Technische Universität München

Technische Universität München · 80290 München

Sylva Sadovska
Faculty of information technology
Brno University of Technology
Božeěchova 2
61266 Brno
Czech Republic

Fakultät für Informatik
Institut für Informatik
Lehrstuhl VII

Prof. Dr.
Javier Esparza

Boltzmannstraße 3
85748 Garching b. M"unchen
Germany

Tel +49.89.289.17204
Fax +49.89.289.17207

esparza@in.tum.de
www7.in.tum.de

Garching, 8. Juni 2015

Report on the PhD thesis of Ondřej Lengál

Topic and Summary

Automatic program verification is a mature area of computer science which draws its toolbox from fundamental research on logic and automata theory. Indeed, some of the most important advances in automatic verification have been obtained by adapting and applying logic and automata technology to verification problems. Classical examples are the use of Binary Decision Diagrams to compactly represent sets of program states, or the use of SAT-solvers to find inputs that execute a certain program path.

The topic of the thesis of Ondřej Lengál is the design, optimization, and implementation of logic and automata technology for the automatic verifciation of programs that manipulate a heap. This is a very active and competitive are in formal verification, which has produced many contributions to top conferences and journals in automatic verification during the last years. It is also a very challenging topic, due to the difficulty of finding good data structures and algorithms for representing and manipulating infinite sets of heap configurations.

Content

The thesis is divided into three parts. Part I shows how to map verification problems for programs acting on a heap to abstract logical and automata-theoretic problems. Part II provides new algorithms for these abstract problems. Finally, Part III

Technische Universität München

Fakultät für Informatik
Institut für Informatik
Lehrstuhl VII

designs, implements, and evaluates algorithmic techniques and symbolic data structures for the efficient execution of the algorithms.

Part I proposes to use arrays of tree automata to encode infinite families of labeled graphs, each representing a heap configuration. A graph is encoded as an array of trees, together with information showing how to glue the trees together to yield the graph. While it is relatively easy to encode families whose members can be split into a fixed number of trees, many families, like the graphs underlying double linked lists, do not satisfy this condition. This problem is adressed by means of a clever hierarchical construction, in which alphabet letters of a tree automaton act as "calls" to other tree automata. A first automaton can then, for instance, generate an arbitrarily long path with edges labeled by calls to a second automaton, which transforms each edge into a small circuit, yielding a doubly linked list. The thesis presents a fixed point algorithm that automatically generates such a hierarchical array. In a final chapter, the approach is extended beyond pure shape analysis by considering the data stored in heap cells.

Part II presents novel decision algorithms for two different logics. The first one is the fragment of separation logic consisting of existentially quantified conjunctions of inductively defined spacial atoms. The fragment allows one to specify structures like singly and doubly linked lists, or skip lists. The thesis presents a two-step decision procedure that first conducts some logical preprocessing, and then applies the graph-to-tree decomposition approach developd in Part I. The second logic is WS1S, weak monadic second-order logic of one successor. For this logic there exists a classical automata theoretic approach that, given a formula with free variables, constructs a finite automaton recognizing the models of the formula; validity of closed formulas without free variables is obtained as a subproduct. The thesis presents an optimized procedure for closed formulas when one is only interested in the validity of the formula, and not in the set of models of any subformula. The thesis transforms this case into the problem of deciding whether an automaton accepts the empty string or not, and develops an interesting specialized algorithm.

Part III designs and implements efficient algorithms for the manipulation of tree automata. The first contribution is an elegant algorithm for the inclusion problem that processes trees in a top-down way. While the algorithm is inspired in previous work by Hosoya, the thesis introduces important optimizations by applying antichain and simulation-based subsumptions. The second contribution presents algorithms for tree automata over very large alphabets, extending MONA's word automata whith transitions labeled by BDDs. The final chapter of the thesis reports on VATA, a library for handling tree automata, which implements all the algorithms.

Technische Universität München

Fakultät für Informatik
Institut für Informatik
Lehrstuhl VII

Evaluation

The thesis of Ondřej Lengál is remarkable for its large scope. Most theses on automatic verification concentrate on one of the three areas corresponding to the three parts of the thesis: mapping of verification problems to logical or automata-theoretic problems, design of decision procedures for these problems, and design and implementation of efficient data structures and algorithms for the decision procedures. A few theses contribute to two of the areas. This thesis, however, presents relevant contributions to all three. This requires a wide range of skills: deep knowledge of logic and automata theory, mathematical maturity, and excellent algorithmic, programming, and tool building abilities.

The author also exhibits very good expository skills. The chapters are well documented, with good examples and well presented proofs.

The results of the thesis have been published in a contribution to Acta Informatica, a very good journal, and in six contributions to international conferences, including one paper in CAV and two in TACAS, both of them among the top conferences on automatic verification. Even if we take into account that the papers have an average of about 4 authors, this is an excellent record.

I recommend the Faculty of Information Technology to accept the thesis without any hesitation.

Prof. Dr. Javier Esparza