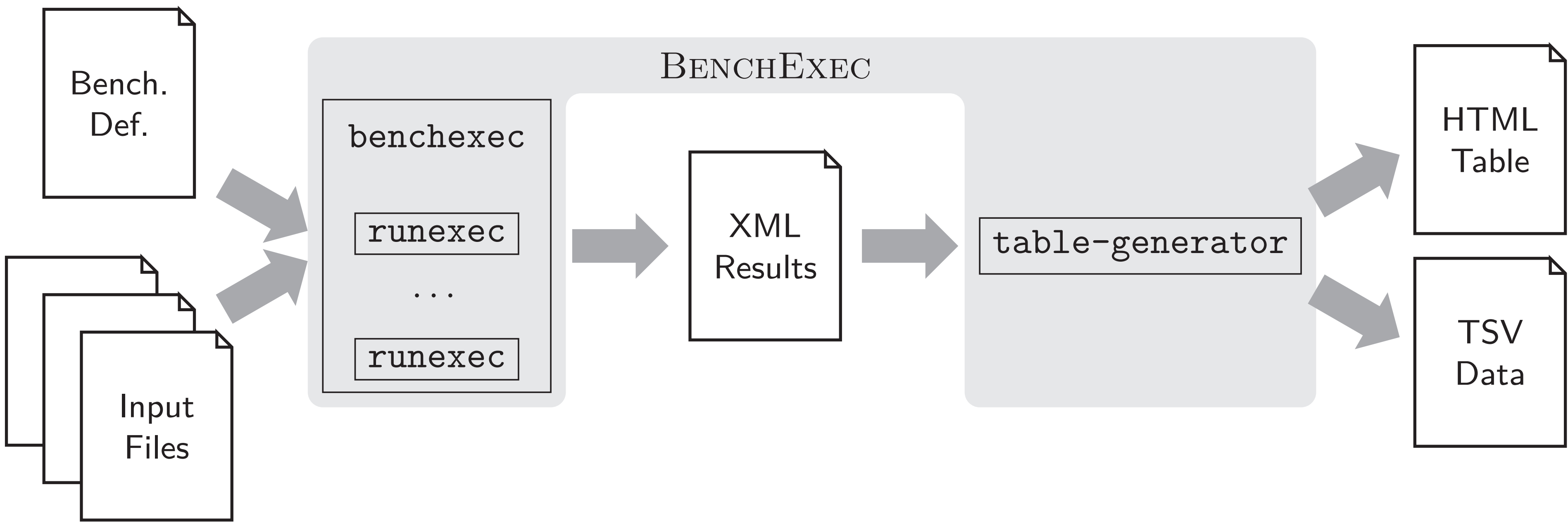


Dirk Beyer, Stefan Löwe, and Philipp Wendler

OVERVIEW



BENCHEXEC: A Framework for Reliable Benchmarking and Resource Measurement

BENCHMARKING REQUIREMENTS

1. Measure and Limit Resources Accurately
2. Terminate Processes Reliably
3. Assign Cores Deliberately
4. Respect Nonuniform Memory Access
5. Avoid Swapping
6. Isolate Individual Runs

SCOPE

- Linux systems
 - CPU-bound tool (negligible I/O)
 - No use of other resources such as GPUs
 - No networking / distributed execution
 - No user interaction
 - No malicious intent
- ⇒ **Great for solvers, verifiers, etc.!**

USE CASES

- Low-level command for isolated, limited, and measured execution of a tool
- Integration in other benchmarking frameworks via command line and Python API (**used by StarExec**)
- Benchmarking with large number of runs
- Competition execution (**used e.g. by SV-COMP since 2016**)
- Regression testing

TECHNIQUES AND FEATURES

- Benchmarking containers** implemented with Linux features such as
- **Control groups (cgroups)** for resource limitation and measurements (compatible with cgroups v1 and v2)
 - **Namespaces** for isolation
 - **Overlay filesystem (overlayfs)** for intercepting file writes (same techniques as used by Docker, etc.)
- Parallel execution of tools
 - Automatic calculation of distribution of cores and memory regions
 - Knows about NUMA and hyper threading
 - Configurable file-system layout in container (hide directories, allow write access, etc.)

EXECUTION

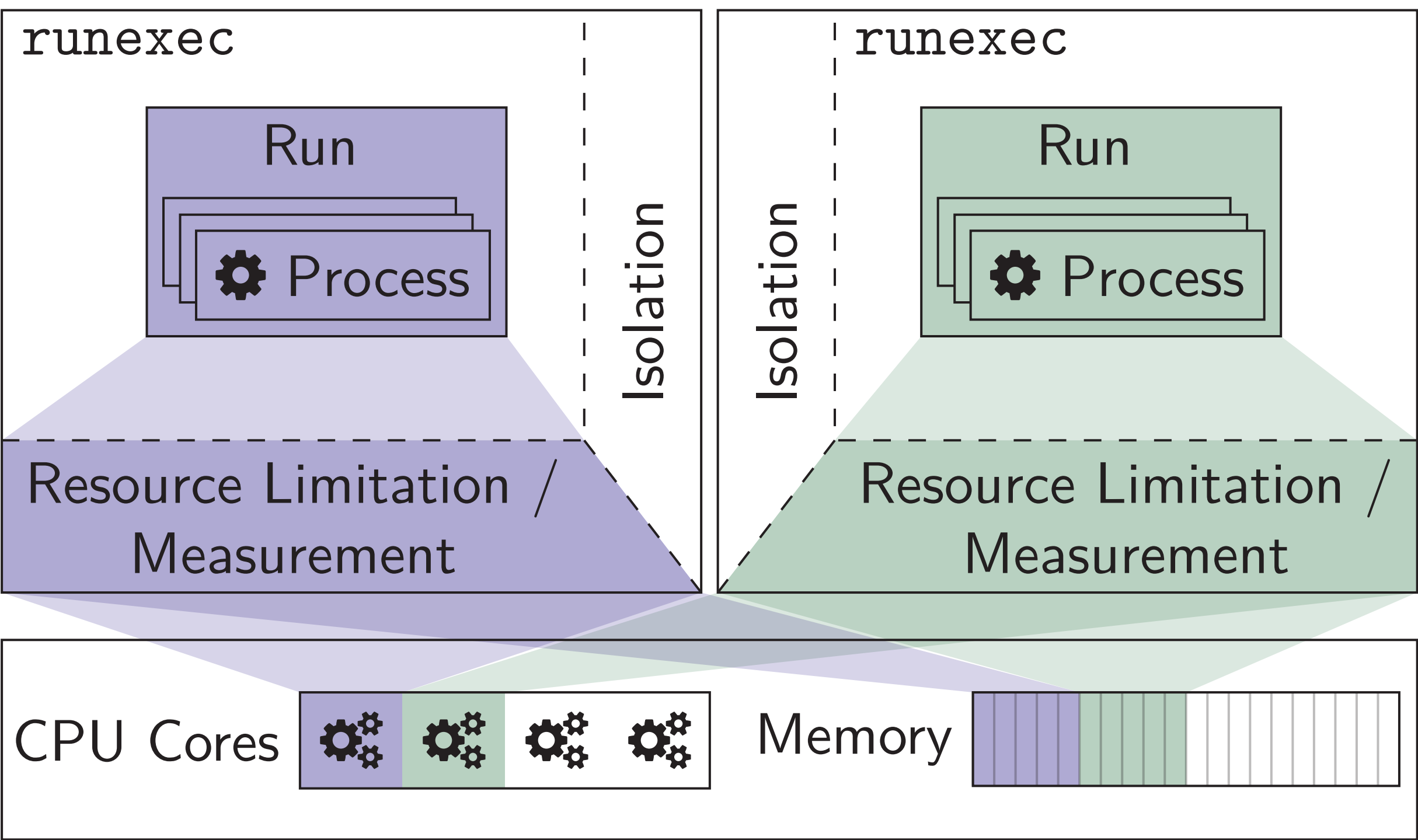


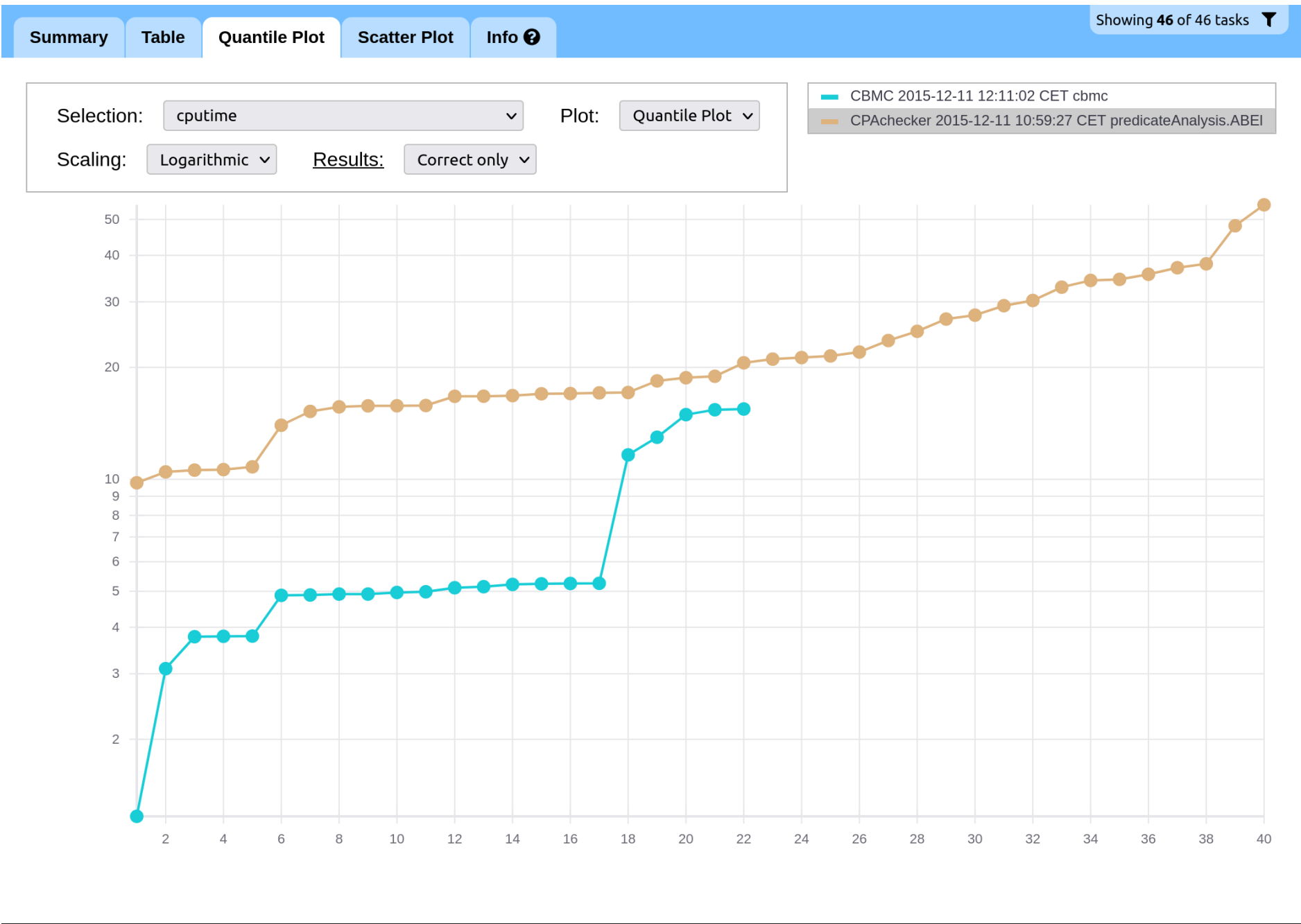
TABLE-GENERATOR

- Combine results from several executions
- Define table layout
- Select and filter results
- Compute statistics
- Export raw data as TSV
- Generate interactive tables as stand-alone HTML files
- Quantile and scatter plots
- Live analysis of data



Interactive
online example

task	status	cputime (s)	walltime (s)	memUsage (MB)
ntdrivers/cdaudio_false-unreach-call.i	TIMEOUT	60.1	60.1	987
ntdrivers/diskperf_false-unreach-call.i	false(reach)	1.24	1.26	104
ntdrivers/floppy_false-unreach-call.i	false(reach)	3.10	3.11	140
ntdrivers/kbfiltr_false-unreach-call.i	TIMEOUT	60.1	60.1	573
ntdrivers/parport_false-unreach-call.i	false(reach)	13.0	13.3	676
ntdrivers/cdaudio_true-unreach-call.i	TIMEOUT	60.1	60.4	993
ntdrivers/diskperf_true-unreach-call.i	TIMEOUT	60.2	60.1	3500
ntdrivers/floppy2_true-unreach-call.i	TIMEOUT	60.1	60.0	989
ntdrivers/floppy_true-unreach-call.i	TIMEOUT	60.0	60.4	391
ntdrivers/parport_true-unreach-call.i	TIMEOUT	61.1	61.7	2180
sshs3_cint_blast.01_false-unreach-call.i	false(reach)	11.6	12.0	411
sshs3_cint_blast.02_false-unreach-call.i	false(reach)	3.79	4.15	182
sshs3_cint_blast.03_false-unreach-call.i	false(reach)	3.78	3.82	182
sshs3_cint_blast.04_false-unreach-call.i	false(reach)	3.77	3.79	180
sshs3_srvr_blast.01_false-unreach-call.i	false(reach)	4.91	4.94	242
sshs3_srvr_blast.02_false-unreach-call.i	false(reach)	4.87	4.90	229
sshs3_srvr_blast.03_false-unreach-call.i	false(reach)	4.89	4.90	229



PAPER



- STTT 2017
- Open Access
- DOI 10.1007/s10009-017-0469-y
- Important aspects for benchmarking, hardware influence, how to present results, ...

THANKS TO ALL CONTRIBUTORS!

Aditya Arora, Laura Bschor, Thomas Bunk, Montgomery Carter, Saransh Chopra, Andreas Donig, Karlheinz Friedberger, Peter Häring, Florian Heck, Hugo van Kemenade, George Karpenkov, Mike Kazantsev, Michael Lachner, Thomas Lemberger, Sebastian Ott, Stephan Lukasczyk, Alexander von Rhein, Alexander Schremmer, Dennis Simon, Andreas Stahlbauer, Thomas Stieglmaier, Martin Yankov, Ilja Zakharov, and more (100 in total)!

TOOL BENCHEXEC

- License Apache 2.0
- No root access required for benchmarking
- Available on PyPI and github.com/sosy-lab/benchexec

