An Exercise in Applying Daubert Evaluation Criteria to Security Fuzz Testing

Phil Koopman

January 23, 2015
Carnegie Mellon University
koopman@cmu.edu

Electrical & Computer ENGINEERING

NATIONAL ROBOTICS NREC ENGINEERING CENTER

Overview

- Rule 702 & Daubert criteria for US Federal Courts
  - Specifically intended to reject “junk science”
- How does Fuzz Testing measure up?
  - This is just an example; feel free to follow along with your favorite technique in mind
- What are the implications for security?
  - (And other related properties?)

Rule 702: Testimony By Experts

- Expert witness
  - Qualified by knowledge, skill, experience, training, or education
  - In US, generally hired by parties to a lawsuit
    - Unsurprisingly, opposing experts may disagree
- Testimony acceptable if all of:
  - Must assist "trier of fact" in understanding
  - Based on sufficient facts or data
  - Product of reliable principles & methods
  - Witness has reliably applied principles and methods to the facts of the case

Daubert Criteria

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993)

1. Theory or technique has been tested
2. Subject to peer review and publication
3. What is known/potential error rate?
4. Standards controlling method's operation
5. Widespread scientific community acceptance

- Judge is gate-keeper for testimony validity
  - Flexible application; not all criteria need be met

Hypothetical Situation

- Company C being sued for insecure system
  - E.g., class action suit by credit card customers
  - Data released, but no "smoking gun" for how it happened
    - Plaintiffs can't name a concrete bug/vulnerability
- Defense expert D has a report that says:
  - "System S is secure because fuzz testing found no exploitable vulnerabilities"
- Should D be allowed to testify?
  - Or is D using "junk science"?

Subject To Peer Review and Publication

- Typical fuzz testing papers:
  - We fuzz tested (or robustness tested, fault injected, etc.) and…
  - Found ways to crash the software
- But, need something more…
  - Find papers that did manual analysis to show that fuzz testing found exploitable vulnerabilities

Theory or Technique Has Been Tested

- Is the theory falsifiable (scientific method)?
  - Or refutable; or testable
- A typical fuzzing paper:
  - We found something with fuzz testing
  - Further analysis showed it was exploitable

- But, Expert D's hypothesis is:
  - "Finding nothing with fuzzing means the system is secure"
  - What kind of publication does D need to find?

An Aside: "Dauberting" an Expert

- Plaintiffs (the class action lawyers) can challenge Defendant expert opinion admissibility
  - They file a report explaining why opinion is junk science
  - Judge decides based on Daubert criteria
  - Can exclude some or all of report
    - Excluding a report on either side can essentially terminate the case (summary judgment for prevailing side)

- What academic paper are Plaintiffs looking for?
  - What is they key argument you'd pursue if you were helping them?

What Is Known/Potential Error Rate?

- Origins in applying statistical analysis
  - E.g., “toxic tort” such as exposure causing cancer
- What are the chances the analysis is correct?
  - US civil standard is “more likely than not” = 51%
  - Some flaws in analysis “go to weight, not admissability”
- How reliable is fuzzing at finding security vulnerabilities?
  - In absence of further analysis – just fuzzing results
  - If you find nothing, does that correlate with security?
  - If you find something, does that correlate with insecurity?

Standards Controlling Method's Operation

- Are there standards for applying fuzzing?

- Is it practical to create a "standard" for fuzzing?
  - (See the SIGDeB:
    IFIP WG 10.4 Dependability Benchmarking SIG)

Widespread Scientific Community Acceptance

- Are there any papers advocating the technique?
- Are there many papers supporting the technique?
- Are there lots of credible papers both for and against?

- Is the technique actually being used the way that the papers say it should be used to be acceptable?
  - Rule 702: reliably applying the method to the specifics of the case

Practical Issues

- Are judges (and PC reviewers) adequately trained to evaluate security publications?
  - Is there an accepted list of criteria that makes such publications “good?”
  - Are the lists of security “snake oil” scientifically proven to be predictive of junk science?
  - What should judge do if the academic community is split as to validity of technique?

- Beyond Daubert, at trial, things get complicated
  - This just determines whether someone gets to speak
  - The jury decides the outcome