

CIPAFilter Whitepaper

Web Content Filter

Revised: 2008-01-05

CIPAFilter's content filtering system was designed to provide school district technology coordinators with a simplified and more effective approach to web filtering. The system was developed based on feedback from schools, and offers a set-and-forget system that does not over block web-sites, prevents use of anonymous proxy servers and circumvention, and uses a very strict context-sensitive pornography filtering algorithm.

The web content filtering provides the following features and benefits.

- **Enhanced pornography filtering** that is not as vulnerable to filter circumvention methods and catches in appropriate content regardless of source IP/URL, due to then context-sensitive system that looks at each page
- **Advanced anonymous proxy server detection** that combined with proper configuration virtually eliminates the use of proxies to circumvent the filter
- **False Positives and over blocking extremely minimal** due to lack of lots of categories and a purely black-list based filtering, but instead a fine-tuned system designed to eliminate this problem and help minimize required management and user intervention. A "set-and-forget" filtering philosophy.

The Web Usage Reporting, available as an additional feature, provides the tools needed to deal with students and users that still are attempting to access inappropriate websites. We believe this is necessary if you wish to discipline and/or enforce the Internet use policy agreement in place.

Pornography Filtering

The pornography filtering is a unique context-sensitive algorithm that was designed to block as much pornography as possible without blocking legitimate web-sites. These results are attained by using a balance between "good", "bad" and "cancel" words. By carefully examining and balancing these variables, we have built a filter that displays complex emergent behavior. If CIPAFilter finds a "bad" word, such as "breast" or "tit", it looks for cancel words on the page such as "chicken" or "cancer", "wildlife" or "bird" – finding in what context these words were used on that particular page, and ignores words as necessary when they are used in good context. It then further examines how frequently words are used, and in conjunction with what other potentially "bad" words are used.

With this technique we can focus on the inappropriate content, while avoiding the ever present collateral damage that other filters cause, that simply use black-lists. Pornography is different from many other categories due to the language and words used. CIPAFilter can identify inappropriate and vulgar language from many other sources such as cached web-sites, web-mail, and anonymous proxy servers, personal web-pages and more.

The filtering system is automatic and does not need to be updated. It also includes an **automatic strict safe-search enforcement for common search engines**, which will aid in preventing inappropriate search words or images on Google Images for example.

Search Engines that currently have safe search enforcement are:

Google, MSN, Alltheweb, Altavista, Ask, Dogpile, Metacrawler, Metaspy, Webfetch

Yahoo, Webcrawler, Hotbot, Infospace

Anonymous Proxy Server Detection

CIPAFilter has implemented our new anti-proxy solution as a two-prong system. The first technique we call fingerprinting. Using unique strings of syntax from a particular website, a trained analyst can create a fingerprint that can be used to detect that website regardless of how it's being rendered. For the students to view the website, certain portions of the code of that site must remain intact on the way to the student's workstation. We can isolate and create pattern matching systems to detect these fragments of code. **With a CIPAFilter configured for high security¹, students will not be able to reach Myspace, Facebook, or any other fingerprinted website, from your network.** The fingerprinted web-sites consist of current popular web-sites that students frequently attempt to access.

Our first technique detects these fingerprints in real time, immediately blocking the student from reaching the fingerprinted website (e.g. Myspace), but our second technique takes this one step further. If Myspace is on the blacklist, and a user has just accessed it, they must have done so through a proxy website. We immediately add the newly detected proxy site they were using to your local blacklist.

Additionally, the hostname of that proxy website will then be transmitted back to CIPAFilter's corporate offices with the URL for confirmation and a digest hash for security. These newly discovered blacklist entries will then be confirmed and distributed twice a day to all CIPAFilter customers. **This means that when any student uses a proxy server to access a fingerprinted website in the morning, the proxy server they used will be blacklisted nation-wide by the afternoon.**

This two-pronged approach of completely eliminating "hot button" websites with fingerprinting and aggressive automatic detection and distribution of web proxy servers will give CIPAFilter customers the edge, ensuring that all students attempting to circumvent the filter, are stopping proxy servers with a global "trap" that is immediate and allows all CIPAFilter units to work together to put an end to anonymous proxies.

1. A CIPAFilter configured for high security is a unit configured as a drop firewall to block proxy software like tor, and not allowing SSL website access for students, or only allowing SSL access to a preselected group of websites.

URL Blacklist

CIPAFilter's URL content filter black-list consists of selected categories that can successfully be blocked, without causing false positives. CIPAFilter focuses on key categories that relate to education, but does not attempt to create a "G-rating", essentially having too many categories with much management to follow. This is a major philosophical difference between our content filter and competitors in the marketplace.

Category black-lists are updated weekly automatically, and there is also a user-defined white-list and black-list so that each administrator can select which web-sites to block in addition to what is already in the database.

Current categories include:

Gamling, Games, Image-searches, Social Networking, Web-mail, Shopping

Authentication and Customized Filtering Privileges

In order to setup separate categories and access privileges by group membership and/or IP address subnet, CIPAFilter supports various levels of configuration.

CIPAFilter currently authenticates against Windows Active Directory, Novell EDirectory, and MAC Open Directory. Usually, this is provided to provide Web Usage Reports (*covered in separate document), but it also aids in configuring different levels of access. Different IP address networks may also be used, to separate filtering by building for example, instead of by username or group membership. A combination of these may also be used.

With authentication enabled, each user-group may have a separate white-lists and black-lists and category selection. You can allow shopping and web-mail access to teachers for example, and not for students, or block a specific web-site for students or a group of students.

With authentication enabled, you can allow/prevent use by group membership for:

- **Separate categories/black-lists/white-lists**
- **Download Privileges**
- **Internet Access On/Off**

This access is managed on the network server, and CIPAFilter communicates to the server to verify group membership and access level. There are two different types of authentication:

Transparent Authentication (Windows AD and Novell): This method will allow CIPAFilter to grab the username and authenticated credentials of the user currently logged into the computer. With Windows it currently requires a simple software client (which also happens to be our anti-virus software) to be loaded on the computer, and for Novell networks it simply requires that the user has logged in to the computer.

Manual Authentication (Windows AD, Novell, MAC): This method requires a user to login once more time when launching a web-browser. It does not require that a computer be a member of a domain or network, simply that a proxy setting is in the browser. Many schools prefer this method because it lets the user know they are being monitored.

Each CIPAFilter can only authenticate against one server/unit. If using a SSL proxy setting in the browsers, CIPAFilter can provide additional HTTPS proxy services, forced FTP virus scanning, or can be a stand-alone proxy device should you not want the unit to sit in-line with the traffic.

When the content filter is tripped, you may be notified with an e-mail, the trip is logged to the Web Usage Reporting system, and a page of your choice may be loaded and/or redirected to a web-server, to customize what the student will experience once they have attempted to access a blocked web-site.

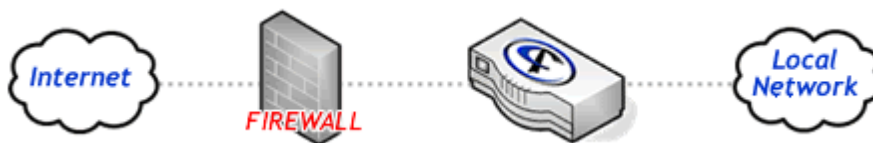
Installation & Setup

CIPAFilter will filter web traffic as a manual or transparent proxy server. As long as we can "see" the web-traffic, or web-browsers are pointed to the CIPAFilter, it will automatically scan each page and the URL/IP. The following three configurations are currently supported.

As Firewall



As Transparent Bridge



As Proxy Server



"Our school has purchased and tried various filter systems in order to comply with CIPA. On several occasions those solutions were circumvented and required too much maintenance. CIPAFilter's solution was not only easy to install, but there were no licenses to purchase and it has worked flawlessly. Their product is excellent! Thank you CIPAFilter!"

"The CIPAFilter is outstanding and the imagination and drive to keep it current makes it a long-term investment of high value. Your support teams is excellent and the cost-savings has been great"

- UT Township Schools, IL

"Even our district consulting engineer has done a complete turn-around! He was skeptical at first, but he is so impressed with the CIPA appliance, he is now recommending CIPAFilter to his other school districts."

Fort Stockton ISD, TX