You think you're not a target?
A tale of 3 developers…

Chris Lamb
Debian Project Leader
@lolamby

foss-backstage.de
Berlin, Germany
13th June 2018

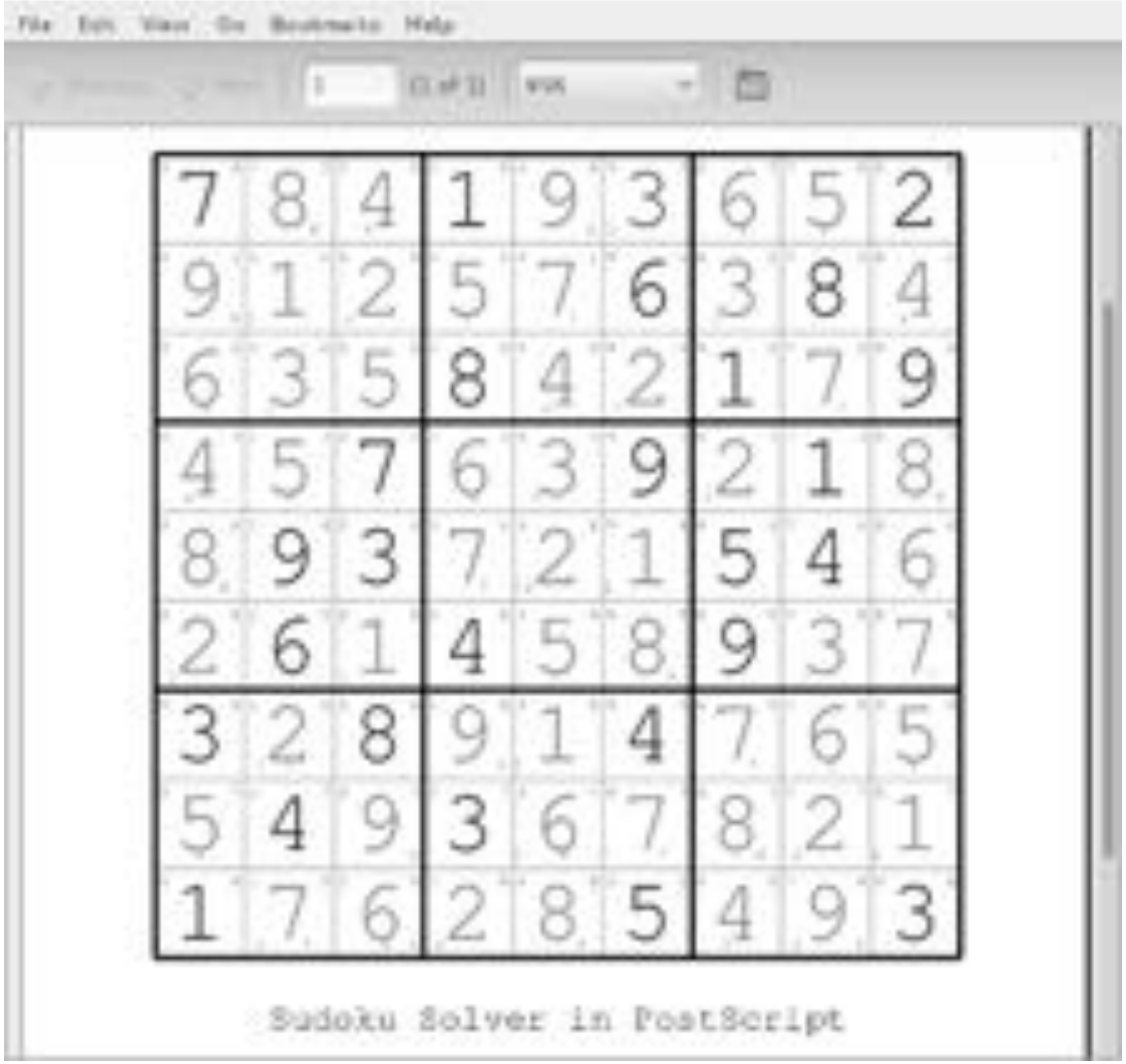
Cambridge
Analytica

Debian Project Leader 2017—

OpenSource.org board director

Free software developer for 10+ years

Freelance software developer



7	8	4	1	9	3	6	5	2
9	1	2	5	7	6	3	8	4
6	3	5	8	4	2	1	7	9
4	5	7	6	3	9	2	1	8
8	9	3	7	2	1	5	4	6
2	6	1	4	5	8	9	3	7
3	2	8	9	1	4	7	6	5
5	4	9	3	6	7	8	2	1
1	7	6	2	8	5	4	9	3

Sudoku Solver in PostScript

```
< zed0> can you get cp to give a progress bar like wget?
```

Damn right you can.

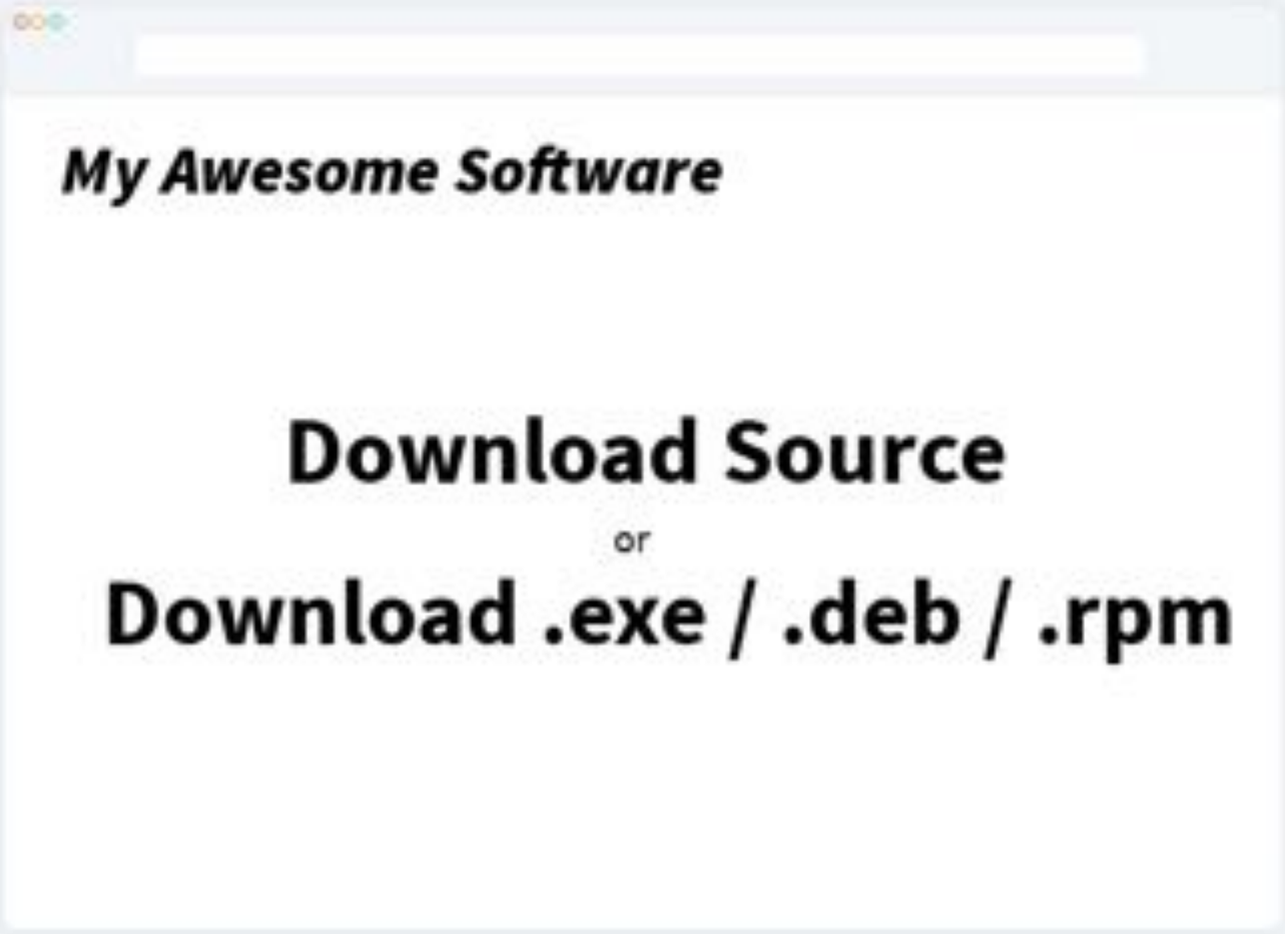
```
#!/bin/sh
cp_p()
{
   strace -q -ewrite cp -- "${1}" "${2}" 2>&1 \
      | awk '{
        count += $NF
            if (count % 10 == 0) {
               percent = count / total_size * 100
               printf "%3d%% [", percent
               for (i=0;i<=percent;i++)
                  printf "="
               printf ">"
               for (i=percent;i<100;i++)
                  printf " "
               printf "]\r"
            }
         }
         END { print "" }' total_size=$(stat -c '%s' "${1}") count=0
}
```

In action:

```
% cp_p /mnt/raid/pub/iso/debian/debian-2.2r4potato-i386-netinst.iso /dev/null
76% [============================================>                    ]
```

Three developers…

Alice

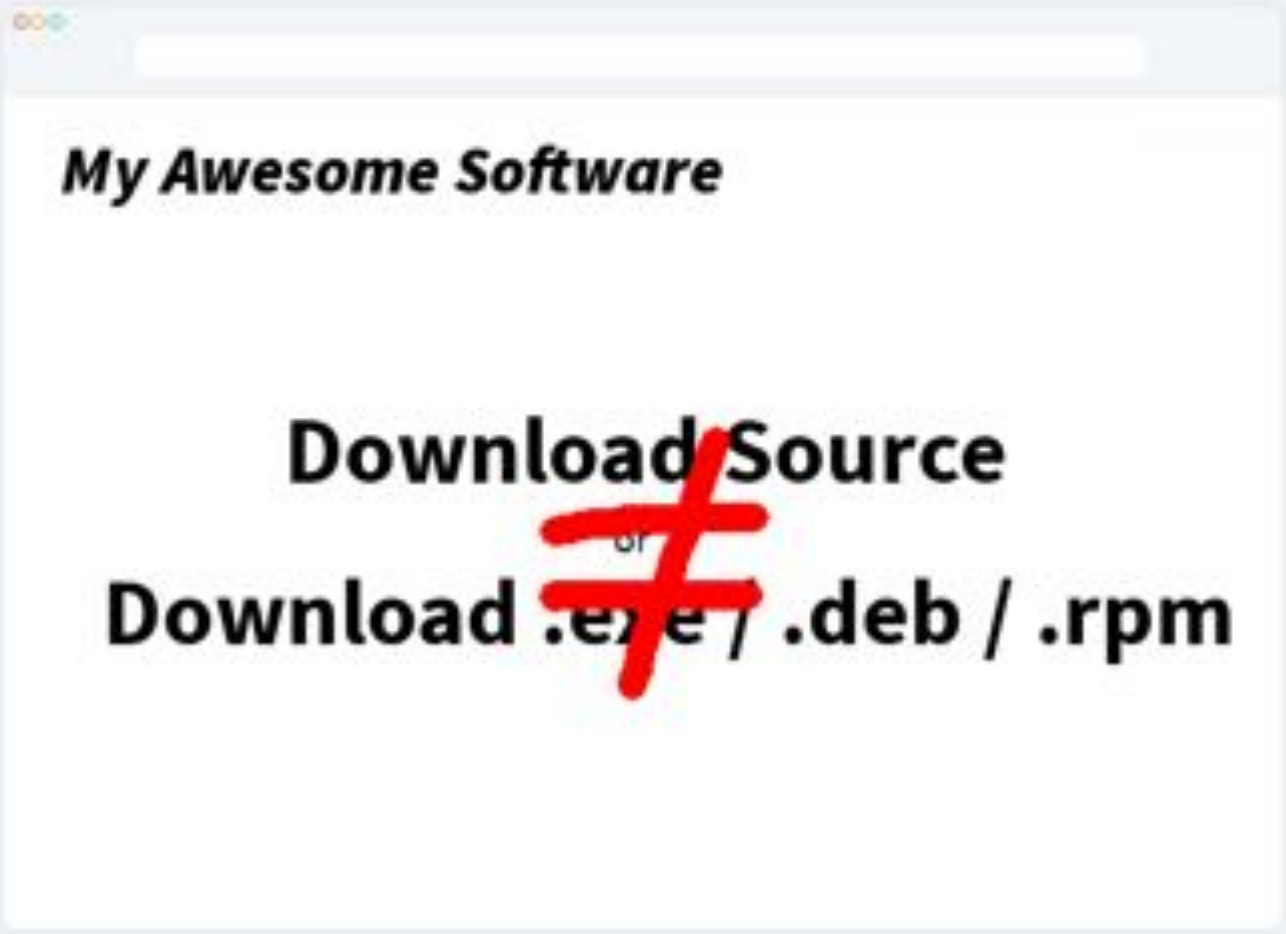


My Awesome Software

Download Source

or

Download .exe / .deb / .rpm



My Awesome Software

Download Source

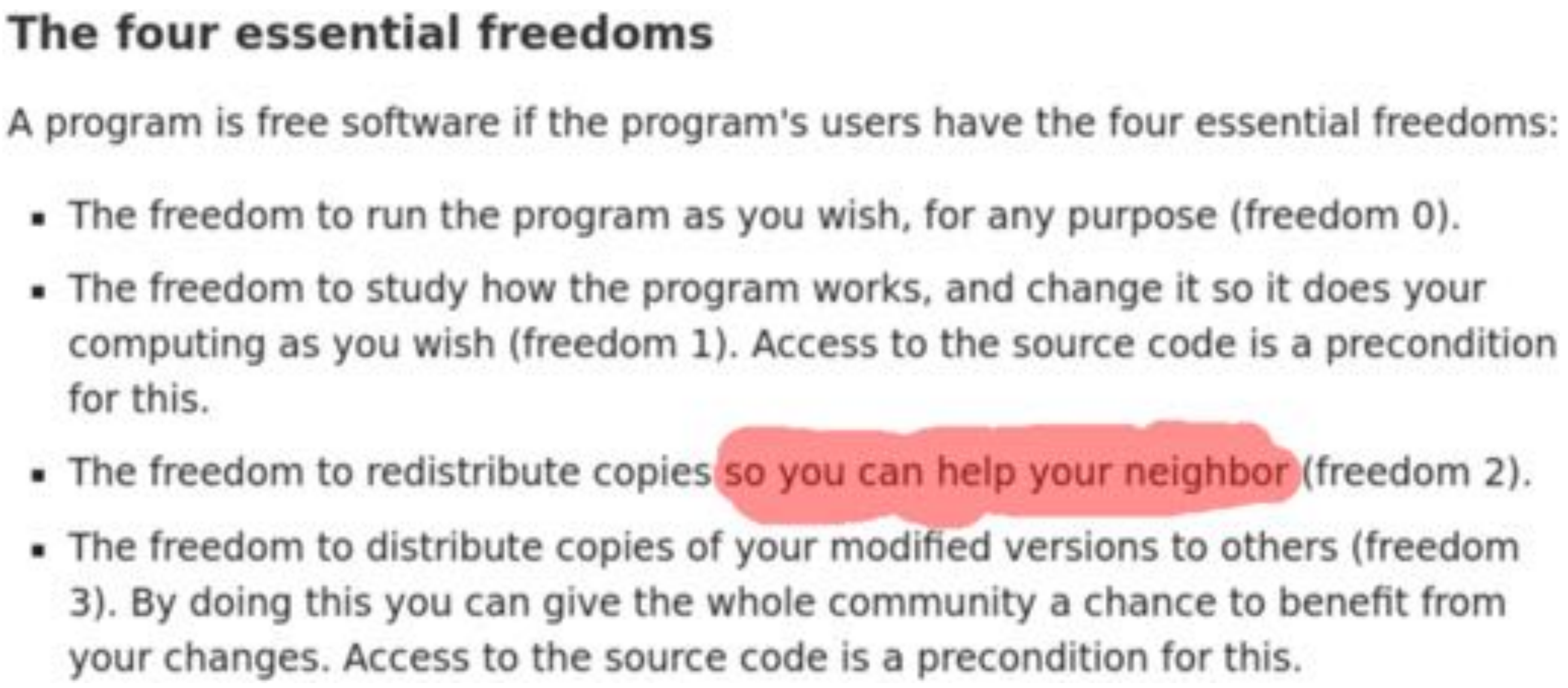
or ≠

Download .exe / .deb / .rpm

Bob

← Caro

Eve →



The four essential freedoms

A program is free software if the program's users have the four essential freedoms:

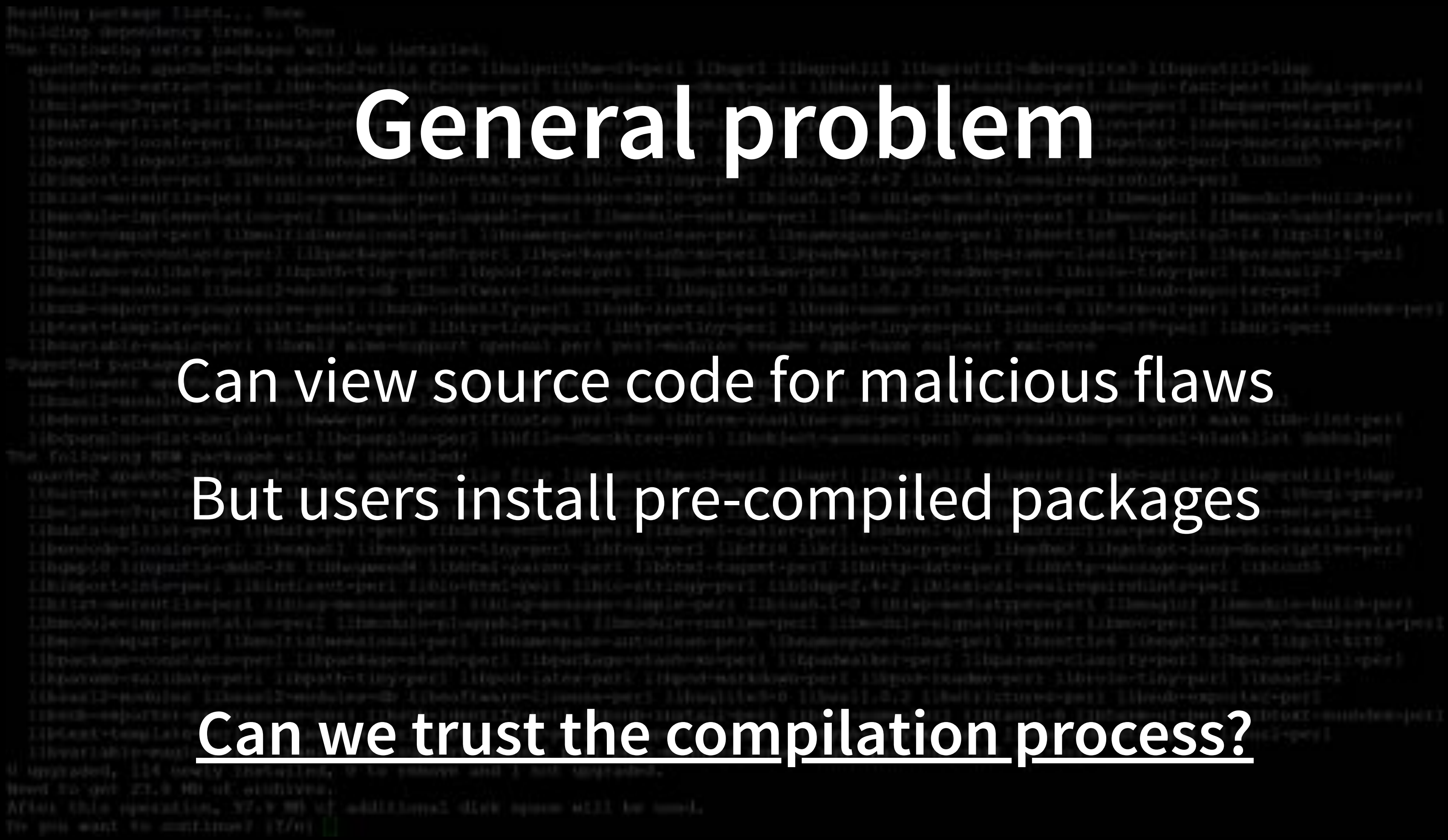
- The freedom to run the program as you wish, for any purpose (freedom 0).
- The freedom to study how the program works, and change it so it does your computing as you wish (freedom 1). Access to the source code is a precondition for this.
- The freedom to redistribute copies so you can help your neighbor (freedom 2).
- The freedom to distribute copies of your modified versions to others (freedom 3). By doing this you can give the whole community a chance to benefit from your changes. Access to the source code is a precondition for this.

General problem

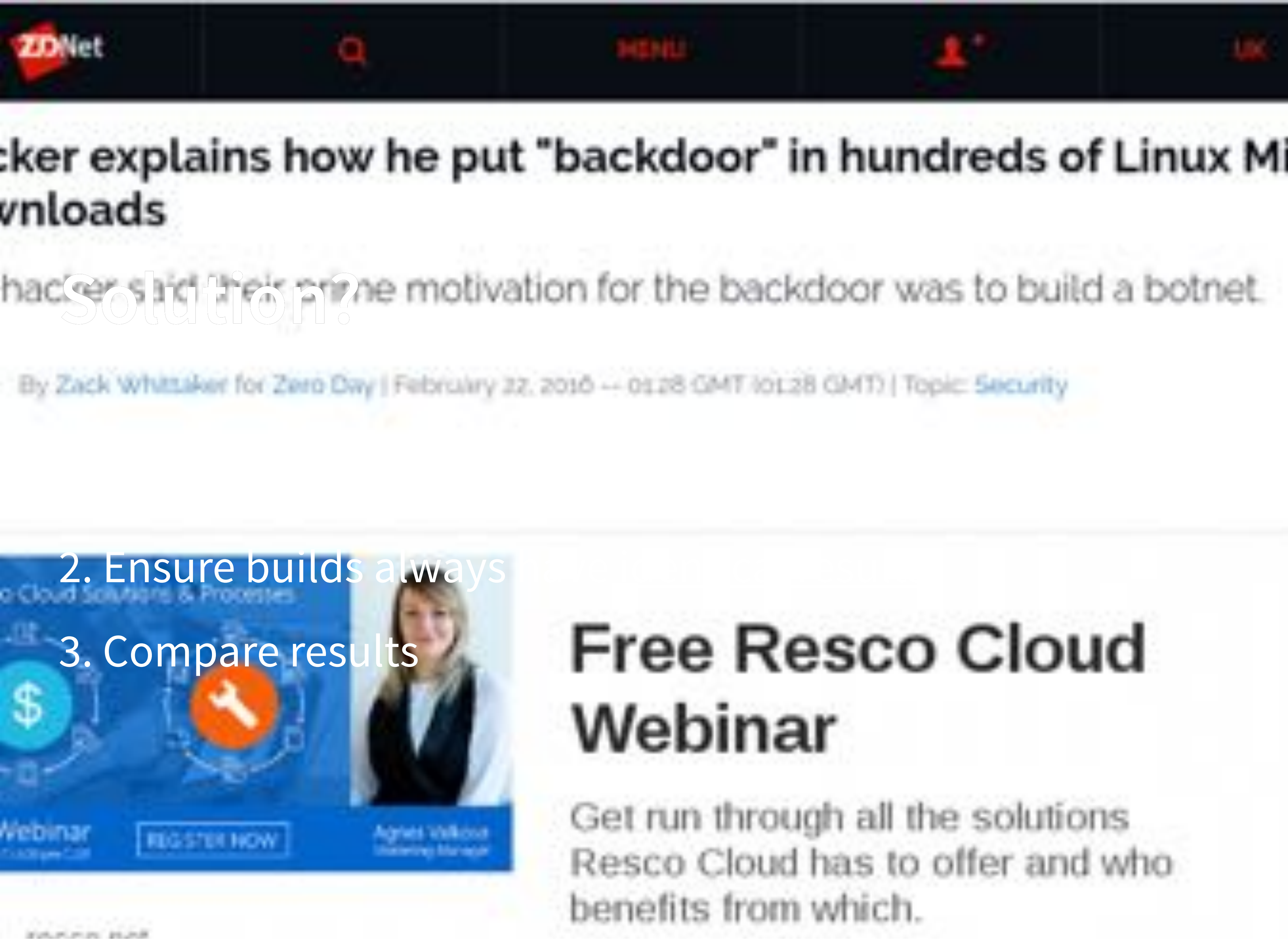
Can view source code for malicious flaws

But users install pre-compiled packages

Can we trust the compilation process?



```
Reading package lists... Done
Building dependency tree... Done
The following extra packages will be installed:
  [illegible]
Suggested packages:
  [illegible]
The following NEW packages will be installed:
  [illegible]
0 upgraded, 114 newly installed, 0 to remove and 1 not upgraded.
Need to get 23.9 MB of archives.
After this operation, 37.9 MB of additional disk space will be used.
Do you want to continue? [Y/n]
```



Solution?

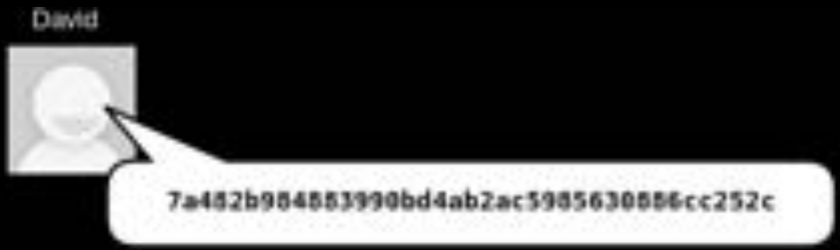
2. Ensure builds always

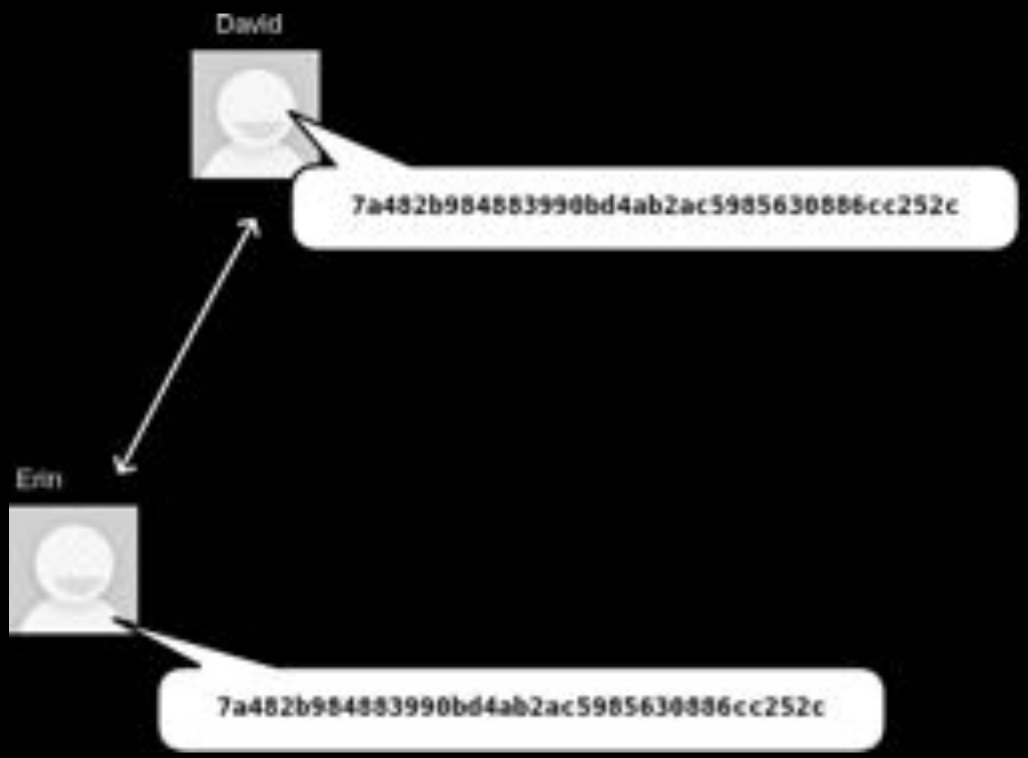
3. Compare results

cker explains how he put "backdoor" in hundreds of Linux Mi
wnloads

hacker said their prime motivation for the backdoor was to build a botnet.

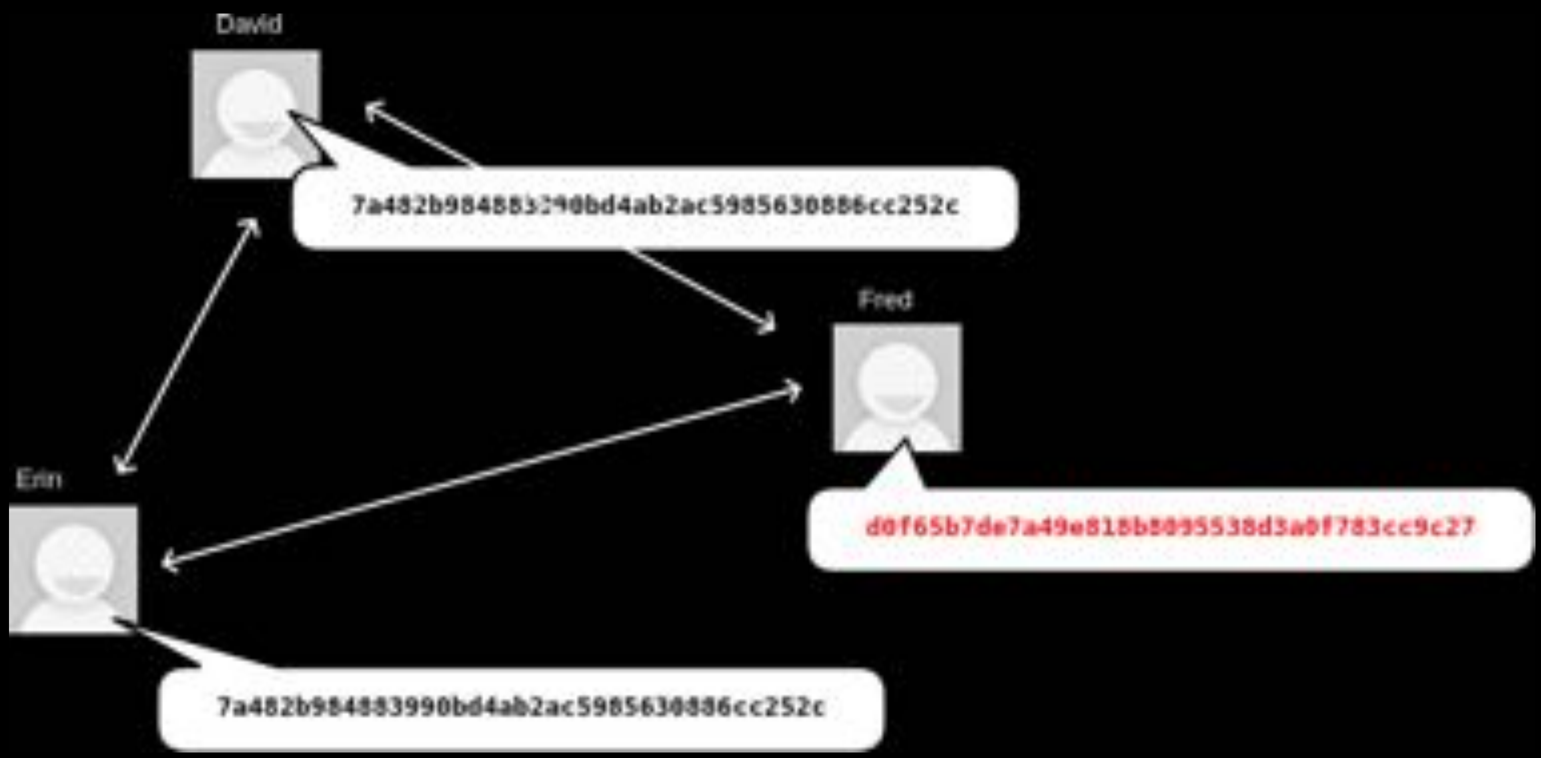
By Zack Whittaker for Zero Day | February 22, 2016 -- 01:28 GMT (01:28 GMT) | Topic: Security


David
7a482b984883990bd4ab2ac5985630886cc252c



7a482b984883990bd4ab2ac5985630886cc252c

7a482b984883990bd4ab2ac5985630886cc252c


David
7a482b984883990bd4ab2ac5985630886cc252c
Fred
d0f65b7de7a49e818b8095538d3a0f783cc9c27
Erin
7a482b984883990bd4ab2ac5985630886cc252c

How does this help?

Alice → Blackmail will be uncovered

Bob → Compromise detected

Carol → Tampered laptop will be discovered

Reduces incentive to attack in the first place

"Builds with the same dependencies"... ✖

"Reliable" builds... ✖

Identical build results

Wait…

Dictionary/hash/database ordering

Parallelism in builds

Timestamps

Build paths

Non-deterministic file ordering

Users, groups, `umask`, environment variables, etc.

Other advantages?

Minimal diffs on "deliberate" changes

Cache ratio — save time, money & CO_2

Remove build-dependencies

Finds bugs!

Predictable OpenID secret

```
# Build.PL
$build->config_data(OpenIDConsumerSecret=>int(1e15*rand()));
```

```
# /usr/share/perl5/GBrowse/ConfigData.pm
{
 'OpenIDConsumerSecret' => '639098210478536',
 'cgibin' => '/usr/lib/cgi-bin/gbrowse',
 'conf' => '/etc/gbrowse',
 [..]
},
```

Every installation of this build shares the same secret.

Random characters in manpages?

```
-This manual page documents the usageoof WikipediaFS.
+This manual page documents the usage of WikipediaFS.
```

```
memcpy(&buf[1], &buf[2], strlen(buf)-1);
```

```
memcpy(3): The memory areas must not overlap
```

```
- memcpy(&buf[1], &buf[2], strlen(buf)-1);
+ memmove(&buf[1], &buf[2], strlen(buf)-1);
```

Fails to build 0.46% of the time?

```
x = f(u('abc'), 16)
y = f(u('abc'), 16)
self.assertEqual(sorted(set(x)), [u('a'), u('b'), u('c')])
```

```
AssertionError: Lists differ: [u'a', u'b'] != [u'a', u'b', u'c']
```

$$(3C_2)*(2/3)^{16} - (3C_1)*(1/3)^{16} =\sim 0.46\%$$

Debian & Reproducible Builds

"Torture test"

Time & date

Hostname & domain name

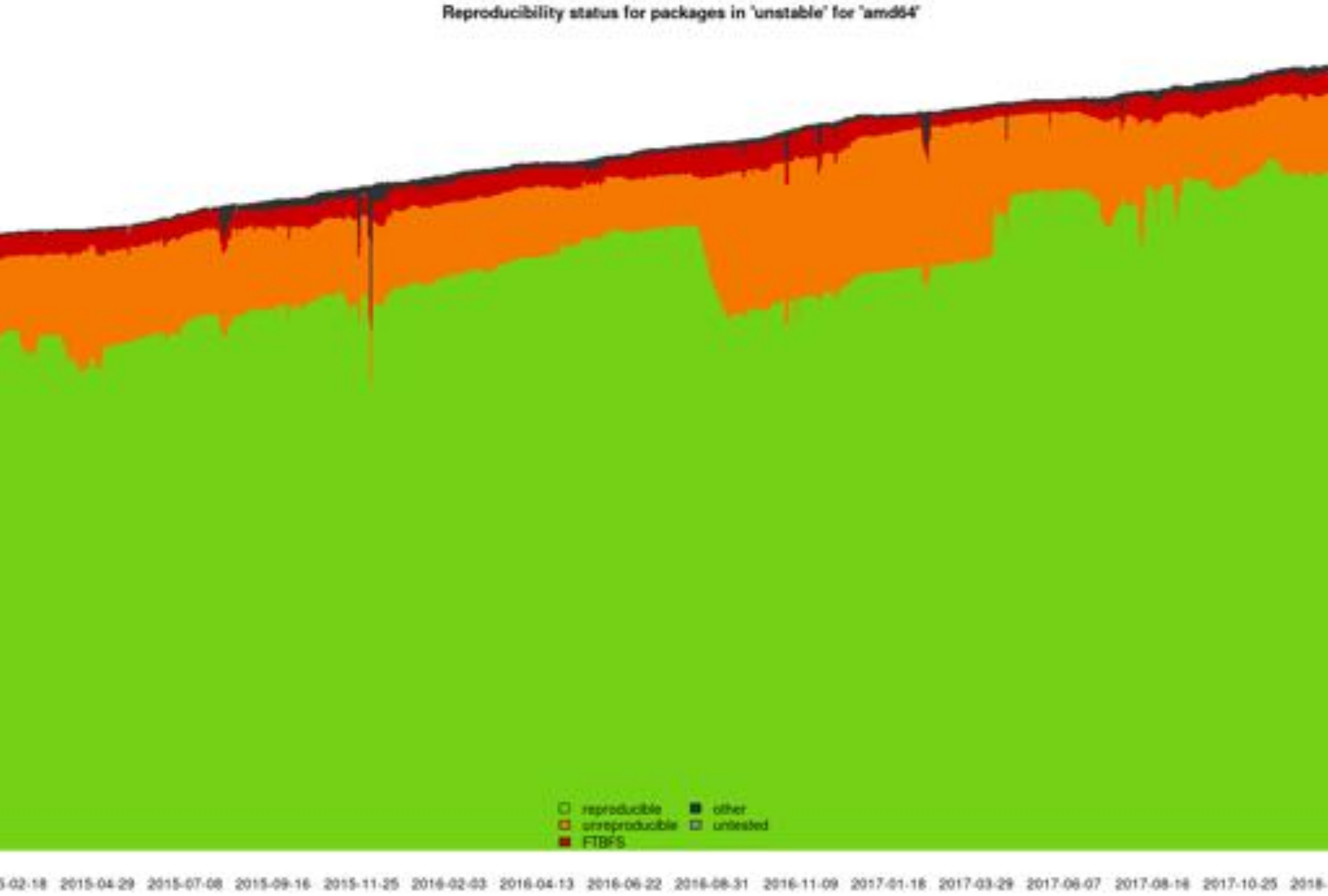
Filesystem (`disorderfs`)

Timezone & locale

`uid` & `gid`

Kernel & CPU type

First rebuild in 2013	24% packages reproducible
March 2018	93% packages reproducible



Reproducibility status for packages in 'unstable' for 'amd64' isdebianreproducibleyet.com

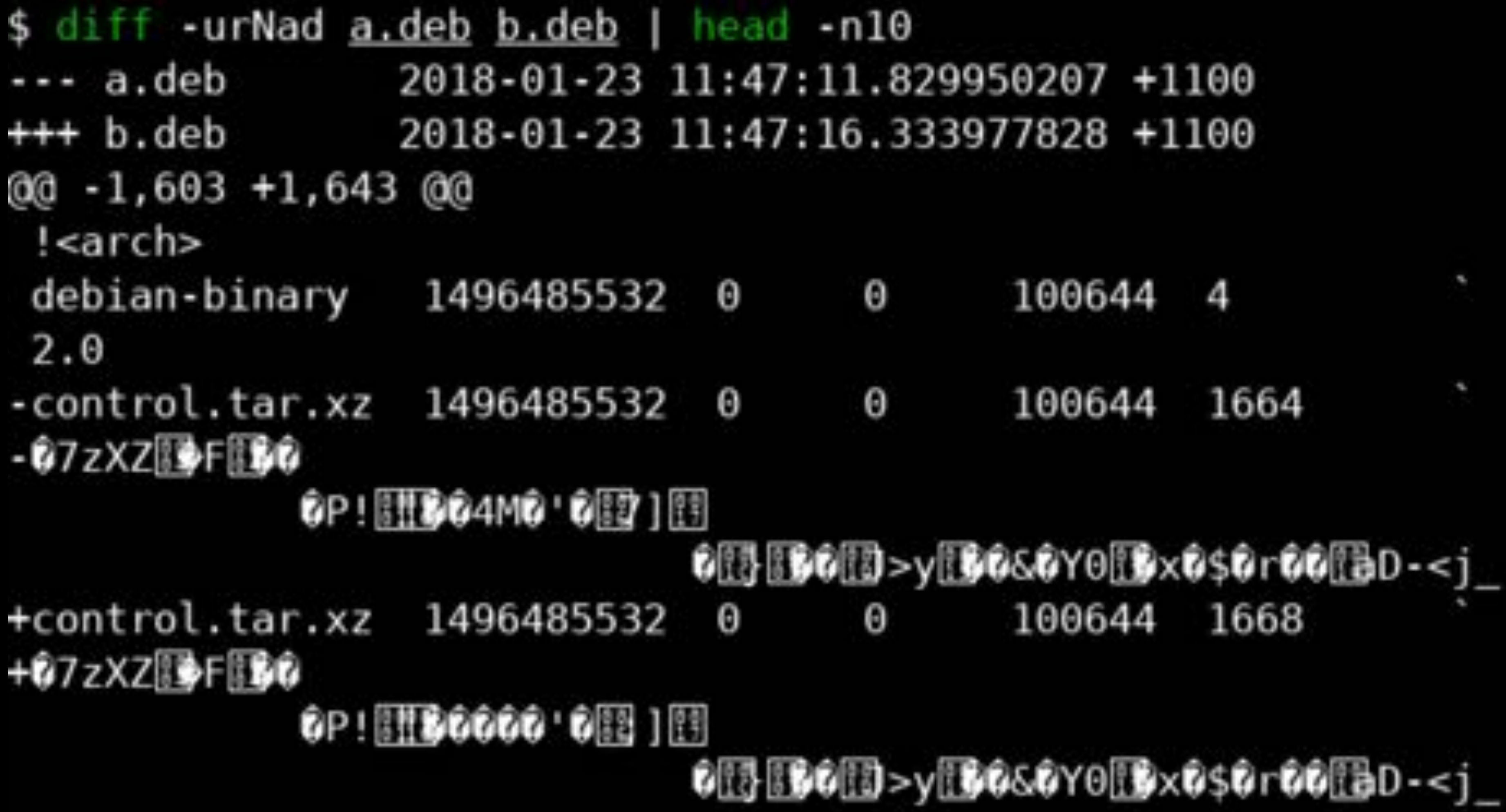
Beyond Debian…

coreboot, Fedora, LEDE, OpenWRT, NetBSD, FreeBSD, Archlinux, Qubes, F-Droid, NixOS, Guix, Meson, etc.

Other projects using "Debian"'s testing framework

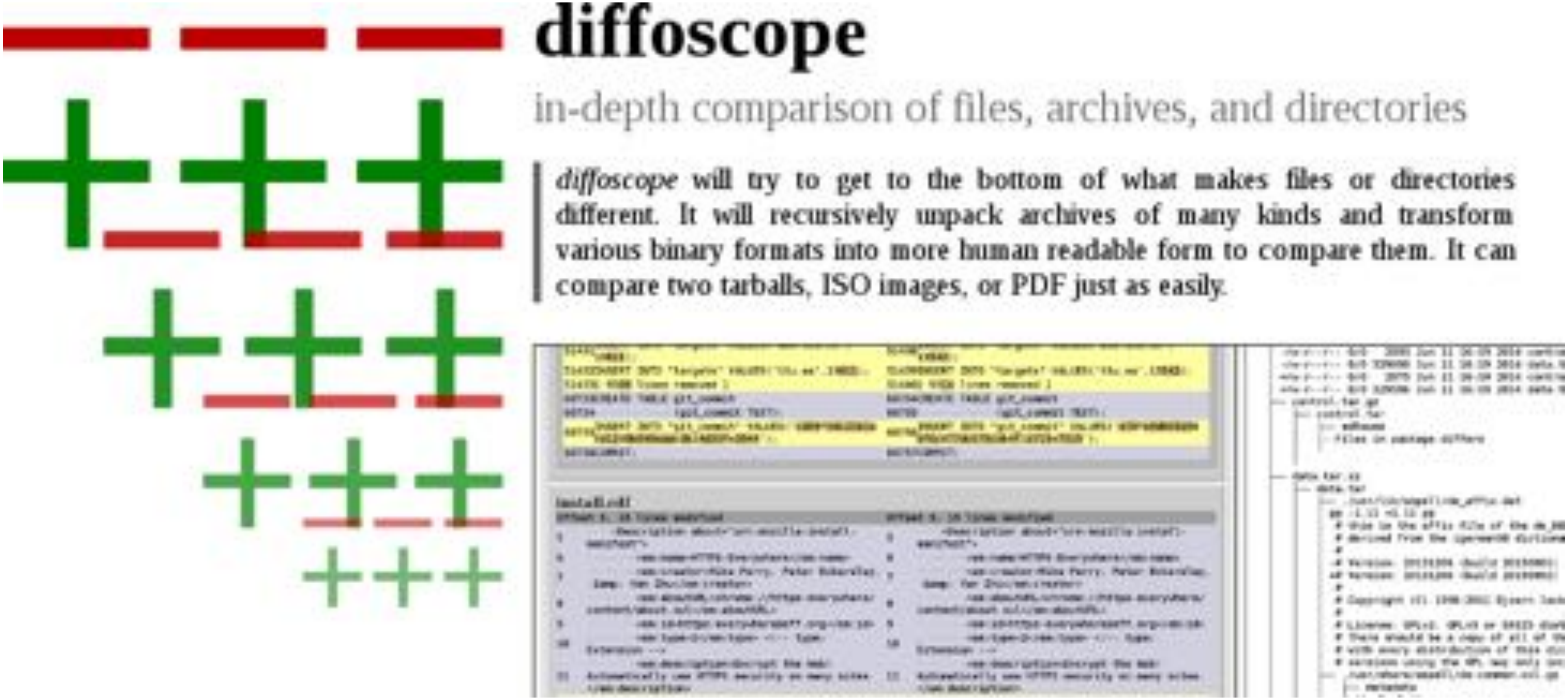
Reproducible Builds summits (Athens, Berlin)

```
# diff -urNad file1 file2
--- file1  2017-06-18 12:37:03.179186661 +0800
+++ file2  2017-06-18 12:37:04.811193648 +0800
@@ -1 +1 @@
-This is the first file.
+This is the second file.
```



```
$ diff -urNad a.deb b.deb | head -n10
--- a.deb       2018-01-23 11:47:11.829950207 +1100
+++ b.deb       2018-01-23 11:47:16.333977828 +1100
@@ -1,603 +1,643 @@
 !<arch>
 debian-binary   1496485532  0     0     100644  4         `
 2.0
-control.tar.xz  1496485532  0     0     100644  1664      `
-�7zXZ▯�F▯��
            �P!▯▯��4M�'�▯7]▯
                            �▯}▯��▯>y▯��&�Y0▯�x�$�r��▯aD-<j_
+control.tar.xz  1496485532  0     0     100644  1668      `
+�7zXZ▯�F▯��
            �P!▯▯�����'�▯ ]▯
                            �▯}▯��▯>y▯��&�Y0▯�x�$�r��▯aD-<j_
```

I SHOULD BUILD A BETTER DIFF



diffoscope

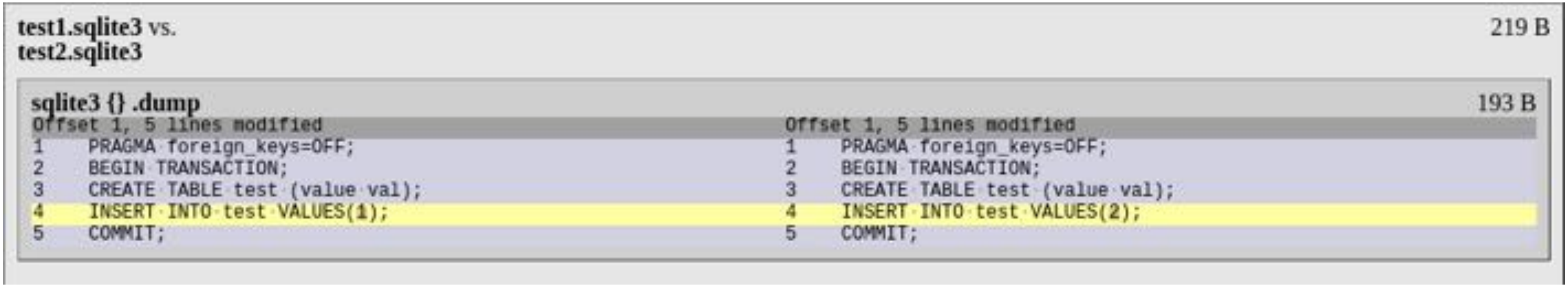
in-depth comparison of files, archives, and directories

> *diffoscope* will try to get to the bottom of what makes files or directories different. It will recursively unpack archives of many kinds and transform various binary formats into more human readable form to compare them. It can compare two tarballs, ISO images, or PDF just as easily.

https://diffoscope.org/

```
├── aspell-de_20131206-5_all.deb
│   ├── metadata
│   │  rw-r--r-- 0/0      4 Jun 11 16:19 2014 debian-binary
│   │ -rw-r--r-- 0/0   2893 Jun 11 16:19 2014 control.tar.gz
│   │ -rw-r--r-- 0/0 329600 Jun 11 16:19 2014 data.tar.xz
│   │ +rw-r--r-- 0/0   2875 Jun 11 16:19 2014 control.tar.gz
│   │ +rw-r--r-- 0/0 329596 Jun 11 16:19 2014 data.tar.xz
│   ├── control.tar.gz
│   │   ├── control.tar
│   │   │   ├── md5sums
│   │   │   │┄ Files in package differ
│   ├── data.tar.xz
│   │   ├── data.tar
│   │   │   ├── ./usr/lib/aspell/de_affix.dat
│   │   │   │  #
│   │   │   │ -# Version: 20131206 (build 20150801)
│   │   │   │ +# Version: 20131206 (build 20150802)
│   │   │   │  #
│   │   │   ├── ./usr/share/aspell/de-common.cwl.gz
│   │   │   │   ├── metadata
│   │   │   │   │ -gzip compressed data, last modified: Sat Aug  1 18:21
│   │   │   │   │ +gzip compressed data, last modified: Sat Aug  1 18:24
```

HTML output



test1.sqlite3 vs.
test2.sqlite3 219 B

sqlite3 {} .dump 193 B

```
Offset 1, 5 lines modified                              Offset 1, 5 lines modified
1    PRAGMA foreign_keys=OFF;                           1    PRAGMA foreign_keys=OFF;
2    BEGIN TRANSACTION;                                 2    BEGIN TRANSACTION;
3    CREATE TABLE test (value val);                     3    CREATE TABLE test (value val);
4    INSERT INTO test VALUES(1);                        4    INSERT INTO test VALUES(2);
5    COMMIT;                                            5    COMMIT;
```

Android APK files, Android boot images, Ar(1) archives, Berkeley DB database files, Bzip2 archives, Character/block devices, ColorSync colour profiles (.icc), Coreboot CBFS filesystem images, Cpio archives, Dalvik .dex files, Debian .buildinfo files, Debian .changes files, Debian source packages (.dsc), Device Tree Compiler blob files, Directories, ELF binaries, Ext2/ext3/ext4/btrfs filesystems, FreeDesktop Fontconfig cache files, FreePascal files (.ppu), Gettext message catalogues, GHC Haskell .hi files, GIF image files, Git repositories, GNU R database files (.rdb), GNU R Rscript files (.rds), Gnumeric spreadsheets, Gzipped files, ISO 9660 CD images, Java .class files, JavaScript files, JPEG images, JSON files, LLVM IR bitcode files, MacOS binaries, Microsoft Windows icon files, Microsoft Word .docx files, Mono 'Portable Executable' files, Ogg Vorbis audio files, OpenOffice .odt files, OpenSSH public keys, OpenWRT package archives (.ipk), PDF documents, PGP signed/encrypted messages, PNG images, PostScript documents, RPM archives, Rust object files (.deflate), SQLite databases, SquashFS filesystems, Statically-linked binaries, Symlinks, Tape archives (.tar), Tcpdump capture files (.pcap), Text files, TrueType font files, XML binary schemas (.xsb), XML files, XZ compressed files, etc.

Try diffoscope now...

diffoscope is a tool to get to the bottom of what makes files or directories different. It recursively unpacks archives of many kinds and transforms various binary formats into more human readable forms to compare them.

File #1 (max: 60MB) Choose file No f....sen

File #2 (max: 60MB) Choose file No f....sen

try.diffoscope.org

Show differences in security uploads

diffoscope ≠ definition of reproducible!

Binary blobs (eg. router images)

What's left to do?

Source code

Programming errors

Backdoors / obfusticated code

Weak algorithms

Code with "testing" modes

```
$ apt install python-pywt-doc
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
  python-pywt-doc
0 upgraded, 1 newly installed, 0 to remove and 4 not upgrad
Need to get 102 kB of archives.
After this operation, 978 kB of additional disk space will
WARNING: The following packages are not reproducible!
  python-pywt-doc
Install these packages anyway? [y/N]
```

Toolchain fixes (GCC, Go, R)

Infrastructure changes

Improving developer tools

Mandating Debian packages be reproducible?

Defeating *Trusting Trust*…?

Get involved!

Visit:	`reproducible-builds.org`
Follow:	**@ReproBuilds** on Twitter
Join:	`#reproducible-builds` (OFTC)

Danke schön!

@lolamby
lamby@debian.org

chris-lamb.co.uk
reproducible-builds.org