A REVIEW OF PH.D. THESIS

Title: Harnessing Forest Automata for Verification of Heap Manipulating Programs

Author: Ing. Jiří Šimáček

Reviewer: prof. RNDr. Mojmír Křetínský, CSc.

Automatic verification of current software systems often needs to model them as infinite-state systems, i.e. systems with an evolving structure and/or operating on unbounded data types. The submitted thesis deals with the problem of verification of infinite-state systems in the form of programs with dynamic memory allocation, manipulating complex dynamic linked data structures on the heap. To address this problem several quite different techniques have been suggested, however, none of them provides a satisfactory solution. Hence this area is a subject of current intensive study, and the topic of the thesis is very up-to date.

One of promising ways for dealing with infinite-state verification is to use symbolic verification in which infinite sets of reachable configurations are represented finitely using a suitable formalism. This approach is taken in the thesis , which in particular builds on tree automata and their generalization, so called forest automata introduced by the author. The approach is also inspired by methods using separation logic as the heap graphs are represented via their (canonical) tree decompositions.

Content. The thesis consists of eight chapters. First, I briefly comment on their respective content.

Chapter 1 introduces a reader to the subjected area and presents relevant related approaches and related work in a concise way. Further, goals of the thesis are stated and the contributions that have been achieved within the particular areas are mentioned. Finally, the plan of the thesis is outlined.

Chapter 2 recalls some basic notions, namely labelled transition systems, tree automata, and simulations over tree automata. A description of the regular tree model checking method presented here is taken from the thesis of Lukáš Holík with his permission.

Chapter 3 introduces the notion of a forest automaton (FA) which serves as a theoretical basis for the presented verification technique for programs manipulating dynamically linked data structures. It employs an idea of a canonical decomposition of a heap into tree components—in fact, an FA is a tuple of tree automata. The notion of an FA is generalized to so called hierarchical FA which increases the expressive power of FA but some makes some operations, e.g., language inclusion checking, much more complicated.

Chapter 4 provides a thorough description of the verification procedure based on FA as a basic tool. It is focus on safety properties of programs manipulating various forms of lists and trees. This chapter also includes an experimental evaluation of the proposed method done by means of the prototype tool Forester based on this procedure.

In Chapter 5 it is pointed out that the suggested verification procedure heavily depends on efficiency of some automata operations. It turns out that a crucial point is to compute simulations on NTA. Hence NTA are transformed onto LTSs and the author suggests and describes in detail an optimized algorithm for performing this task on LTS. Its experimental evaluation is provided as well.

Chapter 6 examines several variants of so called top-down algorithms for language inclusion checking. A basic version is introduced together with its several extension which can greatly improve its performance. Again, an experimental evaluation is presented.

In Chapter 7, a nondeterministic tree automata library (VATA) based on the proposed algorithms for language inclusion checking and simulation computation is described. Various lower-level optimizations of the implementation of the algorithms introduced in Chapter 5 and Chapter 6. are discussed as well.

Finally, Chapter 8 concludes the thesis. It summarized the results archived, outlines further directions of research in the area, and lists publications and tools related to the thesis.

Evaluation. The submitted thesis brings a number of new results. These results are quite involved and their relevance and quality have been appreciated by community. This can be demonstrated by the fact that most of the results have been published on very well know and recognized international conferences and journals. It deserves to mention conferences CAV 2011, ATVA 2011 and TACAS 2012, and the journal Formal Methods in System Design. From my point of view most valuable are the results presented in Chapters 3 and 4 on new verification techniques, an improved top-down algorithm for language inclusion checking, and the general purpose tree automata library VATA.

The text of the thesis is well organized and properly structured. A reader surely appreciates some examples of techniques, e.g. on heap decomposition, and informal presentation of concepts introduced, e.g. forest automata, cut point types, etc. Anyway, in my opinion, some other concepts would deserve to be exemplified as well, for example, hierarchical FA.

I have not found any serious mistakes neither factual nor formal ones. To mention some small drawbacks, I would point to the definitions of downward and upward simulations where all the quantifiers are missed (the fact that they can be deduced from what follows is not a reason why to do so). From a formal point of view, I would like to see all the propositions and theorems equipped with their explicit proofs. For example Propositions 1 to 4 on pages 19, 23, 27, and 31, respectively as well as Theorem 4 on page 98 are missing proofs. Please note that, contrary to conference papers, there is no page limits for a thesis.

Last but not least, the author deserves credits for making experimental evaluations of the suggested methods and algorithms. I also appreciate overviews of related work and further research at the ends of nearly all chapters.

The above mentioned results and publications together with the other publications of the author of the thesis clearly demonstrate his potential and ability to take an active part in current research in computer science as well as to publish the achieved results in an adequate way.

Conclusion. To sum up, the submitted thesis elaborates on a topical area of current research in informatics, brings new involved results which are presented in a way which is standard within this area, most of the results have been already published in recognized conferences and journals. In my judgement, the submitted work fulfills all the commonly recognized criteria to be accepted as a doctoral thesis in informatics. Hence, I fully recommend to accepted this thesis as a doctoral ones.

Brno, 4. 10. 2012 Mojmír Křetínský