Making use of snapshot.debian.org for fun and profit

Johannes Schauer Marin Rodrigues    Frédéric Pierret

August 26, 2021
Johannes Schauer Marin Rodrigues, PhD (@josch)

- upstream author of mmdebstrap, sbuild, botch, img2pdf, dose3, plakativ
- service maintainer of bootstrap.d.n, binarycontrol.d.n, metasnap.d.n
- Contact: josch@debian.org, johannes.schauer@uni-wuerzburg.de
Frédéric Pierret, PhD (@fepitre)

- **Research and development:**
  - Modeling mathematical dynamical systems,
  - Building tools for operating systems or packages.

- **Contact:**
  - github.com/fepitre
  - frederic.pierret@qubes-os.org
Timeline

- **2014** buildinfo2snapshot
- **2015** srebuild (#774415)
- **2016** debrebuild (Perl)
- **2020** debbisect, metasnap.d.n, debrebuild (Python)
- **2021** debootsnap, package-rebuilder, https://debian.notset.fr/snapshot

---

5. https://salsa.debian.org/debian/devscripts/-/merge_requests/231
Overview of the talk

1. Current situation
2. What we want to have
3. The problems
4. Our solutions
Current situation

- started collecting in 2005
- more than 90TB of data in more than 35 million files (as of 2019)
- main archive plus backports, debug, ports, security, volatile
- all architectures we ever had (including ports)
- main interface via debsnap tool
What we want to have

- GPG verified downloads via signed Release file
- reproducing binary package builds using buildinfo files
- bisecting Debian to track down bugs
- create Debian chroots with particular package versions to reproduce bugs
The problems

- binary packages are not associated to signed Release files
- `first_seen` attribute from the API is unreliable
- the network problems...
This will not work:

```
mmdebstrap \ 
--include=libreoffice \ 
--aptopt='Acquire::Check-Valid-Until "false"' \ 
--variant=standard stable /dev/null \ 
https://snapshot.debian.org/archive/debian/20210104T111800Z/
```
J. Schauer, Marin Rodrigues, F. Pierret
Making use of snapshot.debian.org for fun and profit

An error similar to this one will be produced:

Err:245 https://snapshot.debian.org/archive/debian/20210104T111800Z
stable/main amd64 glib-networking-services amd64 2.58.0-2+deb10u2 Could not
connect to snapshot.debian.org:443 (185.17.185.185). - connect (111:
Connection refused) Cannot initiate the connection to
snapshot.debian.org:443 (2001:1af8:4020:b030:deb::185). - connect (101:
Network is unreachable) Cannot initiate the connection to
(101: Network is unreachable) [IP: 185.17.185.185 443]
Other funny errors:

E: Failed to fetch [...] Error reading from server. Remote end closed connection
E: Failed to fetch [...] Hash Sum mismatch
E: Failed to fetch [...] Bad header line Bad header data
E: Failed to fetch [...] Undetermined Error
Err:118 [...] Connection timed out
Solution #1

- Bugs: #838561, #960304, #969906, #977653
- issue might be the varnish cache
- Julien Cristau is working on a solution
Solution #2

Acquire::Retries 3
Acquire::http::Dl-Limit "1000"
Acquire::https::Dl-Limit "1000"
Solution #3

- download with no more than 1000 kB/s
- timeout connections after X seconds of inactivity
- download no more than 2000 files per hour
- retry up to X times in case of:
  - zero size files
  - interrupted partial downloads
  - E_COULDN'T_CONNECT
  - inactivity
  - all HTTP errors that are not 200, 206 or 404 (most likely HTTP 500)
- back-off by $4^{\#\text{retries}} \text{s}$
Solution #4

- https://debian.notset.fr/snapshot
- will be introduced later in this talk
Solution #5

- building a local package cache
- used by the debbisect tool
like git bisect for Debian
proxies apt downloads from snapshot.d.o to not hit rate limits
script is run without chroot, inside a chroot or inside qemu
example to find package responsible for bug #912935:

```
$ ./debbisect --depends=botch "2018-11-17" "2018-11-22" > 'chroot "$1" botch-dose2html --packages=/dev/null --help'

[...] test upgrading python3-minimal 3.6.7-1 -> 3.7.1-2...
upgrading python3-minimal triggered the problem
```
cannot skip timestamps because snapshot.d.o does not export a list of known timestamps (see #969603)
cannot (yet) restrict tested timestamps to those where a given package set changed (can be solved using metasnap.d.n)
The first_seen attribute

- problem known since 2014 (#774279)
- only records the timestamp when a given file (package) was first seen
- a package can be in the /pool directory but not referenced by a Packages file
- does not store which suite the package appears in (works 99% for unstable)
- does not record when a package was last seen
- does not account for “holes” in the seen interval
public webservice with machine readable API

given an package, version, architecture tuple, find the Debian suites and timestamp ranges that reference it

e.example:

curl -F 'buildinfo=<foo.buildinfo' https://metasnap.debian.net/cgi-bin/api

used by debbisect, debootsnap, debrebuild

all archives, timestamps and architectures from snapshot.d.o

information is 12 hours old because of #979115
- create a chroot with exact package versions
  
  $ dpkg-query --showformat '${binary:Package}=${Version}\n' --show | \
  debootsnap --architecture=amd64 --pkgs=- > chroot.tar

  $ debootsnap --buildinfo=foo.buildinfo > chroot.tar

- uses metasnap.d.n to compute minimal set of snapshot timestamps
- uses mmdebstrap instead of debootstrap because
  - we need support for multiple mirrors (multiple snapshot timestamps) and
  - building a chroot without apt (buildinfo files do not contain apt)
- allows reproducing bugs, for example upgrade issues
- currently on the third rewrite
- rebuild a source package in a chroot with package versions from a buildinfo file
- used to use fragile `first_seen` attribute and `base-files` upgrade heuristic
- will become a thin wrapper around `debootsnap` to create a chroot in which the source package is built
- [https://salsa.debian.org/debian/devscripts/-/merge_requests/231/](https://salsa.debian.org/debian/devscripts/-/merge_requests/231/)
● Similar tool than *devscripts/debrebuild.pl* written in Python,

● Attempt to reproduce Debian package according to *buildinfo* file,

● With respect to *devscripts/debrebuild.pl*:
  ● Supports multiple repository sources (e.g. Qubes OS and Debian),
  ● Generates *in-toto* metadata\(^7\).

● Can be used as a:
  ● standalone shell tool,
  ● python client to be imported in another Python code.

\(^7\)https://in-toto.io/examples/debian/
- Standalone rebuildor orchestrator written in Python,
- Global architecture:

- Runs as multi-container Docker/Podman applications.

package-rebuilder – https://github.com/fepitre/package-rebuilder

Monitor and rebuild latest packages available for Debian, Qubes OS and Fedora.

Use rebuilder softwares (*debrebuild* and *rpmreproduce*):

Publish rebuilders artifacts:

- **in-toto** metadata\(^9\), \(^10\),
- logs, statistics, etc.

Serves repository for *apt-transport-in-toto*\(^11\) (also *dnf-plugin-in-toto*\(^12\)).

---

\(^8\)https://github.com/fepitre/rpmreproduce
\(^9\)https://debian.notset.fr/rebuild/
\(^10\)https://qubes.notset.fr/rebuild/
\(^11\)https://github.com/in-toto/apt-transport-in-toto
\(^12\)https://github.com/fepitre/dnf-plugin-in-toto
New snapshot service written from scratch to handle scalability\textsuperscript{13,14,15}

Independant project from original \texttt{snapshot.debian.org}.

With respect to \texttt{snapshot.d.o}:
- Same set of timestamps values (for compatibility and debug):
- Available data from 2017 to today,
- unstable, buster and bullseye suites,
- all, source and amd64 architectures,

Up to past data, it runs on it’s own.

Base URL: \url{https://debian.notset.fr/snapshot}

API: \url{https://debian.notset.fr/snapshot/mr}

\textsuperscript{13} #977653 Please document rate limits on snapshots.debian.org
\textsuperscript{14} #960304 snapshot.debian.org: Snapshot repo repeatedly cutting off connection, returning partial content
\textsuperscript{15} #969906 snapshot.debian.org: error 500 internal server error after some requests via Python
Get `debian` archive available timestamps:\footnote{\url{http://debian.notset.fr/snapshot/mr/timestamp/debian}}:

```
{
    "api": "0.3",
    "comment": "notset",
    "result": [
        "20170101T032652Z",
        "20170101T092432Z",
        (...)
        "20210719T031839Z",
        "20210719T090459Z"
    ]
}
```
Get source files info for `python-designateclient` package version 2.3.0-2:

```json
{
    "_api": "0.3",
    "_comment": "notset",
    "package": "python-designateclient",
    "version": "2.3.0-2",
    "result": [
        {
            "hash": "240d86861138fbf8a34c1bf96412bf290dc8eaed4a660473b0ecee605b8d1288f"
        },
        {
            "hash": "d65b4d861612c0bed42cdeceadb0c32d886fc27bdc5642399ed410de042ed85"
        },
        {
            "hash": "ffb63b9b69d579fabd05d81a84c679dc396c29a663fcd244b0e8c600257478f3"
        }
    ]
}
```

"fileinfo": { 
   "240d86861138ffbf8a34c1bf96412bf290dc8eae4a560473b0ece605b8d1288f": [ 
      { 
         "name": "python-designateclient_2.3.0-2.dsc", 
         "path": "/pool/main/p/python-designateclient", 
         "size": 3417, 
         "archive_name": "debian", 
         "suite_name": "buster", 
         "component_name": "main", 
         "timestamp_ranges": [ 
            ["20170618T072316Z", "20170821T035341Z"], 
            ["20170822T154312Z", "20170922T035316Z"], 
            ["20170924T042402Z", "20171024T092932Z"], 
            ["20171025T221056Z", "20171106T213509Z"] 
         ] 
      }, 
      { 
         "name": "python-designateclient_2.3.0-2.dsc", 
         "path": "/pool/main/p/python-designateclient", 
         "size": 3417, 
         "archive_name": "debian", 
         "suite_name": "unstable", 
         "component_name": "main", 
         "timestamp_ranges": [ 
            ["20170101T032652Z", "20171101T160520Z"] 
         ] 
      } 
   ], 
   "d65b4d861612c0bed42dcedbcb0c32d886fc27bdc5642399ed410de042ed85": [ 
      (...) 
   ]}
Compute a minimal set of timestamps containing all package versions referenced in a buildinfo file (à la metasnap)\(^\text{18}\):

\[
\begin{array}{l}
\{ \\
\quad "\_api": "0.3", \\
\quad "\_comment": "notset: This feature is currently very experimental!", \\
\quad "\_results": [ \\
\quad \{ \\
\quad \quad "archive\_name": "debian", \\
\quad \quad "suite\_name": "buster", \\
\quad \quad "component\_name": "main", \\
\quad \quad "architecture": "amd64", \\
\quad \quad "timestamps": [ \\
\quad \quad \quad "20210705T151228Z" \\
\quad \quad ] \\
\quad \} \\
\quad ] \\
\}
\end{array}
\]

\(^{18}\)https://buildinfos.debian.net/buildinfo-pool/b/bash/bash_5.1-2_amd64.buildinfo