発表概要

サンプリング適用による動的データ構造推定の精度・効率の評価

小泉 雄太[1,a)]　荒堀 喜貴[1]　権藤 克彦[1]

2018年11月1日発表

Howard に代表されるメモリアクセスパターンに基づくデータ構造推定は正確であり，バイナリレベルの脆弱性解析等に有効な技術である．しかし，従来のデータ構造推定手法は全メモリアクセスの監視を行うため実行オーバヘッドが大きく，手法の適用ケースを制限している．そこで本発表では，アクセスパターンに基づくデータ構造推定に対し，監視対象のメモリアクセスをサンプリングする手法を提案し，実行オーバヘッドの削減を試みる．我々は動的バイナリ解析器（Intel Pin）を用いて監視対象をヒープオブジェクトに限定した Howard 実装を再現し，この再現実装に対し Adaptive Bursty Tracing によるサンプリングを適用する．ここで，サンプリングによって欠落したアクセス情報を，同一 calling context でまとめられたグループ単位で集約することによりデータ構造推定精度を保つ工夫を行う．さらに，集約した情報を比較することで calling context 間における関係性の類推を試みる．小規模ベンチマークを用いた実験を通して，提案するサンプリングの有無によるデータ構造推定精度の変化と実行オーバヘッドの変化を評価した結果を報告する．

Presentation Abstract

Measuring the Accuracy and Efficiency of Sampling-based Inference of Dynamic Data Structures

Yuta Koizumi[1,a)]　Yoshitaka Arahori[1]　Katsuhiko Gondow[1]

Presented: November 1, 2018

Techniques, such as Howard, for inferring dynamic data structures based on memory access patterns are precise, and have been known to be effective for analyzing software vulnerabilities at binary level. However, existing techniques for dynamic data structure inference monitor all memory accesses by the target program and thus incur high runtime overheads, which limits their wide adoption in many practical cases. In this talk, we present the application of a memory-access sampling method to the problem of dynamic data structure inference in order to reduce runtime overheads. We use a dynamic binary analysis tool (Intel Pin) to implement a modified Howard, which focuses memory-access monitoring on heap and samples the heap monitoring. Our initial experiments on small benchmarks report the effects of the application of sampling on the accuracy and efficiency of dynamic data structure inference.

[1] 東京工業大学
Tokyo Institute of Technology, Meguro, Tokyo 152–8550, Japan
[a)] zikuomi@sde.cs.titech.ac.jp