allvm - Binary Decompilation

Sandeep Dasgupta
University of Illinois Urbana Champaign

March 25, 2016

Outline

Research Goal

- Research Goal
  - Obtain “richer” LLVM IR than native machine code.
- Motivation
  - Absence of source-code
  - What-you-see-is-not-what-you-execute
  - End-user security enforcement
  - Platform aware optimizations

Outline

Possible Directions

- Decompile `Machine Code` $\longrightarrow$ `LLVM IR`
  - Challenge: Quality
    - Reconstructing code and control flow - Much researched.
    - Variable recovery
    - Function & ABI rules recovery
- Add annotations on `Machine Code` and then Decompile `Machine Code` $\longrightarrow$ `LLVM IR`
  - Challenge: Annotations must be “minimal” & sufficient.
- Ship `LLVM IR`
  - Challenges: Adoption, risks to itellectual property

Outline

Our Approach

- Studied state of the art strategies for Variable Recovery
- Function & ABI rules recovery

Using Mcsema

mcsema: current support

mcsema: limitations

mcsema: demo