

SV-COMP'25 Reproduction* Report

*: Parallel Execution

Levente Bajczi,
Zsófia Ádám, Zoltán Micskei

May 4, 2025



Reproduction Goals

Goal 1: Evaluate reproducibility across different environments

- ▶ How portable are tools and their configurations to other environments?

Reproduction Goals

Goal 1: Evaluate reproducibility across different environments

- ▶ How portable are tools and their configurations to other environments?

Goal 2: Ensure that SV-COMP can endure changes in organization

- ▶ An independent guide on reproduction / competition execution

Reproduction Goals

Goal 1: Evaluate reproducibility across different environments

- ▶ How portable are tools and their configurations to other environments?

Goal 2: Ensure that SV-COMP can endure changes in organization

- ▶ An independent guide on reproduction / competition execution

Goal 3: Enhance trust in results by reaffirming their reliability

- ▶ Do scores or ranks (and medals) change?

Reproduction Goals

Goal 1: Evaluate reproducibility across different environments

- ▶ How portable are tools and their configurations to other environments?

Goal 2: Ensure that SV-COMP can endure changes in organization

- ▶ An independent guide on reproduction / competition execution

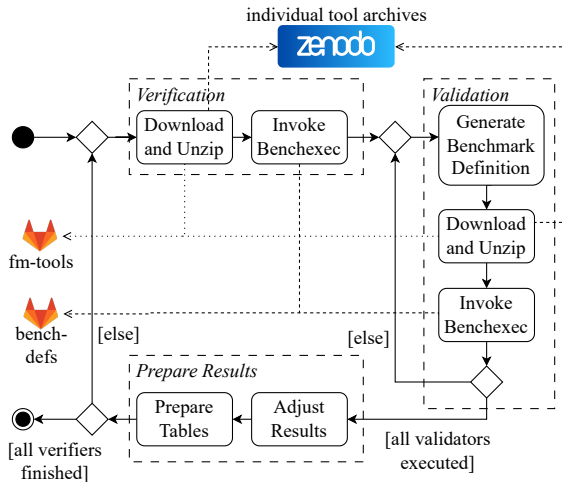
Goal 3: Enhance trust in results by reaffirming their reliability

- ▶ Do scores or ranks (and medals) change?

Goal 4: Evaluate usability of SV-COMP tools **beyond SV-COMP**

- ▶ Are the tools only suitable for competitions, or can they be used independently as well?

Reproduction Workflow



Modified Benchexec (to run on SLURM)

- ▶ 184 node
- ▶ 2x AMD EPYC™ 7763 (64-core)
- ▶ 256 GB RAM



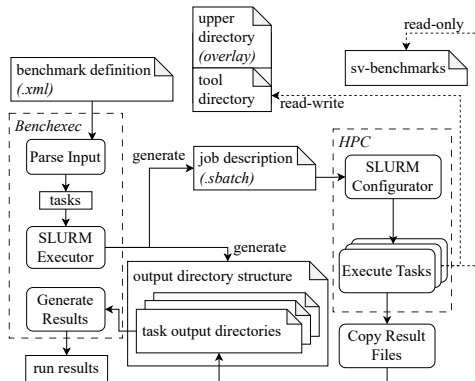
`hpc.kifu.hu/komondor`

Modified Benchexec (to run on SLURM)

- ▶ 184 node
- ▶ 2x AMD EPYC™ 7763 (64-core)
- ▶ 256 GB RAM



hpc.kifu.hu/komondor



Reproduction Results in Numbers

- ▶ *Active* participants: 33 verifiers, 33 validators

Reproduction Results in Numbers

- ▶ *Active* participants: 33 verifiers, 33 validators
- ▶ 495 828 tasks ran

Reproduction Results in Numbers

- ▶ *Active* participants: 33 verifiers, 33 validators
- ▶ 495 828 tasks ran
- ▶ 38 358 tasks with **mismatching verdicts** (7.74%)
 - ▶ 19 222: known problems (see later...)
 - ▶ 19 136: other problems, mostly resource limits

Reproduction Results in Numbers

- ▶ *Active* participants: 33 verifiers, 33 validators
- ▶ 495 828 tasks ran
- ▶ 38 358 tasks with **mismatching verdicts** (7.74%)
 - ▶ 19 222: known problems (see later...)
 - ▶ 19 136: other problems, mostly resource limits
- ▶ 1 result **flipped to wrong** (mlb)
- ▶ 24 result **flipped to correct** (mlb, swat)

Reproduction Results in Numbers

- ▶ *Active* participants: 33 verifiers, 33 validators
- ▶ 495 828 tasks ran
- ▶ 38 358 tasks with **mismatching verdicts** (7.74%)
 - ▶ 19 222: known problems (see later...)
 - ▶ 19 136: other problems, mostly resource limits
- ▶ 1 result **flipped to wrong** (mlb)
- ▶ 24 result **flipped to correct** (mlb, swat)
- ▶ 24% less walltime, 32% less cputime, and 11% more memory

Reproduction Results in Numbers

- ▶ *Active* participants: 33 verifiers, 33 validators
- ▶ 495 828 tasks ran
- ▶ 38 358 tasks with **mismatching verdicts** (7.74%)
 - ▶ 19 222: known problems (see later...)
 - ▶ 19 136: other problems, mostly resource limits
- ▶ 1 result **flipped to wrong** (mlb)
- ▶ 24 result **flipped to correct** (mlb, swat)
- ▶ 24% less walltime, 32% less cputime, and 11% more memory
- ▶ **No medal change** (for tools without “known problems”)

Reproduction Results in Numbers

- ▶ *Active* participants: 33 verifiers, 33 validators
- ▶ 495 828 tasks ran
- ▶ 38 358 tasks with **mismatching verdicts** (7.74%)
 - ▶ 19 222: known problems (see later...)
 - ▶ 19 136: other problems, mostly resource limits
- ▶ 1 result **flipped to wrong** (mlb)
- ▶ 24 result **flipped to correct** (mlb, swat)
- ▶ 24% less walltime, 32% less cputime, and 11% more memory
- ▶ **No medal change** (for tools without “known problems”)

`zenodo.org/records/14913865`
`home.mit.bme.hu/~bajczi/sv-comp-repro-25`

“Known Problems”

Problem 1: Missing packages (15 tools)

Against rules

“Known Problems”

Problem 1: Missing packages (15 tools)

Against rules

Problem 2: Assumed read-write access

Not resolved in repro

- ▶ Overwrote input program (**high** prio) / created files outside workdir (**low** prio)

“Known Problems”

Problem 1: Missing packages (15 tools)

Against rules

Problem 2: Assumed read-write access

Not resolved in repro

- ▶ Overwrote input program (**high** prio) / created files outside workdir (**low** prio)

Problem 3: *Resultfiles* pattern mismatch

Against rules

- ▶ `error-witness.graphml` output, declared `**/witness.*`

“Known Problems”

Problem 1: Missing packages (15 tools)

Against rules

Problem 2: Assumed read-write access

Not resolved in repro

- ▶ Overwrote input program (**high** prio) / created files outside workdir (**low** prio)

Problem 3: *Resultfiles* pattern mismatch

Against rules

- ▶ `error-witness.graphml` output, declared `**/witness.*`

Problem 4: Working dir must be tool dir

No effect on repro

- ▶ Starting script assumes benchexec was started from the tool directory

"Known Problems"

Problem 1: Missing packages (15 tools)

Against rules

Problem 2: Assumed read-write access

Not resolved in repro

- ▶ Overwrote input program (**high** prio) / created files outside workdir (**low** prio)

Problem 3: *Resultfiles* pattern mismatch

Against rules

- ▶ `error-witness.graphml` output, declared `**/witness.*`

Problem 4: Working dir must be tool dir

No effect on repro

- ▶ Starting script assumes benchexec was started from the tool directory

Problem 5: Missing tool metadata in schema

No effect on repro

- ▶ Cgroups access, need for containerization, etc. not in fm-tools entries

"Known Problems"

Notice: no problems with competition scripts!

Problem 1: Missing packages (15 tools)

Against rules

Problem 2: Assumed read-write access

Not resolved in repro

- ▶ Overwrote input program (**high** prio) / created files outside workdir (**low** prio)

Problem 3: *Resultfiles* pattern mismatch

Against rules

- ▶ `error-witness.graphml` output, declared `**/witness.*`

Problem 4: Working dir must be tool dir

No effect on repro

- ▶ Starting script assumes benchexec was started from the tool directory

Problem 5: Missing tool metadata in schema

No effect on repro

- ▶ Cgroups access, need for containerization, etc. not in fm-tools entries

Recommendations

Problem 1: Missing packages (15 tools)

Against rules

- ▶ Enforce package declaration (*via CI?*) – [!539 \(bench-defs\)](#)

Recommendations

Problem 1: Missing packages (15 tools)

Against rules

- ▶ Enforce package declaration (*via CI?*) – [!539 \(bench-defs\)](#)

Problem 2: Assumed read-write access

Not resolved in repro

- ▶ Declare where tools may assume *rw* access – [!538 \(bench-defs\)](#), *maybe not a good idea?*

Recommendations

Problem 1: Missing packages (15 tools)

Against rules

- ▶ Enforce package declaration (*via CI?*) – [!539 \(bench-defs\)](#)

Problem 2: Assumed read-write access

Not resolved in repro

- ▶ Declare where tools may assume *rw* access – [!538 \(bench-defs\)](#), *maybe not a good idea?*

Problem 3: *Resultfiles* pattern mismatch

Against rules

- ▶ Enforce the result files pattern – [!704 \(benchcloud\)](#), *known issue, will be fixed*

Recommendations

Problem 1: Missing packages (15 tools)

Against rules

- ▶ Enforce package declaration (*via CI?*) – [!539 \(bench-defs\)](#)

Problem 2: Assumed read-write access

Not resolved in repro

- ▶ Declare where tools may assume *rw* access – [!538 \(bench-defs\)](#), *maybe not a good idea?*

Problem 3: *Resultfiles* pattern mismatch

Against rules

- ▶ Enforce the result files pattern – [!704 \(benchcloud\)](#), *known issue, will be fixed*

Problem 4: Working dir must be tool dir

No effect on repro

- ▶ Tools should be able to be executed not from their tool dir – [!541 \(bench-defs\)](#)

Recommendations

Problem 1: Missing packages (15 tools)

Against rules

- ▶ Enforce package declaration (*via CI?*) – [!539 \(bench-defs\)](#)

Problem 2: Assumed read-write access

Not resolved in repro

- ▶ Declare where tools may assume *rw* access – [!538 \(bench-defs\)](#), *maybe not a good idea?*

Problem 3: *Resultfiles* pattern mismatch

Against rules

- ▶ Enforce the result files pattern – [!704 \(benchcloud\)](#), *known issue, will be fixed*

Problem 4: Working dir must be tool dir

No effect on repro

- ▶ Tools should be able to be executed not from their tool dir – [!541 \(bench-defs\)](#)

Problem 5: Missing tool metadata in schema

No effect on repro

- ▶ FM-tools schema update – [!20 \(fm-tools\)](#), *current proposal too vague – should we care?*