The Numbers

- 11 teams participated
- **Solvers:**
  - Main track: 21
  - Application track: 10
  - 2 non-competitive
  - 3 non-competitive
- **Logics:**
  - Main track: 40
  - Application track: 14
- **Benchmarks:**
  - Main track: 154238
  - Application track: 9852

Record numbers of solvers, logics, and benchmarks!
Job Pairs

- 1,028,615 job pairs executed (+ some repeats)
- ≈ 5 days × 150 nodes × 2 processors/node of compute time

More than 3 times as many job pairs as in 2014!
StarExec

- All job pairs executed on StarExec
- Over 9,000 job pairs/hour completed

StarExec worked great

- Thanks to Aaron Stump for prompt help when problems or questions arose
- \( \sim 20 \) feature requests and (minor) bug reports submitted to the StarExec developers
Machine Specifications

Hardware:
- Intel Xeon CPU E5-2609 @ 2.4 GHz, 10 MB cache
- 2 processors per node, 4 cores per processor
- Main memory capped at 60 GB per job pair

Software:
- Red Hat Enterprise Linux Workstation release 6.3
- Kernel 2.6.32-431, gcc 4.4.6, glibc 2.12 (∼ 2009-2011)
- Virtual machine image available before the competition

Problems with missing libraries (due to dynamic linking) in several solvers resolved during pre-competition testing in early June.
Almost 60,000 new benchmarks added to SMT-LIB, thanks to $\text{CONTRIBUTORS}$:

- Six new logics, including two new floating-point logics
- Thanks to Clark Barrett for curation and uploading
Benchmark Curation

- Sanity checks
  - One satisfiability check per benchmark in main track
  - Status information set before satisfiability check
- Verify benchmark signature against logic set
- Remove unused symbols
- Improve logic settings
Eligible Benchmarks

All eligible benchmarks were used for the competition. There was no further selection.
Competition Tools Improved

- Fixed an issue where the trace executor would sometimes not count correct solver responses on partially solved incremental benchmarks. (Thanks to Kshitij Bansal for reporting this.)

- Fixed several issues in the benchmark scrambler that caused invalid output in the presence of variable shadowing.
Evolution of Benchmarks: Breakdown

Tier 1 (> 1000 Benchmarks)
Evolution of Benchmarks: Breakdown

Tier 2 (< 1000 Benchmarks)
Evolution of Tool Participation: Breakdown

Quantifier-Free Logics

![Graph showing the evolution of tool participation for different logics in 2014 and 2015.](image-url)
Evolution of Tool Participation: Breakdown

Logics with Quantifiers
Further Thoughts

Benchmarks:
- Still more benchmarks needed, especially for small divisions
- Resolve semantics of partial operations, e.g., bvdiv, fp.min

Solvers:
- Parallelism

Competition:
- Relative weight of benchmarks and benchmark families
- Separate measure of performance on quick jobs
- Additional tracks, e.g., unsat-core, proofs

Teams:
- Congratulations on your accomplishments!
- Thanks for your participation!