Empirical Study on Applying Program Analysis and Testing Tools to Student Code

Frederico Ramos, Filipe Marques, Nuno Santos, Pedro Adão, José Fragoso Santos

INESC-ID / Instituto Superior Técnico, Universidade Lisboa /Instituto Telecomunicações

KLEE Workshop 2022
How well do existing testing and verification tools perform on student code?

➔ **RQ1**: Number of false positives (*Precision*)

➔ **RQ2**: Number of false negatives (*Recall*)

➔ **RQ3**: Resource usage (*Memory / Time*)

**Goal**: Make a case for the introduction of testing and verification tools in undergraduate courses
Contributions

➔ A curated data set consisting of **1160 student projects (405k LoC)** annotated with bug locations
  ◆ Types of bugs detected: heap-overflows, stack-overflows invalid pointers, uninitialised-memory, stack-underflow, memory Leaks

➔ An empirical study characterizing how **9 state-of-the-art testing and verification tools** perform on our curated data set
  ◆ Selected tools: FuseBMC, LibKluzzer, Verifuzz, Klee, Symbiotic, CPAchecker, Infer, Pulse

➔ **Preliminary results** obtained for: Infer, Pulse, KLEE, Symbiotic
Methodology

1. Annotator: AddressSanitizer
   - Input: Data set
   - Output: Annotated Data set

2. Instrumentor: PycParser
   - Input: Annotated Data set
   - Output: Instrumented Data set

3. Tool: config.
   - Input: Instrumented Data set
   - Output: Tool config.

4. Executor
   - Input: Tool config.
   - Output: Bug Reports

5. Validator
   - Input: Expanding Dataset
   - Output: Output

6. Json Bug Reports
   - Input: Annotated Data set
   - Output: Json Bug Reports

Project specific input generators

1160 Student Projects in C
Tool Selection Criteria

➔ **C1:** 5 best-performing tools in the *Cover-Error* category in **Test-Comp 2022**
   - FuseBMC, LibKluzzer, Verifuzz, Klee, Symbiotic

➔ **C2:** Winners of the categories *MemSafety, NoOverflows* and *SoftwareSystems* from **SV-Comp 2022**
   - Symbiotic, CPAchecker

➔ **C3:** Other high-profile static analysis tools
   - Infer, Pulse

If you want us to include your tool, contact us!
All evaluated tools perform well, uncovering most of the memory bugs present in the dataset, with high recall and precision. Updated results in the poster.
Thank You

Collaborate with us

→ The annotated dataset will be open-sourced
→ Do you want us to include your tool in our study? Come talk to us!