



Can your diff(1) do this?!

Improving software review & QA with diffoscope

Chris Lamb
@lolamby
Debian Project Leader 2017—2018

Linux.conf.au
Sydney, Australia
24th January 2018

My name is...



Open source developer for 10+ years

Freelance software developer

Debian Project Leader 2017—2018

< 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% [=====> ]
```

File Edit View Go Bookmarks Help

Previous Next 1 (1 of 1) 85%

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



**Why did we need a
better diff?**

BLOCKCHAIN

DOCKER

Reproducible builds!

Source code available for free software

But everyone runs binaries

How can we trust that they correspond?

Build farms

Developers' machines

Developers' knees

**“Reproducible” builds allows verification
that no flaws have been introduced during
the compilation process**

1. Ensure a build always has identical results
2. Multiple parties compare build results
3. Attacker must infect everybody simultaneously

Identical results?

```
$ sha1sum *.deb
```

```
6f84a299013682203ca2c8345b7b1735e5c575ff a.deb
```

```
85ced2fcdced7a7c3ad31c5e1fa382665e3922f0 b.deb
```



```

$ 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
 -7zXZF
      P! 4M' ]
      >y&Y0x$rD-<j_
+control.tar.xz  1496485532  0      0      100644  1668
+7zXZF
      P! 00000' ]
      >y&Y0x$rD-<j_

```

```

$ diff -urNad <(xxd a.deb) <(xxd b.deb) | head -n20
--- /proc/self/fd/13      2018-01-23 11:51:56.299697293 +1100
+++ /proc/self/fd/14      2018-01-23 11:51:56.299697293 +1100
@@ -5,10 +5,10 @@
 00000040: 2020 600a 322e 300a 636f 6e74 726f 6c2e      ` .2.0.control.
 00000050: 7461 722e 787a 2020 3134 3936 3438 3535  tar.xz  14964855
 00000060: 3332 2020 3020 2020 2020 3020 2020 2020  32  0    0
-00000070: 3130 3036 3434 2020 3136 3634 2020 2020  100644 1664
+00000070: 3130 3036 3434 2020 3136 3638 2020 2020  100644 1668
 00000080: 2020 600a fd37 7a58 5a00 0004 e6d6 b446      ` ..7zXZ.....F
-00000090: 04c0 bf0c 8050 2101 1600 0000 0000 0000  ....P!.....
-000000a0: 8adc 344d e027 ff06 375d 0017 0bbc 1c7d  ..4M.'..7].....}
+00000090: 04c0 c30c 8050 2101 1600 0000 0000 0000  ....P!.....
+000000a0: e2fe ae8d e027 ff06 3b5d 0017 0bbc 1c7d  ....'..;].....}
 000000b0: 0195 c01d 4a3e 7915 c2cc 26a3 5930 12c7  ....J>y...&.Y0..
 000000c0: 788e 24a0 72a5 b416 6144 2d0f 073c 6a5f  x.$.r...aD-..<j_
 000000d0: 0ad7 2255 a796 3502 6c20 fa2a 3248 5c80  .."U..5.l .*2H\
@@ -52,10105 +52,10099 @@
 00000330: f7c4 ab30 9b57 c0fb 161f 96fd 789f 8cc0  ...0.W.....x...
 00000340: 3683 9e02 b7de e805 98ea a09f a4af aaa1  6.....
 00000350: 6575 7d72 d91b 1252 9e39 d172 025c d429  eu}r...R.9.r.\.)

```

```
$ ar x a.deb && ar x b.deb
```

```
$ diff -urNad <(xxd a-control.tar.xz) <(xxd b-control.tar.xz) | head -n10
```

```
--- /proc/self/fd/13      2018-01-23 12:24:53.316388212 +1100
```

```
+++ /proc/self/fd/14      2018-01-23 12:24:53.316388212 +1100
```

```
@@ -1,6 +1,6 @@
```

```
-00000000: fd37 7a58 5a00 0004 e6d6 b446 04c0 bf0c  .7zXZ.....F....
```

```
-00000010: 8050 2101 1600 0000 0000 0000 8adc 344d  .P!.....4M
```

```
-00000020: e027 ff06 375d 0017 0bbc 1c7d 0195 c01d  .'..7].....}....
```

```
+00000000: fd37 7a58 5a00 0004 e6d6 b446 04c0 c30c  .7zXZ.....F....
```

```
+00000010: 8050 2101 1600 0000 0000 0000 e2fe ae8d  .P!.....
```

```
+00000020: e027 ff06 3b5d 0017 0bbc 1c7d 0195 c01d  .'..;].....}....
```

```
 00000030: 4a3e 7915 c2cc 26a3 5930 12c7 788e 24a0  J>y...&.Y0..x.$.
```

```
$ unxz a-control.tar.xz && unxz b-control.tar.xz
```

```
$ diff -urNad <(xxd a-control.tar) <(xxd b-control.tar) | head -n10
```

```
--- /proc/self/fd/13      2018-01-23 12:26:28.564994785 +1100
+++ /proc/self/fd/14      2018-01-23 12:26:28.564994785 +1100
@@ -158,8 +158,8 @@
 000009d0: 0000 0000 0000 0000 0000 0000 0000 0000  .....
 000009e0: 0000 0000 0000 0000 0000 0000 0000 0000  .....
 000009f0: 0000 0000 0000 0000 0000 0000 0000 0000  .....
-00000a00: 3031 3063 3964 6534 6461 6434 3562 6137  010c9de4dad45ba7
-00000a10: 3661 6265 6231 6539 3932 3032 6164 3132  6abeb1e99202ad12
+00000a00: 3436 3135 3165 3139 3236 3631 6162 3663  46151e192661ab6c
+00000a10: 6366 6234 3136 3737 3833 3166 3639 3534  cfb41677831f6954
```

```
$ tar xf a-control.tar && tar xf b-control.tar
```

```
$ diff -urNad a-md5sums b-md5sums
```

```
--- a-md5sums      2017-06-03 20:25:32.000000000 +1000
```

```
+++ b-md5sums      2017-06-03 20:25:32.000000000 +1000
```

```
@@ -1,4 +1,4 @@
```

```
-010c9de4dad45ba76abeb1e99202ad12  usr/bin/pnmixer
```

```
+46151e192661ab6ccfb41677831f6954  usr/bin/pnmixer
```

```
7a14b41780e873903f62d14dd16edc25  usr/share/applications/pnmixer.desktop
```

```
3d244715045bc4e9f2417fa4fd652cb8  usr/share/doc/pnmixer/README.md.gz
```

```
b9e4e837497f0f1b41bbcb26553eb8a4  usr/share/doc/pnmixer/changelog.Debian.gz
```


I SHOULD BUILD A BETTER DIFF



```
$ diffoscope a b
```

```
--- a
```

```
+++ b
```

```
@@ -1 +1 @@
```

```
-foo
```

```
+bar
```


Archive formats

```
$ tar cf a.tar a
```

```
$ tar cf b.tar b
```

```
$ diffoscope a.tar b.tar
```

```
--- a.tar
```

```
+++ b.tar
```

```
— file list
```

```
@@ -1 +1 @@
```

```
--rw-r--r-- 0 lamby (1000) lamby (1000) 4 2018-01-22 22:31:51.000000 a
```

```
+rw-r--r-- 0 lamby (1000) lamby (1000) 4 2018-01-22 22:31:50.000000 b
```

```
--- a
```

```
— +++ b
```

```
@@ -1 +1 @@
```

```
-foo
```

```
+bar
```

Compressed formats

```
$ gzip a.tar
```

```
$ gzip b.tar
```

```
$ diffoscope a.tar.gz b.tar.gz
```

```
--- a.tar.gz
```

```
+++ b.tar.gz
```

```
└─ filetype from file(1)
```

```
@@ -1 +1 @@
```

```
-gzip compressed data, was "a.tar", last modified: Mon Jan 22 22:32:04 2018, from Unix
```

```
+gzip compressed data, was "b.tar", last modified: Mon Jan 22 22:32:08 2018, from Unix
```

```
--- a.tar
```

```
└─ +++ b.tar
```

```
└─ file list
```

```
@@ -1 +1 @@
```

```
--rw-r--r-- 0 lamby (1000) lamby (1000) 4 2018-01-22 22:31:51.000000 a
```

```
+rw-r--r-- 0 lamby (1000) lamby (1000) 4 2018-01-22 22:31:50.000000 b
```

```
--- a
```

```
└─ +++ b
```

```
@@ -1 +1 @@
```

```
-foo
```

```
+bar
```

Recursive

```
├─ hunspell-de-de_20131206-5_all.deb
│   ├── control.tar.gz
│   │   ├── control.tar
│   │   │   ├── md5sums
│   │   │   └─ Files in package differs
│   └─ data.tar.xz
│       ├── data.tar
│       │   └─ ./usr/share/hunspell/de_DE.aff
│       │       @@ -1,11 +1,11 @@
│       │       # this is the affix file of the de_DE Hunspell dictionary
│       │       # derived from the igerman98 dictionary
│       │       #
│       │       -# Version: 20131206 (build 20150801)
│       │       +# Version: 20131206 (build 20150802)
│       │       #
```

ISO images

a.iso vs.
b.iso

74.6 KB

live/filesystem.squashfs

74.3 KB

unsquashfs -s {}

470 B

Offset 1, 10 lines modified	Offset 1, 10 lines modified
1 Found a valid SQUASHFS 4:0 superblock	1 Found a valid SQUASHFS 4:0 superblock
2 Creation or last append time Wed Jan 17 16:08:29 2018	2 Creation or last append time Thu Jan 25 16:46:09 2018
3 Filesystem size 1139612.52 Kbytes (1112.90 Mbytes)	3 Filesystem size 1139612.62 Kbytes (1112.90 Mbytes)
4 Compression xz	4 Compression xz
5 » Dictionary size 1048576	5 » Dictionary size 1048576
6 » Filters selected: x86	6 » Filters selected: x86
7 Block size 1048576	7 Block size 1048576
8 Filesystem is not exportable via NFS	8 Filesystem is not exportable via NFS
9 Inodes are compressed	9 Inodes are compressed
10 Data is compressed	10 Data is compressed

unsquashfs -d " -lls {}

7.72 KB



Tails

the **amnesic** incognito **live** system

Line ordering

```
$ cat a          $ cat b
These           These
lines          order
are            are
in             in
order         lines
```

```
$ diffoscope a b
--- a
+++ b
|... ordering differences only
@@ -1,5 +1,5 @@
   These
-  lines
+  order
   are
   in
-  order
+  lines
```

JSON

```
$ cat a.json  
{"key": "value", "other-key": "after"}
```

```
$ cat b.json  
{"other-key": "before", "key": "value"}
```

```
$ diffoscope a.json b.json  
--- a.json  
+++ b.json  
|  
|   --- a.json  
|   +++ b.json  
|@@ -1,4 +1,4 @@  
|   {  
|       "key": "value",  
|   -   "other-key": "after"  
|   +   "other-key": "before"  
|   }
```

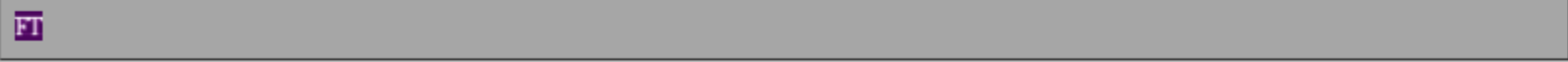
Image content

5.79 KB

Pixel difference



Flicker difference



Offset 1, 36 lines modified

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36

```

Offset 1, 36 lines modified

```

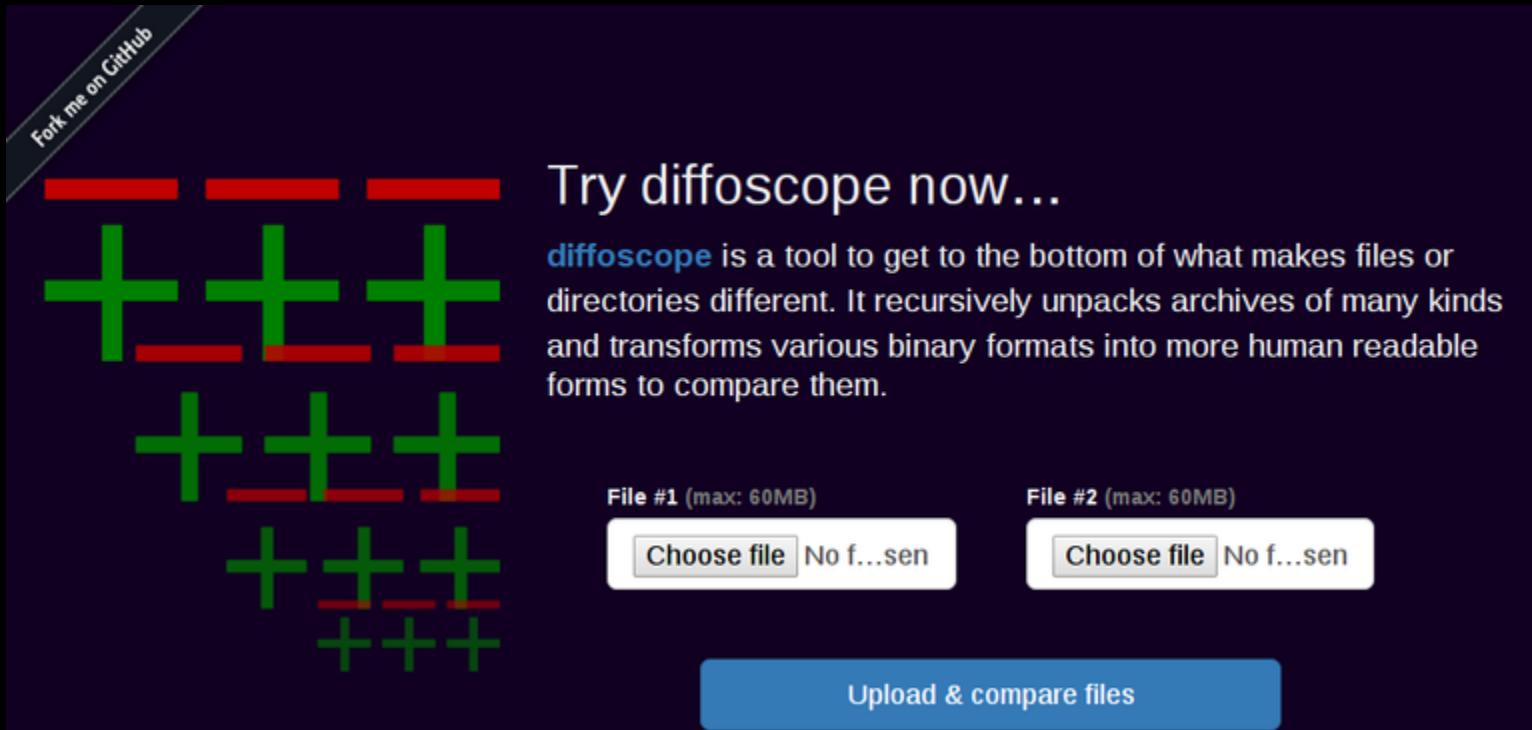
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36

```




Android boot ROMs, Android APKs, ar archives, bzip2 files, Coreboot cbfs, CPIO archives, Dalvik dex file, Debian packages, device (mknod), device tree globs, directories, docx, ELF binaries, ext filesystem images, fontconfig cache dirs, fonts, gettext catalogues, GIF images, Git repositories, GZip files, Haskell binaries, ICC colour profiles, ISO images, Java, bytecode, Javascript, JSON, LLVM binaries, Macho binaries, Mono executables, odt documentens, Ogg borbis, OpenSSH keys, Pascal binaries, PDF files, PGP/GPG keys, PNG images, PostScript, RPM images, R statistical data, Rust binaries, SQLite databases, squashfs images, symlinks, tar archives, tcpdump dumps, text files, XML documents, XZ files, ZIP archives ...

try.diffoscope.org



Fork me on GitHub

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) No files selected

File #2 (max: 60MB) No files selected

The interface features a dark purple background. On the left, there is a graphic consisting of red horizontal bars and green plus signs arranged in a descending staircase pattern. The text 'Fork me on GitHub' is written in white on a dark diagonal banner in the top-left corner. The main heading 'Try diffoscope now...' is in white. Below it, a paragraph describes the tool's functionality. Two file upload sections are present, each with a label 'File #1 (max: 60MB)' and 'File #2 (max: 60MB)', a 'Choose file' button, and the text 'No files selected'. At the bottom center, there is a large blue button labeled 'Upload & compare files'.

```
$ pip install trydiffoscope # or: sudo apt-get install trydiffoscope
$ trydiffoscope --webbrowser /path/to/file1 /path/to/file2
```

Using diffractometer for QA

Just see your changes

Seeing "no" changes

Security "releases"


```
@@ -119,15 +119,15 @@
```

```
Buffer buf;  
Packet packet;
```

```
SocketFactory socket_factory=null;
```

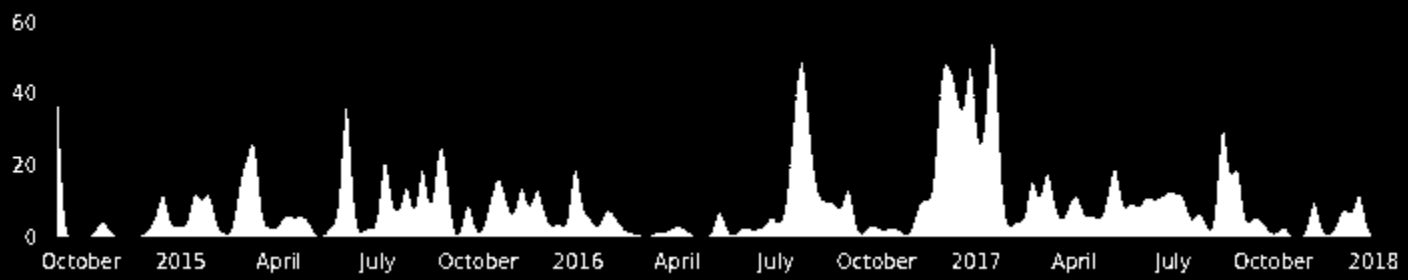
```
static final int buffer_margin = 32 + // maximum padding length  
- 20 + // maximum mac length  
+ 64 + // maximum mac length  
32; // margin for deflater; deflater may inflate data
```

```
private java.util.Hashtable config=null;
```

Index of /firmware/netgear

	<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
	<u>Parent Directory</u>		-	
	<u>GS7xxTP_V5.2.0.9.ros</u>	2013-02-18 10:20	3.7M	
	<u>GS105E_v1.02.04.HEX</u>	2013-02-03 21:25	2.4M	
	<u>GS108EV2_V1.00.12.bin</u>	2013-06-11 17:14	256K	
	<u>GS108Tv2_GS110TP_V5.4.2.9.stk</u>	2013-09-02 22:48	3.9M	
	<u>GS110T_v5.4.0.6.stk</u>	2012-10-12 22:49	4.0M	
	<u>gsm72xxv2-V8.0.1.36.stk</u>	2012-05-12 00:49	7.6M	

Current status



Debian & reproducible builds

Secure | <https://tests.reproducible-builds.org/debian/rb-pkg/unstable/i386/diffoscope-results/nose2.html>

nose2

Test Details

notes

Test history

Filter by: amd64 i386
arm64 armhf

Suite and Architecture

amd64

- 0.6.5-2 in stretch
- 0.6.5-2 in buster
- 0.7.3-1 in unstable

i386

- 0.6.5-2 in stretch
- 0.6.5-2 in buster
- 0.6.5-2 in unstable

unreproducible
at 2017-12-03 21:47 UTC

differences (txt)

buildinfo
rbuild (22KB) build2
(19KB) diff

```
10 quiet]
11 [\-\-log\.-level LOG_LEVEL] [\-\-coverage PATH]
12 [\-\-coverage\.-report TYPE] [\-\-coverage\.-config FILE] [\-\-
C] [\-B] [\-D]
13 [\-F] [\-h]
14 [testNames [testNames ...]]
15 .SS "positional arguments:"
16 .IP
17 testNames
18 .SS "optional arguments:"
19 .TP
20 \fB\.-s\fr START DIR, \fB\.-start\.-dir\fr START DIR
Offset 62, 15 lines modified
21 .TP
22 \fB\.-\coverage\.-config\fr FILE
23 Config file for coverage, default: .coveragerc
24 .TP
25 \fB\.-C\fr, \fB\.-\with\.-coverage\fr
26 Turn on coverage reporting
27 .TP
28 \fB\.-B\fr, \fB\.-\output\.-buffer\fr
29 Enable output buffer
30 .TP
31 \fB\.-D\fr, \fB\.-\debugger\fr
32 Enter pdb on test fail or error
33 .TP
34 \fB\.-F\fr, \fB\.-\fail\.-fast\fr
35 Stop the test run after the first error or failure
36 Enable output buffer
```

./usr/share/man/man1/nose2-3.5.1.gz 1.25 KB

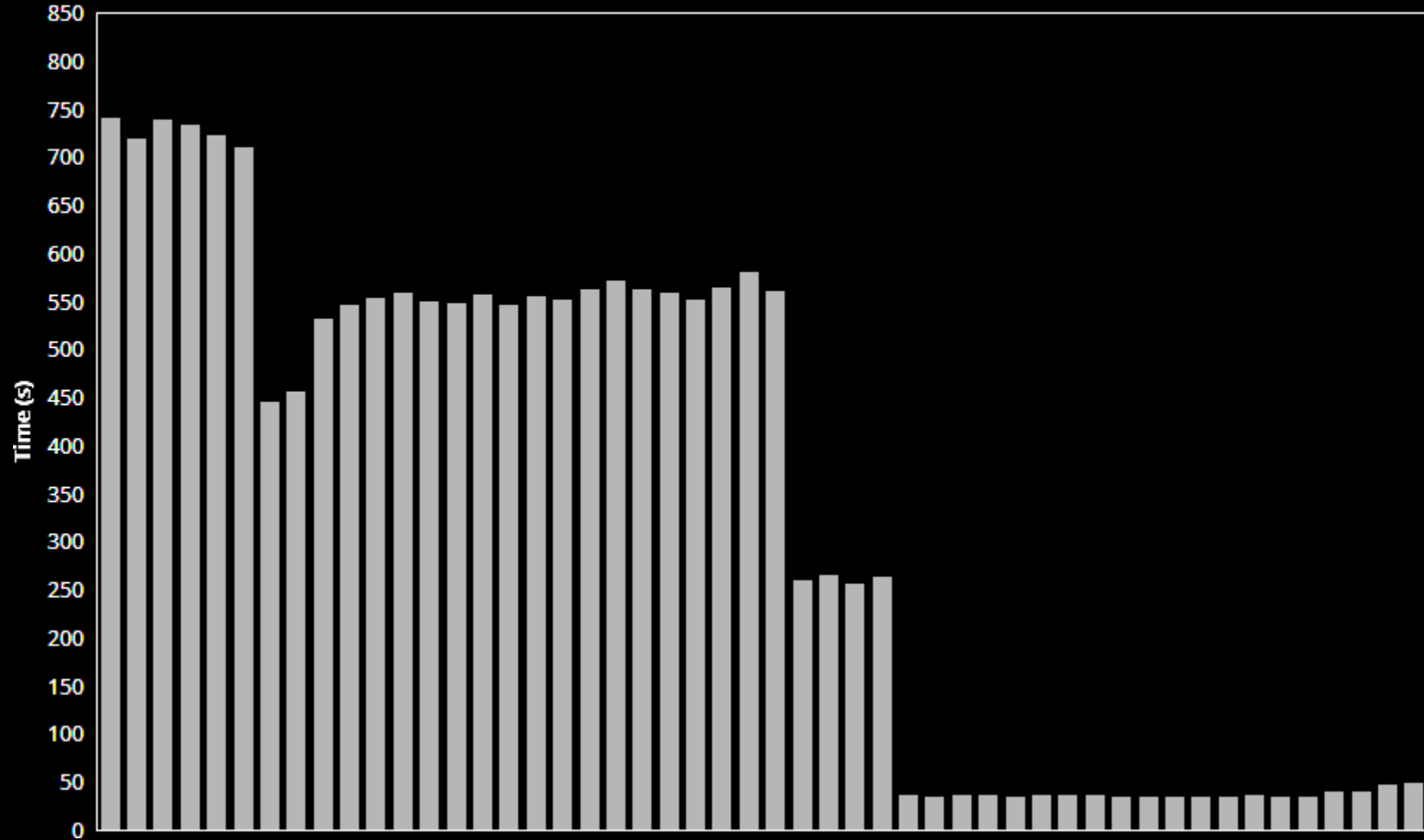
nose2-3.5.1 1.19 KB

```
Offset 4, 16 lines modified
4 nose2-3 \.- nose2-3
5 .SH DESCRIPTION
6 usage: nose2 [\.-s START DIR] [\.-t TOP_LEVEL_DIRECTORY] [\-\-
config [CONFIG]]
7 .IP
8 [\-\-no\.-user\.-config] [\-\-no\.-plugins] [\-\-plugin PLUGINS]
9 [\-\-exclude\.-plugin EXCLUDE_PLUGINS] [\-\-verbose] [\-\-
quiet]
10 [\-\-log\.-level LOG_LEVEL] [\-\-coverage PATH]
11 [\-\-coverage\.-report TYPE] [\-\-coverage\.-config FILE] [\-\-
C] [\-B] [\-D]
12 [\-F] [\-h]
13 [testNames [testNames ...]]
14 .SS "positional arguments:"
15 .IP
16 testNames
```

Hacking!

Optimisations

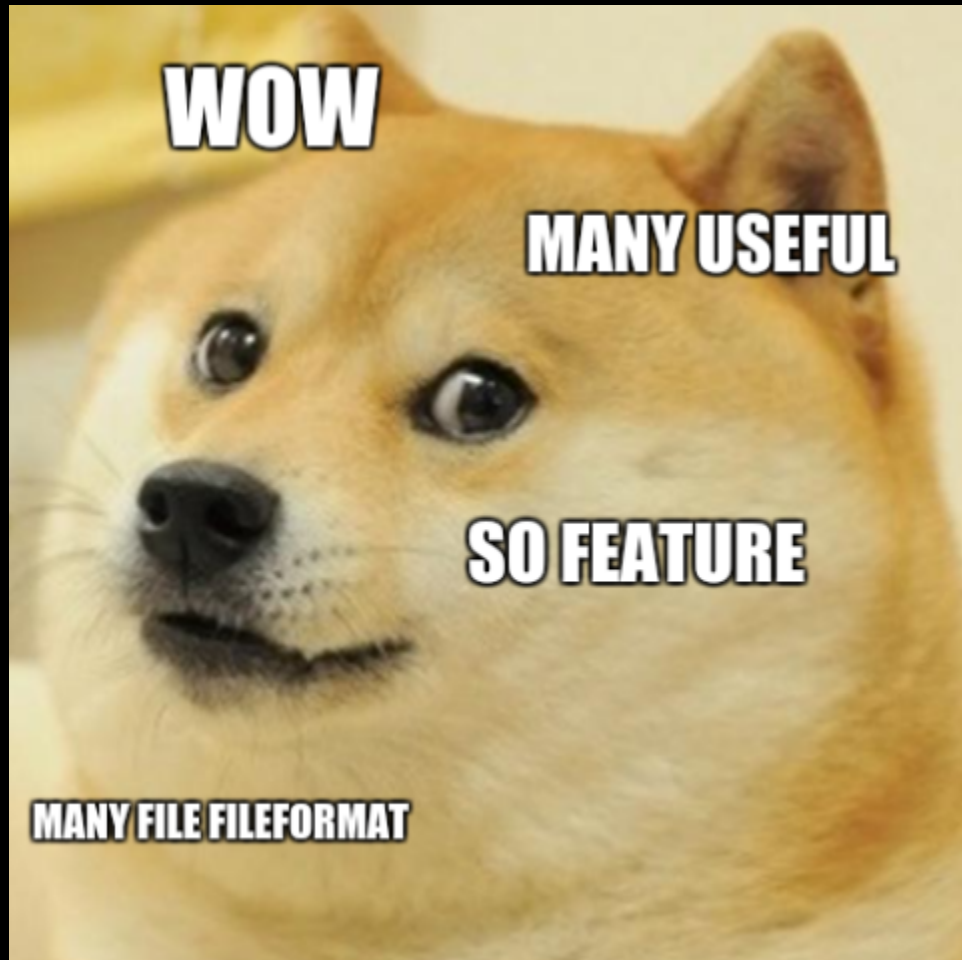
Comparing two NetBSD release tarballs



Parallel processing

Multiple previous attempts...

... Juliana Oliveira working on this now



WOW

MANY USEFUL

SO FEATURE

MANY FILE FILEFORMAT

diffoscope.org

Thanks!



@lolamby
lamby@debian.org

diffoscope.org
chris-lamb.co.uk