Experiments with String Analysis

Kostyantyn Vorobyov, Yang Zhao,
Raghavendra Ramesh and Padmanabhan Krishnan

October, 2019

Static program analysis tools need to use *string analysis* to detect string manipulation errors that lead to security vulnerabilities such as SQL injections or cross-site scripting. *String analysis* is a program analysis technique that computes concrete values occurring as a result of string expressions that allows consumers of the analysis to reason about string manipulation performed and detect errors.

An important issue in string analysis is the tradeoff between performance and scalability. String analysers aiming to compute regular expressions as deterministic finite automata lead to more precise results but fail to scale to large programs. Lightweight techniques, such as constant propagation, can handle large programs but often are less precise.

In this talk we present results of an empirical evaluation comparing scalability and precision of different string analysis techniques: JSA[1] (an approach using regular approximation to compute deterministic finite automata) and our internal string analyser. This experiment uses 303 test programs of the JSA test suite and 17 programs selected from DaCapo datasets.

Our findings indicate that for a problem of reflection analysis in DaCapo programs our tool was considerably faster than JSA and computed values for all 17 programs in slightly over 2 seconds, whereas JSA could analyse only 5 programs in over 17 minutes (analysis of the rest of the programs failed). Precision of JSA and our tool in the 5 programs that JSA could analyse was similar. Further experiments with our tool using different configurations indicate that it scales to large programs even in the extreme case where all possible string values need to be computed. When computing values in JSA unit tests our tool was less precise than JSA because of several presently unsupported features such as field-sensitivity and alias analysis.

References

[1] Aske Simon Christensen, Anders Møller, and Michael I. Schwartzbach. Precise analysis of string expressions. In *Proceedings of the International Symposium on Static Analysis*, pages 1–18. Springer, June 2003.