Towards Learning From Linux Kernel Configurations' Failures with Clang
(for improving Clang support with the kernel)

Prof. Mathieu Acher
@acherm
(University of Rennes 1, Inria/IRISA)
Research: Learning the Linux Configuration space

- 15K+ options; $10^{6000}$ configurations?
- R1: Finding/fixing/preventing configuration bugs
- R2: Predicting kernel size out of a configuration
  - yes, without compiling it and whatever kernel version (and perhaps compiler version)
  - hint: vmlinux size from 7Mb to 1,8Gb
- R3: Predicting compilation time out of a configuration
- TUXML to build lots of configurations
  - Could well be KernelCI, ongoing “bridge” coming
  - https://github.com/TuxML/tuxml
Sometimes configurations fail

- make randconfig with `CONFIG_X86=y` and `CONFIG_X86_64=y`
- compiler: gcc 6.3

<table>
<thead>
<tr>
<th>version</th>
<th>nbuild</th>
<th>nfailures</th>
<th>percentage_failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.13.3</td>
<td>95854</td>
<td>3164</td>
<td>3.3</td>
</tr>
<tr>
<td>4.15</td>
<td>44154</td>
<td>500</td>
<td>1.13</td>
</tr>
<tr>
<td>4.20</td>
<td>23830</td>
<td>330</td>
<td>1.38</td>
</tr>
<tr>
<td>5.0</td>
<td>20398</td>
<td>446</td>
<td>2.19</td>
</tr>
<tr>
<td>5.4</td>
<td>26281</td>
<td>434</td>
<td>1.65</td>
</tr>
<tr>
<td>5.7</td>
<td>22343</td>
<td>822</td>
<td>3.68</td>
</tr>
<tr>
<td>5.8</td>
<td>25151</td>
<td>806</td>
<td>3.2</td>
</tr>
</tbody>
</table>

Disclaimer 1: not all failures are due to an actual bug in the kernel...
Disclaimer 2: some failures are due to the same bug
Disclaimer 3: haven't done yet neither an in-depth analysis nor bug reports or data sharing (if you're interested in, ping me!) for all versions
Sometimes configurations fail

Classification problem: predict the class (BUILD/FAILURE) out of options values

<table>
<thead>
<tr>
<th>X86_64</th>
<th>TRACING</th>
<th>DRM_VBOXVIDEO</th>
<th>GENERIC_ALLOCATOR</th>
<th>DRM_TLM</th>
<th>BUILD?</th>
<th>SIZE (Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>...</td>
<td>1</td>
<td>0</td>
<td>BUILD?</td>
<td>8.3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>...</td>
<td>1</td>
<td>0</td>
<td>FAILURE</td>
<td>78.9</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>...</td>
<td>1</td>
<td>1</td>
<td>BUILD?</td>
<td>97.3</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>...</td>
<td>1</td>
<td>1</td>
<td>BUILD?</td>
<td>22.4</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>2</td>
<td>FAILURE</td>
<td>34.8</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>...</td>
<td>1</td>
<td>1</td>
<td>BUILD?</td>
<td>176.8</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>...</td>
<td>1</td>
<td>0</td>
<td>BUILD?</td>
<td>54.3</td>
</tr>
</tbody>
</table>

Do you recognize a pattern?

(matrix is 95K rows and 12K columns)
Sometimes configurations fail

Classification tree

Configuration bug location/understanding/prevention: TUXML can help to pinpoint faulty (combinations of) options

(even more effective when combined with failures clustering, see details in research papers)
Sometimes configurations fail

- A failure raises two questions:
  - Is it a bug of the Linux kernel?
    - Avoid false positives or duplication of bug reports
    - Pinpoint responsible options (usually it's not one individual option)
    - Preventive fix at Kconfig level to avoid build of configurations that will fail anyway
    - Synthesize a fix or wait for a fix (humans to the rescue!)
  - Is it a bug of the compilation environment?
    - No it's the Linux doc ;)
    - CI service bug e.g., TUXML, not KernelCI ;)
    - **Compiler “bug”** e.g., gcc, not clang ;)

LINUX PLUMBERS CONFERENCE
August 24-28, 2020
Claim: Configuration failures can... help!

- To pinpoint Clang-specific issues
  - compare with gcc: same build status?
- Identify problematic combinations of options for Clang
  - Avoid duplication of bug reports
  - Preventive fix to jump to other problematic cases
  - Help understanding/fixing Clang-related bugs
    - You can know where the problem potentially comes from
  - Help improving Clang
Design Study (brainstorm)

- Large-scale compilation with Clang of random configs.
- Which clang version? (and other specificities: clang options?)
- Which architecture? (eg x86_64?)
- Which Linux version? (eg 5.4 LTS?)
- Which configurations?
  - randconfig
  - which options should be (de-)activated anyway?
- Which observations? (warnings? build status? size? compilation time?)
- Hypothesis (poll)
  - % build failures with clang?
  - (much) more than gcc?
Construct Validity (brainstorm)

- Too costly?
  - ~260 seconds per config (with gcc/x86_64) on 16-core, recent machines

- Too many failures? (let's be provocative)
  - Overwhelming: “we” have already tons of bugs to resolve and fix…
  - Irritating: fixing “esoteric” configurations may decrease motivation or complicate the task

- We don’t know after all... so let’s try!

- **Seeking volunteers to help “designing” and technically realizing this experiment**
  - Can well be done with KernelCI/kcidb
  - Need Clang-specific knowledge
Supplementary material
Research papers
https://hal.inria.fr/hal-02147012/

Learning From Thousands of Build Failures of Linux Kernel Configurations
- Mathieu Acher, Hugo Martin, Juliana Alves Pereira, Arnaud Blouin, Djamal Eddine Khelladi, Jean-Marc Jézéquel
- https://hal.inria.fr/hal-02147012

Learning Very Large Configuration Spaces: What Matters for Linux Kernel Sizes
- Mathieu Acher, Hugo Martin, Juliana Pereira, Arnaud Blouin, Jean-Marc Jézéquel, Djamel Eddine Khelladi, Luc Lesoil, Olivier Barais
- https://hal.inria.fr/hal-02314830

https://hal.inria.fr/hal-02314830

Talk at Embedded Linux Conference Europe (2019, Lyon)
https://www.youtube.com/watch?v=UBghs-cwQX4&feature=youtu.be
Team

- DiverSE research team [http://diverse-team.fr](http://diverse-team.fr)
  - Hugo Martin, Juliana Alves Pereira, Arnaud Blouin, Jean-Marc Jézéquel, Djamel Eddine Khelladi, Luc Lesoil, Olivier Barais to name a few

- TUXML team at ISTIC / University of Rennes 1
  - Paul Saffray, Alexis Le Masle, Michaël Picard, Corentin Chédotal, Gwendal Didot, Dorian Dumange, Antonin Garret, Erwan Le Flem, Pierre Le Luron, Mickaël Lebreton, Fahim Merzouk, Valentin Petit, Julien Royon Chalendard, Cyril Hamon, Luis Thomas, Alexis Bonnet

- Georges Aaron Randrianaina (ENS Rennes)

- IGRIDA [http://igrida.gforge.inria.fr](http://igrida.gforge.inria.fr)
  - Morillon Pascal, Deltel Charles