Mirroring FreeBSD

Jun Kuriyama <kuriyama@FreeBSD.org >
Valentino Vaschetto <logo@FreeBSD.org >
Daniel Lang <dl@leo.org >
Ken Smith <kensmith@FreeBSD.org >
##: 43184

FreeBSD # FreeBSD########

CVSup # John D. Polstra ######

2013-11-13 # hrs.

##

###############(###############)#### mirror FreeBSD#

####

1.

Mirror ######### email # <mirror-admin@FreeBSD.org >######### FreeBSD mirror sites ####s#

2. ## FreeBSD mirrors

2.1.

################ Depending on the set of releases, architectures, and degree of completeness you want to mirror, a huge amount of disk space may be consumed. Also keep in mind that *official* mirrors are probably required to be complete. The CVS repository and the web pages should always be mirrored completely. Also note that the numbers stated here are reflecting the current state (at 9.2-RELEASE/10.0-RELEASE). Further development and releases will only increase the required amount. Also make sure to keep some (ca. 10-20%) extra space around just to be sure. Here are some approximate figures:

- ## FTP Distribution: 126 GB
- CVS repository: 2.7 GB
- CTM deltas: 1.8 GB
- Web pages: 300 MB

2.2.

########## Internet# ############### mirror ##### #### mirror #### FreeBSD ####### # intranet ###### ###################### #### official mirror ################### ##################

- Local site, no public access: basically no minimum, but < 2 Mbps could make syncing too slow.
- Unofficial public site: 34 Mbps is probably a good start.
- Official site: > 100 Mbps is recommended, and your host should be connected as close as possible to your border router.

2.3. #####CPU#RAM

One thing this depends on the expected number of clients, which is determined by the server's policy. It is also affected by the types of services you want to offer. Plain FTP or HTTP services may not require a huge amount of resources. Watch out if you provide CVSup, rsync or even AnonCVS. This can have a huge impact on CPU and memory requirements. Especially rsync is considered a memory hog, and CVSup does indeed consume some CPU. For AnonCVS it might be a nice idea to set up a memory resident file system (MFS) of at least 300 MB, so you need to take this into account for your memory requirements. The following are just examples to give you a very rough hint.

For a moderately visited site that offers Rsync, you might consider a current CPU with around 800MHz - 1 GHz, and at least 512MB RAM. This is probably the minimum you want for an *official* site.

For a frequently used site you definitely need more RAM (consider 2GB as a good start) and possibly more CPU, which could also mean that you need to go for a SMP system.

You also want to consider a fast disk subsystem. Operations on the CVS repository require a fast disk subsystem (RAID is highly advised). A SCSI controller that has a cache of its own can also speed up things since most of these services incur a large number of small modifications to the disk.

2.4. Services to offer

Every mirror site is required to have a set of core services available. In addition to these required services, there are a number of optional services that server administrators may choose to offer. This section explains which services you can provide and how to go about implementing them.

2.4.1. FTP (required for FTP fileset)

This is one of the most basic services, and it is required for each mirror offering public FTP distributions. FTP access must be anonymous, and no upload/download ratios are allowed (a ridiculous thing anyway). Upload capability is not required (and *must* never be allowed for the FreeBSD file space). Also the FreeBSD archive should be available under the path `/pub/FreeBSD` .

There is a lot of software available which can be set up to allow anonymous FTP (in alphabetical order).

- `/usr/libexec/ftpd` : FreeBSD's own ftpd can be used. Be sure to read ftpd(8).
- ftp/ncftpd: A commercial package, free for educational use.
- ftp/oftpd: An ftpd designed with security as a main focus.
- ftp/proftpd: A modular and very flexible ftpd.
- ftp/pure-ftpd: Another ftpd developed with security in mind.
- ftp/twoftpd: As above.
- ftp/vsftpd: The “very secure” ftpd.
- ftp/wu-ftpd: The ftpd from Washington University. It has become infamous, because of the huge amount of security issues that have been found in it. If you do choose to use this software be sure to keep it up to date.

FreeBSD's ftpd, proftpd, wu-ftpd and maybe ncftpd are among the most commonly used FTPds. The others do not have a large userbase among mirror sites. One thing to consider is that you may need flexibility in limiting how many simultaneous connections are allowed, thus limiting how much network bandwidth and system resources are consumed.

2.4.2. Rsync (optional for FTP fileset)

Rsync is often offered for access to the contents of the FTP area of FreeBSD, so other mirror sites can use your system as their source. The protocol is different from FTP in many ways. It is much more bandwidth friendly, as only differences between files are transferred instead of whole files when they change. Rsync does require a significant amount of memory for each instance. The size depends on the size of the synced module in terms of the number of directories and files. Rsync can use `rsh` and `ssh` (now default) as a transport, or use its own protocol for stand-alone access (this is the preferred method for public rsync servers). Authentication, connection limits, and other restrictions may be applied. There is just one software package available:

- net/rsync

2.4.3. HTTP (required for web pages, optional for FTP fileset)

If you want to offer the FreeBSD web pages, you will need to install a web server. You may optionally offer the FTP fileset via HTTP. The choice of web server software is left up to the mirror administrator. Some of the most popular choices are:

- www/apache13: Apache is the most widely deployed web server on the Internet. It is used extensively by the FreeBSD Project. You may also wish to use the next generation of the Apache web server, available in the ports collection as www/apache22.
- www/thttpd: If you are going to be serving a large amount of static content you may find that using an application such as thttpd is more efficient than Apache. It is optimized for excellent performance on FreeBSD.
- www/boa: Boa is another alternative to thttpd and Apache. It should provide considerably better performance than Apache for purely static content. It does not, at the time of this writing, contain the same set of optimizations for FreeBSD that are found in thttpd.

2.4.4. CVSup (desired for CVS repository)

CVSup is a very efficient way of distributing files. It works similar to rsync, but was specially designed for use with CVS repositories. If you want to offer the FreeBSD CVS repository, you really want to consider offering it via CVSup. It is possible to offer the CVS repository via AnonCVS, FTP, rsync or HTTP, but people would benefit much more from CVSup access. CVSup was developed by John Polstra. It is a bit tricky to install on non-FreeBSD platforms, since it is written in Modula-3 and therefore requires a Modula-3 environment. John Polstra has built a stripped down version of M3 that is sufficient to run CVSup, and can be installed much easier. See Ezm3 for details. Related ports are:

- net/cvsup: The native CVSup port (client and server) which requires lang/ezm3 now.
- net/cvsup-mirror: The CVSup mirror kit, which requires net/cvsup-without-gui, and configures it mirror-ready. Some site administrators may want a different setup though.

There are a few more like net/cvsup-without-gui you might want to have a look at. If you prefer a static binary package, take a look here. This page still refers to the S1G bug that was present in CVSup. Maybe John will set up a generic download-site to get static binaries for various platforms.

It is possible to use CVSup to offer any kind of fileset, not just CVS repositories, but configuration can be complex. CVSup is known to eat some CPU on both the server and the client, since it needs to compare lots of files.

2.4.5. AnonCVS (optional for CVS repository)

If you have the CVS repository, you may want to offer anonymous CVS access. A short warning first: There is not much demand for it, it requires some experience, and you need to know what you are doing.

Generally there are two ways to access a CVS repository remotely: via *pserver* or via `ssh` (we do not consider `rsh`). For anonymous access, *pserver* is very well suited, but some still offer `ssh` access as well. There is a custom crafted wrapper in the CVS repository, to be used as a login-shell for the anonymous ssh account. It does a chroot, and therefore requires the CVS repository to be available under the anonymous user's home-directory. This may not be possible for all sites. If you just offer *pserver* this restriction does not apply, but you may run with more security risks. You do not need to install any special software, since cvs(1) comes with FreeBSD. You need to enable access via `inetd`, so add an entry into your `/etc/inetd.conf` like this:

```
cvspserver stream tcp nowait root /usr/bin/cvs cvs -f -l -R -T /
anoncvstmp --allow-root=/home/ncvs pserver
```

See the manpage for details of the options. Also see the CVS *info* page about additional ways to make sure access is read-only. It is advised that you create an unprivileged account, preferably called `anoncvs`. Also you need to create a file `passwd` in your `/home/ncvs/CVSROOT` and assign a CVS password (empty or `anoncvs`) to that user. The directory `/anoncvstmp` is a special purpose memory based file system. It is not required but advised since cvs(1) creates a shadow directory structure in your `/tmp` which is not used after the operation but slows things dramatically if real disk operations are required. Here is an excerpt from `/etc/fstab`, how to set up such a MFS:

```
/dev/da0s1b /anoncvstmp mfs rw,-s=786432,-b=4096,-f=512,-i=560,-c=3,-
m=0,nosuid,nodev 0 0
```

This is (of course) tuned a lot, and was suggested by John Polstra.

3. How to Mirror FreeBSD

Ok, now you know the requirements and how to offer the services, but not how to get it. :-) This section explains how to actually mirror the various parts of FreeBSD, what tools to use, and where to mirror from.

3.1. FTP

The FTP area is the largest amount of data that needs to be mirrored. It includes the *distribution sets* required for network installation, the *branches* which are actually snapshots of checked-out source trees, the *ISO Images* to write CD-ROMs with the installation distribution, a live file system, lots of packages, the ports tree, distfiles, and a huge amount of packages. All of course for various FreeBSD versions, and various architectures.

3.1.1. With FTP mirror

You can use a FTP mirror program to get the files. Some of the most commonly used are:

- ftp/mirror
- ftp/ftpmirror
- ftp/emirror
- ftp/spegla
- ftp/omi
- ftp/wget

ftp/mirror was very popular, but seemed to have some drawbacks, as it is written in perl(1), and had real problems with mirroring large directories like a FreeBSD site. There are rumors that the current version has fixed this by allowing a different algorithm for comparing the directory structure to be specified.

In general FTP is not really good for mirroring. It transfers the whole file if it has changed, and does not create a single data stream which would benefit from a large TCP congestion window.

3.1.2. With rsync

A better way to mirror the FTP area is rsync. You can install the port net/rsync and then use rsync to sync with your upstream host. rsync is already mentioned in # 2.4.2, "Rsync (optional for FTP fileset)". Since rsync access is not required, your preferred upstream site may not allow it. You may need to hunt around a little bit to find a site that allows rsync access.

> ##
>
> Since the number of rsync clients will have a significant impact on the server machine, most admins impose limitations on their server. For a mirror, you should ask the site maintainer you are syncing from about their policy, and maybe an exception for your host (since you are a mirror).

A command line to mirror FreeBSD might look like:

```
% rsync -vaz --delete ftp4.de.FreeBSD.org::FreeBSD/ /pub/FreeBSD/
```

Consult the documentation for rsync, which is also available at http://rsync.samba.org/, about the various options to be used with rsync. If you sync the whole module (unlike subdirectories), be aware that the module-directory (here "FreeBSD") will not be created, so you cannot omit the target directory. Also you might want to set up a script framework that calls such a command via cron(8).

3.1.3. With CVSup

A few sites, including the one-and-only `ftp-master.FreeBSD.org` even offer CVSup to mirror the contents of the FTP space. You need to install a CVSup client, preferably from the port net/cvsup. (Also reread # 2.4.4, "CVSup (desired for CVS repository)".) A sample `supfile` suitable for `ftp-master.FreeBSD.org` looks like this:

```
          #
          # FreeBSD archive supfile from master server
          #
          *default host=ftp-master.FreeBSD.org
          *default base=/usr
          *default prefix=/pub
          #*default release=all
          *default delete use-rel-suffix
          *default umask=002

          # If your network link is a T1 or faster, comment out the ↺
following line.
          #*default compress

          FreeBSD-archive release=all preserve
```

It seems CVSup would be the best way to mirror the archive in terms of efficiency, but it is only available from few sites.

##

Please have look at the CVSup documentation like cvsup(1) and consider using the `-s` option. This reduces I/O operations by assuming the recorded information about each file is correct.

3.2. Mirroring the CVS repository

There are various ways to mirror the CVS repository. CVSup is the most common method.

3.2.1. Using CVSup

CVSup is described in some detail in # 2.4.4, “CVSup (desired for CVS repository)” and # 3.1.3, “With CVSup”.

It is very easy to setup a CVSup mirror. Installing net/cvsup-mirror will make sure all of the needed programs are installed and then gather all the needed information to configure the mirror.

##

Please do not forget to consider the hint mentioned in this note above.

3.2.2. Using other methods

Using other methods than CVSup is generally not recommended. We describe them in short here anyway. Since most sites offer the CVS repository as part of the FTP fileset under the path `/pub/FreeBSD/development/FreeBSD-CVS` , the following methods could be used.

- FTP
- Rsync
- HTTP

##

AnonCVS cannot be used to mirror the CVS repository since CVS does not allow you to access the repository itself, only checked out versions of the modules.

3.3. Mirroring the WWW pages

The best way is to check out the *www* distribution from CVS. If you have a local mirror of the CVS repository, it is as easy as:

```
% cvs -d /home/ncvs co www
```

and a *cronjob*, that calls `cvs up -d -P` on a regular basis, maybe just after your repository was updated. Of course, the files need to remain in a directory available for public WWW access. The installation and configuration of a web server is not discussed here.

If you do not have a local repository, you can use CVSup to maintain an “up to date copy” of the www pages. A sample supfile can be found in `/usr/share/examples/cvsup/www-supfile` and could look like this:

```
        #
        # WWW module supfile for FreeBSD
        #
        *default host=cvsup3.de.FreeBSD.org
        *default base=/usr
        *default prefix=/usr/local
        *default release=cvs tag=.
        *default delete use-rel-suffix

        # If your network link is a T1 or faster, comment out the ↺
following line.
        *default compress

        # This collection retrieves the www/ tree of the FreeBSD ↺
repository
        www
```

Using ftp/wget or other web-mirror tools is not recommended.

3.3.1. Mirroring the FreeBSD documentation

Since the documentation is referenced a lot from the web pages, it is recommended that you mirror the FreeBSD documentation as well. However, this is not as trivial as the www-pages alone.

First of all, you should get the doc sources, again preferably via CVSup. Here is a corresponding sample supfile:

```
        #
        # FreeBSD documentation supfile
        #
        *default host=cvsup3.de.FreeBSD.org
        *default base=/usr
        *default prefix=/usr/share
        *default release=cvs tag=.
        *default delete use-rel-suffix

        # If your network link is a T1 or faster, comment out the ↺
following line.
        #*default compress

        # This will retrieve the entire doc branch of the FreeBSD ↺
repository.
        # This includes the handbook, FAQ, and translations thereof.
        doc-all
```

Then you need to install a couple of ports. You are lucky, there is a meta-port: textproc/docproj to do the work for you. You need to set up some environment variables, like `SGML_CATALOG_FILES` . Also have a look at your `/etc/make.conf` (copy `/usr/share/examples/etc/make.conf` if you do not have one), and look at the `DOC_LANG` variable. Now you are probably ready to run `make` in your doc directory (`/usr/share/doc` by default) and build the documentation. Again you need to make it accessible for your web server and make sure the links point to the right location.

> ##
>
> The building of the documentation, as well as lots of side issues, is documented itself in the FreeBSD Documentation Project Primer. Please read this piece of documentation, especially if you have problems building the documentation.

3.4. How often should I mirror?

Every mirror should be updated on a regular basis. You will certainly need some script framework for it that will be called by cron(8). Since nearly every admin does this his own way, we cannot give specific instructions. It could work like this:

1. Put the command to run your mirroring application in a script. Use of a plain `/bin/sh` script is recommended.

2. Add some output redirections so diagnostic messages are logged to a file.

3. Test if your script works. Check the logs.

4. Use crontab(1) to add the script to the appropriate user's crontab(5). This should be a different user than what your FTP daemon runs as so that if file permissions inside your FTP area are not world-readable those files can not be accessed by anonymous FTP. This is used to “stage” releases —— making sure all of the official mirror sites have all of the necessary release files on release day.

Here are some recommended schedules:

- FTP fileset: daily
- CVS repository: hourly
- WWW pages: daily

4. Where to mirror from

This is an important issue. So this section will spend some effort to explain the backgrounds. We will say this several times: under no circumstances should you mirror from `ftp.FreeBSD.org` .

4.1. A few words about the organization

Mirrors are organized by country. All official mirrors have a DNS entry of the form `ftpN.CC.FreeBSD.org` . *CC* (i.e. country code) is the *top level domain* (TLD) of the country where this mirror is located. *N* is a number, telling that the host would be the *Nth* mirror in that country. (Same applies to `cvsupN.CC.FreeBSD.org` , `wwwN.CC.FreeBSD.org` , etc.) There are mirrors with no *CC* part. These are the mirror sites that are very well connected and allow a large number of concurrent users. `ftp.FreeBSD.org` is actually two machines, one currently located in Denmark and the other in the United States. It is *NOT* a master site and should never be used to mirror from. Lots of online documentation leads “interactive”users to `ftp.FreeBSD.org` so automated mirroring systems should find a different machine to mirror from.

Additionally there exists a hierarchy of mirrors, which is described in terms of *tiers*. The master sites are not referred to but can be described as *Tier-0*. Mirrors that mirror from these sites can be considered *Tier-1*, mirrors of *Tier-1*-mirrors, are *Tier-2*, etc. Official sites are encouraged to be of a low *tier*, but the lower the tier the higher the requirements in terms as described in # 2, “## FreeBSD mirrors #####”. Also access to low-tier-mirrors may be restricted, and access to master sites is definitely restricted. The *tier*-hierarchy is not reflected by DNS and generally not documented anywhere except for the master sites. However, official mirrors with low numbers like 1-4, are usually *Tier-1* (this is just a rough hint, and there is no rule).

4.2. Ok, but where should I get the stuff now?

Under no circumstances should you mirror from `ftp.FreeBSD.org` . The short answer is: from the site that is closest to you in Internet terms, or gives you the fastest access.

4.2.1. I just want to mirror from somewhere!

If you have no special intentions or requirements, the statement in # 4.2, "Ok, but where should I get the stuff now?" applies. This means:

1. Look at available mirrors in your country. The FreeBSD Mirror Database can help you with this.

2. Check for those which provide fastest access (number of hops, round-trip-times) and offer the services you intend to use (like rsync or CVSup).

3. Contact the administrators of your chosen site stating your request, and asking about their terms and policies.

4. Set up your mirror as described above.

4.2.2. I am an official mirror, what is the right site for me?

In general the description in # 4.2.1, "I just want to mirror from somewhere!" still applies. Of course you may want to put some weight on the fact that your upstream should be of a low tier. There are some other considerations about *official* mirrors that are described in # 5, "Official Mirrors".

4.2.3. I want to access the master sites!

If you have good reasons and good prerequisites, you may want and get access to one of the master sites. Access to these sites is generally restricted, and there are special policies for access. If you are already an *official* mirror, this certainly helps you getting access. In any other case make sure your country really needs another mirror. If it already has three or more, ask the "zone administrator" (<`hostmaster@CC.FreeBSD.org` >) or FreeBSD mirror sites ####s first.

Whoever helped you become, an *official* should have helped you gain access to an appropriate upstream host, either one of the master sites or a suitable Tier-1 site. If not, you can send email to <`mirror-admin@FreeBSD.org` > to request help with that.

There are three master sites for the FTP fileset and one for the CVS repository (the web pages and docs are obtained from CVS, so there is no need for master).

4.2.3.1. ftp-master.FreeBSD.org

This is the master site for the FTP fileset.

`ftp-master.FreeBSD.org` provides rsync and CVSup access, in addition to FTP. Refer to # 3.1.2, "With rsync" and # 3.1.3, "With CVSup" how to access via these protocols.

Mirrors are also encouraged to allow rsync access for the FTP contents, since they are *Tier-1*-mirrors.

4.2.3.2. cvsup-master.FreeBSD.org

This is the master site for the CVS repository.

`cvsup-master.FreeBSD.org` provides CVSup access only. See # 3.2.1, "Using CVSup" for details.

To get access, you need to contact the CVSup Mirror Site Coordinator <cvsup-master@FreeBSD.org>. Make sure you read the FreeBSD CVSup Access Policy first!

Set up the required authentication by following these instructions. Make sure you specify the server as `freefall.FreeBSD.org` on the `cvpasswd` command line, as described in this document, even when you are contacting `cvsup-master.FreeBSD.org`

5. Official Mirrors

Official mirrors ######### mirror #

- a) # `FreeBSD.org` DNS entry (### CNAME ##)#
- b) # FreeBSD #### offical mirror #(## handbook)#

############# official mirrors #### Official mirrors ###### *Tier-1*-mirrors ## ###### *Tier-1*-mirror ######### official mirrors ##

5.1. Special Requirements for official (tier-1) mirrors

It is not so easy to state requirements for all official mirrors, since the project is sort of tolerant here. It is more easy to say, what *official tier-1 mirrors* are required to. All other official mirrors can consider this a big *should*.

> **##**
>
> The following applies mainly to the FTP fileset, since a CVS repository should always be mirrored completely, and the web pages are a case of its own.

Tier-1 mirrors are required to:

- carry the complete fileset
- allow access to other mirror sites
- provide FTP and rsync access

Furthermore, admins should be subscribed to the FreeBSD mirror sites ####s. See this link for details, how to subscribe.

> ##
>
> It is *very* important for a hub administrator, especially Tier-1 hub admins, to check the release schedule for the next FreeBSD release. This is important because it will tell you when the next release is scheduled to come out, and thus giving you time to prepare for the big spike of traffic which follows it.
>
> It is also important that hub administrators try to keep their mirrors as up-to-date as possible (again, even more crucial for Tier-1 mirrors). If Mirror1 does not update for a while, lower tier mirrors will begin to mirror old data from Mirror1 and thus begins a downward spiral... Keep your mirrors up to date!

5.2. How to become official then?

An interesting question, especially, since the state of being official comes with some benefits, like a much higher bill from your ISP as more people will be using your site. Also it may be a key requirement to get access to a master site.

Before applying, please consider (again) if another official mirror is really needed for your region. Check first with your zone administrator (<`hostmaster@CC.FreeBSD.org`>) or, if that fails, ask on the FreeBSD mirror sites ####s.

Ok, here is how to do it:

1. Get the mirror running in first place (maybe not using a master site, yet).
2. Subscribe to the FreeBSD mirror sites ####s.
3. If everything works so far, contact the DNS administrator responsible for your region/country, and ask for a DNS entry for your site. The admin should able to be contacted via <`hostmaster@CC.FreeBSD.org`>, where *CC* is your country code/TLD. Your DNS entry will be as described in # 4.1, “A few words about the organization”.

If there is no subdomain set up for your country yet, you should contact <mirror-admin@FreeBSD.org >, or you can try the FreeBSD mirror sites ####s first.

4. Whoever helps you get an official name should send email to <mirror-admin@FreeBSD.org > so your site will be added to the mirror list in the FreeBSD Handbook.

That is it.

6. Some statistics from mirror sites

Here are links to the stat pages of your favorite mirrors (a.k.a. the only ones who feel like providing stats).

6.1. FTP site statistics

- ftp.is.FreeBSD.org - <hostmaster@is.FreeBSD.org > - (Bandwidth) (FTP processes) (HTTP processes)
- ftp.cz.FreeBSD.org - <cejkar@fit.vutbr.cz > - (Bandwidth) (FTP processes) (rsync processes)
- ftp2.ru.FreeBSD.org - <mirror@macomnet.ru > - (Bandwidth) (HTTP and FTP users)

6.2. CVSup site stats

- cvsup[23456].jp.FreeBSD.org - <kuriyama@FreeBSD.org > - (CVSup processes)
- cvsup.cz.FreeBSD.org - <cejkar@fit.vutbr.cz > - (CVSup processes)