allvm - Binary Decompilation

Sandeep Dasgupta
University of Illinois Urbana Champaign

October 9, 2017

Preliminary Work

- Improving the IR extracted by McSema.
  - Stack frame deconstruction
  - Stack Variable promotion
  - Type extraction
    - Tools 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 validation the translation (decompilation) of binary to LLVM IR.[1]

[1] Missing pieces like 1. Operational semantics of LLVM IR in K & 2. Language independent program equivalence checker, KEQ, are already in progress by other team.