Mirroring FreeBSD

Abstract

An in-progress article on how to mirror FreeBSD, aimed at hub administrators.

Table of Contents

1. Contact Information . . . 1
2. Requirements for FreeBSD Mirrors . . . 1
3. How to Mirror FreeBSD . . . 4
4. Where to Mirror From . . . 6
5. Official Mirrors . . . 7
6. Some Statistics from Mirror Sites . . . 8

We are not accepting new community mirrors at this time.

1. Contact Information

The Mirror System Coordinators can be reached through email at mirror-admin@FreeBSD.org. There is also a FreeBSD mirror sites mailing lists.

2. Requirements for FreeBSD Mirrors

2.1. Disk Space

Disk space is one of the most important requirements. 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 web pages should always be mirrored completely. Also note that the numbers stated here are reflecting the current state (at 12.0-RELEASE/11.3-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:

- Full FTP Distribution: 1.4 TB
- CTM deltas: 10 GB
- Web pages: 1GB

The current disk usage of FTP Distribution can be found at ftp://ftp.FreeBSD.org/pub/FreeBSD/dir.sizes.

2.2. Network Connection/Bandwidth

Of course, you need to be connected to the Internet. The required bandwidth depends on your intended use of the mirror. If you just want to mirror some parts of FreeBSD for local use at your site/intranet, the demand may be much smaller than if you want to make the files publicly available. If you intend to become an official mirror, the bandwidth required will be even higher. We can only give rough estimates here:

- 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. System Requirements, 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 rsync. This can have a huge impact on CPU and memory requirements as it is considered a memory hog. 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 SVN 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.

FreeBSD's `ftpd`, `proftpd` 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/apache24: Apache is still one of the most widely deployed web servers on the Internet. It is used extensively by the FreeBSD Project.
- www/boa: Boa is a single-tasking HTTP server. Unlike traditional web servers, it does not fork for each incoming connection, nor does it fork many copies of itself to handle multiple connections. Although, it should provide considerably great performance for purely static content.
- www/cherokee: Cherokee is a very fast, flexible and easy to configure web server. It supports the widespread technologies nowadays: FastCGI, SCGI, PHP, CGI, SSL/TLS encrypted connections, vhosts, users authentication, on the fly encoding and load balancing. It also generates Apache compatible log files.
- www/lighttpd: lighttpd is a secure, fast, compliant and very flexible web server which has been optimized for high-performance environments. It has a very low memory footprint compared to other web servers and takes care of cpu-load.

- www/nginx: nginx is a high performance edge web server with a low memory footprint and key features to build a modern and efficient web infrastructure. Features include a HTTP server, HTTP and mail reverse proxy, caching, load balancing, compression, request throttling, connection multiplexing and reuse, SSL offload and HTTP media streaming.
- 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 others. It is also optimized for excellent performance on FreeBSD.

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. Mirroring the FTP Site

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, and a snapshot of the ports tree. All of course for various FreeBSD versions, and various architectures.

The best 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 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 -vaHz --delete rsync://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).

3.2. Mirroring the WWW Pages

Since doc migration to Hugo/Asciidoctor on 2021-01-25, mirroring the website with rsync no longer works.

There are ongoing studies to implement a website mirror with the official infrastructure.

For the former website mirrors, a way to achieve mirroring the website today is building the website locally with the corresponding address it will be hosted.

```
% cd website && env HUGO_baseURL="https://www.XX.freebsd.org/" make
```

Check for more details about the build tools on FreeBSD Documentation Project Primer for New Contributors book.

Notice the website was split into www.FreeBSD.org and docs.FreeBSD.org, and there are links between them; plus, at this moment, `HUGO_baseURL` variable won't cover all links, this way, mirroring the website is discouraged.

3.3. Mirroring Packages

Due to very high requirements of bandwidth, storage and administration the FreeBSD Project has decided not to allow public mirrors of packages. For sites with lots of machines, it might be advantageous to run a caching HTTP proxy for the pkg(8) process. Alternatively specific packages and their dependencies can be fetched by running something like the following:

```
% pkg fetch -d -o /usr/local/mirror vim
```

Once those packages have been fetched, the repository metadata must be generated by running:

```
% pkg repo /usr/local/mirror
```

Once the packages have been fetched and the metadata for the repository has been generated, serve the packages up to the client machines via HTTP. For additional information see the man pages for pkg(8), specifically the pkg-repo(8) page.

3.4. How Often Should I Mirror?

Every mirror should be updated at a minimum of once per day. Certainly a script with locking to prevent multiple runs happening at the same time will be needed to run from cron(8). Since nearly every admin does this in their own way, specific instructions cannot be provided. It could work something 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 cannot 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
- 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 `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 Requirements for 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 Ok, but Where Should I get the Stuff Now? applies. This means:

1. Check for those which provide fastest access (number of hops, round-trip-times) and offer the services you intend to use (like rsync).
2. Contact the administrators of your chosen site stating your request, and asking about their terms and policies.
3. 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 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 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 mailing lists 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 is one master site for the FTP fileset.

ftp-master.FreeBSD.org

This is the master site for the FTP fileset.

`ftp-master.FreeBSD.org` provides rsync access, in addition to FTP. Refer to Mirroring the FTP Site.

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

5. Official Mirrors

Official mirrors are mirrors that

- a) have a `FreeBSD.org` DNS entry (usually a CNAME).
- b) are listed as an official mirror in the FreeBSD documentation (like handbook).

So far to distinguish official mirrors. Official mirrors are not necessarily *Tier-1*-mirrors. However you probably will not find a *Tier-1*-mirror, that is not also official.

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*.

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 mailing lists. 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?

Please contact the Cluster Administrators as documented at https://www.FreeBSD.org/administration/#t-clusteradm.

6. Some Statistics from Mirror Sites

Here are links to the stat pages of your favorite mirrors (aka 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)
- ftp2.ru.FreeBSD.org - mirror@macomnet.ru - (Bandwidth) (HTTP and FTP users)