



Ad filters for your browser BANISHING BANNERS

ARMIN AND ALEXANDER ROEHRL, STEFAN SCHMIEDL

Intelligent proxies are handy things to have – they filter out banners, help to protect your privacy and let you surf faster. We introduce six blocking proxies for Linux.

Banner advertising on the Internet is not nearly as successful as marketing strategists had once hoped. That is why banners keep growing and are placed in increasingly intrusive positions – even right in the middle of the page. Although most portal pages like Yahoo limit the initial image size to 30K, four more files of up to 50K can be loaded subsequently. All together, that's a whopping 230K.

Filtering proxies stem this unwanted stream of data which, after all, users pay for with connection charges. Under Linux, you can get standalone solutions as well as extensions for Squid – every Linux user's favourite proxy.

Proxies and browsers

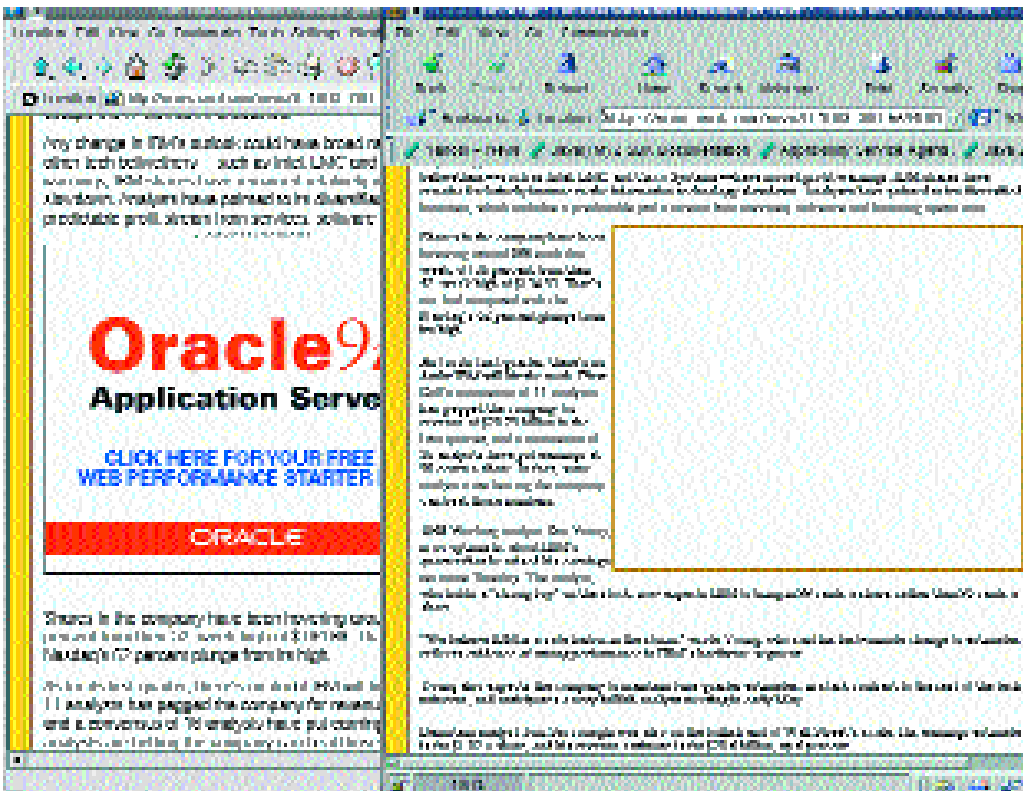
Blocking software works in the same way for all browsers. The browser is instructed in the settings not to send any request directly to the Net, but rather to the proxy. If this is running on your own machine, it is sufficient to specify *localhost* or the IP address *127.0.0.1* and the proxy port. Blockers tend to use different standard ports (see Table 1).

For security reasons, proxies should not be accessible from outside. Programs like Nessus Security Scanner indicate weak points in this respect. Incorrect requests will slow a system down considerably.

In Netscape and Mozilla, click on *Edit* (top left), then *Preferences*, double-click *Advanced*, then *Proxies* and *Manual Proxy Configuration*. For Konqueror the settings can be found under *Settings, Configure Konqueror, Proxies*. With Opera the path is *File, Preferences, Network* to the *Proxy Server*.

What is filtered?

In principle, anything that is sent from the browser or returned from the Internet can be filtered. Pornography and cookies, which people like to block to protect their privacy, can be stopped, as can advertising banners or animated images. Most anti-ad proxies replace adverts with a transparent image or broken image icon. The ad image is deliberately not replaced with a 1x1 pixel



News.cnet.com page: Left in Konqueror without and right in Netscape with Adzapper. For ten Cnet articles the proportion of advertising to content in terms of data volume is on average 60 to 40

transparent GIF, so as not to interfere with the layout of the Web page.

What happens to HTTPS?

Secure Hypertext Transfer Protocol (HTTPS, RFC 2660) is a standard protocol for the encrypted transmission of sensitive data like credit card information. As soon as the browser sends sensitive data to, or receives it from, `https://-Address`, the proxy could still filter the data, but it would not be able to recognise ad images, as these are now also encrypted. Proxies therefore always transmit HTTPS requests unfiltered.

Naturally, it is possible to change the port as required for all programs, either through the configuration file or using a parameter. The port should be set higher than 1024 if the proxy does not run as root, although that's generally a bad idea.

The filter tables provided should be treated with caution, as there is no problem in teaching most programs to filter Flash animations for instance, even if this cannot be done using the standard setting. The crux is in the detail – for instance, whether it is possible to treat cookies differently according to domain and page.

Adzapper

Of the tested proxies, we liked Adzapper (current version 0.7) best, as it is simple to use and delivers good results. Unfortunately, it hung once in a while. Perhaps this was because we were using version 2.0 of Python in the testing and Adzapper's author recommends 1.5.2. The picture is rounded off by

Adzapper's simple Web user interface `http://adzapper/`. It can even be used to create zaplets very easily. Zaplets are small XML files that define what constitutes an advertising nuisance. They can apply to one domain or to a whole domain class.

Adzapper comes with a lot of demo zaplets for popular sites like Slashdot, CNN or Yahoo – fortunately mostly English language ones. Zaplets can even include regular expressions in Perl syntax. A good introduction to this is provided by `python.org`. To install, unzip `adzapper-0.3.0.tar.gz` and run the `install` or `install-daemon` script. The program is started with `adzapper` or `adzapperd` (daemon version).

Here, as an example, theregister.co.uk zaplet:

```
<?xml version="1.0"?>
<zaplet>
  <version>0.9</version>
  <host>theregister.co.uk</host>
  <block_url type="regex">/media/(\\d)*\\.gif?
f</block_url>
</zaplet>
```

Table 1: Blocking Proxies

Program	Filters	Port
Adzapper	banners, popups, JavaScript, Java applets, cookies	51966
Squid with Adzap	banners, popups, cookies	3128
Muffin	animations, Java applets and JavaScript, cookies, referers, HTTP user agent headers, HTTP server headers	51966
Rabbit	banners, background images, cookies	9666
Junkbuster	banners, cookies, referers	5865
Webwasher	banners, popups, JavaScript, Java applets, cookies, referers	8080
Guidescope	banners, cookies, referers	8000



If you want to filter out all GIF images ending in



The Register page: In Netscape without and in Konqueror with built-in Adzapper. 52 percent of the data volume on this page consists of advertising. The total volume exceeds 60K

ad.gif, use regular expressions: *images!.*ad.gif*. Other examples of popular websites can be found at www.zaplet.org.

The creation of the zaplet is simple. First, load *theregister.co.uk* page. The trained eye will immediately recognise the ads. Select *View Image* in the browser, confirm through *Reload* and go to *http://adzapper*.

Adzapper now asks which parts of the file name to block. In the case of *theregister*, this is particularly easy, as the string *media* occurs in all ad banner links.

Squid with Adzap

Squid is probably the best general proxy for Linux. It is fast and robust, but not really suitable for use as an ad filter. Squid is intended as a robust cache for whole companies or the networked home and it performs this task very well. However, it also has a redirect option that can read request information

and then deliver it suitably changed. The entry

```
redirect_program /user/local/bin/squid_redirect
```

in the Squid configuration file (*squid.conf*) would filter all URLs through *squid_redirect*, with *squid_redirect* being, for example, a Perl script. We will be using Adzap for this purpose.

If Squid hasn't been installed yet this is quickly rectified using *rpm -ih squid.rpm*.

Adzap is available as a tar file. The file contains the redirector, the images to be displayed instead of the adverts, and a short wrapper script. After unpacking it with

```
tarx -xvzf adzap-20010606.tar.gz;
```

the redirector can be copied to */usr/local/bin/squid_redirect*, for example, and made executable with *chmod +x*. A relevant line needs to be added to *squid.conf* and finally Squid has to be restarted.

To adapt Adzap to UK Web pages, it is advisable to use the wrapper script. It must then be entered as a redirector in the Squid configuration. The path of the actual Adzap program then has to be set correctly in the wrapper script itself. The syntax is more like the *glob* expressions of the Bourne Shell than Perl-type regular expressions.

Muffin: Coffee and cookies

Muffin is a sophisticated Java program, currently available in version 0.9.3a. After a bit of fine-tuning Muffin works readily and quickly.

The package contains a number of useful utilities: Secretary can complete Web forms with pre-defined answers, dispensing with the need to constantly enter your own email address. Painter automatically changes the background colour of all transmitted Web pages. Document Info shows hidden information in the HTTP header of HTML pages, for example, the change date or the actual URL of the page.

The installation of Muffin is very simple. Download the jar file and start it with *java*:

```
muffin-0.9.3a.jar (binary only; right click to download)
java -cp muffin-0.9.3a.jar:$CLASSPATH Muffin &
```

Rabbit

Rabbit is available in version 2.0.8 and is distinguished by the fact that it compresses text and images, which can be especially interesting on large networks. Text is compressed by about 75 per cent using *gzip* and images are converted to a lower resolution with JPEG. Rabbit does this using Image Magick's *convert* utility.

Warning: Five surfers can easily push an 800MHz machine to its limits with this.

Rabbit allows you to view pages in their original

form by putting *noproxy* in front of the server address: *http://noproxy.www.altavista.digital.com/*. Install Rabbit with

```
gunzip RabbIT2.tar.gz; tar -xvf RabbIT2.tar;
```

it is started with *jr*. As for all other proxies, a restart is necessary after any change in the configuration. In *.conf/rabbit.conf* you can see what Rabbit regards as superfluous and consequently ignores.

```
# A filter to remove advertising from pages.
[rabbit.filter.AdFilter]
# The list of evils. Comma separated.
adlinks=.fsh,yimg.com,/ad/,/ads/,/ads2/,/net2
adsrv./,ad./,ads./,ads2./,ad./,ads./,ads1./,
.ads1./,adv./,cadv./,adserver./,adlog./,adba2
nnercenter./,nsads./,adclick./,banner./,
..
```

The first entry, *.fsh*, ensures that no file is loaded that has *.fsh* in its name. This should protect you from any more Flash ad banners.

Junkbuster

Junkbuster (version 2.0.2) is a classic. There are even providers who let their customers use Junkbuster directly. The program has no GUI, but it is very stable. A typical Junkbuster logfile looks like this:

```
www.altavista.com/
www.altavista.com/images/img_avsubtitle_0002
.gif crunch!
www.altavista.com/images/pix.gif crunch!
www.altavista.com/images/img_avlogo_000.gif2
crunch!
ad.uk.doubleclick.net/ad/.../sponsor-button2
/toptext/lefta;sz=1x1 crunch!
```

The addresses marked *crunch!* are blocked by Junkbuster and are never accessed.

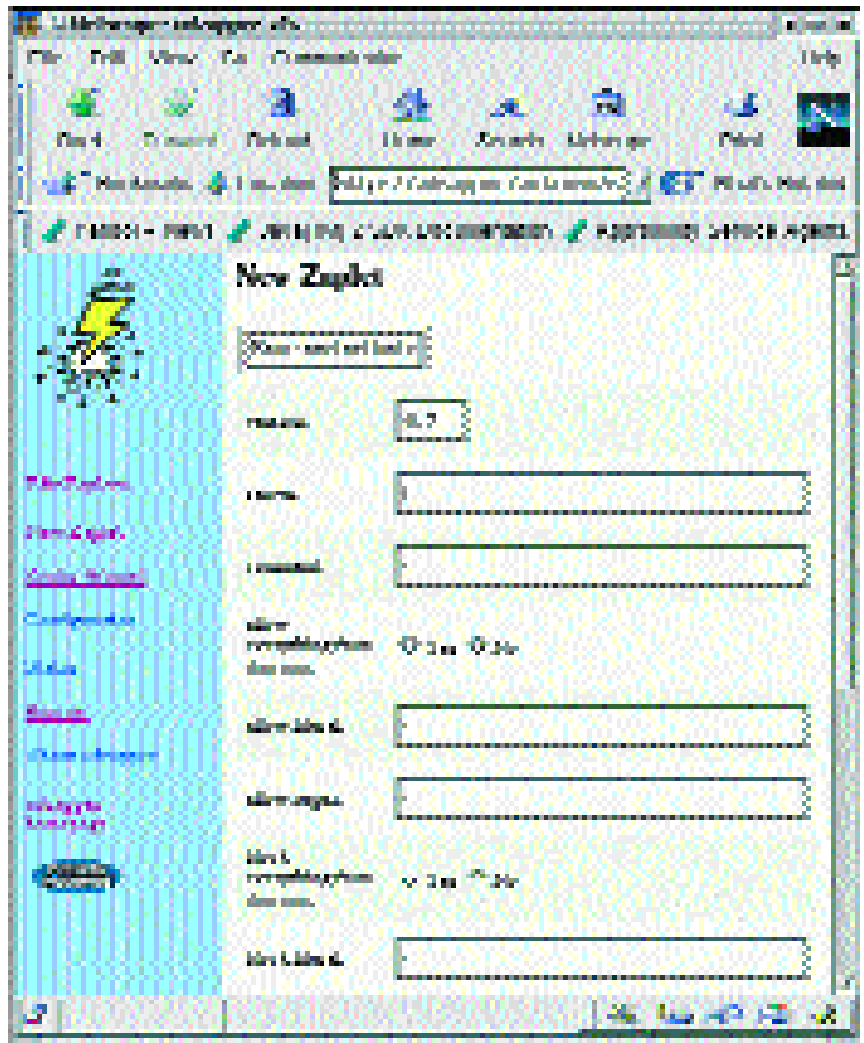
The heart of Junkbuster is the block file. The entries determine directly which addresses are called and which ones aren't. The sequence of entries is optional, but they should be divided into individual sections for the sake of clarity.

To start with, a ready-made block file will do, but for our test there was a great need for adaptation. We aimed for as few entries as possible and as many entries as necessary. This is an example blocking all image files from Altavista.com:

```
# URLs:
altavista.com/*.*(gif|jpg|jpeg|png)
```

Guidescope

Guidescope is a semi-commercial product by former Junkbuster developers. At the time of testing, the price had not yet been set. Guidescope sends each request to a central database, which returns a list of images to be blocked. In theory, Guidescope can record who is looking at which pages, but the



manufacturer emphasises that there is no intention of doing this.

The user can extend the database with local filters and is also able to pass identified ads to Guidescope so that they can be included in the database. Guidescope is planning to provide central company databases for large commercial customers.

During testing, the installation instructions on the website turned out to be wrong. Guidescope does work well for popular pages like Cnn.com.

Ironically, in its default setting it displays ad links – which supposedly makes it easier to get to good Web pages. This feature can be switched off.

Webwasher

Webwasher (version 3.0) is free for non-commercial use, a commercial license costs 29 Euro or \$29 after a 30-day test phase. Webwasher.com was founded by Siemens in 1998.

The manufacturer asserts that server-based Internet filters enable companies to provide their staff with Internet access while still being able to control which areas they can access. This is the only effective way of preventing abuse and unproductive use of the Internet, such as the occasional little

Adzapper's very easy-to-use configuration menu invites you to declare war on ad banners



game. Web filters protect productivity by making life difficult for permanent surfers. According to a study by the Meta Group, employees spend about 40 per cent of their time on the Internet on things that are nothing to do with their work.

The documentation is insufficient for a commercial product. However, with a bit of patience it is possible to find solutions to problems such as Webwasher under Wine using search

The authors

Armin Roehrl and Stefan Schmiedl run the software forge Approximity. Their main interests are high load and high availability systems.

Info

Cnet: <http://news.cnet.com>

Yahoo Fusion Marketing: http://fusion.yahoo.com/resource_center/adspecs/flash.html

Nessus Security Scanner: <http://www.nessus.de>

RFC 2660: <http://www.ietf.org/rfc/rfc2660.txt>

Adzapper: <http://www.zaplet.org/adzapper/>

Regular expressions: <http://www.python.org/doc/current/lib/module-re.html>

Example Zaplets: <http://www.zaplet.org/adzapper/updates.html>

Squid: <http://www.squid-cache.org>

Adzap: <http://www.zip.com.au/~cs/adzap>

Muffin: <http://muffin.doit.org/>

Rabbit: http://www.nada.kth.se/projects/prup98/web_proxy/

Image Magick: <http://www.imagemagick.org/>

Junkbuster: <http://internet.junkbuster.com/>

ISP with Junkbuster: <http://www.junkbusters.com/ht/en/ijbfaq.html#does>

Guidescope: <http://www.guidescope.com>

Webwasher: <http://www.webwasher.com/>

Webwasher and Wine: <http://www.miba.auc.dk/~bme/stuff/webwasher.html>

Clearway Technologies stops ad blocker: <http://news.cnet.com/news/0-1005-200-336086.html>

engines like Google.

Clearway proxy undesirable

In December 1998 Clearway Technologies were planning to launch an ad-blocking proxy program, but the outcry, particularly from many small website operators making their living from advertising – the manufacturer's customer base for their other products – led to Clearway withdrawing the product. The official explanation was that the Internet is, above all, about democracy and free, open websites, and that the idea of filtering was not compatible with this.

Conclusion

Of all the programs tested Webwasher is the best solution for companies who need a filter version that is always up to date. In addition, the Enterprise edition offers the Dyna Blocator, a filtering mechanism that is constantly updated by its own search engine and also checks the content of images. Apart from the standalone solution Adzapper, Squid with filters like Adzap, is probably the most stable and the fastest solution for home use and company networks.

An overall evaluation is difficult, as a basic knowledge of scripting languages is sufficient to write your own filters quickly for any of these programs. Unless you are planning to set up a proxy gateway, speed is of no consequence either. ■