Cross-language program analysis for dynamic web applications

Hung Viet Nguyen

February 2016

Web applications have become one of the most important and prevalent types of software. In modern web applications, the display of any web page is usually an interplay of multiple languages and involves code execution at different stages (the server side, the database side, and the client side). These characteristics make it hard to write and maintain web applications. Much of the existing research and tool support often deals with one single language and therefore is still limited in addressing those challenges. To fill in this gap, this dissertation is aimed at developing *an infrastructure for cross-language and cross-stage program analysis for dynamic web applications* to support creating reliable and robust web applications with higher quality and lower costs. To reach that goal, we have developed the following research components. First, to understand the client-side code that is embedded in the server-side code, we develop an *output-oriented symbolic execution engine* that approximates all possible outputs of a server-side program. Second, we use variability-aware parsing, a technique recently developed for parsing conditional code in software product lines, to parse those outputs into a compact tree representation (called *VarDOM*) that represents all possible DOM variants of a web application. Third, we leverage the VarDOM to extract semantic information from the server-side code. Specifically, we develop novel concepts, techniques, and tools (1) to build *call graphs for embedded client code* in different languages, (2) to compute *cross-language program slices*, and (3) to compute a novel test coverage criterion called *output coverage* that aids testers in creating effective test suites for detecting output-related bugs. The results have been demonstrated in a wide range of applications for web programs such as *IDE services*, *fault localization*, *bug detection*, and *testing*.