allvm - Binary Decompilation

Sandeep Dasgupta
University of Illinois Urbana Champaign

October 10, 2017

Preliminary Work

- Improving the IR extracted by McSema.
  - Stack frame deconstruction
  - Stack variable promotion
  - Type extraction

    Tools Developed: Augment IDA Type &
    Dwarf Type Reader

- Improving McSema Applicability.
  - Vector instruction support
  - Translating unknown instructions into inline assembly

Current Work

Problem: Semantic translation from extracted CFG (from binary) to LLVM IR can be buggy & difficult to extend.

Approach:

- Learn the semantics rules automatically using Strata.
- Use K framework to define the learned semantics.
  - Help in validating the translation (decompilation) of binary to LLVM IR.[1]

[1] Missing pieces like a) Operational semantics of LLVM IR in K & b) Language independent program equivalence checker, KEQ, are already in progress by other members.