12th Competition on Software Verification

Proc. TACAS 2023, doi:10.1007/978-3-031-30820-8_29

Dirk Beyer (Competition Chair)

Verification Problem

Input:

- C program → GNU/ANSI C standard
- Property
  → Reachability of error label, of overflows
  → Memory safety (inv-deref, inv-free, memleak)
  → Termination

Output:

- TRUE + Witness (property holds)
- FALSE + Witness (property does not hold)
- UNKNOWN (failed to compute result)

Scoring Schema

Common principles: Ranking measure should be

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

SV-COMP:

- Ranking measure is the quality of verification work
- Expressed by a community-agreed score
- Tie-breaker is CPU time

Scoring Schema (2023, unchanged)

Reported result	Points	Description
UNKNOWN	0	Failure, out of ressources
FALSE correct	$+1$	Error found and confirmed
FALSE incorrect	-16	False alarm (imprecise analysis)
TRUE correct	$+2$	Proof found and confirmed
TRUE incorrect	-32	Missed bug (unsound analysis)

Fair and Transparent

Jury:

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

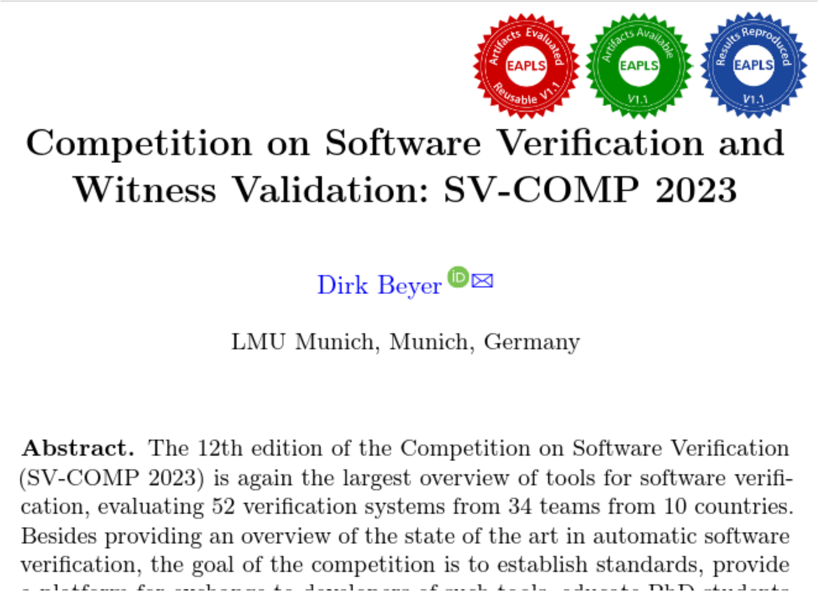
Systems:

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

Benchmark Sets

- Everybody can submit benchmarks (conditions apply)
- Eight categories when closed (scores normalized):
  - Reachability: 9814 tasks
  - Memory Safety: 4543 tasks
  - Concurrency: 5295 tasks
  - NoOverflows: 10200 tasks
  - Termination: 3103 tasks
  - Software Systems: 5132 tasks
  - Overall: 38644 tasks
  - Java: 827 tasks

Reproducibility



Artifacts Evaluated EAPLS Reusable V1.1

Artifacts Available EAPLS V1.1

Results Reproduced EAPLS V1.1

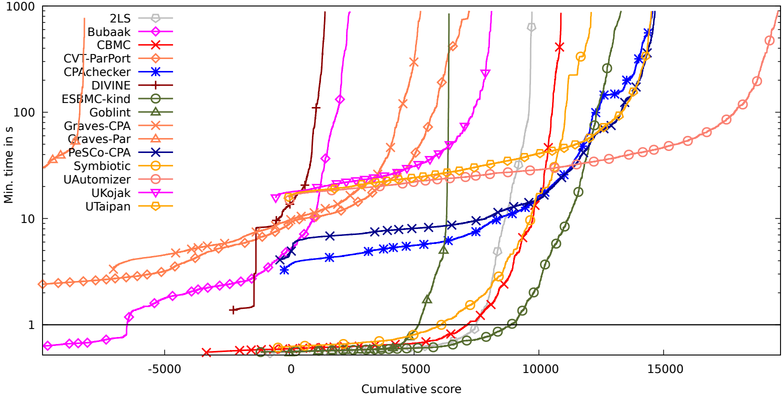
Competition on Software Verification and Witness Validation: SV-COMP 2023

Dirk Beyer

LMU Munich, Munich, Germany

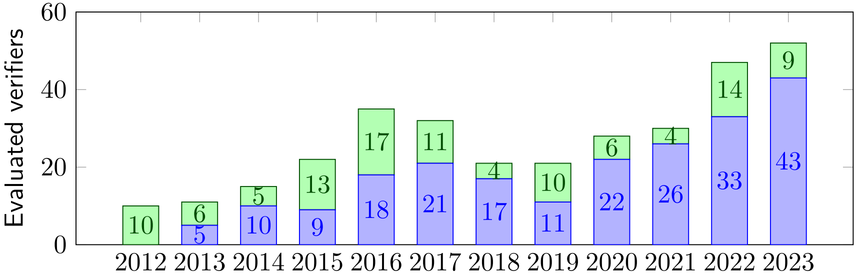
Abstract. The 12th edition of the Competition on Software Verification (SV-COMP 2023) is again the largest overview of tools for software verification, evaluating 52 verification systems from 34 teams from 10 countries. Besides providing an overview of the state of the art in automatic software verification, the goal of the competition is to establish standards, provide a platform for exchange to developers of such tools, educate PhD students

Results – Example: Overall



Number of Participants

Number of evaluated verifiers for each year
(first-time participants on top)


Evaluated verifiers