Blocking advertising with your browser or proxy

Banner-Free Surfing

Advertising on websites makes it difficult to find your way around, and consumes bandwidth. In this article we will be looking at tools that allow you to block advertising banners on your own machine or the local network. BY OLIVER FROMMEL

There is hardly a website today that does not use banner advertising. In addition to banners that cover the full width of the screen, there is an increasing tendency toward large images that replace headings, making it difficult for users to find their way around.

Although this may be understandable from the content provider's point of view, most users are anything but amused when confronted with banner ads, especially as they impact surfing speed. In fact, it is quite common for garish ads in image or flash formats to take up more memory than the actual content.

These issues have led to the development of a number of programs that prevent unwanted banners from appearing on Websites, and even prevent the ads from loading in the first place. Blocking tools filter the datastream allowing interesting content to pass, and rejecting unsolicited advertising. This works because the browser retrieves the requested Web page from the server first.

Intelligent software checks the page for advertising images, and removes the offending sections from the HTML file. Only then will the browser retrieve the elements to render the page.

There are several approaches. Some browsers have an integrated feature, or can use a plug-in. Filter programs that run as **proxy servers** independently of the browser, but on the same machine are an alternative. They can also run on a separate machine if required, providing a proxy service to the computers on your home or office network.

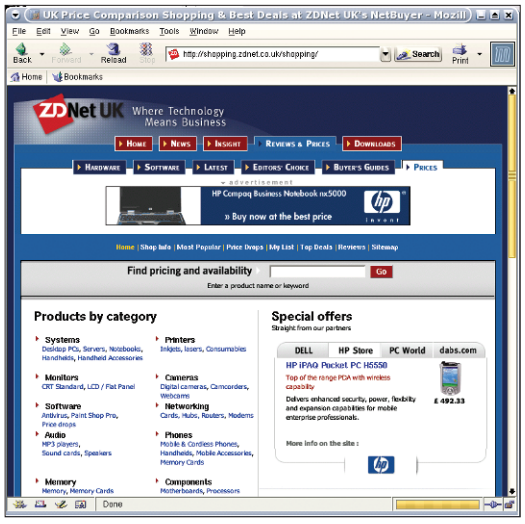

Figure 1: A Web page full of advertising makes it hard to see the actual content.

Browser-Integrated

Mozilla users have a simple option. The Mozilla browser has a plug-in that blocks advertising. The appropriately named AdBlock tool filters on the basis of the **URL**. It is easy to install the plug-in – just click the link on the project homepage [1]. You do not need administrative privileges to do so, as the AdBlock plug-in installs below the current user's home directory in *~/.mozilla*.

You need to re-launch your browser to enable the plug-in. The plug-in menu is available in the *adBlock* section below *Tools*. If you add the patterns shown in Figure 2, the plug-in will block an impressive number of ads. The same menu allows you to view the elements on the current page. If you discover an unsolicited ad, you can use its URL to create a new filtering rule.

You can also right-click an ad, and select *Adblock Image* in the drop-down menu to access the same

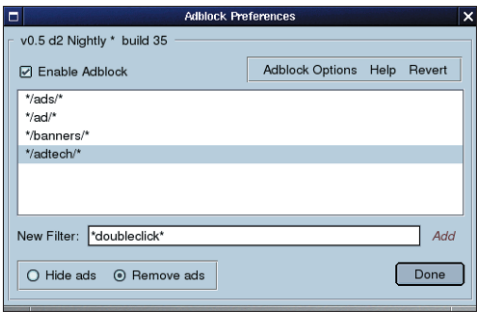

Figure 2: Using the Mozilla menu to configure AdBlock.

features. Doing so pops up a small dialog showing the address of this image. You can use a **wildcard** (*) to modify the URL to match similar addresses. Assuming that the following URL is displayed:

```
http://img-cdn.mediaplex.com/
ads/2399/9556/DE_DE_mofg_dim
4600_dhs_q1w0304_300x200_15k_
FL_gif.gif
```

You can remove the section following the server name and the *ads* directory, and type an asterisk as a wildcard instead:

```
http://img-cdn.mediaplex.com
/ads/*
```

This line tells the plug-in to block any files from the *ads* directory on the Mediaplex server. Now click the *Reload* button, and, hey presto, the ad disappears. Likely candidates are easy to find using the menu mentioned previously, or by inspecting the HTML file source code (see Figure 3). Users with the Mozilla Firefox browser (previously known as Firebird) will be pleased to hear that the plug-in will work with their browser too (we tested this with Firefox 0.8).

Some ads can be blocked without the plug-in. The easiest way of doing this is to enable the *Load Images | for the originating site only* option below *Options | Web Features*. This works quite well, because most banner advertising is not actually served up by the Web server itself, but supplied by third parties that handle ad click invoicing for the content provider. To use this feature while surfing, right-click an ad, and select *Block images from server*.

Note that this method might backfire on you, as it will prevent you from downloading any other images which are not stored directly on the original server. Also, this approach will not block ads from the original site. The adBlock plug-in definitely provides more granular settings. If you are still not satisfied, you might prefer to use a more flexible and powerful proxy that will run with other browsers such as Konqueror or Opera.

Tasty Java

Muffin is one such proxy. It resides between the Web server and your browser. As a Jar package, Muffin can be run directly with Java, and does not need installing or compiling. You do need the Java Runtime Environment (JRE), but most distributions install JRE by default. When downloading from [2], make sure that you right-click the link to the Jar file, and then select *Save link to disk*. Otherwise, your browser may decide to launch the Jar file directly.

If the *java* program is not in your path, add the directory where the program is stored, for example:

```
export PATH=$PATH:/usr/java
/j2sdk1.4.2_02/bin
```

You can then launch Muffin by typing *java -jar muffin-0.9.3a.jar* (see Figure 4). Muffin uses the *NoThanks* filter by default. The filter applies simple rules to block banner advertising. The proxy includes a few extra filters, which are not enabled by default, such as *Animationkiller* for removing gif animations, and *Cookiemonster*, which takes care of the ubiquitous browser **Cookies**. The *NoThanks* filter will not run out-of-the-box. Instead, it needs to load a small configuration file, the so-called killfile, first. A sample killfile is available from the Muffin website below *Samples*.

To load the killfile, select the entry for the *NoThanks* filter below *Enabled Filters* Click *Preferences*. In the dialog box that appears, select *Browse* below *Kill File* and locate the killfile in your Muffin directory. Click *Apply*, and then *Load*, to tell Muffin to use the killfile. The *Save* button stores this setting permanently,

Figure 3: You can inspect the HTML source code to identify the advertising addresses. Many sites actually use "ad" as part of their names.

GLOSSARY

Proxy servers *(or proxies for short) reside between a client (e.g. a Web browser) and a server. Viewed from the client's perspective, the proxy is a server, whereas the proxy is a client from the server's viewpoint. Proxies cache websites thus improving access speeds. In some cases, a proxy is needed to allow clients without a direct connection to access the Internet.*

URL: *A Uniform Resource Locator comprises a service acronym (http, ftp, ..), the address of an Internet server, and optional directory and file names. This allows documents on the Internet to be uniquely identified, for example, http://www.linux-magazine.com/issue/34/KDETricks.pdf.*

Cookies: *Small snippets of text-based information which a browser associates with a website. Content providers use cookies to store user-specific information on a user's Internet surfing behavior between two visits to a website ("When did the user last visit the website?").*

Wildcard: *Many Linux programs (such as the shell) use specific characters to represent one or more letters. For example, the asterisk (*) typically represents an arbitrary string, although it can mean any number of repetitions of a certain character (in regular expressions). When you type ls *.jpg, the shell will display all filenames with the .jpg suffix, no matter what letters the filenames start with.*

Port: *As multiple server programs can run on a single machine, a combination of the port number and the IP address is used to uniquely identify a connection. The Internet standards assign well-known ports to specific services, for example port 80 for the World Wide Web (HTTP), and port 25 for email (SMTP).*

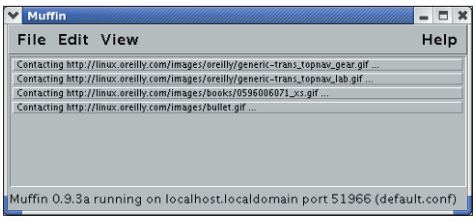

Figure 4: The Java-based Muffin program has its own GUI, and tells you what files it is handling.

and is a good idea, unless you want to repeat this procedure each time you launch Muffin. The filter program creates a *Muffin* directory below your home directory, and uses this directory to store its configuration and logfile.

You need to modify your browser settings, by entering the new tool as a proxy, for this to work. This applies to any other proxy tool you might use. If you have the Mozilla browser, open the settings in (*Edit* | *Preferences*) and click on the small triangle labeled *Advanced*. Look for the *Proxies* item; this is where you can enable the manual configuration and enter the correct values for *HTTP Proxy* and *Port* (see Figure 5). The first field is typically *localhost*. Enter the port number for your proxy in the second field (see Table 1). Opera has similar settings below *Network* | *Proxy Servers*; Firefox uses *Tools* | *Options*, and finally *Connection Settings*.

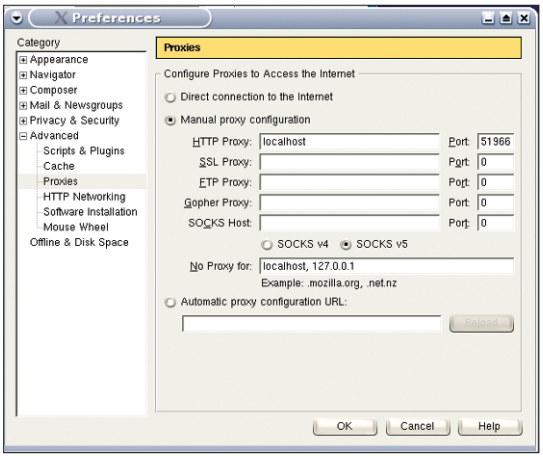

Figure 5: Proxy configuration in Mozilla showing the settings for the Muffin Web filter running on port 51966 of the same machine (localhost).

Filter with a History

This software is based on a classic tool, Junkbuster [3]. Privoxy [4] is still under active development, in contrast to other programs discussed in this article. The Privoxy website has packages for several distributions. By default, the package will install globally using the administrative account.

On Red Hat, the operating system launches the filter on booting, along with other server programs. You can do this manually using */etc/rc.d/init.d/privoxy start*. The configuration files are located in */etc/privoxy*, the main configuration file being */etc/privoxy/config*. Use this file for more granular settings, but Privoxy will perform quite well using the defaults.

After you set the proxy port to 8118 in your browser, as described previously, Privoxy will start filtering advertising out of the Web pages you visit. The software has a useful feature for tagging locations where it has removed ads, and can display both the filtered image, and the filter rule that it matched. This allows you to check if the tool is just removing advertising, or also preventing access to images you would prefer to view. You can also set up Privoxy directly from your browser using a special address, *http://p.p* (see Figure 6).

Whitewashed

Home users might like to check out Webwasher [5], a free tool by the Webwasher company. This version is restricted to two users, or a maximum of 20 simultaneous connections. The commercial version has special filters for Javascript which are not available in the free version. Webwasher needs administrative privileges to install, and there is no easy workaround. Webwasher provides two package formats, a RPM and a gzipped tar archive, which contains an installation script.

The software immediately launches, without being asked to do so, and listens for requests on port 9090. The configuration files are located in */etc/wwasher*, the logfiles in */var/log/wwasher*. Logging is disabled by default, but can be enabled using the Web-based front-end, which you can access via a special address, *http://-web.washer-/*. The default user name is *admin*, and the password is *webwasher*.

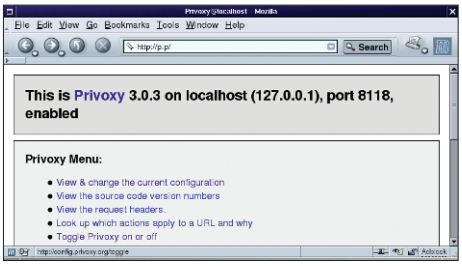

Figure 6: The Privoxy configuration page in a browser, where you can click to enable or disable the filter.

Spoilt for Choice

Your choice of filtering tool is basically a question of taste. If you don't like Java, you won't like Muffin. If you insist on entirely free software, you can rule out Webwasher. Mozilla and Firefox users can load the adBlock plug-in, although it is not as powerful as other filters. Privoxy is a mature tool which is under active development, and available as a package for many distributions. It is not difficult to install any of the programs we have looked at. The DVD with this issue provides a compact installation guide. ■

Table 1: Web Filter Overview

	Adblock	Muffin	Privoxy	Webwasher
Non-administrative install possible	yes	yes	yes	no
Proxy	no	yes	yes	yes
Standard port	–	51966	8118	9090
Under development	yes	no	yes	yes
Pre-configured	no	no	yes	yes
License	free	free	free	Restricted, free for private use

INFO

[1] AdBlock: *http://adblock.mozdev.org*
[2] Muffin: *http://muffin.doit.org*
[3] Junkbuster: *http://internet.junkbuster.com/*
[4] Privoxy: *http://www.privoxy.org*
[5] Webwasher: *http://www.webwasher.com/client/download/private_use/linux*