



TOOLympics II: competitions on formal methods

A special issue for TOOLympics 2019

Dirk Beyer¹ · Marieke Huisman² · Fabrice Kordon³ · Bernhard Steffen⁴

Accepted: 20 May 2021
© The Author(s) 2021

Abstract

This is the second issue in the new “Competitions and Challenges” (CoCha) theme of the International Journal on Software Tools for Technology Transfer. The new theme was established to support competitions and challenges with an appropriate publication venue. The first issue presented the competition on software testing Test-Comp 2019, which was part of the TOOLympics 2019 event. In this second issue for TOOLympics, we present selected competition reports. The TOOLympics event took place as part of the 25-years celebration of the conference TACAS. The goal of the event was to provide an overview of competitions and challenges in the area of formal methods.

Keywords Competitions · Challenges · TOOLympics · Formal verification · Software tools · Technology transfer · Comparative evaluation

1 TOOLympics 2019

The growing importance of computerized systems in our society raises the need for new modeling, analysis, and verification methods, as well as associated automated tools, able to cope with the increasing complexity of such systems. The existence of such tools and techniques allows system designers to evaluate the correctness of such systems, with regards to their requirements, and to ensure their quality, i.e., the absence of bugs in the system.

The various research communities working on the verification of systems (software, hardware, and the underlying involved mechanisms) have considered the problem of developing tools that implement new theoretical results for

decades. Thus, naturally, special emphasis is put on the evaluation of new scientific contributions, by bringing together the involved communities to compare the state of the art, in order to identify progress of and new challenges in the research area. Competitions are a suitable way to achieve that.

Competitions became part of the research community, and it is necessary to continuously discuss and improve the way the competitions are operated, and to share experience between organizers. In 2019, two events were organized, having this goal in mind:

- A Lorentz workshop took place in Leiden and was dedicated to “Advancing Verification Competitions as a Scientific Method” [8]. This event gathered people involved in about 25 competitions who shared their experiences in organizing competitions.
- The TOOLympics 2019 event was part of the celebration of 25 years of TACAS at ETAPS 2019 [2]. This event gathered 16 competitions during the ETAPS week, allowing organizers and participants to get an overview and learn from each other.

✉ Dirk Beyer
dirk.beyer@sosy-lab.org

¹ LMU Munich, Oettingenstr. 67, D-80538 Munich, Germany

² University of Twente, P.O. Box 217, NL-7500, AE Enschede, Netherlands

³ Sorbonne Université, CNRS, LIP6, 4 Place Jussieu, F-75005 Paris, France

⁴ TU Dortmund, Otto-Hahn-Straße 14, D-44227 Dortmund, Germany

The present special issue in the “Competitions and Challenges” theme of STTT is the second issue for TOOLympics. It presents contributions that describe competitions that participated at TOOLympics 2019.

2 This special issue

After several evaluation rounds, five articles were selected for this special issue. The articles illustrate that competitions can be different in their objectives (evaluation of software, evaluation of methodologies, etc.) and in the way they are organized. The competitions are either on-site (the main activities take place at the conference location) or off-site (the main activities take place in research institutions and the results are presented at the conference location). Benchmarking large sets of problem instances usually requires (non-laptop) computing machines and infrastructure to run the benchmarks, such as BenchExec [3], BenchKit [6], or StarExec [11]. The TOOLympics report provides more details [1].

VerifyThis 2019: a program verification competition [4]

VerifyThis is a program-verification competition in which participants prove expressive input/output properties of small programs with complex behavior. The first edition of this event took place aside the 2nd International Conference on Formal Verification of Object-Oriented Software in Torino, Italy in October 2011. In 2019, the 8th edition of this competition was organized.

In contrast to most other competitions, VerifyThis aims at evaluating the capability of a group of people using tools (possibly those they develop) to solve dedicated verification problems on programs. This is an on-site event in which competing groups have to work on a set of problems for a day and provide conclusions in the afternoon. A jury evaluates the methodology as well as the capability of groups in coping in a non-automatic way with the proposed verification problems.

This report analyzes how the participating teams dealt with the proposed verification challenges, discusses what makes a verification challenge more or less suitable for the typical VerifyThis participants, and outlines the difficulties of comparing the work of teams using wildly different verification approaches in a competition focused on the human aspect.

SL-COMP: Competition of solvers for separation 86 logic: report on the third edition [10]

SL-COMP aims at evaluating solvers for separation logic (SL), an established and fairly popular extension of Hoare logic for imperative, heap-manipulating programs. The first

edition of this event took place during the Vienna Summer of Logic in July 2014. In 2019, the 3rd edition of this competition was organized. Its main interest is to evaluate the necessary heuristics required to solve formulas which do not have nice decidability properties.

This competition is an off-site event. Submitted tools are operated against a benchmark during a 3 months average period, and then results are presented during the event it is associated to.

This paper presents the way this edition was operated and how competitors were able to cope with the more than 1000 satisfiability and entailment problems proposed this year.

CoCo 2019: report on the 8th confluence competition [9]

The Confluence Competition evaluates software tools that aim at proving or disproving confluence and related (undecidable) properties of rewrite systems automatically. The first edition of this event took place aside the First International Workshop on Confluence in 2012 (co-located with the 23rd International Conference on Rewriting Techniques and Applications).

This competition is an off-site event. Submitted tools are uploaded into the cross-community competition platform StarExec a few weeks before the events. A test run on a few selected problems for each category of problems allows developers to fix last-minute problems before the steering committee of CoCo operates the competition on StarExec. Results are usually presented during the International Workshop on Confluence, but exceptionally this time they were presented during TOOLympics.

This paper presents how the competition was operated and how tools faced the 100 problems proposed in the 2019 edition.

The RERS challenge: towards controllable and scalable benchmark synthesis [5]

The Rigorous Examination of Reactive Systems challenge (RERS) is a verification challenge that focuses on temporal and reachability properties of reactive systems. RERS was founded at ISoLA 2010, and since its first instance in 2012, it has been a yearly event. The results have been presented during the following conferences: ISoLA (2012, 2014, 2016, and 2018), ASE (2013), RV (2015), ISSTA/SPIN (2017), and TOOLympics (2019).

RERS is an off-site event. Every year, problems in different categories are proposed and participants have to cope with such problems on their own, using the methodology and toolset of their choice (usually, they use the tool they develop). Then, a report is presented and evaluated by the organizing committee. The presentation of results, as well as

discussions between participants, take place at the associated event.

This paper deals with the main challenge of organizing RERS: the definition of problems for which properties are known in advance. To do so, the organizing team has developed dedicated techniques to synthesize hard benchmarks. In the present article, the most recent developments are reported.

Study of the efficiency of model-checking techniques using results of the MCC from 2015 to 2019 [7]

The Model-Checking Contest is a tool competition dedicated to model-checking tools. It focuses on model checking of asynchronous systems (in contrast to the Hardware Model-Checking Competition, which is dedicated to synchronous hardware systems). The first edition took place at the Petri-Net Conference in 2011 and it is a yearly event associated to this conference. In 2019, it was exceptionally participating in TOOLympics.

MCC is an off-site event. Every year, participants upload their tool in a virtual machine that is operated by the organizers using a dedicated execution environment, BenchKit. Tools participate in one or more of the following categories: State-space generation, evaluation of global properties, computation of upper bounds on specifications, evaluation of reachability formulas, evaluation of CTL formulas, evaluation of LTL formulas. Every year, new models are added to the benchmark set (more than 1000 in 2019) and new (more complex) formulas are generated.

This paper briefly presents the contest itself but focuses on a pluriannual analysis of MCC results, using the results of the five editions from 2015 to 2019. The objective is to sketch some trends of the evolution of model-checking techniques (or combination of techniques) as they are operated by the participating tools, and in particular by the tools which were on the podium (gold, silver, and bronze medals). The overall benefits of this contest to the targeted communities were also investigated.

Acknowledgements We are grateful to all the authors for their contributions. We also thank the attendees of the workshop at Leiden [8] for their discussion contributions and the reviewers who helped us evaluating and selecting the papers for this special issue (both groups listed together in alphabetic order): Armin Biere, Nikolaj Bjorner, Doina Bucur, Tom van Dijk, Carlo A. Furia, Hubert Garavel, Jan Friso Groote, Arnd Hartmanns, Keijo Heljanko, Marijn Heule, Lom Messan Hillah, Holger Hoos, Francis Hulin-Hubard, Sebastian Junges, Alfons Laarman, Jeroen Meijer, Aart Middeldorp, Rosemary Monahan, Aina Niemetz, Luca Pulina, Martina Seidl, Mihaela Sighireanu, Kate Smith-Miles, Mattias Ulbrich, Jaco Van de Pol, Tjark Weber, Thomas Wies, Mathias Preiner, Tim Quatmann, and Hans Zantema.

Funding Open Access was funded by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Bartocci, E., Beyer, D., Black, P.E., Fedyukovich, G., Garavel, H., Hartmanns, A., Huisman, M., Kordon, F., Nagele, J., Sighireanu, M., Steffen, B., Suda, M., Sutcliffe, G., Weber, T., Yamada, A.: TOOLympics 2019: An overview of competitions in formal methods. In: Proc. TACAS (3), LNCS 11429, pp. 3–24. Springer (2019). https://doi.org/10.1007/978-3-030-17502-3_1
2. Beyer, D., Huisman, M., Kordon, F., Steffen, B. (eds.): Proc. TACAS (3): 25 Years of TACAS: TOOLympics, LNCS 11429. Springer (2019). <https://doi.org/10.1007/978-3-030-17502-3>
3. Beyer, D., Löwe, S., Wendler, P.: Reliable benchmarking: requirements and solutions. *Int. J. Softw. Tools Technol. Transf.* **21**(1), 1–29 (2019). <https://doi.org/10.1007/s10009-017-0469-y>
4. Dross, C., Furia, C.A., Huisman, M., Monahan, R., Müller, P.: VerifyThis 2019: A program verification competition. *Int. J. Softw. Tools Technol. Transf.* (2021). <https://doi.org/10.1007/s10009-021-00619-x>
5. Howar, F., Jasper, M., Mues, M., Schmidt, D., Steffen, B.: The RERS challenge: Towards controllable and scalable benchmark synthesis. *Int. J. Softw. Tools Technol. Transf.* (2021). <https://doi.org/10.1007/s10009-021-00617-z>
6. Kordon, F., Hulin-Hubard, F.: BENCHKIT: A tool for massive concurrent benchmarking. In: Proc. ACS'D, pp. 159–165. IEEE (2014). <https://doi.org/10.1109/ACSD.2014.12>
7. Kordon, F., Jézéquel, L., Paviot-Adet, E., Hulin-Hubard, F., Hillah, L.M.: Study of the efficiency of model checking techniques using results of the MCC from 2015 to 2019. *Int. J. Softw. Tools Technol. Transf.* (2021). <https://doi.org/10.1007/s10009-021-00615-1>
8. Kordon, F., Seidl, M., van de Pol, J.: Lorentz workshop on advancing verification competitions as a scientific method. Tech. rep., Lorentz Center (2019)
9. Middeldorp, A., Nagele, J., Shintani, K.: CoCo 2019: Report on the 8th confluence competition. *Int. J. Softw. Tools Technol. Transf.* (2021). <https://doi.org/10.1007/s10009-021-00620-4>
10. Sighireanu, M.: SL-COMP: Competition of solvers for separation logic: Report on the third edition. *Int. J. Softw. Tools Technol. Transf.* (2021). <https://doi.org/10.1007/s10009-021-00628-w>
11. Stump, A., Sutcliffe, G., Tinelli, C.: StarExec: A cross-community infrastructure for logic solving. In: Proc. IJCAR, LNCS 8562, pp. 367–373. Springer (2014). https://doi.org/10.1007/978-3-319-08587-6_28

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.