

A Replay Approach to Software Validation Using Regular Expressions in C++11 and Python

James Pascoe

james@jamespascoe.com

www.jamespascoe.com

This is a story in two halves ...

1. A 'replay' approach to software validation
 - Replay in what sense?
 - Why is this useful?
 - Replay for The Android Hardware Composer
2. Why I love Regular Expressions and why you should too
 - When (and when not) to use them
 - Regular expressions in C++11 and Python
 - Building a Replay Tool

Replay

- Most software produces some form of textual logging:
 - Readable by humans (often for debug)
 - Already in use by customers, validation engineers, triage teams
 - Ideal for further processing ...
- Instead of viewing logs purely as output, we can:
 1. Use C++11 Regular Expressions to parse them
 2. Interpret time-stamps to recreate event timing
 3. Generate the original input stimuli
 4. Call the underlying API at the 'correct' times

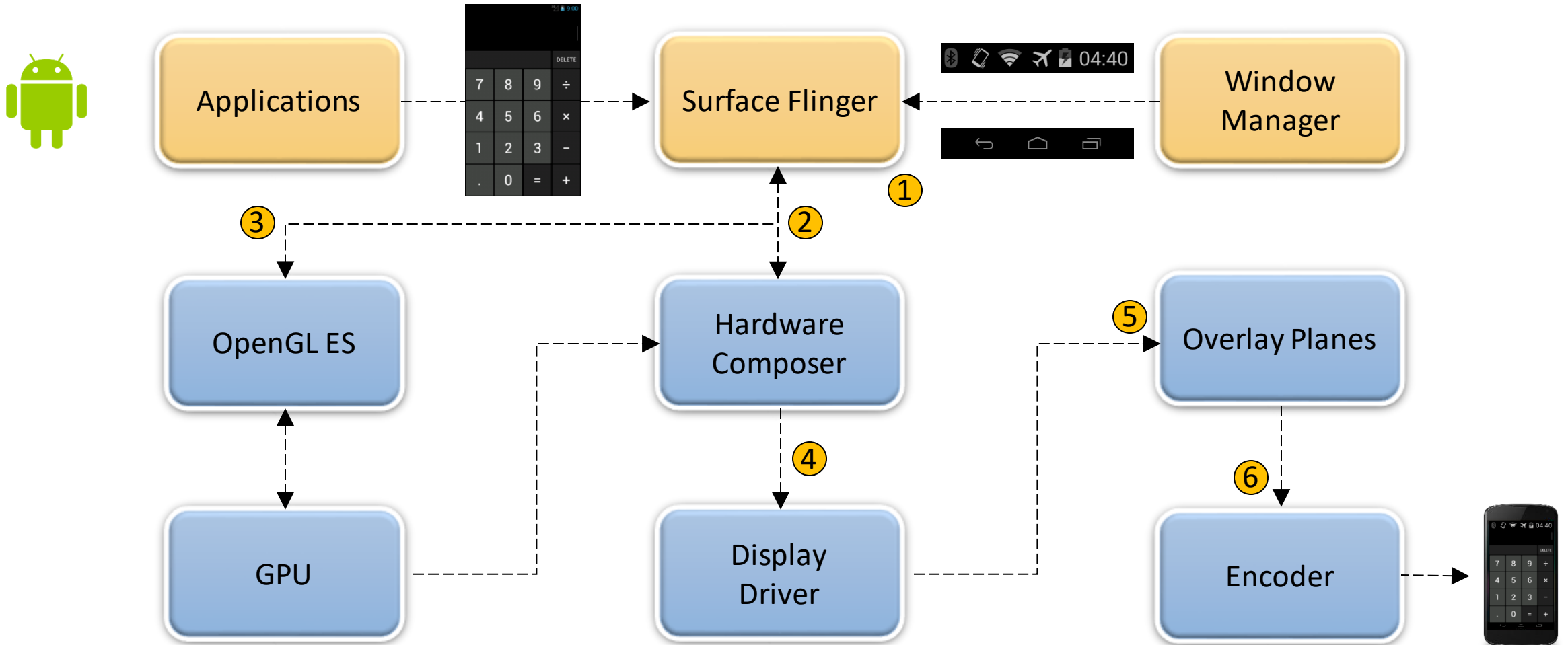


Replay Benefits

- Reduces the length of time for validation cycles
- Recreates bugs faster (and deterministically)
- Much easier to share experimental setups
- More effective bug triage
- Add automated test cases quickly
- Generate stimulus cases for benchmarking
- Create pathological test-cases
- Support TDD/BDD development activities

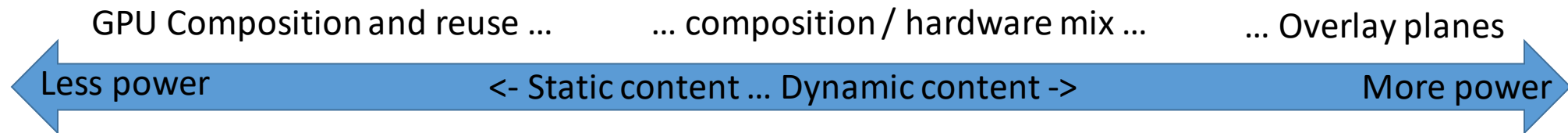


Android Surface Composition



The Hardware Composer

- Complex, highly dynamic, multi-threaded power optimisation engine
- Called by SurfaceFlinger. Determines which 'layers' should be:
 1. Composed in software (GPU / OpenGL)
 2. Mapped to hardware directly



- Next buffer has to be ready in 16ms (i.e. 60hz refresh):
 - All HWC decision making, compositions, surface flinger, buffer writes ...

Hardware Composer API

```
/* (*prepare)() is called for each frame before composition  
and is used by SurfaceFlinger to detect the compositions that  
the HWC can handle. */
```

```
int (*prepare)(struct hwc_composer_device_1 *dev,  
    size_t numDisplays, hwc_display_contents_1_t** displays);
```

```
/* (*set)() is called to update the displays with the content  
of their work lists. */
```

```
int (*set)(struct hwc_composer_device_1 *dev,  
    size_t numDisplays, hwc_display_contents_1_t** displays);
```

https://github.com/android/platform_hardware_libhardware/blob/master/include/hardware/hwcomposer.h

Hardware Composer Log File Extract

94257s 259ms 311115ns TID:16192 SF0 onPrepare Exit frame:6 Fd:-1 outBuf:0x0 outFd:-1 flags:0

0 OV 0x7f8d49e38e70:21:0 60 BL:1.00 RGBA:X 1920x1200 0.0, 0.0,1920.0,1200.0
0,0,1920,1200 -1 -1 V: 0, 0,1920,1200 U:00000b00 Hi:0 Fl:0 A BL

1 OV 0x7f8d49e38fb0:25:0 60 BL:1.00 RGBA:X 600x400 0.0, 0.0, 600.0, 400.0
300,200,900,600 -1 -1 V: 300, 200, 900, 600 U:00000b00 Hi:0 Fl:0 A BL

2 TG 0x7f8d49e38880:23:0 60 BL:1.00 RGBA:X 1920x1200 0.0, 0.0,1920.0,1200.0
0,0,1920,1200 -1 -1 V: 0, 0,1920,1200 U:00001a00 Hi:0 Fl:0 A BL

Timestamp

Thread Id

Individual layer fields

Parsing an HWC Log File – Incumbent Solution

```
bool Parse(std::string const& str)
{
    int layer_num;
    char layer_flag[3], layer_handle[13];

    int num_parsed = std::sscanf(str.c_str(), " %d %s 0x%s:",
        &layer_num, layer_flag, layer_handle);
    int layer_handle_num = atoi(layer_handle);

    if (num_parsed == 3)
    { ...
```

I know ... I know ...



Drawbacks

- No atomicity
 - The code breaks (subtly) when the log format changes (and it does often)
 - Code is always 'slightly' broken – no confidence in results
- Verbosity
 - `sscanf` format string is not flexible enough to cover all variations
 - Lots of duplication - required over 3 kloc just for parsing !
- Fragmentation
 - Number of leaf functions is huge !
 - All require unit-tests, causes code bloat, overall logic is obscured ...

What we need is ...

Std::regex!

Regular Expression History

- Started in the 1940s with two Neurophysiologists:¹
 - Warren McCulloch and Walter Pitts
- Stephen Kleene described these models in algebra - 'regular sets'
 - Kleene devised a notation to express these sets – 'regular expressions'
- In 1968, Ken Thompson wrote a regular expression compiler that produced IBM 7094 object code:
 - This led to work on `qed` - the editor which became `ed` on Unix
 - `ed` had a command to display lines of a file that matched a regular expression
 - `g/Regular Expression/p` was read 'Global Regular Expression Print' and became `grep` (which was later extended into `egrep`)

[1] [Mastering Regular Expressions, 3rd Edition, Jeffrey E.F. Friedl, O'Reilly, 2006.](#)

Regular Expressions in C++11

- Full match, any match and replacement:
 - `regex_match()`, `regex_search()`, `regex_replace()`
- Iterate over matches / tokenize:
 - `regex_iterator()`, `regex_token_iterator()`
- Constants (defined in `std::regex_constants`):
 - Syntax options – case insensitivity, which regular expression grammar ...
 - Match options – defines whether the first character matches ^ ...
 - Error types – thrown when parsing badly formed regular expressions

C++11 ECMAScript Regex Refresher

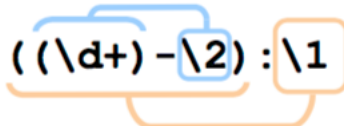
.	Match any character
^	Match beginning of input
\$	Match end of input
\b	Match word boundary
\B	Match anything other than a word boundary
	Or operator

Capture groups

Denoted with parentheses

Referred to as **\1**, **\2** etc.

Counted in order of left parentheses:


`((\d+) - \2) : \1`

Repetition

Symbol	Repeats matched
?	≤ 1
*	≥ 0
+	≥ 1
{n}	n
{n,}	$\geq n$
{n, m}	$\geq n \ \&\& \ \leq m$

Classes

alpha	punct
digit or d	lower
alnum or w	upper
	graph
space or s	print
blank	xdigit
cntrl	

Sets

Symbol	Matches
[abc]	Any of the characters included
[^abc]	Any of the characters NOT included
[a-z]	Any characters in the range
[a-zA-Z]	Any characters in the ranges
[=c=]	Equivalence class for the character
[.ae.]	Specified collating element



cpprocks.com

Regular Expressions for Replay

94257s 166ms 673083ns TID:16192 SF0 onSet Entry frame:0 Fd:-1 outBuf:0x0 outFd:-1 flags:1

```
const std::string onset_string =
```

```
    R"REGEX(^\d+s \d+ms (?:\d+ns )?(?:TID:\d+ )?SF\d onSet Entry )REGEX"
```

```
    R"REGEX(?:frame:\d+ )?Fd:(-?\d{1,2}) )REGEX"
```

```
    R"REGEX(outBuf:0x.{1,8} outFd:(-?\d{1,2}) [fF]lags:(\d+)(:?.*)$)REGEX";
```


Regular Expressions for Replay

```
1 OV 0x7f8d49e38f60:22:0 60 BL:1.00 RGBA:X 600x400 0.0, 0.0, 600.0, 400.0
300, 200, 900, 600 -1 -1 V: 300, 200, 900, 600 U:00000b00 Hi:0 Fl:0 A BL
```

```
const std::string layer_string_hdr =
    "^\\s*(\\d+) (\\w{2}) *0x(.{1,12}): ?(?:--|(\\d{1,3}))?: ?(\\d) ?(\\d+) "
    "(\\w{2}): ?(.{1,4}) (\\w{1,5}) *:[XLY] *(\\d{1,4})x(\\d{1,4}) * "
    " *(-?\\d+\\.?.?\\d*), *(-?\\d+\\.?.?\\d*), *(-?\\d+\\.?.?\\d*), *(-?\\d+\\.?.?\\d*)"
    " *(-?\\d{1,4}), *(-?\\d{1,4}), *(-?\\d{1,4}), *(-?\\d{1,4}) "
    "(-?\\d+) (-?\\d+) V: *(\\d{1,4}), *(\\d{1,4}), *(\\d{1,4}), *(\\d{1,4}) ";
```

```
const std::string layer_string_trl =
    " *U:({1,8}) * Hi:(\\d+)(:?[[:alpha:]]*)? Fl:(\\d+)(:?[[:alnum:]]*).*";
```

Matching and Searching

```
std::regex onset_regex(onset_string);  
.  
.  
.  
auto start = high_resolution_clock::now();  
while (getline(infile, line))  
{  
    ++lines;  
    if (std::regex_match(line, onset_regex) ||  
        (std::regex_search(line, layer_hdr_regex) && std::regex_search(line, layer_trl_regex)))  
        ++matches;  
}  
auto end = high_resolution_clock::now();
```

```
std::cout << "Took " << duration_cast<milliseconds>(end-start).count() << " ms for " << matches <<  
    " matches (" << lines << " lines processed)" << std::endl;
```

➤ Took 2906 ms for 2605 matches over 7269 lines
(Averaged over 10,000 runs) ~ 0.4 ms per line

Clang++-3.8 running on an I7 / 32 Gb (Ubuntu 14.04)

Capture Groups

Values are held in specialisations of `std::match_results`:

```
std::smatch onset_matches; // String specialization of std::match_results
if (std::regex_match(line, onset_matches, onset_regex))
{
    std::cout << "Saw onSet Entry with timestamp " <<
        onset_matches[1] << "s " <<
        onset_matches[2] << "ms " <<
        std::stoi(onset_matches[3]) << "ns " << std::endl;

    // Note: onset_matches[0] is the whole match !
}
```

Advice

- Regular expressions are dense (with lots of different grammars)
 - POSIX, ECMAScript, Python, Perl, JavaScript, awk, grep, egrep
 - (C++11 regular expressions and Python `re` are compatible 😊 !)
- When writing Regular Expressions, start small and build
 - Start with: `^ . * $` and not `^ (? : (? : (? : 0 ? [1 3 5 7 8] | 1 [0 2]) ...`
- Use visualisation websites:
 - regexpal.com and regexexplained.co.uk
- Raw String Literals can be useful for Regexs
- Use counters to provide checks that are easy to verify with `grep`

How Do We Validate All of This?

- **Question:** how do you validate a Replay tool?
- **Answer:** compare the replayed log to the original log
- C++11 Regular Expressions and Python are compatible!
- We can develop validation / visualisation tools 😊 !
- PyReplay:
 - Compares two HWC log files. Identifies subtle mismatches
 - ~250 lines of Python (with comments). Based on TkInter
 - Python 3 regular expressions (`re`) are the same as C++11 !



Regular Expressions in Python

```
onset_string = \
    r'^(\d+)s (\d+)ms (?: (\d+)ns )?(?:TID:(\d+) )?SF(\d) onSet Entry ' \
    r'(?:frame:(\d+) )?Fd:(-?\d{1,2}) ' \
    r'outBuf:0x(.{1,8}) outFd:(-?\d{1,2}) [fF]lags:(\d+)(:?.*)$'
```

```
layer_string_hdr = \
    "^\\s*(\\d+) (\\w{2}) *0x(.{1,12}): ?(?:--|(\\d{1,3}))?: ?(\\d) ?(\\d+) " \
    "(\\w{2}): ?(.{1,4}) (\\w{1,5}) *:[XLY] *(\\d{1,4})x(\\d{1,4}) *" \
    " *(-?\\d+\\.?.\\d*), *(-?\\d+\\.?.\\d*), *(-?\\d+\\.?.\\d*), *(-?\\d+\\.?.\\d*)" \
    " *(-?\\d{1,4}), *(-?\\d{1,4}), *(-?\\d{1,4}), *(-?\\d{1,4}) (-?\\d+) (-?\\d+) " \
    "V: *(\\d{1,4}), *(\\d{1,4}), *(\\d{1,4}), *(\\d{1,4}) ";
```

```
layer_string_trl = \
    " *U: (.{1,8}) * Hi:(\\d+)(:?[[:alpha:]]*)? Fl:(\\d+)(:?[[:alnum:]]*)*";
```

```

124654: 1616s 514ms 548198ns TID:2320 Fence: dup Fd:31 -> Fd:47
124655: 1616s 514ms 551775ns TID:2320 Fence: close Fd:35
124656: 1616s 514ms 558892ns TID:2320 Fence: Duping retire fence 31 - Layer 0 fd 34 Layer 1 fd 47
124657: 1616s 514ms 593504ns TID:2320 D0 onSet Exit frame:2309 Fd:31 outBuf:0x0 outFd:0 flags:1
124658: 0 0V 0xb93bedb0: 0:4 60 OP:FF NV12 1280x736 1.0, 0.0,1279.0, 720.0 622, 0,1299,1200 -1 -1 V: 622, 0,1299,1200
0 U:00000900 Hi:0 Fl:1312d00 V S SO SC
124659: 1 FB 0x0: 0:4 60 BL:FF ??? 0x0 1.0, 1.0,1199.0,1848.0 71, 0,1920,1200 -1 -1 V: 71, 0,1920,1200
0 U:00000000 Hi:0 Fl:0 DISABLE S SO
124660: 2 FB 0x0: 0:4 60 BL:FF ??? 0x0 1.0, 0.0,1199.0, 71.0 0, 0, 73,1200 -1 -1 V: 0, 0, 73,1200
0 U:00000000 Hi:0 Fl:0 DISABLE S SO
124661: 3 TG 0xb9207560:23:0 60 BL:FF RGBA 1920x1200 0.0, 0.0,1920.0,1200.0 0, 0,1920,1200 -1 34 V: 0, 0,1920,1200
0 U:00001a00 Hi:0 Fl:0 A B
124662: 1616s 514ms 618325ns TID:2320 HWCVAL: checkSetExit mOnSetSequence=2311
124663: 1616s 514ms 634009ns TID:2326 Consume WorkItem:0xb920359c frame:2309 [timeline:2310]
124664: 1616s 514ms 639897ns TID:2326 drmIoctl( DRM_IOCTL_I915_GEM_WAIT[ boHandle 4, timeout 3000000000 ] )
124665: 1616s 514ms 659928ns TID:2326 drmIoctl( DRM_IOCTL_I915_GEM_WAIT[ boHandle 10, timeout 3000000000 ] )
124666: 1616s 524ms 469426ns TID:2324 drm Flip Crtc 3 completed flip to frame:2308 [timeline:2309]
124667: 1616s 524ms 550018ns TID:2326 Fence: Dm Crtc 3 issuing drm updates for frame frame:2309 [timeline:2310]
124668: 1616s 524ms 554128ns TID:2326 Crtc:3 Panel fitter scaling Disabled Skipped (No Change)
124669: 1616s 524ms 557728ns TID:2326 Crtc:3 Zorder:4,SAPASBCA Skipped (No Change)
124670: 1616s 524ms 561920ns TID:2326 Plane 5 Disabled (No Change)
124671: 1616s 524ms 564523ns TID:2326 PLANE 4 H:0xb9369d50 TX:0 S:0.0,0.0,677.0x1200.0 F:622,0,677x1200 Skipped (No Change)
124672: 1616s 524ms 581952ns TID:2326 drmModePageFlip( crtc_id 3, fb 23, flags 1, user_data 0xb9203648 )
124673: 1616s 524ms 617943ns TID:2326 CRTC 3 H:0xb9207560 TX:0 S:0.0,0.0,1920.0x1200.0 F:0,0,1920x1200 :FLIPEVENT
124674: 1616s 536ms 111775ns TID:2320 HWCVAL:E Layer @ 0xb93a4500 has no buffer.
124675:
124676:
124677: 1616s 536ms 172193ns TID:2320 D0 onPrepare Entry frame:2310 Fd:-1 outBuf:0x0 outFd:0 flags:1
124678: 0 FB 0xb9232c60: 0:4 60 OP:FF NV12 1280x736 1.0, 0.0,1279.0, 720.0 622, 0,1299,1200 -1 -1 V: 622, 0,1299,1200
0 U:00000900 Hi:0 Fl:1312d00 V S SO SC
124679: 1 FB 0xb92326e0:33:4 60 BL:FF RGBA 1200x1920 1.0, 1.0,1199.0,1848.0 71, 0,1920,1200 -1 -1 V: 71, 0,1920,1200
0 U:00000000 Hi:0 Fl:0 A B S SO SC
0x0: 0:4 60 BL:FF ??? 0x0 1.0, 0.0,1199.0, 71.0 0, 0, 73,1200 -1 -1 V: 0, 0, 73,1200
DISABLE S SO
60:23:0 60 BL:FF RGBA 1920x1200 0.0, 0.0,1920.0,1200.0 0, 0,1920,1200 -1 -1 V: 0, 0,1920,1200
A B
441ns TID:2320 D0 InputAnalyzer::onPrepare Frame:2310 1920x1200 60Hz RGBA Enabled Geometry Video
60: 0:4 60 OP:FF NV12 1280x736 1.0, 0.0,1279.0, 720.0 622, 0,1299,1200 -1 -1 V: 622, 0,1299,1200
12d00 V S SO SC
60:33:4 60 BL:FF RGBA 1200x1920 1.0, 1.0,1199.0,1848.0 71, 0,1920,1200 -1 -1 V: 71, 0,1920,1200
A B S SO SC
0x0: 0:4 54 BL:FF ??? 0x0 1.0, 0.0,1199.0, 71.0 0, 0, 73,1200 -1 -1 V: 0, 0, 73,1200
DISABLE S SO
608ns TID:2320 D0 SurfaceFlingerComposer Frame:2310 1920x1200 60Hz RGBA Enabled Geometry Video
60: 0:4 60 OP:FF NV12 1280x736 1.0, 0.0,1279.0, 720.0 622, 0,1299,1200 -1 -1 V: 622, 0,1299,1200
12d00 V S SO SC
60:33:4 60 BL:FF RGBA 1200x1920 1.0, 1.0,1199.0,1848.0 71, 0,1920,1200 -1 -1 V: 71, 0,1920,1200
0 U:00000900 Hi:0 Fl:0 A B S SO SC
124689: 2 0x0: 0:0 1000000000 BL:FF RGBA 0x0 0.0, 0.0,1920.0,1200.0 0, 0,1920,1200 -1 -1 V: 0, 0,
1920,1200 U:00000000 Hi:0 Fl:0 A B DISABLE
124690: 1616s 536ms 350453ns TID:2320 D0 onPrepare Exit frame:2310 Fd:-1 outBuf:0x0 outFd:0 flags:1
124691: 0 0V 0xb9232c60: 0:4 60 OP:FF NV12 1280x736 1.0, 0.0,1279.0, 720.0 622, 0,1299,1200 -1 -1 V: 622, 0,1299,1200
0 U:00000900 Hi:0 Fl:1312d00 V S SO SC
124692: 1 0V 0xb92326e0:33:4 60 BL:FF RGBA 1200x1920 1.0, 1.0,1199.0,1848.0 71, 0,1920,1200 -1 -1 V: 71, 0,1920,1200
0 U:00000900 Hi:0 Fl:0 A B S SO SC
124693: 2 FB 0x0: 0:4 60 BL:FF ??? 0x0 1.0, 0.0,1199.0, 71.0 0, 0, 73,1200 -1 -1 V: 0, 0, 73,1200
0 U:00000000 Hi:0 Fl:0 DISABLE S SO
124694: 3 TG 0xb9207560:23:0 60 BL:FF RGBA 1920x1200 0.0, 0.0,1920.0,1200.0 0, 0,1920,1200 -1 -1 V: 0, 0,1920,1200
0 U:00001a00 Hi:0 Fl:0 A B
124695: 1616s 541ms 540213ns TID:2324 drm Flip Crtc 3 completed flip to frame:2309 [timeline:2310]
124696: 1616s 548ms 807186ns TID:2320 D0 onSet Entry frame:2310 Fd:-1 outBuf:0x0 outFd:0 flags:1
124697: 0 0V 0xb9232c60: 0:4 60 OP:FF NV12 1280x736 1.0, 0.0,1279.0, 720.0 622, 0,1299,1200 -1 -1 V: 622, 0,1299,1200
0 U:00000900 Hi:0 Fl:1312d00 V S SO SC
124698: 1 0V 0xb92326e0:33:4 60 BL:FF RGBA 1200x1920 1.0, 1.0,1199.0,1848.0 71, 0,1920,1200 -1 -1 V: 71, 0,1920,1200
0 U:00000900 Hi:0 Fl:0 A B S SO SC
124699: 2 FB 0x0: 0:4 60 BL:FF ??? 0x0 1.0, 0.0,1199.0, 71.0 0, 0, 73,1200 -1 -1 V: 0, 0, 73,1200
0 U:00000000 Hi:0 Fl:0 DISABLE S SO
124700: 3 TG 0xb9207560:23:0 60 BL:FF RGBA 1920x1200 0.0, 0.0,1920.0,1200.0 0, 0,1920,1200 -1 -1 V: 0, 0,1920,1200
0 U:00001a00 Hi:0 Fl:0 A B
124701: 1616s 548ms 831950ns TID:2320 Fence: check complete Fd:48
124702: 1616s 548ms 836345ns TID:2320 Fence: close Fd:48
124703: 1616s 548ms 874901ns TID:2320 VppComposer Video
124704: 0 0xb9232c60: 0:4 60 OP:FF NV12 1280x736 1.0, 0.0,1279.0, 720.0 0,
```


Conclusions

- Replay:
 - Found numerous real bugs and often very subtle
 - Had some great cultural benefits and insights
 - Particularly when combined with Jenkins / CI
 - Low-cost to implement and maintain
 - Compatibility of `std::regex` and Python `re` is very useful
- Limitations:
 - Replay: multithreading can disrupt bug reproducibility
 - Better to use fuzz-testing + coverage + sanitisers for that
 - Regular expressions: not the right tool for sophisticated lexical analysis
 - Probably want to build something like an Abstract Syntax Tree for that ...




```
#include <iostream>

int main()
{
    std::string questions;
    while (1)
    {
        std::cout << "Questions?" << std::endl;

        if (std::cin >> questions && questions == "Y")
            std::cout << "Answers" << std::endl;
        else
            goto cornubia;
    }

    cornubia:
        std::cout << "Thank you for coming !" << std::endl;
}
```

