

# “Keep the bad stuff out!”

Linux Web Filtering in 7 steps

How to install a transparent Squid proxy server  
with real-time HTTP virus scanning on Mandrake 10.0 using  
**DansGuardian** and **ClamAV**?

By Frank Neugebauer

<http://www.linux-tip.net>

linux-tip@web.de

Last Modified: 24/08/2004 4:42 PM



## **0. Introduction**

People quickly and easily access volumes of research on the Internet and correspond with a mouse click. For more and more companies, content filtering is part of the large battle to combat all kinds of online threats, including hackers, worms and viruses.

Linux content filtering allows administrators to configure and manage Internet access across the entire network and to block unwanted Web content like pornography, shopping Web sites, games and gambling.

The combination with different anti virus software makes it even more powerful and helps to protect our own system against the common threats. This guide contains all the necessary information for installing and understanding the architectural layout of the implementation. It was written with the assumption that you understand how to install programs and have a basic understanding of Linux Mandrake. This includes installing Linux Mandrake and RPM packages, editing files, making directories, compiling software and understanding general UNIX commands. This guide doesn't explain how to use or configure Squid, DansGuardian and ClamAV but information on where to obtain this information can be found in the "Additional information" section.

### Getting the software

To get the whole thing running we need the following software:

- Linux Mandrake 10.0
- <http://www.mandrakelinux.com/en/ftp.php3>
- Squid Proxy Server  
<http://www.squid-cache.org/>
- Clam Anti Virus  
<http://www.clamav.net/>
- DansGuardian – Web Content Filter  
<http://dansguardian.org/>
- Anti-Virus plugin  
<http://www.harvest.com.br/asp/afn/dg.nsf>

The software version used in this guide is always mentioned in the belonging chapter.

## **Step 1: Mandrake 10.0 Installation**

I don't want to explain how to install Mandrake. It is very easy these days. If you need help, please use the following link:

<http://www.mandrakelinux.com/en/fdoc.php3>

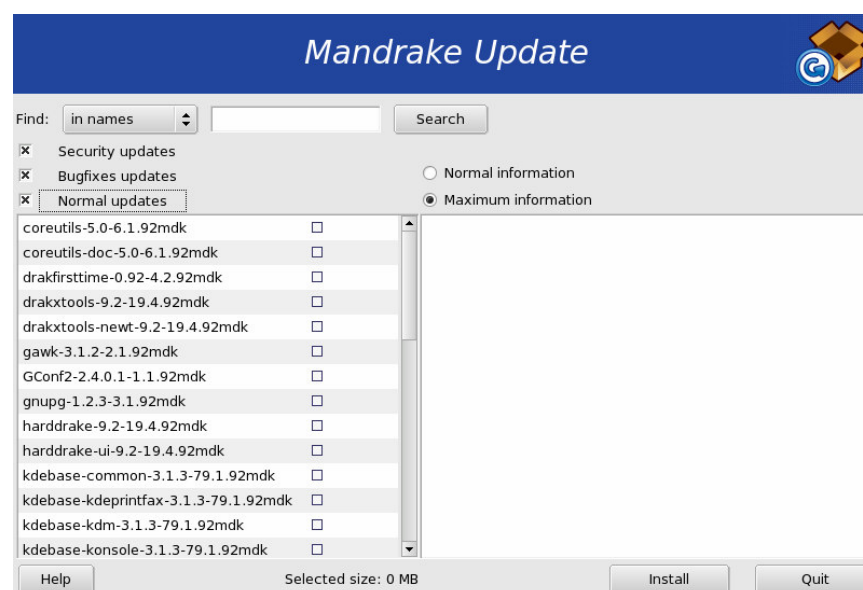
It is now time to specify which programs you wish to install on your system. There are thousands of packages available for *Mandrake Linux*, and to make it simpler to manage the packages have been placed into groups of similar applications. We just need a basis system. For that reason you should select the following groups:

- Console Tools
- Development
- KDE Workstation (or Gnome)

Aditonalty you need the following packages:

- Squid
- Webmin
- Iptables
- Perl
- zlib and zlib-devel
- libesmtplib and libesmtplib-devel

Please use the Mandrake Control Center to perform an update of your software. By clicking on "Mandrake Update" the system will be connected to the nearest FTP server and will get security updates, bugfixes and normal updates.



## **Step 2: Webmin installation and configuration**

It is time to get Webmin running. Webmin is a web-based interface for system administration for UNIX. Using any browser that supports tables and forms (and Java for the File Manager module), you can setup user accounts, Apache, DNS, MySQL, file sharing and so on.

Webmin consists of a simple web server, and a number of CGI programs which directly update system files like /etc/inetd.conf and /etc/passwd. The web server and all CGI programs are written in Perl version 5, and use no non-standard Perl modules. Please get more information about Webmin here:

<http://www.webmin.com/>

Honestly, we really do not need Webmin to get everything running, but it is a wonderful tool for a LINUX system administrator and it will help us to configure Squid and DansGuardian

After the installation please check if webmin is already running:

```
/etc/init.d/webmin status
```

If not, please start it like this:

```
/etc/init.d/webmin start
```

You can now use the Webmin interface with your favourite browser via the following URLs:

```
https://localhost:10000 or https://IP-address:10000
```

### **Step 3: Squid installation and configuration**

Squid is a high-performance proxy caching server for web clients, supporting FTP, gopher, and HTTP data objects. The software is designed to operate on any modern Unix system. The current stable version is 2.5.

If you started this workshop correctly, you installed Squid already in **Step 1**. Some basic configuration is to be done using the file `/etc/squid/squid.conf`. Please make sure doing the necessary changing as user root. Use the **su** command to become a “Good” user. We just need the following lines to get Squid running as a transparent proxy:

```
# set cache_dir to an area that has a large amount of hard disk space
cache_dir ufs /var/spool/squid 5796 16 256
```

```
# server hostname
visible_hostname myserver
```

```
#http port to use
http_port 3128
```

```
#by default http_access is denied to all. Please set your own rules here
http_access allow all
```

```
#This user (squid) should have the permissions to read and write the cache directory and the
#log file. Please make sure to create it first.
```

```
cache_effective_user squid
cache_effective_group squid
```

```
#Getting transparent caching to work requires the following entries:
```

```
httpd_accel_host virtual
httpd_accel_port 80
httpd_accel_with_proxy on
httpd_accel_uses_host_header on
```

With the following command you should create user and group squid:

```
groupadd -r squid
useradd -g squid -d /var/spool/squid -s /bin/false -r squid
```

Additionally you have to configure your server to accept the redirected packets – any IP address, on port 80 – and deliver them to your proxy application. This will be done in **Step 6** with IP forwarding features built into the kernel. We will use iptables to get this part running.

After you’ve finished editing the configuration file, you can start Squid for the first time to create the cache directories. This command will do the trick:

```
/usr/sbin/squid -f /etc/squid/squid.conf -z
```

If everything is working fine, you can use the following commands to start or stop the service:

**service squid start**  
**service squid status**  
**service squid stop**

Check the **cache.log** file in your logs directory. Squid generates run time error messages you can find here if something is not working. Additionally you can prove if squid is really listening on port **3128** with the following command:














**netstat -ln | grep 3128**

```
tcp      0      0 0.0.0.0:3128        0.0.0.0:*          LISTEN
```

Webmin makes it easy to configure and maintain Squid. See Step 2.

[Webmin Index](#) [Help](#) [Module Config](#) **Squid Proxy Server** [Apply Changes](#) [Stop Squid](#) [Search Docs](#)  
Squid version 2.5

---

 <a href="#">Ports and Networking</a>	 <a href="#">Other Caches</a>	 <a href="#">Memory Usage</a>	 <a href="#">Logging</a>
 <a href="#">Cache Options</a>	 <a href="#">Helper Programs</a>	 <a href="#">Access Control</a>	 <a href="#">Administrative Options</a>
 <a href="#">Authentication Programs</a>	 <a href="#">Delay Pools</a>	 <a href="#">Miscellaneous Options</a>	 <a href="#">Cache Manager Statistics</a>
 <a href="#">Clear and Rebuild Cache</a>			

## **Step 4: Installing and configuring anti-virus software ClamAV**

Clam AntiVirus is an anti-virus toolkit for UNIX. The main purpose of this software is integration with mail servers (attachment scanning). The package provides a flexible and scalable multi-threaded daemon, a command line scanner, and a tool for automatic updating via Internet.

We will later use the software to check HTTP content for viruses. ClamAV needs the user **clamav** installed.

Additionally you will need the **zlib** and **zlib-devel** package installed. If you didn't do this during the installation process, please complete it now using the **rpm** command or the Mandrake Control Center.

Please download the anti virus software here:

<http://www.clamav.net/>

Create user and group clamav as root::

```
groupadd clamav
useradd -g clamav -s /bin/false -c "Clam AntiVirus" clamav
```

Compile the downloaded software:

```
tar -zxvf clamav-0.75.tar.gz
cd clamav-0.75
./configure --sysconfdir=/etc
make
su
make install
```

I didn't bother changing anything in /etc/clamav.conf. I ran **freshclam** to update the virus database and created a root cron entry with **crontab -e**

```
0 * * * * /usr/local/bin/freshclam --quiet -l /var/log/clam-
update.log
```

That's it. To test our installation, please try to scan recursively the source directory:

```
clamscan -r -l scan.txt clamav-0.75
```

It should find some test viruses in the clamav-0.75/test directory. The scan result is saved in the scan.txt log file.

You also need to add the following line into **/etc/clamav.conf**  
**TemporaryDirectory /var/tmp**

## **Step5: Installing and configuring DansGuardian with antivirus plugin**

First of all let's download the software. You can use the following links:

<http://dansguardian.org>  
<http://www.pcxperience.org/dgvirus/>

You can also use this link to download DansGuardian including the necessary anti-virus patch. You do not need to patch the software because everything you'll need is already included:

<http://www.harvest.com.br/asp/afn/dg.nsf>

I download the following file:

`dansguardian-2.8.0.2-antivirus-6.3.1.tar.gz`

Before we start to compile DansGuardian we have to check if `libesmtplib` and `libesmtplib-devel` are installed on our system. If not, please install it by using the rpm package on CD 3 or Mandrake Control Center.

Time to compile it:

```
tar xvzf dansguardian-2.8.0.2-antivirus-6.3.1.tar.gz
cd dansguardian-2.8.0.2-antivirus-6.3.1
./configure      --sysconfdir=/etc/dansguardian/
                  --cgidir=/var/www/cgi-bin/
                  --runas_usr=squid
                  --runas_grp=squid

make
su
make install
```

Then add the following or something similar to root's cron with `crontab -e`.

```
59 23 * * sat /etc/dansguardian/logrotation
```

This will run logrotation at 23:59 every Saturday.

Configuring DansGuardian is our next step. You should find it in the subdirectory `/etc/dansguardian` on your server. We just have to edit **`dansguardian.conf`**.

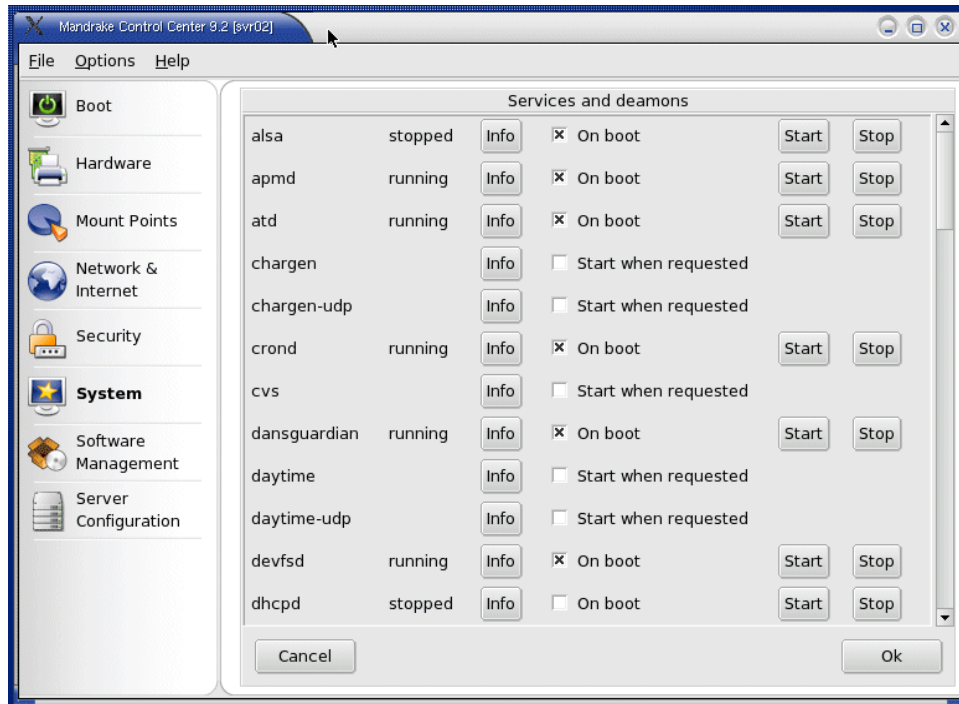
Here are my settings:

```
reportinglevel = 3
filterip = 127.0.0.1
filterport = 8080
proxyip = 127.0.0.1
proxyport = 3128
daemonuser = 'squid'
daemongroup = 'squid'
```

```
accessdeniedaddress = 'http://myserver/cgi-bin/dansguardian.pl'  
maxcontentfiltersize = 0
```

Before running dansguardian, edit `/etc/ld.so.conf` to include `/usr/local/lib`. Then run `/sbin/ldconfig`. Otherwise, it's going to complain that libclamav libraries are missing.

Start DansGuardian using Mandrake Control centre.



## Step 6: Redirect traffic

I recommend writing a small script (called **start\_transp.sh**) with the commands below. Please add it to your startup file.

**/etc/rc.local**

That will give you the possibility to start this script automatically during the boot process.

```
#!/bin/sh
#This command will redirect traffic
#Please add the file to your /etc/rc.local file
iptables -t nat -A OUTPUT -p tcp --dport 80 -m owner --uid-owner squid -j ACCEPT
iptables -t nat -A OUTPUT -p tcp --dport 3128 -m owner --uid-owner squid -j ACCEPT
iptables -t nat -A PREROUTING -i eth0 -p tcp --dport 80 -j REDIRECT --to-port 8080
#That's it. The user should be able to browse the Internet, but there will be a
# problem with web pages from banks or secure gateways working with ssl on
# port 443. Ftp on port 20,21 will also not work properly. This command will fix it:
echo 1 > /proc/sys/net/ipv4/ip_forward
```

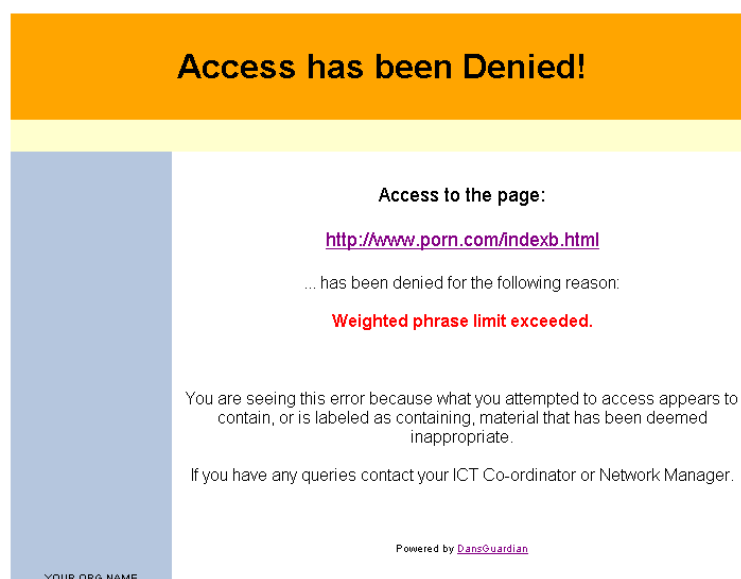
The client machines should be configured like this:

**Default gateway:** IP or name of your proxy server (i.e. 192.168.2.100)

**Your favorite browser proxy settings:** disabled (no proxy)

Now when users enter a forbidden Web address they will receive an "Access Denied" page instead of the aberrant site.

The layout of this page can be customized by editing the template.html file in the appropriate language section located in **/etc/dansguardian/languages**.



## Step 7: Configuring the Web Filter

I recommend using Webmin to configure DanGuardian. First of all we need to download a Webmin Module. Please use the following link to download the file **dg-0.5.10-pr4.wbm** and save it on the proxy server.

<http://sourceforge.net/projects/dgwebminmodule/>

Open Webmin with your favorite browser like described above and find **Webmin Modules** clicking on **Webmin – Webmin Configuration**. Use “From local file”, searching to the downloaded file and install the module.

## Webmin Modules

be added after installation by using the form to the right. Modules are .wbm files, each of which can contain one or more modules. Modules from RPM files if supported by your operating system.

**Install Module**  
 From local file   
 From uploaded file   
 From ftp or http URL   
 Standard module from [www.webmin.com](http://www.webmin.com)   
 Ignore module dependencies when installing  
 Grant access only to users and groups:   
 Grant access to all Webmin users

ore than one copy of the same module with different configurations, the form to the one any existing module. The clone will have a new name, a new module access control options and may be made available to different users.

**Clone Module**  
Module to clone:   
Cloned module name:   
Assign to category:

Please find the new installed Module here:

The screenshot shows the Webmin interface with several tabs: Webmin, System, Servers, Networking, Hardware, Cluster, and Others. Under the 'Webmin' tab, several modules are listed with icons and links:

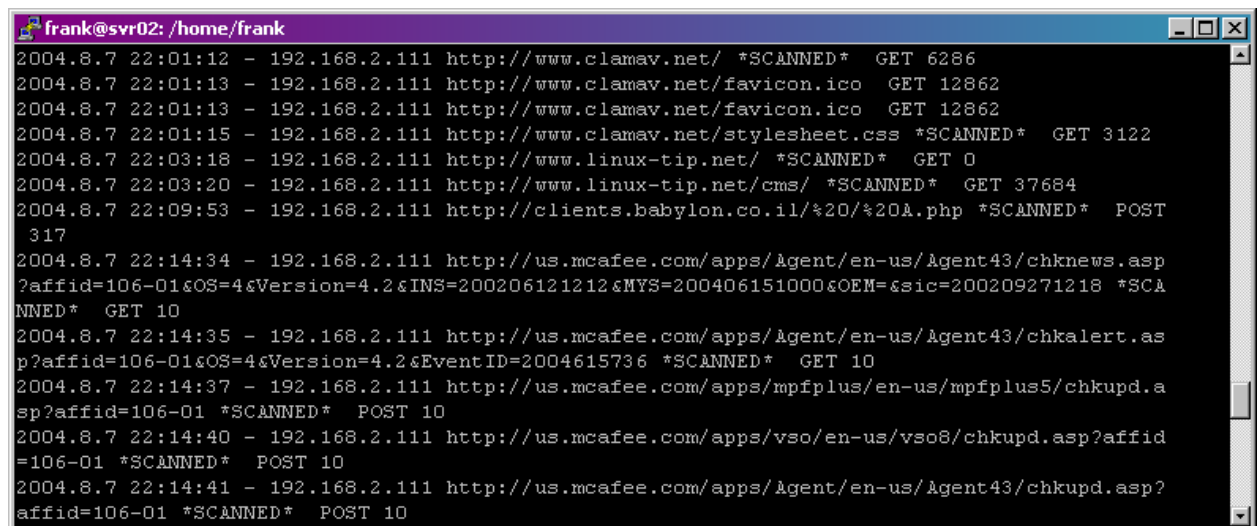
- Apache Webserver
- BIND DNS Server
- DHCP Server
- DansGuardian
- Majordomo List Manager
- MySQL Database Server
- Netatalk

Do the necessary settings using the DansGuardian Module. It will help you to protect your network and block unwanted web content.

## Troubleshooting and additional help

If everything is working fine, you should check the following file. Open a console and watch the entries in `/var/log/dansguardian/access.log` with

```
tail -f /var/log/dansguardian/access.log
```



```
frank@svr02: /home/frank
2004.8.7 22:01:12 - 192.168.2.111 http://www.clamav.net/ *SCANNED* GET 6286
2004.8.7 22:01:13 - 192.168.2.111 http://www.clamav.net/favicon.ico GET 12862
2004.8.7 22:01:13 - 192.168.2.111 http://www.clamav.net/favicon.ico GET 12862
2004.8.7 22:01:15 - 192.168.2.111 http://www.clamav.net/stylesheet.css *SCANNED* GET 3122
2004.8.7 22:03:18 - 192.168.2.111 http://www.linux-tip.net/ *SCANNED* GET 0
2004.8.7 22:03:20 - 192.168.2.111 http://www.linux-tip.net/cms/ *SCANNED* GET 37684
2004.8.7 22:09:53 - 192.168.2.111 http://clients.babylon.co.il/%20/%20A.php *SCANNED* POST
317
2004.8.7 22:14:34 - 192.168.2.111 http://us.mcafee.com/apps/Agent/en-us/Agent43/chknews.asp?affid=106-01&OS=4&Version=4.2&INS=200206121212&MYS=200406151000&OEM=&sic=200209271218 *SCANNED* GET 10
2004.8.7 22:14:35 - 192.168.2.111 http://us.mcafee.com/apps/Agent/en-us/Agent43/chkalert.asp?affid=106-01&OS=4&Version=4.2&EventID=2004615736 *SCANNED* GET 10
2004.8.7 22:14:37 - 192.168.2.111 http://us.mcafee.com/apps/mpfplus/en-us/mpfplus5/chkupd.asp?affid=106-01 *SCANNED* POST 10
2004.8.7 22:14:40 - 192.168.2.111 http://us.mcafee.com/apps/vso/en-us/vso8/chkupd.asp?affid=106-01 *SCANNED* POST 10
2004.8.7 22:14:41 - 192.168.2.111 http://us.mcafee.com/apps/Agent/en-us/Agent43/chkupd.asp?affid=106-01 *SCANNED* POST 10
```

DansGuardian is marking the virus scanned files as **\*SCANNED\***

Use the `netstat` command to check if squid is working (LISTEN) on port 3128 and Dansguardian on port 8080 like this:

```
netstat -ln | grep 3128
```

```
tcp 0 0 0.0.0.0:3128 0.0.0.0:* LISTEN
```

```
netstat -ln | grep 8080
```

```
tcp 0 0 0.0.0.0:8080 0.0.0.0:* LISTEN
```

You can get additional information checking the log file `/var/log/messages`.

Please use the following web pages if you need assistance and helpful suggestions:

Netfilter and firewalling	<a href="http://www.netfilter.org/">http://www.netfilter.org/</a>
DansGuardian Anti-Virus Plugin	<a href="http://www.pcxperience.org/dgvirus/">http://www.pcxperience.org/dgvirus/</a>
ClamAV Documentation	<a href="http://www.clamav.net/doc/">http://www.clamav.net/doc/</a>
Blacklist updater script	<a href="http://www.harvest.com.br/asp/afn/dg.nsf">http://www.harvest.com.br/asp/afn/dg.nsf</a>
squid-vscan (virus scanning with squid)	<a href="http://www.openantivirus.org/projects.php">http://www.openantivirus.org/projects.php</a>