# Reproducible builds for Debian... and more?

Lunar < lunar@debian.org>

2014-02-01 FOSDEM'14

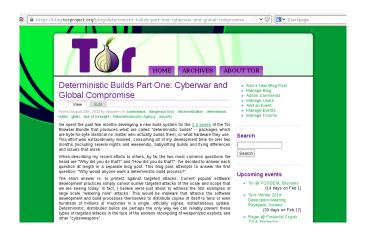
## What are reproducible builds?

"reproducible" builds enable anyone to reproduce the exact same binary packages from a given source

# Why?

- Prevent targeted attacks
- Debugging: ensure known source; create missing debug symbols
- Help building Multi-Arch: same packages (Debian specific)

#### How did this start?



## Nothing new

From: Martin Uecker <muecker@gmx.de>
Cc: debian-devel@lists.debian.org

Date: Sun, 23 Sep 2007 23:32:59 +0200

I think it would be really cool if the Debian policy required that packages could be rebuild bit-identical from source. At the moment, it is impossible to independly verify the integricity of binary packages.

https://lists.debian.org/debian-devel/2007/09/msg00746.html

## Although, reactions were not enthusiastic

From: Neil Williams <codehelp@debian.org>

To: debian-devel@lists.debian.org
Date: Mon, 24 Sep 2007 07:22:30 +0100

> Then third parties can recreate the binaries

> and publish recreated hashes.

Why? I see no benefit.

https://lists.debian.org/debian-devel/2007/09/msg00747.html

## Although, reactions were not enthusiastic

From: Manoj Srivastava <srivasta@debian.org>

To: debian-devel@lists.debian.org
Date: Sun, 23 Sep 2007 23:25:16 -0500

I, for one, think this technically infeasible, but hey, I'll be happy to be proved wrong.

https://lists.debian.org/debian-devel/2007/09/msg00760.html

## BoF during DebConf13

- Planned at the last minute
- 30 attendees
- Kicked off wiki.debian.org/ReproducibleBuilds

## How?

- Record the build environment
- Reproduce the build environment
- Eliminate unneeded variations

#### Record the build environment

Record which versions of the build dependencies (and their dependencies) are installed.

## Reproduce the build environment

snapshot.debian.org

#### Source of variations

- Timestamps
- Build paths
- ▶ File order
- Locale
- **.**..

## Timestamps

#### gzip stores a timestamp.

```
$ file README.txt.gz
README.txt.gz: gzip compressed data, was "README.txt", from Unix,
last modified: Mon Mar 5 00:05:49 2012, max compression
```

## **Timestamps**

ar, tar, zip, jar... store timestamps.

## **Timestamps**

#### javadoc writes timestamps:

```
$ head -n 5 /usr/share/doc/libjaxe-java-doc/api/serialized-form.html
<!DDCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">
<!-- NewPage -->
<html lang="en">
<head>
<!-- Generated by javadoc (version 1.6.0 27) on Sat Jul 13 17:27:51 UTC 2013 -->
```

## Build paths

#### Build path is embedded in debug symbols:

## File order

readdir() returns file in the order of the file system.

#### Locale

b

а

#### Behaviour can change depending on configured locale:

```
$ printf 'a\nà\nb\n' | LC_ALL=C.UTF-8 sort
$ printf 'a\nà\nb\n' | LC_ALL=fr_FR.UTF-8 sort
```

## Misc.

- Hostname
- Uname output
- Username

## Cheat

- Use a VM: same kernel, same user, same build path
- ▶ libfaketime

## The hard path

- ► Configure the toolchain: binutils --enable-deterministic-archives
- Add missing options: javadoc --no-timestamps
- ▶ Patch build systems: gzip -n

- Build and rebuild of many source packages
- Using EC2 VM instances from Amazon Web Services
- Many thanks David Suárez!

- Build packages twice
- Setup clean chroot, unpack source code, install build-deps, build
- ► And again...
- Pass the timestamp of the first build to dpkg through environment variable

#### Variations in this context:

- ▶ Time
- Build path

No changes in hostname, username, uname, file order, locale...

#### Modified packages:

- dpkg: use single timestamp in the archives
- dpkg: re-use timestamp from environment if given
- dpkg: stable file order in the archives
- debhelper: dh\_strip calls debugedit
- dpkg: pass -fno-merge-debug-strings through dpkg-buildflags
- binutils: built with
  --enable-deterministic-archives

- Upon 5151 source packages
- 3196 produced identical binary packages

62%

Waow.

# Already reproducible

source name	popcon insts
findutils	164641
wget	164512
klibc	163312
busybox	161494
installation-report	157494
laptop-detect	157352
python-support	155075
netkit-ftp	145548

## Failures in the remaining packages

```
1017 build-id-mismatch
295 unknown
108 jar-file
106 haskell-prof
 103 haskell-dev
101 php-registry
 101 html-mismatch
  63 same-depends-different-order
  62 r-rds
  52 gzip-timestamp
  46 kde-doc-index
```

## Failures in the remaining packages

```
45 mono
35 specific
33 docbook-to-man-timestamp
23 do-not-use-dpkg-buildflags
21 debugedit-not-run-or-failed
16 puredata
13 perl-manpage
11 rdoc-timestamp
10 zip-file
 8 ocaml-md5sums
 7 fonts
 7 erlang
```

#### Other distributions

- Fedora
  http://securityblog.redhat.com/2013/
  09/18/reproducible-builds-for-fedora/
- OpenSUSE build-compare https://build.opensuse.org/package/ show/openSUSE:Factory/build-compare
- NixOS http://lists.science.uu.nl/pipermail/ nix-dev/2013-June/011357.html

## Questions? Comments?

7

wiki.debian.org/ReproducibleBuilds

Note: no harm was done to the privacy of any cat for this presentation.

