically block or screen material on the Internet.

(b) A provider is considered to be in compliance with this section if the provider places, on the provider's first page of world wide Web text information accessible to a subscriber, a link leading to the software or a service described by Subsection (a). The identity of the link or other on-screen depiction of the link must appear set out from surrounding written or graphical material so as to be conspicuous.

Sec. 35.103. CIVIL PENALTY.

(a) A person is liable to the state for a civil penalty of $2,000 for each day on which the person provides an interactive computer service for a fee but fails to provide a link to software or a service as required by Section 35.102. The aggregate civil penalty may not exceed $60,000.[13]

(b) The attorney general may institute a suit to recover the civil penalty. Before filing suit, the attorney general shall give the person notice of the person's noncompliance and liability for a civil penalty. If the person complies with the requirements of Section 35.102 not later than the 30th day after the date of the notice, the violation is considered cured and the person is not liable for the civil penalty.

The following are international laws involving content filtering:

- UK: Data Protection Act
- EU: Safer Internet Action Plan
- Many other countries have also enacted legislation

Additionally, the United Kingdom and some other European countries have data retention policies. Under these policies ISPs and carriers are obliged to retain a record of all their clients' Web browsing. The data retention period varies from six months to three years. In the U.K. this retained data is available to a very wide range of public bodies, including the police and security services. Anyone who operates a proxy service of any kind in one of these countries needs to be aware that a record is kept of all Web browsing through their computers. On March 15, 2006, the European Union adopted Directive 2006/24/EC, which requires all member states to introduce statutory data retention. The United States does not have a statutory data retention specifically targeting information in this area, though such provisions are under consideration.

7. CIRCUMVENTING CONTENT FILTERING

Shortly after the first content filter was deployed, an industrious user found a way to bypass it. This next section will be highlighting a few things most people try, but should in no way be considered an exhaustive guide to preventing circumvention. For most of us, the goal shouldn't be to uncover every single circumvention method, but take care of the low hanging fruit. Another way to look at it is to make it hard enough for the common user to give up and don't worry about the über-technical person because the method they find will likely be too hard for the common person to figure out, even with instructions.

Circumvention Technologies

When it comes to circumvention, remember one thing, the user looking to get around your filter needs to get lucky once. You on the other hand to stop them, need to be lucky all the time. See how the cards are stacked against you? There are many ways to bypass a content filter; in fact there are commercial products and services that will help you do it. There are two primary ways to bypass content filtering and they are, using a proxy or tunneling the traffic through another protocol.

Tip: There are a few simple steps that administrators can take that will make it extremely difficult for users to bypass their content filters. First, require administrative access to install software on company owned machines, and manage software through a software management and distribution tool. Second, love your logs... There is so much valuable information that can be found in logs and all security devices worth buying will generate access logs.

Proxies

The first method we will explore is a proxy. There are basically five types of proxies:

- Client-based proxies
- Open proxies
- HTTP Web-based proxies
- Secure public and private Web-based proxies
- Secure anonymous Web-based proxies

Client-Based Proxies and Tunneling

These are programs that users download and run on their computers. Many of these programs are run as "portable applications," which means they don't require any installation or elevated privileges, so they can be run from a USB thumb drive by a user with limited privileges. The three most widely used include TorPark, which uses Firefox and the XeroBank network, Google Web Accelerator, and McAfee's Anonymizer.

These programs create a local proxy server using a nonstandard port. Then they configure the browser to use

13. Texas ISP Laws can be found here: www.tlc.state.tx.us/legal/b&ccode/b&c_title10/80C258(3).HTML

the local proxy by changing its proxy server settings to the form localhost:port = 127.0.0.1:9777, for example. Web content requests are then tunneled through the proxy program to an appropriate proxy server using a custom protocol, which is typically encrypted. The content-filtering gateway doesn't see the browser-to-local proxy traffic, because it flies under its content inspection radar. All the gateway may see is the custom protocol that encapsulates the user's Web request.

The network of proxy servers is either static, as is the case with commercial programs such as McAfee's Anonymizer and Google Web Accelerator, or it's private and dynamic, as is the case with Psiphon and XeroBank. In both cases, the proxy server network is typically built by individuals who volunteer their home computers for use by installing the corresponding proxy server software.

There are multiple ways to identify and block this circumvention method. One simple way is through examining firewall logs. The firewall will show unusual amounts of activity on non-standard ports, which is an easy way to identify this type of circumvention.

Currently the UltraSurf proxy client is the most advanced tool available to circumvent gateway security Web content filters. UltraSurf was developed by an organization called UltraReach,[14] which was founded by a group of Chinese political dissidents. UltraReach developers continue to actively maintain and update UltraSurf. They designed UltraSurf specifically to allow Chinese citizens to circumvent the Chinese government's efforts to restrict Internet use in China. The UltraSurf application is a very sophisticated piece of software. It uses a distributed network of proxy servers, installed and maintained by volunteers around the world, much like a peer-to-peer network. It uses multiple schemes to locate the proxy servers in its network, spanning different protocols. It uses port and protocol tunneling to trick security devices into ignoring it or mishandling it. It also uses encryption and misdirection to thwart efforts to investigate how it works.

Ultrasurf is free and requires no registration, which makes it widely distributable. It requires no installation and can be run by a user who doesn't have administrative permissions to his computer, which makes it very portable. It can easily be carried around on a USB thumb drive and run from there.

Another formidable bypass application is Psiphon.[15] This is a distributed "personal trust" style Web proxy designed to help Internet users affected by Internet censorship securely bypass content-filtering systems typically set up by governments. Psiphon was developed by the Citizen Lab at the University of Toronto, building on previous generations of Web proxy software systems. Psiphon's recommended use is among private, trusted relationships that span censored and uncensored locations (such as those that exist among friends and family members, for example) rather than as an open public proxy. Traffic between clients and servers in the Psiphon system is encrypted using the HTTPS protocol.

According to Nart Villeneuve, director of technical research at the Citizen Lab, "The idea is to get them to install this on their computer, and then deliver the location of that circumventor, to people in filtered countries by the means they know to be the most secure. What we're trying to build is a network of trust among people who know each other, rather than a large tech network that people can just tap into."

Psiphon takes a different approach to censorship circumvention than other tools used for such purposes, such as The Onion Router, aka Tor. Psiphon requires no download on the client side and thus offers ease of use for the end user. But unlike Tor, Psiphon is not an anonymizer; the server logs all the client's surfing history. Psiphon differs from previous approaches in that the users themselves have access to server software. The developers of Psiphon have provided the user with a Microsoft Windows platform executable for the Psiphon server. If the server software attains a high level of use, this would result in a greater number of servers being online. A great number of servers online would make the task of attacking the overall user base more difficult for those hostile to use of the Psiphon proxy than attacking a few centralized servers, because each individual Web proxy would have to be disabled one by one.

There are inherent security risks in approaches such as Psiphon, specifically those presented by logging by the services themselves. The real-world risk of log keeping was illustrated by the turnover of the emails of Li Zhi to the Chinese government by Yahoo. Li was subsequently arrested, convicted, and sent to jail for eight years.[16] Some have raised concerns that the IP addresses and the Psiphon software download logs of Psiphon users could fall into the wrong hands if the Citizen Lab computers were to get hacked or otherwise compromised.

These tools are a double-edged sword: They are incredibly powerful tools for allowing political dissidents around the world to evade oppression, but they also provide end users on private, filtered networks with a way to access the Internet that violates acceptable use policies and introduces liability to an organization. One way to block these types of circumventing technologies is by

14. UltraReach Information can be found at www.ultrareach.com/.

15. Psiphon: http://psiphon.civisec.org/. and servers in the Psiphon system is encrypted using the HTTPS protocol. https://s3.amazonaws.com/8qep-lrim-kctj/en.html

16. Yahoo may have helped jail another Chinese user, www.infoworld.com/article/06/02/09/75208_HNyahoohelpedjail_1.html.

using deep packet inspection to identify these applications and block their access.

Open "Explicit" Proxies

Open proxies are services that are offered online and can be accessed by changing the configuration of your browser. Your browser can be modified to send all traffic to a proxy at a specific IP and port. Some organizations deploy their own explicit proxy to enforce content filtering, in which case the browser will already be configured to use a proxy. To locate the proxy settings using Internet Explorer, go to Tools | Internet Options | Connections | LAN Settings. When a proxy is defined in the browser it is called an explicit proxy, because you are explicitly telling the browser where to send your web traffic. Depending on the configuration of your operating system, it may require you to have administrative privileges to change this setting.

When your PC is configured to use an open proxy, the browser simply sends all its Web content requests to the proxy, as opposed to resolving the URL to an IP and sending the request directly to the destination Web site. The open proxy then does the DNS name resolution, connects to the destination Web site, and returns that content to the browser.

There are multiple ways to identify and block this circumvention method. Firewall logs showing unusual amounts of access on non-standard ports could be an easy indicator that someone is using an explicit proxy service. Restricting access to browser settings without administrative access is a great way to prevent someone from using an explicit proxy, which needs to be coupled with restricting users ability to install software. Finally, if you are using a proxy server yourself consider denying clients direct access to the Internet. Remember your proxy will facilitate the clients' requests so if this option is a possibility, you may only need to give your proxy server access to the Internet.

HTTP Web-Based Proxies (Public and Private)

These are Web sites that are purpose-built to proxy Web traffic. They are very simple to use because a user just browses to the "proxy" web site, and types a URL in to a text box on the site. The proxy web site will then display the content from the URL that was submitted. The content inspection gateway only sees traffic to the proxy site.

PHProxy and CGIProxy are the most well-known development efforts used for this purpose. Peacefire's Circumventor is an example of an application using these tools kits. The HTML code request and response from the proxy is specifically engineered to evade filtering. The more difficult it is to reverse engineer the URL of the proxied site from the HTTP traffic that's flowing between the browser and the proxy, the more successful this method of circumvention is.

The biggest benefit for these types of proxies for the circumventer is that they are simple to install on your home or office computer. A nontechnical user can do it in a matter of minutes. Once it is installed, the user has a Web-based proxy running on his home computer, which is presumably not filtered. He can then access his home computer from a filtered network (like an office or school network) using just a browser, and circumvent your carefully crafted Web filtering policy.

At first glance it may seem difficult to detect this type of circumvention, but it's actually pretty easy. In fact, since this is likely accessed through your content filtering gateway there will be a record of these transactions, however it is very unlikely that your content filtering company would have rated this site. Examining your access logs for sites with an "unknown rating" can be a very easy clue that someone has created their own proxy. Additionally, most content filtering companies have a category for proxy avoidance, which should be blocked as a best practice.

Secure Public Web-Based Proxies

These proxies are basically the same as the HTTP Web-based proxies except that they use the SSL encrypted HTTPS protocol. There are two types of HTTPS proxies: public and anonymous. Public HTTPS proxies are built by organizations such as Proxy.org and Peacefire and are publicized via mailing lists and word of mouth. They intentionally look like completely legitimate sites, with properly constructed certificates that have been issued by trusted certificate authorities like VeriSign. These sites can be easily blocked today by using a gateway device that has the ability to intercept SSL traffic and blocking the proxy avoidance category. Additionally, looking for unrated https sites can indicate a user has created a secure version of the home proxy.

There is real risk when using Internet based proxies and some companies choose to explain these risks to users before they consider using this type of tool. There are nefarious individuals on the Internet who build open proxies and publicize them through mailing lists and message boards, etc.. Often they are built with criminal intent, as a way to steal user credentials. An open proxy can see, capture and log everything you are sending and receiving, even HTTPS. Explaining this risk to the end user may be enough to keep them from exploring it on their own.

Process Killing

Some of the more poorly designed PC installed content-filtering applications can be shut down by killing their processes. For example, that would include Microsoft Windows through the Windows Task Manager or in Mac OS using Activity Monitor.

Remote PC Control Applications

Another form of tunneling is to use: Windows RPC, VNC, Citrix GoToMyPc, BeAnywhere, WallCooler, I'm InTouch, eBLVD, BeamYourScreen, PCMobilizr, and Cisco's WebEx to browse the Internet from another machine. Some of these applications are business critical and will not be blocked by corporately deployed content-filtering systems, and their intentional misuse can be a serious risk.

8. ADDITIONAL ITEMS TO CONSIDER: OVERBLOCKING AND UNDERBLOCKING

Overblocking occurs when the content-filtering technology blocks legitimate Web sites that don't violate policy. Underblocking occurs when a content filter doesn't identify content that should be blocked.

Casual Surfing Mistake

A friend sends a link in email, a popup window offers up something interesting, or you mistype a Web site address and get a typo-squatter porn site. All these ways will land you on a Web site that is not approved by your content-filtering system. You better have a way to deal with this reality.

Getting the List Updated

Most content-filtering companies send out very frequent updates or offer real-time access. These must be accessed, downloaded, and incorporated. Scheduled checks should be done to ensure these updates are happening on a regular basis to avoid out of date information. Figure e66.7 below shows the flow of web site categorization and distribution of the updated lists to clients.

Override Authorization Methods

Many content filters have an option that allows authorized people to bypass the content filter. This is especially useful in environments where the computer is being supervised and the content filter is aggressively blocking Web sites that need to be accessed. Usually the company owners and executives claim this privilege.

Warn and Allow Methods

This option allows for the company to state its policy on browsing to certain categories or web sites, but ultimately leaves the decision to continue on to the site to the user. An example is shopping, many companies will warn users that personal shopping during business is a violation of company policy, but if the user is shopping for business

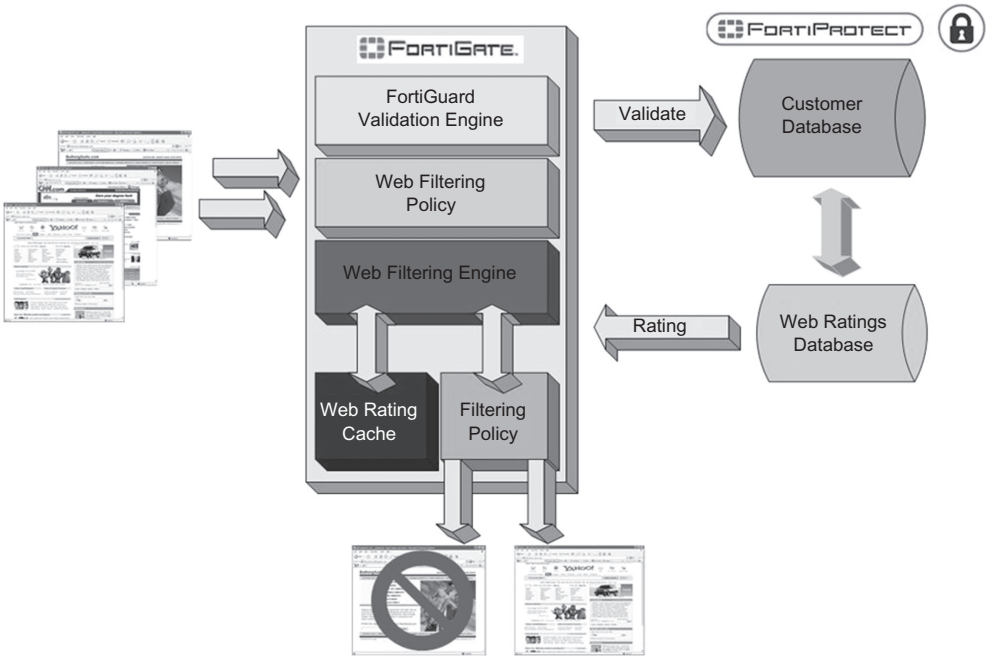

FIGURE e66.7 The flow of web site categorization and distribution of the updated lists to clients.

items, they can click a button that allows them to access the site.

Integration with Spam Filtering tools

Many content-filtering providers have a related component that inspects mail. That also includes the coordinates policy with the content-filtering gateway.

Spyware and Malware Categories

Most new content-filtering technology goes beyond just blocking offensive content. The same technology is looking for and blocking malware and spyware in HTTP data. Don't select an enterprise product without this feature.

Integration with Directory Servers

The easiest way to manages content filtering that requires granular user-lever control is to set up groups within directory servers. For example, Trusted User Groups, Executive User Groups, Owner User Group, and Restricted User Group will have different browsing policies. Some companies create role based access, choosing to create a category for HR and Finance, etc. and assign policies based on these roles.

Language Support

The content-filtering gateway must have support for multiple languages or the surfer will just find Spanish porn sites, for example. Typically a global ratings database will support multiple languages.

Financial Considerations are Important

Don't forget that a content filter project includes some of these items when calculating total cost of ownership for ROI payback:

- Licensing costs: Per user or per gateway.
- Hardware needed to run the solution; don't forget about a server for reporting
- Installation: Can you do it yourself or do you need a consultant or the manufacturer to help?
- Maintenance: Support and updates are necessary to keep your solution current.
- Ongoing administration from your IT staff.
- Patching, scanning, remediation by your IT staff.
- Some content filters need add-on server and license costs, for example:
  - ISA Server
  - MS Server
  - Logging Server
  - Analyzer Server
  - AV Server
  - Firewall
- Some content-filtering systems require integration costs with third-party enforcement points such as a firewall.

Reporting is a Critical Requirement

Reporting is a critical component of any content filtering solution, not only because it is key to compliance for industry specific regulations, but also because there is so much information to learn from logging and reporting. Real-time visibility to Internet usage, historical trending of Web traffic, and detailed forensic reporting help gauge user intent, help in enforcing Internet use policies, and enable retention of archived records to satisfy legal requirements and aid in regulatory compliance.

All web filtering companies will have a reporting module or product you can buy, and most will give you the basics like, who, what, when, where. In addition to the basics, many will have predefined report packages with useful information for management or the security team. Here are a couple of important predefined reports you may want to look for:

- *Management reports:* Management reports tend to give a high level overview of the environment. Most don't get in to detail on who did what, but focus on resource utilization, security overview, and any benefits of having the solution; for example if it caches, the amount of bandwidth that was saved during the reporting period.
- *Summary reports:* Summary reports are exactly what it sounds like, it gives a summary of traffic during the reporting period. Most will include a top talker and top sites, along with resource utilization and security overview.
- *User detail reports:* User level reports give you a complete picture of what someone has been doing over a period of time. Most will get very specific with the destination, category, object URL, time spent, amount downloaded, etc.
- HR reports: Human Resources reports typically contain the same information as the detail reports and also highlight denied requests, recreational categories, etc.
- Additionally this information is logged for reporting purposes:
- Source IP
- Destination IP
- URL
- Policy Action (allow, block, monitor)
- Content Category

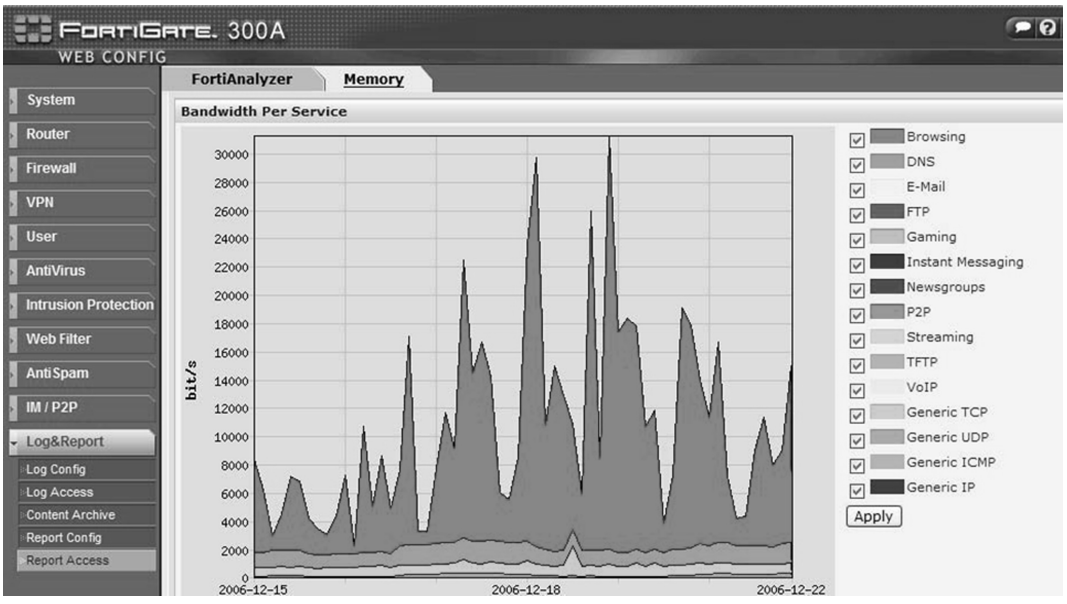

FIGURE e66.8 A bandwidth monitoring report from Fortinet.

Bandwidth Usage

Content-filtering systems should have the ability to display current protocol usage and report on patterns. Bandwidth savings give the most rapid ROI and need to be measurable. Figure e66.8 shows a bandwidth monitoring report from Fortinet.

Precision Percentage and Recall

The accuracy and efficacy of content-filtering systems are measured by precision and recall. Precision is the percentage of the number of relevant Web sites retrieved compared to the total number of irrelevant and relevant Web sites retrieved. Recall is the percentage of the number of relevant records retrieved compared to the total number of relevant records in the database. There is an inverse relationship between these two metrics that cannot be avoided: Maximizing one minimizes the other, and vice versa. Precision and recall must be considered together. A single metric of adding the precision and recall together is a good overall indication of the accuracy and efficacy.

The categorization of Web sites is an information retrieval process whereby each URL or Web page can be considered a record. A correctly categorized URL is a relevant record retrieved, whereas an incorrectly categorized URL is an irrelevant record retrieved. The objective of Web filtering is to block Web pages that are designated to be blocked and allow Web pages that are permitted. Web filtering precision is a measure of underblocking, or letting pages through that should be blocked. Higher precision leads to lower underblocking.

Web filtering recall is a measure of overblocking. Overblocking results from false positives and means blocking pages that should not be blocked. High recall leads to fewer false positives and lower overblocking. A perfect Web filtering system would have 100% precision and 100% recall, or a score of 200% overall.

A customer's Internet access policies dictate the Web sites to block, and typically all Web sites that are potentially liable, objectionable, or controversial are blocked.

Technical Support Challenges

Ensuring that your content filtering solution is not to blame for mysterious issues while surfing can be an ongoing IT support challenge. For example here is the trouble shooting flowchart from Bluecoat. See Figure e66.9.

9. RELATED PRODUCTS

Instant messaging, IRC, FTP, telnet, and especially email are all forms of communication that can be inspected with content-filtering technology. Also, more and more companies are integrating DLP (Data Leakage Prevention) technologies to reduce the risk of confidential, HIPPA, PCI, and PII information from leaking out over Internet connections. DLP inspection and blocking enforce data leakage and encryption policies.

Internet accountability software is a type of computer software that provides detailed reports that account for user behavior, surfing history, chat sessions, and actions on the Internet. Internet accountability software is used for various reasons, court-mandated sanctions, company

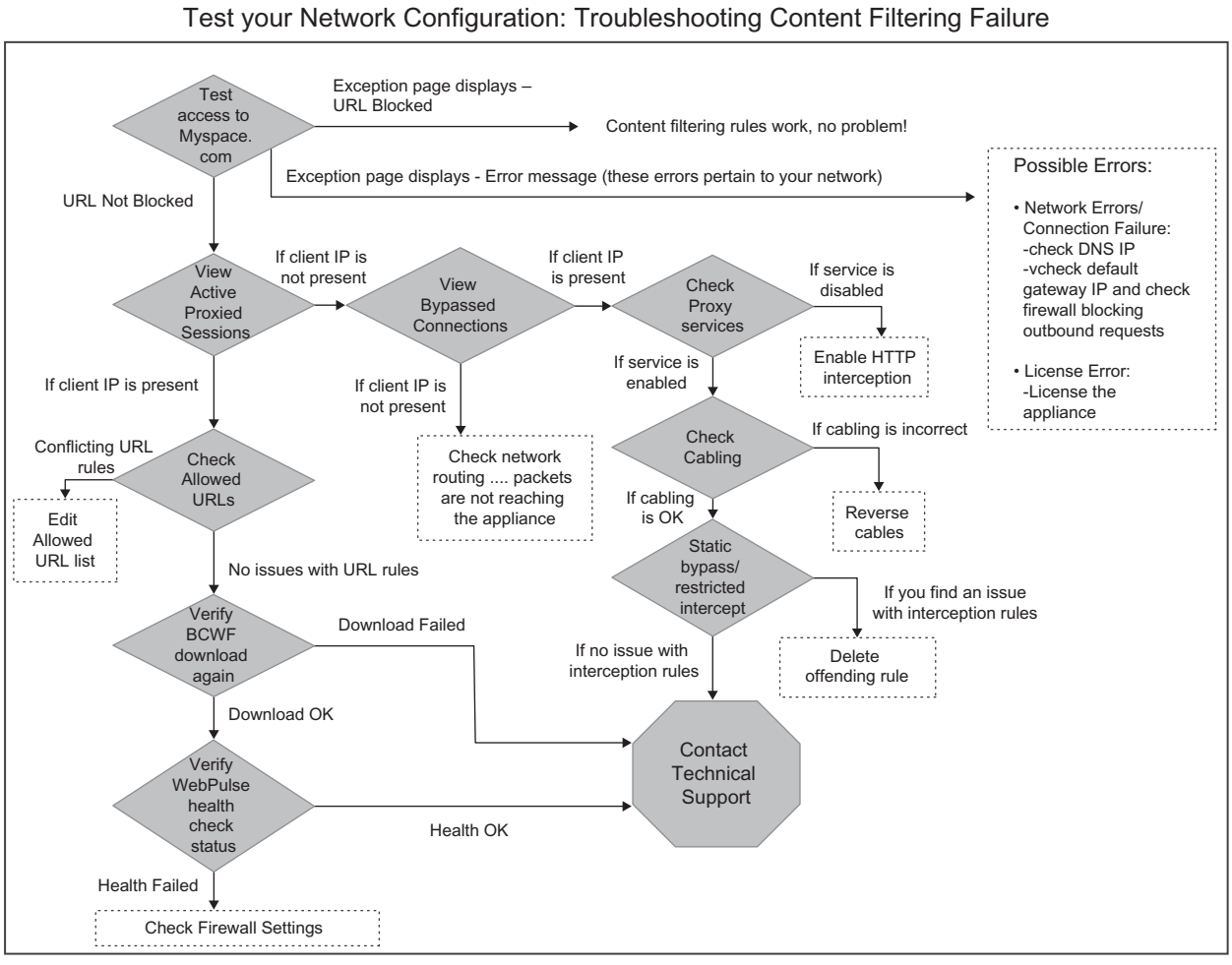

FIGURE e66.9 Trouble Shooting Content Filtering.

policy obligations, and as a recovery step in porn addiction. Versions of accountability software monitor Internet use on a personal computer, or Internet use by a specific user on a computer. These software applications then generate reports of Internet use, monitored by a third party, that account for and manage an individual's Internet browsing.

The first vendor to offer Internet accountability software was Covenant Eyes. Available in March 2000, Covenant Eyes accountability software was developed to provide Internet users with a means of reporting their online activity to one or more "accountability partners." The term "accountability partner" is a well-known concept in addiction-recovery circles and 12-step programs, such as Alcoholics Anonymous and Sexaholics Anonymous. Accountability partners have access to a user's Internet browsing record, which eliminates the anonymity of Internet use, thus providing incentive to not view Internet pornography or other explicit sexual images online. Today there are several accountability software providers: Covenant Eyes, Promise Keepers, K9 Web Protection, and X3watch.

10. SUMMARY

Content filtering is a tool and like any tool, knowing how to use it correctly will help you accomplish your goal. Whether you're a CEO looking to manage productivity or a parent protecting your children from the dangers of the Internet, content filtering is likely something you will have to think about at some point. We've considered the motivations for content filtering, examined the various methods, and highlighted a few market leading products, which should give you a good idea of what you will need to consider when choosing a solution. We have also touched on just a few common methods used to bypass content filtering and how a diligent administrator can use the tools at their disposal to discover this type of activity. Content filtering is a fast-paced battle of new technologies and the relentless trumping of these systems by

subversion and evasion. Altruistic development efforts by passionate programmers on a mission to support citizens in countries that block access to content will win, then lose, and then win again in a never-ending cycle. Other challenges include employees and kids who don't understand all the risks and don't think the abuse of a school-or company-provided computer and network is a big deal. Add new technologies, Web 2.0 applications, YouTube, and streaming sites, and the challenges and arguments for content filtering will not end anytime soon.

As we have explored, content filtering and its three objectives—accuracy, scalability, and maintainability—are at odds with each other. Accurate blocking makes it hard to scale and maintain, and easily scalable and maintainable systems are not as accurate. Companies that make content-filtering technology are attempting to make these challenges easier to manage and maintain.

Content filtering is sometimes controversial, and the law is frequently changing in the U.S. and internationally. IT policies try to cope and are being updated every year to deal with new legal issues.

Content filtering is morphing and aggregating with other technologies to address multifaceted threats. Today companies should be considering a defense in layers strategy, for example managing malware - using content filtering and anti-virus scanning at the gateway and anti-virus software at the desktop. If one control fails, there are backups in place to pick up the slack. In the high-stakes chess game of content filtering, the censors and policy enforcers are always perpetually destined to have the worst move in chess: the second to last one.

Finally, let's move on to the real interactive part of this Chapter: review questions/exercises, hands-on projects, case projects and optional team case project. The answers and/or solutions by chapter can be found in the Online Instructor's Solutions Manual.

CHAPTER REVIEW QUESTIONS/EXERCISES

True/False

1. True or False? Casual business relatad Web surfing has caused many businesses countless hours of lost productivity and occasionally hostile work environments have been created by employees who view and download offensive content.
2. True or False? There are many reasons companies consider implementing content filtering, which can range from improving employee productivity, to blocking web based threats, and even prevent data leaks.
3. True or False? Financial organizations have unique privacy and security concerns due to the fact that they need to protect their customer's personally identifiable information such as credit card numbers, Social Security numbers, and other financial related information which means they have room for error.
4. True or False? ISPs have unique motivations with regards to content filtering.
5. True or False? In the United States, the threat from websites that host malicious software presents a significant risk which cannot be easily managed through content filtering.

Multiple Choice

1. The use of _____ or _____ varies widely in public libraries in the United States, since Internet use policies are established by local library boards.
   A. Reputation
   B. Internet filters
   C. Log
   D. Encrypted
   E. Content-control software
2. There are many ways parents can protect their children from age inappropriate material on the Internet; and, _____ should be one of them?
   A. Opinity
   B. Web content filtering
   C. Scale
   D. Access
   E. Active monitoring
3. There are many technologies that can be used to categorize:
   A. Organizations
   B. Rapleaf
   C. Worms
   D. Content
   E. Security
4. The _____ method allows the creation of a blacklist dictionary that contains keywords or phrases?
   A. Keyword lists
   B. Denial of service attack
   C. Venyo
   D. Port traffic
   E. Taps
5. What contain full and/or partial URLs, which are compared to the URL in an HTTP get request?
   A. Systems security plan
   B. TrustPlus
   C. Denying service
   D. Decision making
   E. URL lists

EXERCISE

Problem

Why does one need content filtering?

Hands-On Projects

Project

Will content filtering slow down an Internet connection?

Case Projects

Problem

Will users who choose not to filter be affected by content filtering implementation?

Optional Team Case Project

Problem

What happens when users are denied access to a site? Will they know that the content filter blocked the site?