In Villetaneuse, April 2, 2024

Review of doctoral thesis “Boolean Satisfiability Modulo Differential Equation Simulations” by Tomáš Kolárik

The thesis by Tomáš Kolárik, entitled “Boolean Satisfiability Modulo Differential Equation Simulations” and submitted to the Faculty of Information Technology, Czech Technical University in Prague, is made of 10 chapters and 167 pages.

Up-to-dateness of the dissertation

The thesis by Tomáš Kolárik is concerned with the formal verification of critical cyber-physical systems in connection to numerical simulations. Such verification is necessary to avoid unwanted behaviors (bugs).

SAT (Boolean satisfiability) modulo ODE (ordinary differential equations) solvers can perform automated verification, but the size of the considered benchmarks is often outperformed by an order of magnitude by the needs in the industry. Another issue is that SAT modulo ODE approaches have a mathematical semantics that differs from the semantics used in simulation and testing, e.g., in the industry.

Therefore, the main contributions of the Ph.D. candidate rely in the proposition of a new approach integrating simulation-based ODEs with SMT solving. These contributions are very timely, and fit into an ongoing trend to use slightly less formal methods within the area of formal verification.

One of the witnesses for the up-to-dateness of the dissertation is the publication of several papers in good to excellent venues, including one of them in the latest edition of the prestigious Formal Methods conference (in 2023).

Publications The list of peer-reviewed publications in the framework of this manuscript is as follows:

1. Tomás Kolárik, Stefan Ratschan: SAT Modulo Differential Equation Simulations. TAP@STAF 2020: 80-99 (described in Chapters 4 and 6)
2. Tomás Kolárik, Stefan Ratschan: Railway Scheduling Using Boolean Satisfiability Modulo Simulations. FM 2023: 56-73 (described in Chapter 8)
3. Tomás Kolárik, Stefan Ratschan, Pavel Surynek: Multi-Agent Path Finding with Continuous Time Using SAT Modulo Linear Real Arithmetic. To appear in the International Conference on Agents and Artificial Intelligence (2024) (described in Chapter 9)

Formal structure and organization of the dissertation

Main contributions The main contributions of the dissertation are:

1. the proposition of a new semantic concept of "SAT modulo ODE", where solving ODEs is based on *numerical simulations*, similar to the ones used in simulation tools, as opposed to mathematical techniques;
2. the design of an algorithmic approach for solving SAT modulo ODE, implemented into a novel toolkit, and compared on several non-trivial benchmarks against two existing tools;
3. a benchmark problem from the railway scheduling community in which the proposed approach makes a difference in efficiency, notably due to the fact that continuous information such as deceleration and velocity are modeled; and
4. a solution to the multi-agent path-finding with continuous time, directly translated into an SMT problem using quantifier-free linear real arithmetic, and in which collision and avoidance of agents is handled based on floating-point simulations.

Structure and organization The first chapter introduces the overall motivation and outlines the contributions before briefly reviewing general related works.

The second chapter introduces the necessary preliminary notions, notably ODEs (with their so-called simulation semantics), hybrid automata, SAT and Satisfiability modulo theory (SMT), and bounded model checking.

The third chapter reviews existing approaches.

The fourth chapter, which is also the first contributing chapter, defines the problems targeted by the thesis, notably satisfiability modulo ODEs. Functions (solutions of ODEs) are handled as first-order objects. The theory of SAT modulo ODEs can be parameterized by the actual domain of floating-point variables, but also by the method solving ODEs. Finally, black-box simulations are made possible within the SAT framework, by using functions the behavior of which is (partially) unspecified, such as Simulink models or, even more interesting, neural networks.

The fifth chapter reviews state-of-the-art tools to solve numeric ODEs, SAT problems, SMT problems, to perform reachability analyses of hybrid systems, and simulation (such as Simulink).

The sixth chapter, which is also the second contributing chapter, combines simulation-based approaches for ODEs with SMT solving techniques. The method might answer "`unknown`" as result, but a syntactical characterization of the inputs for which a real solution (i.e., "`sat`" or "`unsat`") can be provided has been done. The chapter describes in details the implementation of the SMT solver together with the ODE numerical solver, and their integration.

The seventh chapter, which is also the third contributing chapter, studies the proposed implementation against three benchmarks, two of them ("glucose control" and "hormone therapy") being taken from the database of the dReal3 tool, while the two others are "railway scheduling" (detailed in Chapter 8) and "racing car". A discussion compares the proposed approach with dReal3: the proposed approach overall largely outperforms dReal3.

The eighth chapter, which is also the fourth contributing chapter, focuses on the railway scheduling benchmark, briefly studied in the former chapter. This benchmark is an extension of an existing model, and is claimed to exhibit both non-trivial discrete and continuous behaviors, therefore differing from other existing benchmarks. An extension compared to the existing model is that the segments visited by the trains have speed limits, and therefore the trains must accelerate or decelerate along their journey, making the benchmark continuous in nature. The comparison with a concurrent tool *railperfcheck* has both strong and weak points, suggesting improvements.

The ninth chapter, which is also the fifth contributing chapter, aims at solving a continuous-time version of the multi-agent path-finding (MAPF) problem using a translation to SMT. Collision avoidance and detection of agents is handled using floating-point simulations. An implementation has been made by the candidate, on top of the MathSAT5 SMT solver; a comparison is then made using 3 classes of benchmarks against two concurrent tools (CCBS and SMT-CCBS). The tool of the candidate performs well against these competitor tools, notably on a `bottleneck` benchmark designed on purpose.

The tenth chapter summarizes the contributions of the manuscript and sketches very briefly some possible future works.

Evaluation of the results and contributions of the dissertation

Chapters 2 and 3 are illustrated with a number of pedagogical examples, making the reading smooth.

Chapter 4 made the choice to follow a slightly informal ("semi-formal") presentation, rather than the logical theory as in the underlying published article. This makes the manuscript pedagogical (and still formal enough in my opinion), while leaving the original definitions accessible in the associated publication.

While the main approach's drawback is its lack of formal guarantees (due to possible numerical errors), its main advantage is that it is inline with simulation-based tools used in the industry—such as Simulink. In this way, for sufficiently complex systems, it can be seen as closer to the real world than formal mathematical approaches. Also, the fact that it is much more efficient than formal mathematical approaches (that can be even undecidable) is an excellent point towards providing at least some guarantees during the design phase.

The candidate designed both his own SMT solver (with MiniSat2 as the underlying SAT solver), and his ODE simulation engine, implemented in C++ and available Gitlab in an open-source manner. This toolkit, named "UN/SOT", is not just a small prototype, but comes with a (documented) core input language, derived from SMT-LIB, and a preprocessing language enriching the syntax using macros à la C.

The candidate chose to develop his own SMT solver, rather than reusing an existing one, typically with floating-point arithmetic. This choice is carefully motivated, and the reasons seem to be valid to me.

The experimental results in Chapter 7 show the excellent results of the proposed approach on the two considered benchmarks: the proposed approach largely outperforms the existing tool dReal3, sometimes by an order of magnitude. While the proposed approach performs much better than dReal3, the final discussion is interesting, as it also highlights the points where the proposed approach can still be improved further.

Again, the results of Chapter 7 are available on Gitlab, which is an excellent point towards reproducibility and reusability.

The implementation of Chapter 9 is also available publicly.

Finally, I highly appreciated that, in several chapters, limitations are pointed out, showing that the candidate has a good hindsight on his works and this research area. This in turn makes the future works in the final chapter slightly disappointing, as I would have expected a little more precise directions for future research.

Remarks, objections, notes, and questions for the defense

Remarks, objections and notes Page 20, when mentioning Hybrid/timed automata and "symbolic model checking", it must be noted that this symbolic part encodes the continuous part of the state space (i.e., clocks or more permissive continuous variables), while most other "symbolic model checking" algorithms encode symbolically the *discrete* part of the state space (i.e., locations).

In Section 5.4, in the list of tools able to handle linear hybrid automata, PHAVer, and its recent optimized version PHAVerLITE, are missing and could be added, even if absent from the mentioned competition. I also recommend to add to the list the parametric timed model checkers IMITATOR (for timed automata) and Roméo (for time Petri nets), able to manage (very) simple flows of the form $\dot{x} = c$, with c constant (called "multi-rate automata").

The results of Chapter 7 are available on Gitlab, which is an excellent point towards reproducibility and reusability; I would however suggest the candidate to upload them on a (very) long-term archiving repository, such as Zenodo.

Possible questions for the defense

1. The bouncing ball example used as a running example is most probably simple enough to be handled using the classical mathematical semantics—as opposed to the "simulation semantics" introduced in the manuscript. For example, hybrid automata (as noted in Chapter 4) are able to model and reason about this case study. It would be nice if the Ph.D. candidate could comment on this point during the defense.

2. As said above, the candidate chose to develop his own SMT solver as opposed to reusing one. Despite the (nice) motivation, it is unfortunate as the candidate is cutting himself from highly optimized SMT solvers; would it be really impossible to use an existing SMT solver and, if not, what would be the cost for doing so? Please discuss this point.

3. Is it possible to design a simple benchmark for which UN/SOT would return "unknown"?

4. Would it be possible to discretize the train benchmark, e.g., by encoding an acceleration by several iteratively growing (constant) speeds? How imprecise would be such an approach?

5. In Chapter 8, why not comparing the approach with dReal3?

6. Could IMITATOR, the model checker for parametric timed automata extended with simple flows, be used to verify the benchmark in Chapter 8?

Overall evaluation of the dissertation

The thesis is well-written, easy to follow, and is illustrated with numerous examples and interesting benchmarks.

I am quite impressed by the contributions that go from the theoretical level all the way until the implementation, through clearly explained algorithms. The benchmarks are interesting yet far from trivial, notably the railway scheduling one.

The good to excellent venues in which the candidate published his results show the very good recognition by the formal methods community of these results. The nice presentation also shows the candidate good ability to present technical results in a very digest manner.

In spite of the minor objections and remarks pointed out earlier in this report, the thesis manuscript contains new and original results, of high importance as seen from the list of publications in good or even excellent venues. The Ph.D. candidate proved his ability to conduct research and achieve scientific results. For all these reasons, and in accordance with par. 47, letter (4) of the Law Nr. 111/1998 (The Higher Education Act), I **do** recommend the thesis of Tomáš Kolárik for the presentation and defense with the aim of receiving the Ph.D. degree.

Yours faithfully,

É.André

Étienne André

Full professor