

Third International Competition on Runtime Verification (CRV16)

Giles Reger, Sylvain Hallé, Yliès Falcone

RV 2016

History

- ▶ First competition ran in 2014
- ▶ Changing competition organisers
 - 2014 Ezio Bartocci, Borzoo Bonakdarpour, Yliès Falcone
 - 2015 Yliès Falcone, Dejan Nickovic, Giles Reger, Daniel Thoma
 - 2016 Giles Reger, Sylvain Hallé, Yliès Falcone
- ▶ Overall goals have remained the same
 - ▶ Stimulate RV tool development and visibility
 - ▶ Provide community benchmarks
 - ▶ Evaluate RV tools and discuss the metrics used
- ▶ Related to the COST Action (at least supported since 2015)

Design

- ▶ Structure has remained relatively consistent
- ▶ Main change: reduce the number of benchmarks

- ▶ Three tracks
 - ▶ Offline
 - ▶ Online Java
 - ▶ ~~Online C~~ [lack of interest]

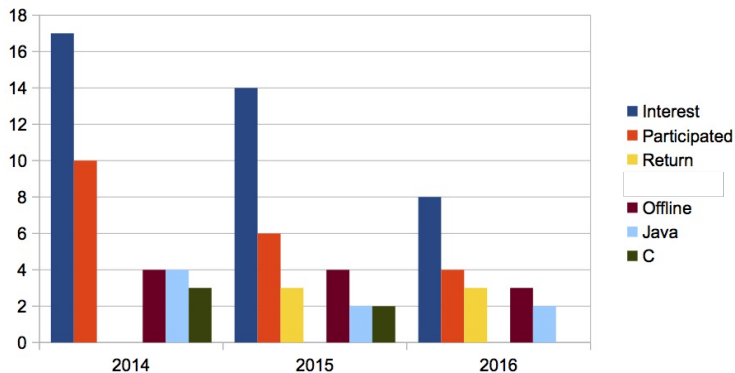
- ▶ Phases
 - ▶ Registration
 - ▶ Benchmark Submission
 - ▶ Clarifications
 - ▶ Monitor Submission
 - ▶ Evaluation
 - ▶ Results

Organisation

- ▶ Registration was completed via a Google form
- ▶ A Wiki for collecting team and benchmark information was hosted in Québec
 - ▶ A page per benchmark
 - ▶ A benchmark page contains all necessary information
 - ▶ It should also contain all clarifications and communication related to that benchmark
- ▶ A server was provided
 - ▶ Each team had a space to upload their trace and source files
 - ▶ Teams installed their system in this space
 - ▶ The server was used for evaluation, allowing teams to test their submissions on the evaluation machine

Participation

- ▶ Both interest and participation has decreased
- ▶ This year we directly contacted all previous participants and potential new participants, as well as advertising on email lists
- ▶ The main reason for not returning was the time commitment



Teams

- ▶ Four teams reached evaluation
- ▶ Only one newcomer (BEEPBEEP 3)

Tool	Affiliation
Java track	
LARVA	University of Malta, Malta
MARQ	University of Manchester, UK
MUFIN	University of Lübeck, Germany
Offline track	
BEEPBEEP 3	Université du Québec à Chicoutimi, Canada
MARQ	University of Manchester, UK

Benchmarks

- ▶ Offline track (6 benchmarks)
 - ▶ 2 business-level properties
 - ▶ 1 system-level property
 - ▶ 3 properties from a video game case study
- ▶ Java track (9 benchmarks)
 - ▶ 3 benchmarks from a finance system case study
 - ▶ 2 business-level properties
 - ▶ 4 system-level properties
- ▶ No benchmarks came from real-world applications

Results

- ▶ MARQ won the Offline track (again, 2014)
- ▶ MUFIN won the Java track (again, 2015)

- ▶ LARVA suffered from time-outs (and lost points for this)
- ▶ Question: should we remove points for time-outs?

Team	Bench.	Correct.	Time	Memory	Total	Average
Offline Track						
BEEP BEEP 3	6	60	14.42	25.51	97.93	16.32
MARQ	6	45	45.58	36.49	127.07	21.18
Java Track						
LARVA	9	45	10.88	15.36	71.24	7.92
MARQ	8	80	20.25	17.30	117.65	14.71
MUFIN	9	90	58.87	57.34	206.21	22.91

Reflection

- ▶ Existing trace formats were not sufficient
 - ▶ BEEP BEEP 3 submitted XML traces with structured data
 - ▶ This were translated into an existing format but it was ugly
- ▶ The C track
 - ▶ What are we doing wrong?
- ▶ General Engagement
 - ▶ Feedback: the competition is too regular and too much work
- ▶ The usual suspects
 - ▶ We are working towards a benchmark repository to export the benchmarks used in the competition to the community in general
 - ▶ We want a general specification language but do not know how to proceed here

The Future

- ▶ Currently, the proposal is to not hold the competition in its current form in 2017
- ▶ This gives us time and space to
 - ▶ Consult widely on changes that need to be made
 - ▶ Announce the competition with enough time for teams to prepare (e.g. develop new techniques)
 - ▶ Allow participants to feel that it has been long enough since they last took part
- ▶ In 2017 we want to hold an alternative activity
- ▶ For example, a showcase or non-competitive challenge
- ▶ Any ideas?