

5th Competition on Software Testing



Proc. FASE 2023,

https://doi.org/10.1007/978-3-031-30826-0_17

Dirk Beyer (Competition Chair)

Supported by:



Motivation - Goals

1. Community suffers from unreproducible results
→ Establish set of benchmarks
2. Publicity for tools that are available
→ Provide state-of-the-art overview
3. Support the development of verification tools
→ Give credits and visibility to developers
4. Establish standards
→ Specification language, Test-suites,
Benchmark definitions, Validators

Schedule of Sessions

Session 1:

- ▶ Competition Report, by organizer
- ▶ System Presentations, 15 min by each team
- ▶ Short discussion

Session 2:

- ▶ Open Jury Meeting, Community Discussion, moderated by organizer

Procedure – Time Line

Three Steps – Three Deadlines:

- ▶ Benchmark submission deadline
- ▶ System submission
- ▶ Notification of results (approved by teams)

Test Problem

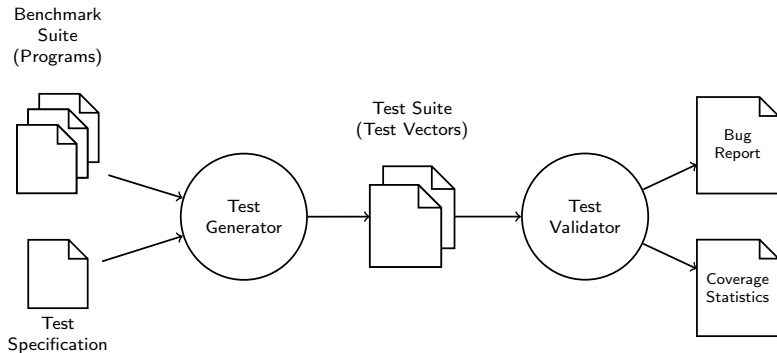
Input:

- ▶ C program → GNU/ANSI C standard
- ▶ Test Specification:
 - Coverage of function call
 - Branch coverage

Output:

- ▶ Test suite

Flow of the Test-Comp execution



Environment

Machines (1000 \$ consumer machines):

- ▶ CPU: 3.4 GHz 64-bit Quad-Core CPU
- ▶ RAM: 33 GB
- ▶ OS: GNU/Linux (Ubuntu 22.04)

Resource limits:

- ▶ 15 GB memory
- ▶ 15 min CPU time

Scoring Schema (since 2019)

Common principles: Ranking measure should be

- ▶ easy to understand
- ▶ reproducible
- ▶ computable in isolation for one tool

Test-Comp:

- ▶ Coverage of call to function:
1 point or 0 points
- ▶ Coverage of branches:
TEST-COV coverage value (between 0 and 1)

Fair and Transparent

Jury:

- ▶ Team: one member of each participating candidate
- ▶ Term: one year (until next participants are determined)

Systems:

- ▶ All systems are available in open GitLab repo
- ▶ Configurations and Setup in GitLab repository
→ Integrity and reproducibility guaranteed

Competition Candidates

Qualification:

- ▶ 13 Qualified (3 hors concours)
- ▶ One person can participate with different tools
- ▶ One tool can participate with several configurations (frameworks, no tool-name inflation)

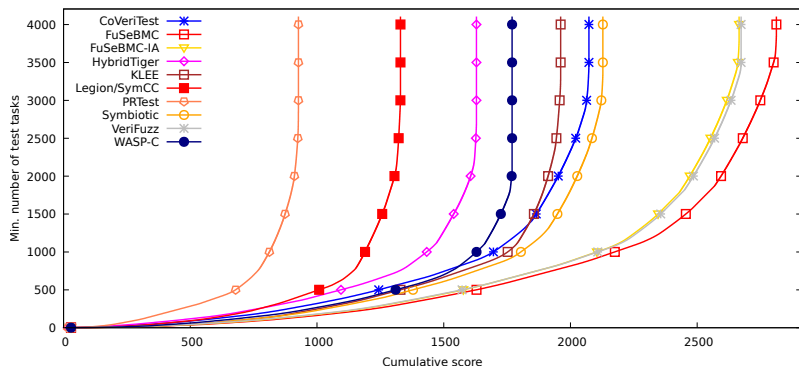
Benchmark quality:

- ▶ Community effort, documented on GitHub

Role of organizer:

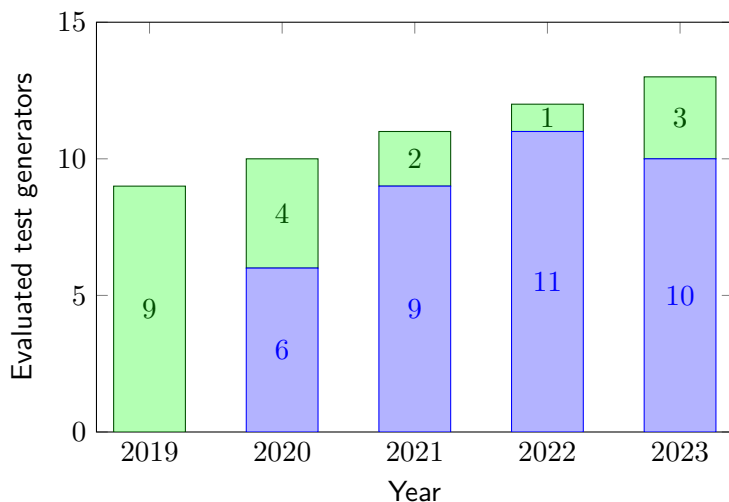
- ▶ Just service: Advice, Technical Help, Executing Runs

Results – Example: Overall



Number of Participants

Number of evaluated test generators for each year (top: number of first-time participants; bottom: previous year's participants)



Thanks to:

- ▶ Jury (14 people)
- ▶ Participants
- ▶ Sponsors: LMU Munich and ETAPS

