

From Internet Access Management to Content Security Management

- the challenge for an efficient use of the Internet

by Roland H.G. Cuny, Chief Technology Officer, webwasher.com AG

Introduction

The evolution of the Internet from a military network to a public resource, as well as the dramatic urbanization and economic growth of the 90s offered the potential for a free, borderless flow of information to nearly every individual. Without question, the new freedom of the Internet inspired pioneers and fuelled droves of investors. Unprecedented opportunities for new business, services and products seemed to overshadow any associated financial, social and technical threats. While the noise of the Internet hype has calmed down, the downside to this free flow of information has emerged in the form of:

- lost workforce productivity due to excessive surfing
- vandalism including cracking, viruses, worms and other malicious codes
- computer-enabled espionage
- personal privacy intrusions
- illegal content like child pornography
- exposure of minors to inappropriate content
- bandwidth congestion

Even so, the Internet should not be characterized as a medium used only by predators, criminals and paedophiles. In general, the Internet is no better or worse than the real world. Comprised of a vastly diverse community of millions of users, the Internet is a reflection of everything that exists in the tangible world; the good stuff and the bad. However, in the case of information and data exchange, freedom does not have to assume anarchy and chaos. Just like in the real world, each person has a right to self-determination and the ultimate responsibility to safeguard him/herself. Luckily, Internet users are not alone. Assisted by an emerging technology called Content Security Management (or CSM), every user has the power to decide which kind of content is appropriate for him/herself. Increasingly, CSM is used as a universal means for users, parents, schools, libraries, and companies alike, to surf freely, safely and efficiently on the Internet. WebWasher Enterprise Edition - the all in one

filtering solution for Content Security Management - helps users *and* institutions to implement a secure, fast and clean Internet access.



*Fig. 1: Packshot of WebWasher Enterprise Edition
– the all in one filtering solution for Content Security Management*

Components of CSM

Content Security Management is a software-based solution designed to protect users and networks from any potential threat (inbound and outbound) and goes far beyond the protections of the classic firewall. It encompasses access control, content filtering, content security, privacy filtering, and email filtering combined with reporting under a single umbrella interface. These six complementary and interdependent components of CSM collaborate to offer absolute security against the threats of the Internet.

1. Access management (access control)

Access control prevents the downloading of entire web pages based on what the user defines as inappropriate content. Often, schools need to block content that is considered to be harmful to children. Based on the school's Internet policy, their CSM software can be configured to prevent students from being exposed to inappropriate content. Employers have additional incentive to deploy access control measures. Beyond the productivity losses associated with excessive personal surfing, companies are also exposed to legal and security risks when inappropriate content is accessed via the corporate infrastructure. Use of access control measures prevents illegal or dangerous content from ever entering the corporate network.

There are two approaches to access control:

- “Block Only”: This approach offers unlimited access to the Internet except for specific content that the user or administrator specifies. For example, when a corporation specifically blocks illegal content but allows access to everything else, they are applying the block only approach.
- “Allow Only”: This approach restricts all access to the internet except for select web pages. For example, if a school were to only allow access to a web page dedicated to biology, they would be

using the allow only approach. Corporations might also limit access for some employees to the company's URLs or a vendor's online procurement page.

Both of these methods rely on technology that is able to analyse and classify web content. While there are a variety of methods to content classification, all of them belong to one of the following basic approaches.

- **Pre-processing:** CSM vendors or third parties use a variety of methods to crawl the Internet, individually analysing and classifying every site they encounter. The result is a vast filter database that may include millions of URLs in dozens of categories. Categorization methods used by these companies include human analysis, automated robots, image recognition technology, neuronal networks or statistical methods. The best results in terms of quality and quantity can only be achieved by applying a combination of all of these methods. Certainly, the automated processes are unmatched in their ability to analyse and classify content quickly. However, due to the diverse and context-rich nature of Internet content, the classification system must be constantly trained and monitored by human beings. Initially the system must "learn" how to classify content from human experts, after which the system becomes "balanced" and can rate content automatically. One example of a successful multi-methodology classification system is the 58 category filter database DynaBLocator of WebWasher Enterprise Edition. Due to performance and accuracy advantages, pre-processed databases have become the preferred solution for the load-intensive environments of corporations, universities and government agencies.
- **Real-time:** This approach requires analysis of individual web page content with each download. Real-time analysis methods include on-the-fly image recognition, content, meta-tag and structural analysis. While designed to eliminate the theoretical potential of inaccurate blocking, real-time analysis introduces several problems. First, since real-time analysis requires that a page be fully downloaded before analysis can begin, latency and poor performance often results. Second, as the analysis algorithms introduce one more process the system must handle, side effects often include high CPU loads and bottlenecks, degrading connection speed. Finally, without the benefit of human quality control, these robots remain imperfect in deciphering sophisticated and context-rich Internet content. To date, successes for real-time content classification have been limited to niche applications. Nevertheless, as computers become faster and technology advances, access control based on real-time analysis might become more widely deployed in the future.

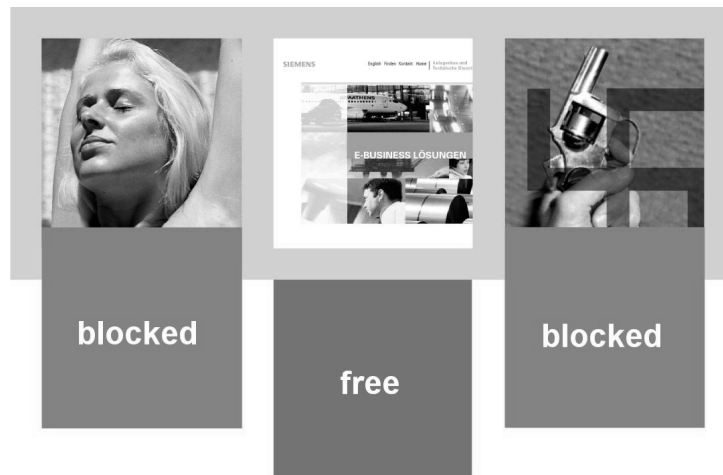


Fig. 2: Internet Access Control blocks all kinds of unwanted information

2. Content filtering

Content filtering is designed to remove specific types of unwanted content from within a Web pages, preventing it from being displayed or executed in the browser. Typical examples include eliminating annoyances like advertising banners, pop-up windows or animated images, all in an effort to simplify browsing, improve performance and save bandwidth. Moreover, content filtering can be an important line of defence against aggressive javascript or activeX code hiding in web pages. Capable of hijacking browser functionality, these scripts are able to create a number of frustrations including changing the browser's start-up page, adding bookmarks, resizing the window, launching new windows upon closing one, and even disabling mouse buttons.

By peeking into the page source code during download content filters can instantly detect and remove any harmful code. The resulting 'cleaned' code is then transferred to the browser, which executes it, fetches only the requested objects, and displays the cleaned Web page.

3. Content security

Content security represents an additional layer of protection over content filtering and is specifically designed to prevent harmful content from being attached to or included in downloaded Web pages. Typical targets for content security filters include viruses or trojans in downloaded files, active code embedded in the page, Java, and activeX executables.

Due to the intentionally elusive nature of these offenders, content security requires a series of sophisticated filters. A mime type filter is able to block file types that typically contain malicious code like Visual Basic Scripts. An added benefit of the mime type filter for any organization is its ability to screen non-work related files like music or videos that quickly devour network bandwidth. Embedded objects filters are designed to seek out potentially harmful active codes that are often hidden in Web pages. Finally, a virus scanner represents the last line of defence, scanning downloaded files for viruses and trojans. Content security, along with traditional firewalls and intrusion detection represent the "big three" of security measures deployed in networks of any size.

4. Privacy filter

Some say the undisclosed collection of user data is just data mining. Others say the gathered data is a minor inconvenience that enables companies to offer extra 'services'. Privacy experts in the European Union just say it is a violation against human rights. Regardless opinion, privacy filters offer the ability to harness privacy infringing content on the Internet and manage its use according to the user's individual preferences. This has very important implications on corporations since profiling employees can result in espionage and leakage of sensitive information or competitive information. While the threats can be enormous they begin with tiny invisible images. These "web bugs" silently communicate the identity and surfing habits of visitors to a Web page.

Other threats are aimed to identify and track users. Web servers are able to initiate the storage of a small piece of information (cookie) on the user's harddrive, containing a unique identification number, similar to a passport number in the real world. When the user enters the site again, the cookie is sent to the Web server and the identification is performed. Another such threat is referer information. Referers pass information to web site owners about URLs visited by each user prior to landing on their page. This way the structure of the corporate network can be silently communicated to anybody who operates a web site. Search engine queries also reveal a lot about a user. The real danger starts when some sites use prefixes in the URLs of search engine results. Then the search engine operator or even a third party is informed about the results the user has chosen to visit.

When collected in isolation, this information is often harmless. However, when used in combination, or in a targeted effort to uncover information, these "harmless" bits of data can offer detailed analysis of a user or a corporation's identity and Internet use. Recently, this danger has been intensified by consolidation of previously disparate databases. This has resulted in large organizations owning all of the pieces necessary to create a rich and detailed analysis of any Internet user. Applying privacy filters to eliminate some or all of this information is a major step users and corporations can take in retaining their online privacy.

5. Email filtering

As one more open door into a corporate network, the SMTP (email) protocol requires many of the same measures used to manage web traffic. Email filtering therefore has two core tasks:

- Blocking of unwanted incoming and outgoing emails and/or their attachments
- Removing of harmful code embedded or attached to mails.

The first task is designed to block unsolicited incoming mails including unwanted advertising mail (Spam) or mail with inappropriate content or attachments such as pornography. In addition it guards against mail bombing, the mass dumping of email garbage into one's inbox, as well as email viruses which can multiply by sending messages to all addresses in the user's address book. The second task scans emails and the attachments for viruses, worms, and embedded malicious code. Both tasks can be coupled with instructions on how to handle violating content. Options include deleting, quarantining, cleaning, returning to sender and archiving. For example, an infected mail can be

cleaned and delivered to the recipient, while also sending warning messages to the administrator or sender.

6. Reporting

While not actively filtering or blocking content, reporting represents a key element of the CSM solution. By logging events associated with all CSM protections, displaying the current status and generating analyses, these reports offer critical insight that allow administrators diagnose and address security breaches. Reporting benefits also extend to a variety of users throughout the organization and therefore need to be customisable for different people. For example, the operator of the system is interested in performance data and efficiency of the filtering. The network administrator is mostly concerned about threats to the network. Management may only want a summary report. Finally, the security officials might need to be informed about illegal content filtered by the CSM software.

As required by employee privacy regulations in the European Union, the informative benefits of reporting need to be carefully balanced with each employees right to privacy. In the EU, employers are not allowed to monitor the surfing habits of employees. Therefore, in order to be deployed globally, the reporting component should be customisable to the laws of each country. In the most conservative situations, rich reports can be created without any user data. In most situations, this amount of information is sufficient. When necessary however, reports can include user data and facilitate analyses of Internet use by each user. This is usually only warranted in serious cases of employee internet violations. Also, to prevent abuses that can be a basis for employee discrimination lawsuits, user data should be logged encrypted, allowing only select personnel to decrypt the logs.

The Content Security Mangement system contains several components, each containing many filters. The complexity and interdependency of the functions requires a single location where administrators can manage and monitor the system. Facilitating system management via a single, browser based interface makes configuration and deployment a snap. Also, remote administration via encrypted HTTP allows the administrator to control the system from their office or any other location. An example for such a one umbrella interface is given in the Fig. 3. With thousands or even tens of thousands of employees, user administration is not always an easy task. Therefore a robust CSM system should leverage existing user authentication databases such as LDAP or NTLM.



Fig. 3: Web-based umbrella interface of WebWasher Enterprise Edition

Outlook

How is filtering done in the future and how will technology evolve?

The European Union is actively researching Internet filtering technologies. In its Safer Internet Action Plan (<http://europa.eu.int/ISPO/iap>) many projects covering a variety of purposes have been addressed. One of these projects has been charged with the creation of a Filtering Software and Benchmarking study. The results will help customers select the right filtering solution based on their requirements. Another project called Worldwide Web Safe Surfing Services (3W3S) (<http://3W3S.eurecom.fr>) was successfully finished last year, delivering a prototype filtering platform that enables plug-in filter modules to run locally or, for the first time, distributed over the network or Internet. This also opens up technology companies to the possibility of easily “product-izing” new content classification algorithms in a lightweight filter module, without the headache of complete filter software deployment.

The future of CSM will bring the development of new machine learning algorithms, enabling the software to detect new types of harmful contents or security threats. Today's CSM solution represents a best of breed approach to network and Internet protection. Considering the enormous risks of going without, Content Security Management systems are good investments in terms of cost of return and increase of security, but also in the sense of meeting future demands. Scalability and expansion to new filtering techniques makes SCM the premium filtering solution for all those who want to keep their Web fast, secure and clean.

Table: The different aspects of Internet filtering

Who is filtering?	What is filtered?					Why is it filtered?								How is it filtered?					
	inappropriate content (by age)	bandwidth sucking content	malicious content	unwanted content	espionage code	legal enquiries	cost savings	protect IT infrastructure	prevent espionage, data leakage	workforce productivity	privacy protection	protect minors	cyberwar defense	access control	content filter	email filter	content security	privacy filter	reporting
Corporate & Govt Agencies	X			X		X	X			X				X	X	X			X
Corporate & Govt Agency Administration Personnel		X	X					X							X		X		X
Corporate & Govt Agency Security Personnel					X			X		X				X	X	X		X	X
Schools	X		X	X	X	X	X	X		X	X	X		X	X	X	X	X	X
Public domain: Libraries, Universities, Internet Cafes	X	X	X			X		X						X	X		X		X
National Communication Regulators	X		X										X	X			X		X
Internet Users	X		X	X	X	X		X		X	X				X		X	X	X
Parents	X		X	X	X	X		X				X		X	X	X	X	X	X