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)

<table>
<thead>
<tr>
<th>Tool</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Java track</strong></td>
<td></td>
</tr>
<tr>
<td>LARVA</td>
<td>University of Malta, Malta</td>
</tr>
<tr>
<td>MarQ</td>
<td>University of Manchester, UK</td>
</tr>
<tr>
<td>Mufin</td>
<td>University of Lübeck, Germany</td>
</tr>
<tr>
<td><strong>Offline track</strong></td>
<td></td>
</tr>
<tr>
<td>BeepBeep 3</td>
<td>Université du Québec à Chicoutimi, Canada</td>
</tr>
<tr>
<td>MarQ</td>
<td>University of Manchester, UK</td>
</tr>
</tbody>
</table>
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
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?

<table>
<thead>
<tr>
<th>Team</th>
<th>Bench.</th>
<th>Correct.</th>
<th>Time</th>
<th>Memory</th>
<th>Total</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Offline Track</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BeepBeep 3</td>
<td>6</td>
<td>60</td>
<td>14.42</td>
<td>25.51</td>
<td>97.93</td>
<td>16.32</td>
</tr>
<tr>
<td>MarQ</td>
<td>6</td>
<td>45</td>
<td>45.58</td>
<td>36.49</td>
<td>127.07</td>
<td>21.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Java Track</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larva</td>
<td>9</td>
<td>45</td>
<td>10.88</td>
<td>15.36</td>
<td>71.24</td>
<td>7.92</td>
</tr>
<tr>
<td>MarQ</td>
<td>8</td>
<td>80</td>
<td>20.25</td>
<td>17.30</td>
<td>117.65</td>
<td>14.71</td>
</tr>
<tr>
<td>Mufin</td>
<td>9</td>
<td>90</td>
<td>58.87</td>
<td>57.34</td>
<td>206.21</td>
<td>22.91</td>
</tr>
</tbody>
</table>
Reflection

- Existing trace formats were not sufficient
  - BeepBeep 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?