A Call for a Public Bug and Tool Registry

Jeff Foster
University of Maryland, College Park

Measuring Results

- # warnings generated is easy, but not that useful
  - Roughly corresponds to effort to review them
- False positives are also easy...maybe...
  - How does a warning correspond to a "bug"?
    - One "bug" could yield multiple warnings or vice-versa
    - Which bugs are important? Null pointer errors? Race conds?
- False negatives are really hard
  - Other tools not available, report different errors, have different design tradeoffs
  - Bug databases and changelogs have low-quality info

Reproducing Results

- Target code may be hard to compile
- Paper descriptions of tools lack lots of details
  - Sometimes those details matter
- Tools are not always publicly available
  - Even if they are available, they may be hard to use
  - Ex: Couldn't get tools for Java 1.3 to run on 1.4 bytecode
- Reimplementing techniques ≠ publication
  - Replicating results is not valued

What do We Need?

- Benchmarks
  - Open-source programs
  - Environment configuration information
  - Detailed information about bugs
    - From within the community
    - From developers
- Comparable results
  - Tools that are available to others
  - And/or detailed tool output to compare against