Making Fedora Linux (more) reproducible

Davide Cavalca
Fedora Contributor

✉️ dcavalca@fedoraproject.org
 يمكنك المطالبة بمزيد من المعلومات حول محتوى الفيدورا أو الاتصال بمطور الفيدورا عن طريق البريد الإلكتروني في عنوان البريد الإلكتروني المذكور. كيف يمكن للمطورين مساعدتك في مشاركتهم في التحسينات والميزات الجديدة؟

🔍 @davide.cavalca.name
 يمكنك متابعة مطور الفيدورا على الـ Twitter للتواصل والتواصل معهم وتزويدهم بالملاحظات أو الأسئلة. كيف يمكن للمطورين فهم وتطبيق الملاحظات المقدمة من قبل المستخدمين؟
Agenda

- Reproducible builds
- Reproducible builds in Fedora
- Irreducibility examples
- Current status
Reproducible builds
Reproducible builds

What

A build is reproducible if given the same source code, build environment and build instructions, any party can recreate bit-by-bit identical copies of all specified artifacts.

– reproducible-builds.org
Reproducible builds

Why

- Security
  - Independent verification
  - Strengthen the supply chain
  - Trusting trust
- Quality
  - Hardware failures
  - Build system issues
  - Packaging bugs
  - Software bugs
Reproducible builds

History

- Started in 2013 within Debian
- By 2017 more than 90% packages reproducible, >98% today
- Automated rebuilds
- Dashboards
- Tools
  - Diffoscope
  - strip-nondeterminism
- [https://reproducible-builds.org](https://reproducible-builds.org)
Reproducible builds in Fedora
Reproducible builds in Fedora

Infrastructure

- Centralized infrastructure managed by Red Hat CPE
- Package sources stored in dist-git: https://src.fedoraproject.org
- Packages built in Koji: https://koji.fedoraproject.org
  - Non-scratch builds only from dist-git
  - Release artifacts only from Koji
- dist-git -> src.rpm -> binary rpms
- Koji also performs signing
Reproducible builds in Fedora

Package signatures

- Signatures stored as tags in the RPM
- Cannot be reproduced by design
  - Private key is private
  - Some singing algorithms introduce randomness
- Detached signatures?
  - https://github.com/rpm-software-management/rpm/issues/1482

<table>
<thead>
<tr>
<th>Tag Name</th>
<th>Value</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dsaheader</td>
<td>267</td>
<td>bin</td>
<td>OpenPGP DSA signature of the header (if thus signed)</td>
</tr>
<tr>
<td>Longsigsig</td>
<td>270</td>
<td>int64</td>
<td>Header+payload size if &gt; 4GB.</td>
</tr>
<tr>
<td>Payloaddigest</td>
<td>5092</td>
<td>string array</td>
<td>Cryptographic digest of the compressed payload.</td>
</tr>
<tr>
<td>Payloaddigestalgo</td>
<td>5093</td>
<td>int32</td>
<td>ID of the payload digest algorithm.</td>
</tr>
<tr>
<td>Payloaddigestalgalt</td>
<td>5097</td>
<td>string array</td>
<td>Cryptographic digest of the uncompressed payload.</td>
</tr>
<tr>
<td>Rsaheader</td>
<td>268</td>
<td>bin</td>
<td>OpenPGP RSA signature of the header (if thus signed).</td>
</tr>
<tr>
<td>Sha1header</td>
<td>269</td>
<td>string</td>
<td>SHA1 digest of the header.</td>
</tr>
<tr>
<td>Sha256header</td>
<td>273</td>
<td>string</td>
<td>SHA256 digest of the header.</td>
</tr>
<tr>
<td>Sigpgp</td>
<td>262</td>
<td>bin</td>
<td>OpenPGP DSA signature of the header+payload (if thus signed).</td>
</tr>
<tr>
<td>Signd5</td>
<td>261</td>
<td>bin</td>
<td>MD5 digest of the header+payload.</td>
</tr>
<tr>
<td>Sigpgp</td>
<td>259</td>
<td>bin</td>
<td>OpenPGP RSA signature of the header+payload (if thus signed).</td>
</tr>
<tr>
<td>Sigsiz</td>
<td>257</td>
<td>int32</td>
<td>Header+payload size.</td>
</tr>
</tbody>
</table>
Reproducible builds

A build is reproducible if given the same source code, build environment and build instructions, any party can recreate bit-by-bit identical copies of all specified artifacts.

– reproducible-builds.org
Reproducible builds in Fedora

A build is reproducible if given the same source code, build environment and build instructions, and metadata from the build artifacts, any party can recreate copies of the artifacts that are identical except for the signatures and parts of metadata.

– https://discussion.fedoraproject.org/t/87469
Reproducible builds in Fedora

Package comparison

- Solution: ignore some tags when comparing
- Use rpmdiff for comparison
  - Suppresses differences that are not interesting
  - [https://github.com/rpm-software-management/rpmlint](https://github.com/rpm-software-management/rpmlint)
- Use diffoscope for in-depth comparisons
Irreproducibility examples
Irreproducibility examples

RPM headers

- Build-environment specific information in tags
  - BUILDHOST
  - BUILDTIME
  - Optflags
  - SPEC
- Could be stripped, but are valuable for debugging
- [https://github.com/rpm-software-management/rpm/issues/2602](https://github.com/rpm-software-management/rpm/issues/2602)
- [https://github.com/rpm-software-management/rpm/issues/2603](https://github.com/rpm-software-management/rpm/issues/2603)
Irreproducibility examples

Source RPMs

- Source RPMs encode specifics of the build environment
  - Arch is set to the builder arch
  - User/group file ownership
  - BuildRequires can vary based on the arch
- [https://github.com/rpm-software-management/rpm/issues/2601](https://github.com/rpm-software-management/rpm/issues/2601)
- [https://github.com/rpm-software-management/rpm/issues/2604](https://github.com/rpm-software-management/rpm/issues/2604)
Irreproducibility examples

**Timestamps**

- Timestamps can be tricky
  - Files modified in `%prep` and during the build
  - Git commits created by `%autosetup -S`

- Solution: clamp timestamps to `$SOURCE_DATE_EPOCH`
  - ... which in turn is set to the timestamp of the last changelog entry
  - F38: [https://fedoraproject.org/wiki/Changes/ReproducibleBuildsClampMtimes](https://fedoraproject.org/wiki/Changes/ReproducibleBuildsClampMtimes)
  - F41: [https://github.com/rpm-software-management/rpm/pull/2930](https://github.com/rpm-software-management/rpm/pull/2930)
Irreproducibility examples

Packaging bugs
- Python “pickle” files
  - Non deterministic object serialization
  - Usually build leftovers that need to be removed
- noarch packages installing files into archful paths
  - e.g. using %{_libdir} or using archful conditionals
  - Individual instances need to be debugged and fixed
Irreproducibility examples

Archives

- Java JAR files
  - Embed build timestamps
    https://pagure.io/fedora-reproducible-builds/project/issue/10
  - When extracted results depends on the local timezone
    https://pagure.io/fedora-reproducible-builds/project/issue/16

- Static library archives
  - “ar” format from gcc-ar
  - Embed timestamps and uid/gid
  - https://pagure.io/fedora-reproducible-builds/project/issue/7
Irreproducibility examples

Known issues

- Python .pyc serialization is arch-specific
  - Functionally equivalent but not bit-by-bit identical
  - [https://pagure.io/fedora-reproducible-builds/project/issue/12](https://pagure.io/fedora-reproducible-builds/project/issue/12)
  - Proposed fix via postprocessing: [https://src.fedoraproject.org/rpms/python-rpm-macros/pull-request/170](https://src.fedoraproject.org/rpms/python-rpm-macros/pull-request/170)

- Golang debuginfo is non deterministic
  - .gdb_index section size is not consistent
  - [https://pagure.io/fedora-reproducible-builds/project/issue/15](https://pagure.io/fedora-reproducible-builds/project/issue/15)
Current status
Current status

What are we doing?

- Distro-wide rebuilds in mock, comparing the results with koji
  - [https://github.com/keszybz/fedora-repro-build](https://github.com/keszybz/fedora-repro-build)
- Tracking exposed issues (and ideally fixing them)
- Writing documentation
Current status

Where are we?

- Reproducibility stats:
  - 55% of source packages are reproducible
  - 78% of binary packages are reproducible
- [https://fedorapeople.org/~zbyszek/builds-2024-02-fc41-filtered.summary.txt](https://fedorapeople.org/~zbyszek/builds-2024-02-fc41-filtered.summary.txt)
Current status

What’s next

- Fixing more reproducibility issues
- Integrating add-determinism
  - Rust reimplementaion of strip-nondeterminism
  - https://github.com/keszybz/add-determinism
- Automated rebuilds
- Integration with reproducible-builds.org dashboards
Current status

Want to help?

- Report and fix reproducibility issues
  - https://pagure.io/fedora-reproducible-builds/project
- Contribute to the documentation
  - https://pagure.io/fedora-reproducible-builds/docs-site
- Participate in upstream discussions
  - https://github.com/rpm-software-management/rpm/discussions/2654
  - https://github.com/rpm-software-management/rpm/discussions/2934
- Join us on Matrix!
  - #reproducible-builds:fedora.im
Questions?