Mirroring the Virus Database

Luca Gibelli

1st March 2004

Some guidelines for people interested in contributing to the distribution of ClamAV virus database.

1 Introduction

1.1 This doc

The latest version of this document is always available at `http://www.clamav.net/doc/mirrors/`. Before going any further, please check that you are reading the latest version.

1.2 Who is responsible for the virus database

The virusdb team take care of reviewing virus signatures, checking for new viruses in the wild and committing changes to the virus database file.

The updates are released quite often (usually no less than three times a week). If you want to be notified whenever the virus database is updated subscribe to clamav-virusdb *at* lists.sourceforge.net .

Every time the virusdb team updates the database, the ChangeLog will be posted to the mailing-list.

Visit `http://lists.sourceforge.net/mailman/listinfo/clamav-virusdb` for the list description and archives.

If you need to contact the virusdb team please write to: virus-team *at* clamav.net

1.3 Virus submission

Whenever you find a new virus which is not detected by ClamAV you should send it to the virusdb team. They will review your submission and update the database so that

the whole ClamAV user community can take benefit from it.

There are two ways to submit a virus sample:

- fill the form at `http://clamav.sourceforge.net/cgi-bin/sendvirus.cgi` (preferred method)
- send it by email to the following address: virus *at* clamav.net

Never send virus samples to ClamAV mailing-lists or developers addresses.

1.4 Getting a copy of the latest virus database

The most important factor for an antivirus's efficiency is to be up to date. ClamAV comes with a tool to update the virus database automatically: its name is *freshclam*.

freshclam reads a list of hostnames from *mirrors.txt* and tries to connect to them. If the first server in the list fails, freshclam will sleep for 10 secs and then try again with the next one, and so on.

If freshclam finds a new database, it downloads it and then sends a notify to clamd (if active) to reload the database.

It is important for the machine running ClamAV to be able to connect to port 80 of external hosts on Internet either directly or through a proxy. There are known problems with some transparent proxies caching what they shouldn't cache. If you should run into this kind of problem, please check your proxy configuration before reporting a bug.

2 Mirroring the database

2.1 The need for mirrors

To prevent the spread of worms it is essential to check for updates frequently. ClamAV users often configure freshclam with a check interval of 10 minutes.

With an exponentially growing number of ClamAV users, the servers hosting the virus database files get easily overloaded.

Without mirrors, the traffic on our main site was 100GB/month (May 2003).

On Feb 2004 the traffic on each mirror (11 in total) reached 120GB/month.
So if you are going to set up a new mirror, you can expect a traffic of about 120*11 = 1320 GB/month divided by the number of already existing mirrors+1.

As you can see traffic is growing fast and we need to always add more and more mirrors.

freshclam downloads the database from `http://database.clamav.net/`.

database.clamav.net is a round robin record that tries to equally balance the traffic between all the database mirrors. The round robin record allows us to alter the mirrors list in real-time, thus if a mirror stops working or ceases to get updates it can be removed immediately from the list without any intervent on the user side.

2.2 Requirements to become a mirror

We need fast reliable mirrors. Servers eligible for becoming mirrors have to provide:

- At least a 10Mbit/s link to the Internet[1]
- Unlimited traffic
- At least 50MB of web space
- Support for our *push-mirroring* system
- The mirror has to be available to all ClamAV users. We DO NOT support private mirrors.

We also appreciate (but do not require) having shell access to the server hosting the mirror. FTP access is no longer accepted.
The virusdb team will use the account *only* to update the virus database.

2.3 How to become a mirror

You have to follow these steps:

1. Set up a virtual host for `http://database.clamav.net/`.
   Additional server aliases for the above virtual host can be added too, but are not required. If you are using name based virtual hosts[2] see
   `http://httpd.apache.org/docs/mod/core.html#serveralias`
   for more information.
   Here is an example for a typical setup:

   <VirtualHost 10.1.2.3>

[1] Traffic is bursty, that's why we request such a large pipe

[2] You can check whether the mirror setup is correct or not, simply by adding a line like this:
your-server-ip database.clamav.net
to the /etc/hosts on your client machine. Then visit http://database.clamav.net and see if you can download files from your mirror's directory.

ServerAdmin john@clamav.foo.com
DocumentRoot /home/users/clamavdb/public_html
ServerName database.clamav.net
ServerAlias clamav.foo.com
</VirtualHost>
An http redirect (e.g. RedirectPermanent) is not enough! freshclam can't handle redirects.

2. Create an account with login "clamavdb" and give it write access to the virtual host's DocumentRoot.
   Disable password authentication:
   # passwd -l clamavdb
   The "clamavdb" user's shell must be /bin/sh or /bin/bash . Otherwise the user won't be able to run the command associated with the ssh public key[3].

3. Download the following files:
   `http://www.clamav.net/doc/mirrors/clam-clientsync.conf`
   `http://www.clamav.net/doc/mirrors/clam-clientsync`
   `http://www.clamav.net/doc/mirrors/authorized_keys_shell`
   `http://www.clamav.net/doc/mirrors/authorized_keys_noshell`
   `http://www.clamav.net/doc/mirrors/authorized_keys_shell.sig`
   `http://www.clamav.net/doc/mirrors/authorized_keys_noshell.sig`
   Verify the signature using:
   $ gpg –verify authorized_keys_noshell.sig authorized_keys_noshell
   $ gpg –verify authorized_keys_shell.sig authorized_keys_shell
   My PGP public key is available on most keyservers and on ClamAV web site. It can eventually be verified by telephone. Contact me by email first.

4. If you don't want to give us shell access, copy *authorized_keys_noshell* to *~clamavdb/.ssh/authorized_keys*:
   $ cp authorized_keys_noshell ~/.ssh/authorized_keys

[3]Take a look at the content of *"authorized_keys_noshell"*: the only command which can be executed by the owner of the corresponding ssh private key is ~/bin/clam-clientsync. We will only be able to trigger the execution of that script and nothing else!
However, shell access is really appreciated. If you are willingly to give us shell access, use *authorized_keys_shell* instead which contains Luca Gibelli and Tomasz Papszun ssh public keys too.

If you want to give us shell access, use *authorized_keys_shell* instead:
$ cp authorized_keys_shell ~clamavdb/.ssh/authorized_keys

5. Copy clam-clientsync to ~clamavdb/bin/
   Copy clam-clientsync.conf to ~clamavdb/etc/
   chmod 600 ~clamavdb/etc/clam-clientsync.conf
   chmod 744 ~clamavdb/bin/clam-clientsync
   Everything must be owned by user clamavdb.
   The clam-clientsync requires the "lockfile" program, which is part of the *procmail* package. Before going any further, please check that "lockfile" is available.

6. Send the server's details (account info, ip address, country, virtual host aliases, available bandwidth and sysadmin's email address) to *luca at clamav.net* .
   If your server meets the above requirements, you'll be given a login and password to access our master server at rsyncX.clamav.net.

7. Edit ~clamavdb/etc/clam-clientsync.conf . If your DocumentRoot (see paragraph 1) is */home/users/clamavdb/public_html* , your login is *foo* and your password *guessme*, then your clam-clientsync.conf will look like this:
   TO=/home/users/clamavdb/public_html
   RSYNC_USER=foo
   RSYNC_PASSWORD=guessme
   EXCLUDE="–exclude logo.png"

8. Reconfigure your packet filter to allow incoming connections on port 22/tcp and outgoing connections to ports 873/tcp and 873/udp.
   You can furtherly restrict access to these ports by only allowing connections from/to the IP addresses associated with rsync.clamav.net.
   rsync.clamav.net is a round robin record which points to our master mirror servers.
   Any changes to this record will be announced on the clamav-mirrors mailing-list.

9. You are welcome to put your company logo on the mirror home page. Just copy it to the DocumentRoot and rename it to "local_logo.png". The index.html is unique for every mirror. Please note that any file in the DocumentRoot whose name doesn't match "local_*" will be deleted at every mirror sync.

10. Subscribe to clamav-mirrors *at* lists.sourceforge.net: see
    `http://lists.sourceforge.net/mailman/listinfo/clamav-mirrors`
    for more info.

Subscribe requests have to be approved. We will approve your subscription request only *after* reviewing your server's info.

When everything is done, your server's IP address will be added to the round robin record and your company will be listed on our mirrors list page.

2.4 Statistics

Although it's not required, we really appreciate if you can make access statistics of your mirror available to us. They should be available at http://your-mirror-host-name/local_stats/ and they **must** be protected with login and password. You should use the same login and password you are using in your ~clamavdb/etc/clam-clientsync.conf file.

If possible, please tell your statistics generator to ignore requests made by the "ClamAV-MirrorCheck" agent.

If you are using Webalizer, you can add the following directive to your conf. file:
HideAgent ClamAV-MirrorCheck
If you are using AWStats, you can add this one instead:
SkipUserAgents="ClamAV-MirrorCheck"
Refer to your stats generator's manual for more info.

2.5 Admin's duty

- Scheduled downtimes should be announced on the clamav-mirrors mailing-list in advance.
- IP address changes should be notified in advance too.
- Changes in the ssh host public key of the mirror host should be announced on the clamav-mirrors mailing-list.
- It is essential to be able to contact the sysadmin responsible for the mirror server and get a quick response. Whenever a problem with a mirror occurs we need to immediately find out its cause and act consequently.

3 Notes for sigmakers

New sigmakers should send their ssh2 public key to *luca at clamav.net* . Their public key will be added to rsyncX.clamav.net authorized_keys (after being verified).

Sigmakers can upload a new database to either rsync1.clamav.net or rsync2.clamav.net using a (scp|sftp|rsync)-only account.

The new database won't be available to other people immediately. First, sigmakers have to notify the rsyncX.clamav.net server that a new database is available.

Here is the step-by-step procedure to release a new database version and propagate it around the world:

1. Assume your ssh private key is ~/.ssh/id_rsa and you've just built a new daily.cvd. Assume you want to use rsync1.clamav.net
2. In order to upload the new database, you have to run:
   $ rsync -tcz –stats –progress -e ssh -i ~/.ssh/id_rsa daily.cvd clamupload@rsync1.clamav.net:public_html/
3. Next, you need to notify rsync1.clamav.net that a new database is available:
   $ ssh rsync1.clamav.net -i ~/.ssh/id_rsa -l clamavdb sleep 1
4. rsync1.clamav.net will verify the digital signature of the newly uploaded database using *sigtool -i*. If it finds an error, it will refuse to distribute the database to other mirrors.
5. rsync1.clamav.net will copy the previously uploaded database to its rsync shared directory.
6. rsync1.clamav.net will generate the legacy viruses.db/viruses.db2 databases (together with their respective MD5 checksums)
7. rsync1.clamav.net will notify every mirror that a new database is available
8. Every mirror will rsync its copy of the database from *rsync1.clamav.net::clamavdb* (only mirrors can access the rsync server at rsync1.clamav.net, it's password protected)

As a fallback, every six hours, either rsync1.clamav.net or rsync2.clamav.net force an update on every mirror.

If rsync1 can't reach rsync2 or viceversa, the automatic update doesn't take place. This is done to avoid propagating an old database.

To avoid conflicts, sigmakers should use rsync1 by default and if it fails, switch to rsync2. Whenever a sigmaker uses rsync2, he should announce it on the clamav-team mailing-list so that every other sigmaker uses rsync2 too, until the issues with rsync1 are over.

4 Mirror status

Every mirror is continously monitored to ensure that every ClamAV user gets the latest virus database.

Every six hours we upload a file called *timestamp* on every mirror. Every hour we choose a random mirror and check that *timestamp* is fresh. If the file is one day old or unavailable, the mirror if marked as "old" and the ClamAV team receive a warning. If the situation persists for two days, the mirror is temporarily removed from the list.

You can view the current status of every ClamAV database mirror at `http://www.clamav.net/mirrors.html`.