CHARLES UNIVERSITY PRAGUE

faculty of mathematics and physics

The Board of Doctoral Studies
Faculty of Mathematics and Physics
Charles University Prague
Ke Karlovu 3
12116 Prague 2

Prague, July 9, 2014

Re: David Hauzar, Doctoral Thesis – Advisor's Reference

The submitted PhD thesis concentrates on the potential of static analysis of languages with dynamic features (DLs for short). The thesis aims at three main goals, namely (1) to design a heap analysis precisely modeling built-in data structures that are common in DLs, (2) to generically define interplay of heap and value analyses for DLs, and (3) to design a framework that allows defining precise static analyses for DLs independently of computing control flow and accessing data structures.

The thesis starts with an overview of existing techniques for static analysis of DLs and with an overview of existing techniques for combing heap and value analyses (Chapter 2). Based on the overview, the most important obstacles for static analysis of DLs are identified and goals of the thesis are articulated (Chapter 3). In Chapter 4, a novel heap analysis modeling associative arrays and prototype objects is described. The novel contribution of the technique is that it takes into account the semantics of non-decomposable multidimensional updates to associative arrays and prototype objects, and it models such updates soundly and precisely even if statically-unknown data from the input are used to specify targets of the updates. In Chapter 5, the framework for static analysis of DLs is described. The novelty of the framework is that it defines how heap and value analyses interplay with each other to allow in value analysis automatically access variables, array indices, and object properties and to allow in heap analysis automatically use theirs values. Next, the thesis shows how the framework coordinates automatic resolution of dynamic features. As a result, the second novel contribution of the framework is that it makes possible to define static analysis independetly of resolving dynamic features. In Chapter 6, an implementation of the framework for analysis of PHP programming language and a tool for analysis of PHP integrated to Eclipse IDE is presented. The main contribution here is that the framework includes implementations of all PHP native operators, native library functions, and implicit conversions, and it includes implementations of two heap analyses. Chapter 7 presents evaluation of the proposed techniques. It shows the scalability of the heap analysis and presents two case studies based on real PHP applications. Importantly, the framework found one serious, previously unknown, security flaw and two previously unknown less serious security flaws. Moreover, the comparison with the state-of-the-art tools shows that the framework is superior both in error coverage and false positive rate.

This work has been partially published at the COMPSAC 2012 workshop (proceedings by IEEE), ESSS 2014, workshop (proceedings by EPTCS), SEFM 2014 (proceedings by Springer).
With respect to all these facts, I recommend the thesis for defense and to grant a PhD degree to David Hauzar.

František Plášil
Advisor