



# Making Fedora Linux (more) reproducible

**Davide Cavalca**

Fedora Contributor

✉ [dcavalca@fedoraproject.org](mailto:dcavalca@fedoraproject.org)

[m] [@davide:cavalca.name](https://matrix.to/#/!d Davide Cavalca)

# Agenda

- Reproducible builds
- Reproducible builds in Fedora
- Irreproducibility 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](https://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>

# 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
- <https://rpm-software-management.github.io/rpm/manual/tags.html>
- Cannot be reproduced by design
  - Private key is private
  - Some signing algorithms introduce randomness
- Detached signatures?
  - <https://github.com/rpm-software-management/rpm/issues/1482>

Tag Name	Value	Type	Description
Dsaheader	267	bin	OpenPGP DSA signature of the header (if thus signed)
Longsigsize	270	int64	Header+payload size if > 4GB.
Payloaddigest	5092	string array	Cryptographic digest of the compressed payload.
Payloaddigestalgo	5093	int32	ID of the payload digest algorithm.
Payloaddigestalt	5097	string array	Cryptographic digest of the uncompressed payload.
Rsaheader	268	bin	OpenPGP RSA signature of the header (if thus signed).
Sha1header	269	string	SHA1 digest of the header.
Sha256header	273	string	SHA256 digest of the header.
Siggpg	262	bin	OpenPGP DSA signature of the header+payload (if thus signed).
Sigmd5	261	bin	MD5 digest of the header+payload.
Sigpgp	259	bin	OpenPGP RSA signature of the header+payload (if thus signed).
Sigsize	257	int32	Header+payload size.

# 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](https://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>
- 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/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/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>
  - F41: <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>
  - Proposed fix via postprocessing:  
<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>

# 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>
- Tracking exposed issues (and ideally fixing them)
  - <https://pagure.io/fedora-reproducible-builds/project/issues?tags=irreproducibility&status=Open>
- Writing documentation
  - <https://docs.fedoraproject.org/en-US/reproducible-builds/>

# 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>

# Current status

## What's next

- Fixing more reproducibility issues
- Integrating add-determinism
  - Rust reimplement 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](https://matrix.to/#/#reproducible-builds:fedora.im)



**Questions?**

