Obfuscation by Partial Evaluation of Distorted Interpreters
(Invited Talk)

Neil D. Jones

Computer Science Department
University of Copenhagen
2100 Copenhagen, Denmark
e-mail: neil@diku.dk

Abstract. How to construct a general *program obfuscator*? We present a novel approach to automatically generating obfuscated code P' from any program P whose source code is given. Start with a (program-executing) interpreter `interp` for the language in which P is written. Then "distort" `interp` so it is still correct, but its specialization P' w.r.t. P is transformed code that is equivalent to the original program, but harder to understand or analyze. Potency of the obfuscator is proved with respect to a general model of the attacker, modeled as an approximate (abstract) interpreter. A systematic approach to distortion is to make program P obscure by transforming it to P' on which (abstract) interpretation is *incomplete*. Interpreter distortion can be done by making residual in the specialization process sufficiently many interpreter operations to defeat an attacker in extracting sensible information from transformed code. Our method is applied to: code flattening, data-type obfuscation, and opaque predicate insertion. The technique is language independent and can be exploited for designing obfuscating compilers.

Keywords: Obfuscation, semantics, partial evaluation, program transformation, program interpretation, abstract interpretation.

The talk is based on joint work with Roberto Giacobazzi and Isabella Mastroeni [1].

References

1. Roberto Giacobazzi, Neil D. Jones, and Isabella Mastroeni. Obfuscation by partial evaluation of distorted interpreters. In *Proceedings of the ACM SIGPLAN 2012 workshop on Partial evaluation and program manipulation*, PEPM '12, pages 63–72, New York, NY, USA, 2012. ACM.