Vulnerability Discovery

David Warren

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DM-0004078

Vulnerability Discovery Project

Increase **assurance** of DoD software through **enhanced vulnerability discovery techniques**

Team

- Edward Schwartz, PhD, CERT
- David Warren, CERT
- Allen Householder, CERT

- David Brumley, PhD
- Thanassis Avgerinos, PhD
- Tyler Nighswander

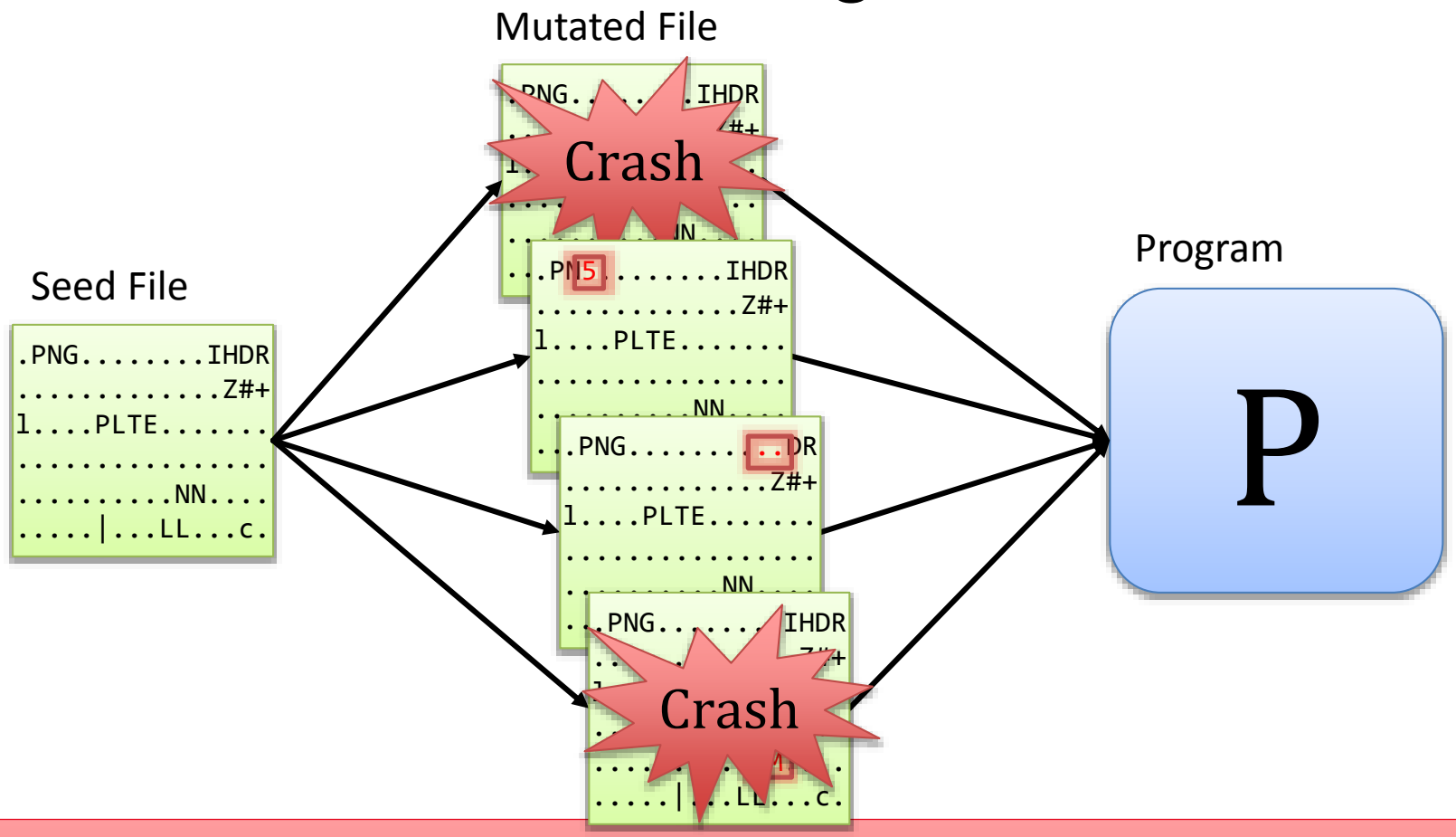
Agenda

Revisiting Widely Held Beliefs about Black-Box Fuzzing

SMART Fuzzing: How to Intelligently Combine a Fuzzer with a Concolic Executor

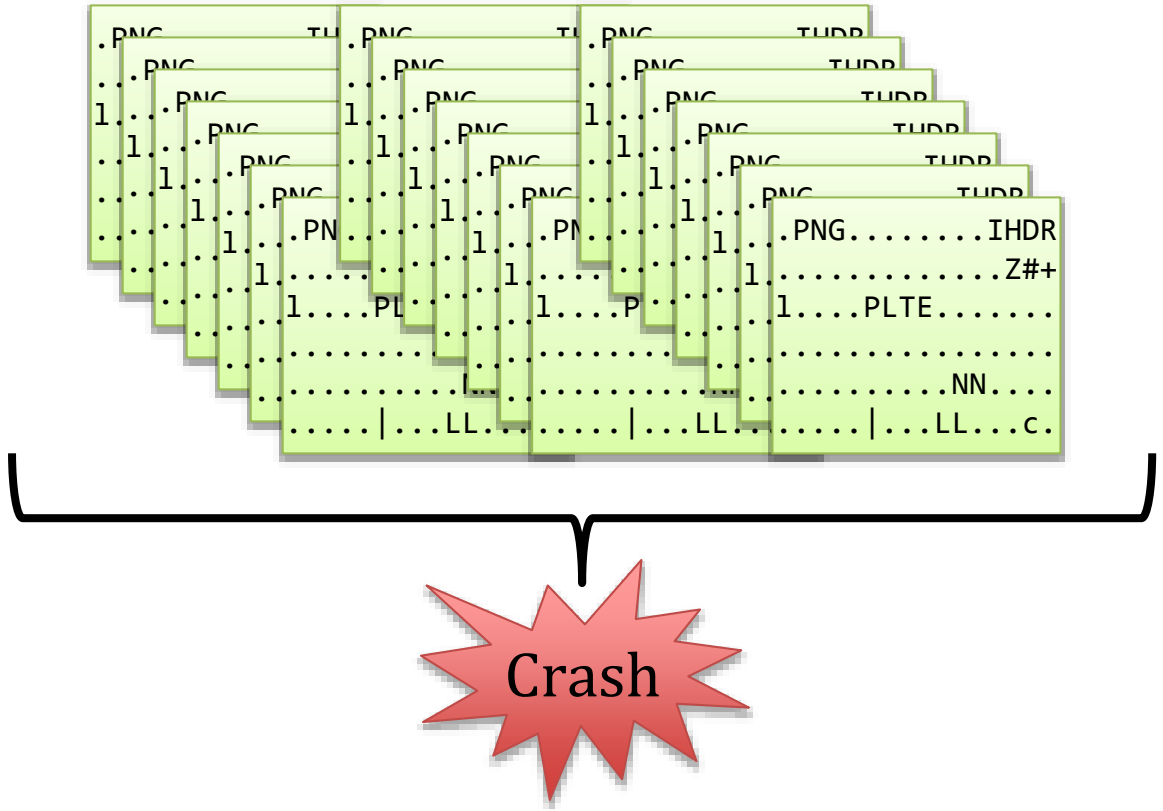
Revisiting Widely Held Beliefs About Black-Box Fuzzing

Background: Mutational Fuzzing of Software

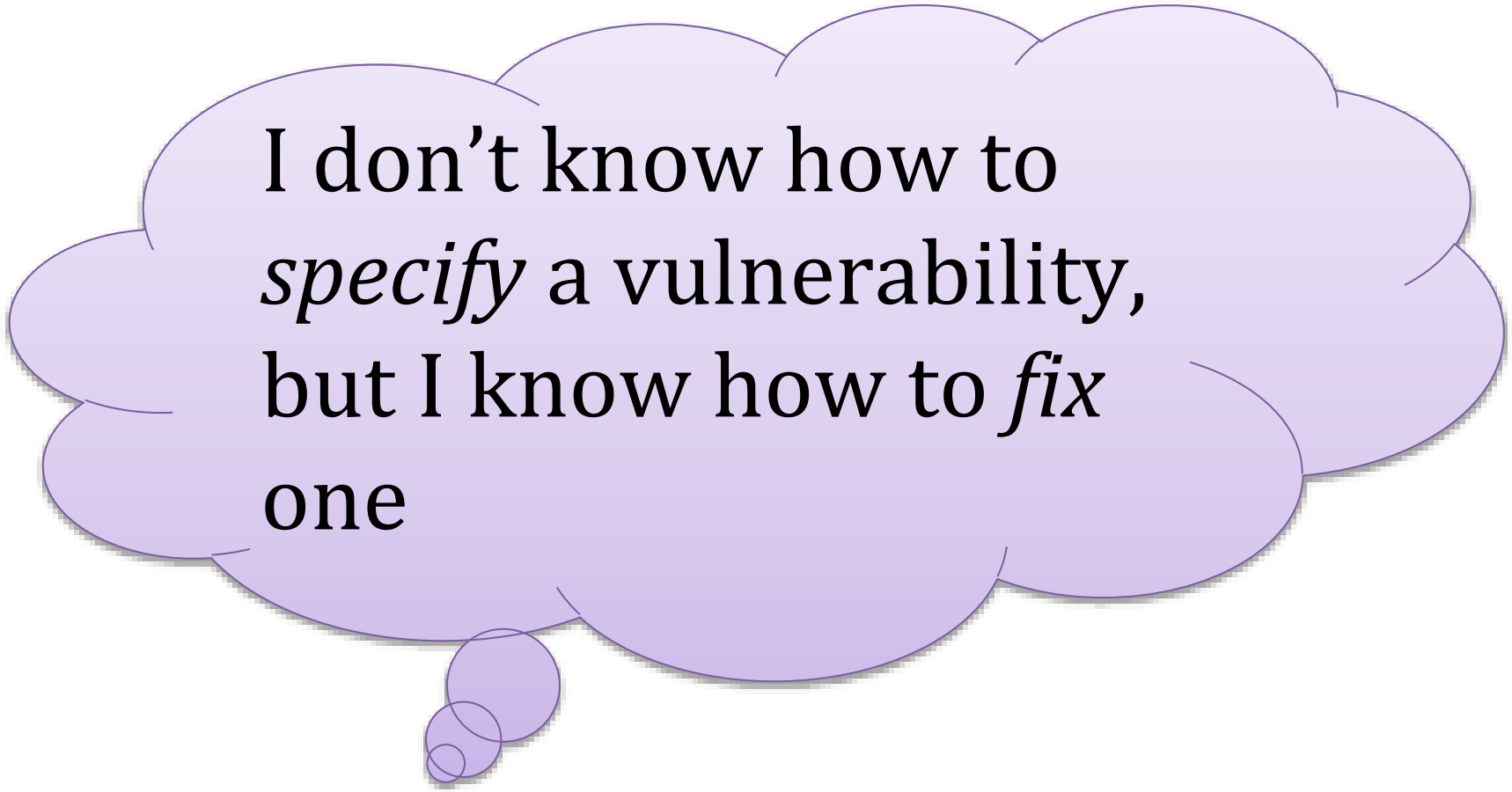


Testing of programs by randomly mutating program inputs (seeds)

Challenge: How Many Software Vulnerabilities Are There?



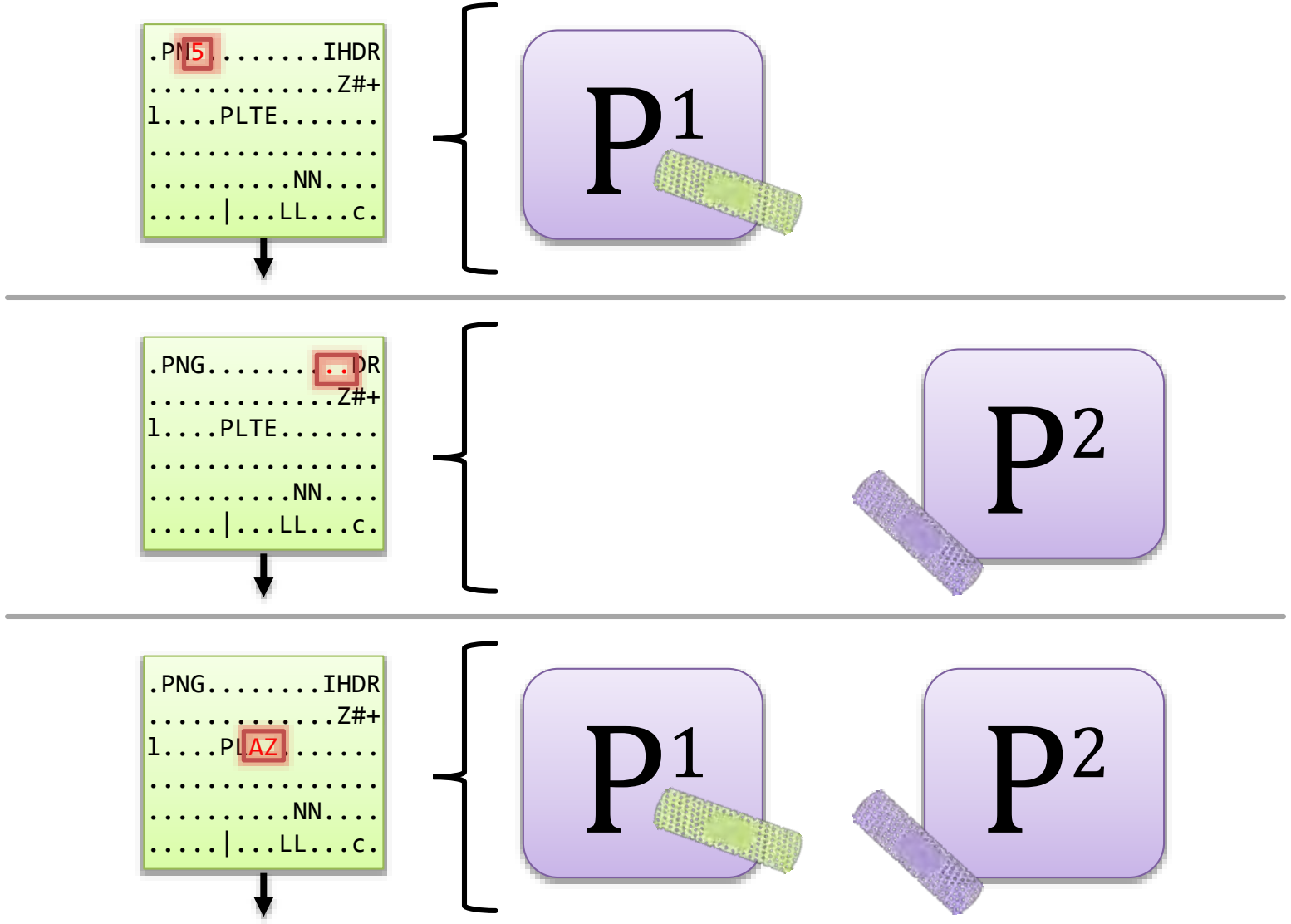
Problem: Distinguishing One Vulnerability From Another



The Idea: Patches Define Vulnerabilities

Any crash that is fixed by the patch is also affected by vulnerability V

Example Ground Truth



Patching software

- Fuzzed each program for 1 week with BFF fuzzer
- Manually patched all vulnerabilities to collect ground truth

	Flasm	ImageMagick	Jasper	OpenJpeg
Crashes (BFF)	253	64	93	145
Vulnerabilities	6	31	12	36

Revisiting Common Beliefs about Black-Box Fuzzing

Belief 1: Stack backtrace hashing always accurately counts vulnerabilities.

- Used by BFF and other fuzzers.

Belief 2: Sanitization never harms fuzzing performance.

- Detects vulnerabilities that do not cause a crash.

Belief 3: The AFL fuzzer always finds more vulnerabilities than non-guided fuzzers.

- Newer is better.

Belief 1: Stack Backtrace Hashing Always Accurately Counts Vulnerabilities

Undercount (UC)

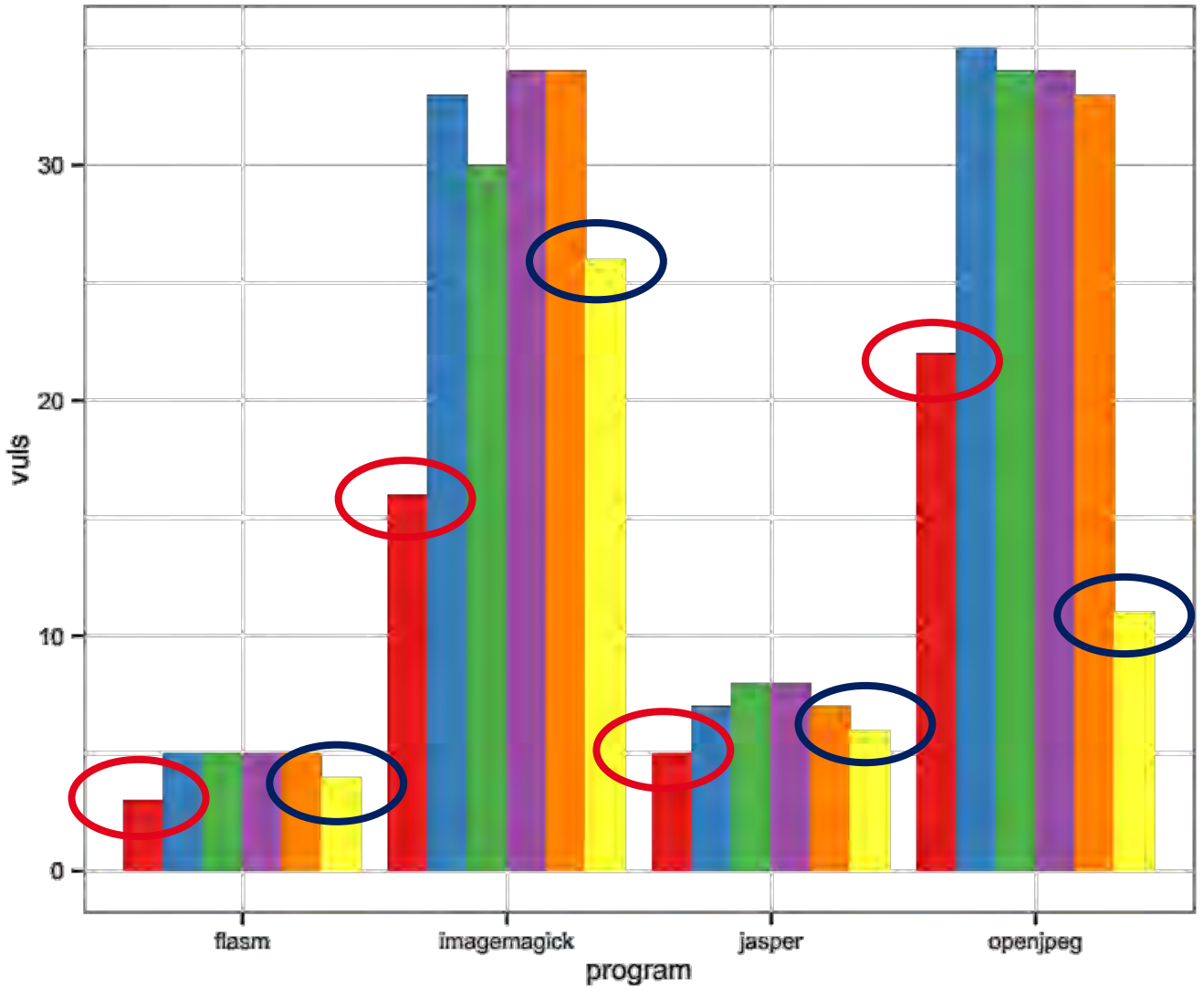
- # vulnerabilities missed by stack backtrace hashing on average
- We never see these vulnerabilities

Overcount (OC)

- # times a vulnerability is included more than once by stack backtrace hashing on average
- We see these vulnerabilities more than once

Program	# Vuls	UC	%	OC	%
Flasm	6	1.8	29%	410.9	6,848%
ImageMagick	31	1.9	6%	67.9	219%
Jasper	12	0.0	0%	226.4	1,887%
OpenJpeg	36	0.1	0%	267.5	743%

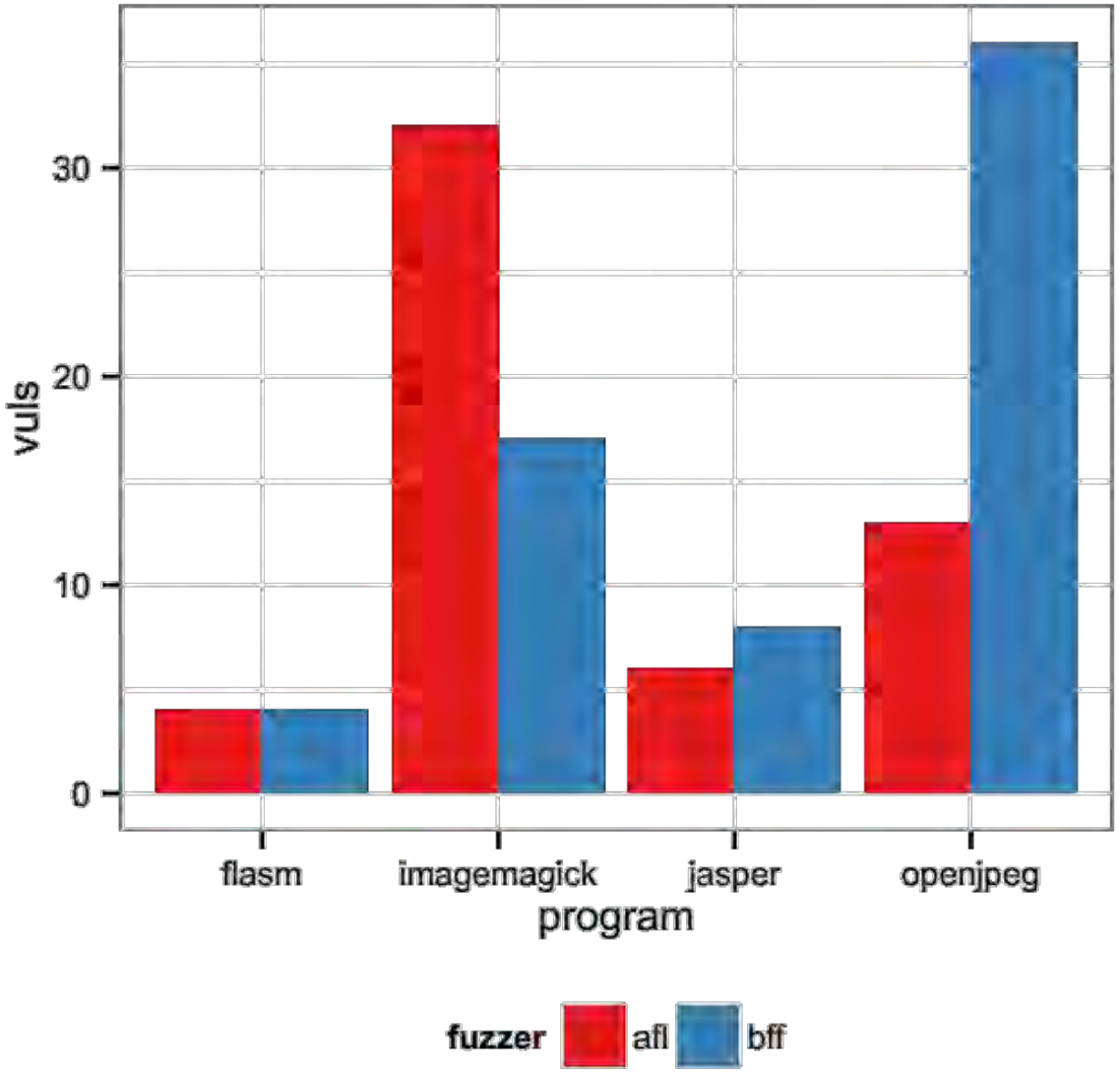
Belief 2: Sanitization Never Harms Fuzzing Performance



Sanitizer

Sanitizer

Belief 3: The AFL Fuzzer Always Finds More Vulnerabilities Than Non-Guided Fuzzers



Collaboration with ForAllSecure

ForAllSecure

DARPA DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

EXPLORE BY TAG

ABOUT US / OUR RESEARCH / NEWS / EVENTS / WORK WITH US /

Defense Advanced Research Projects Agency > News And Events > DARPA Celebrates Cyber Grand Challenge Winners

DARPA Celebrates Cyber Grand Challenge Winners

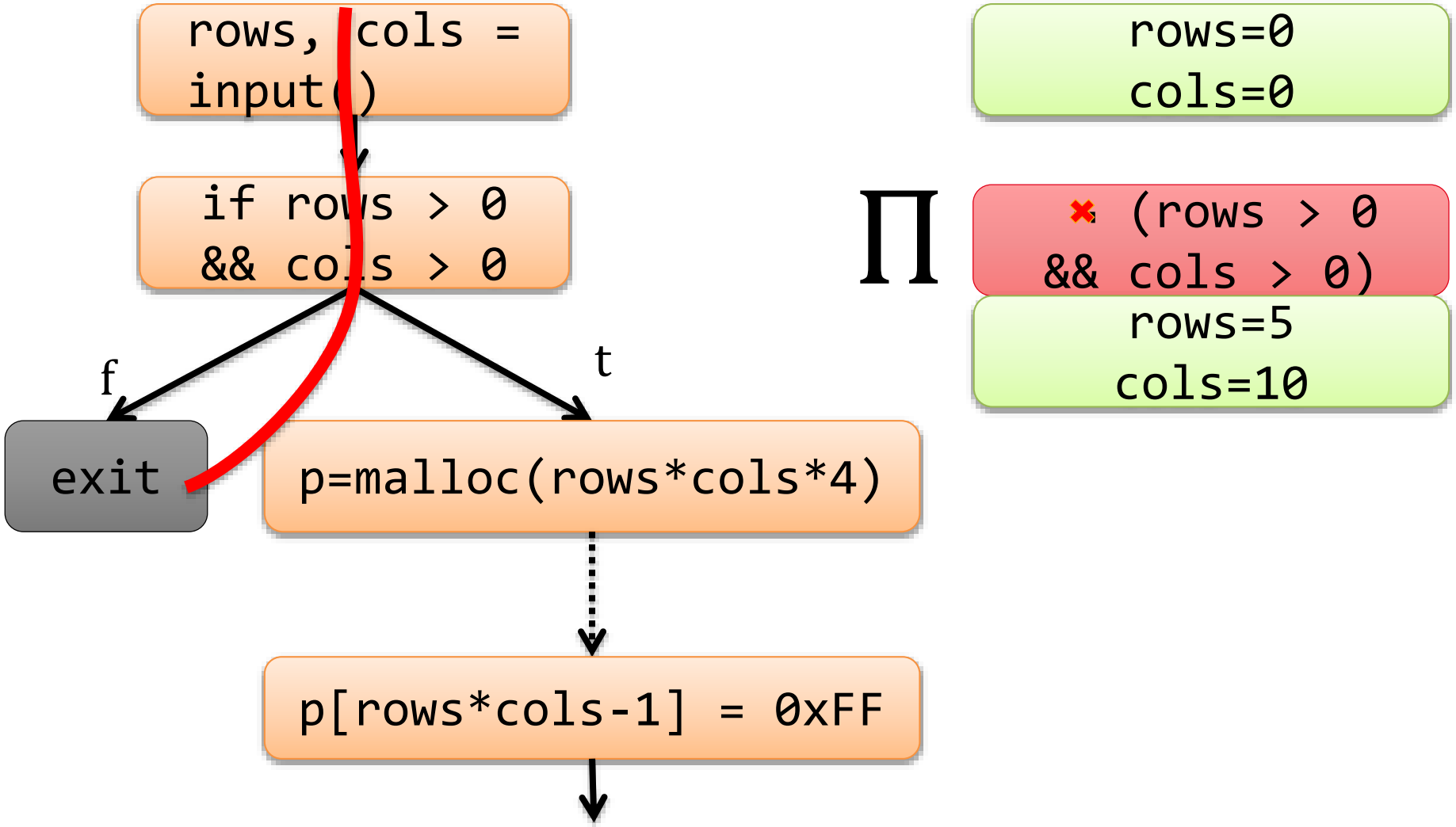
Top-scoring cyber reasoning system becomes first machine to be invited to participate in DEF CON Capture the Flag tournament

OUTREACH@DARPA.MIL
8/5/2016

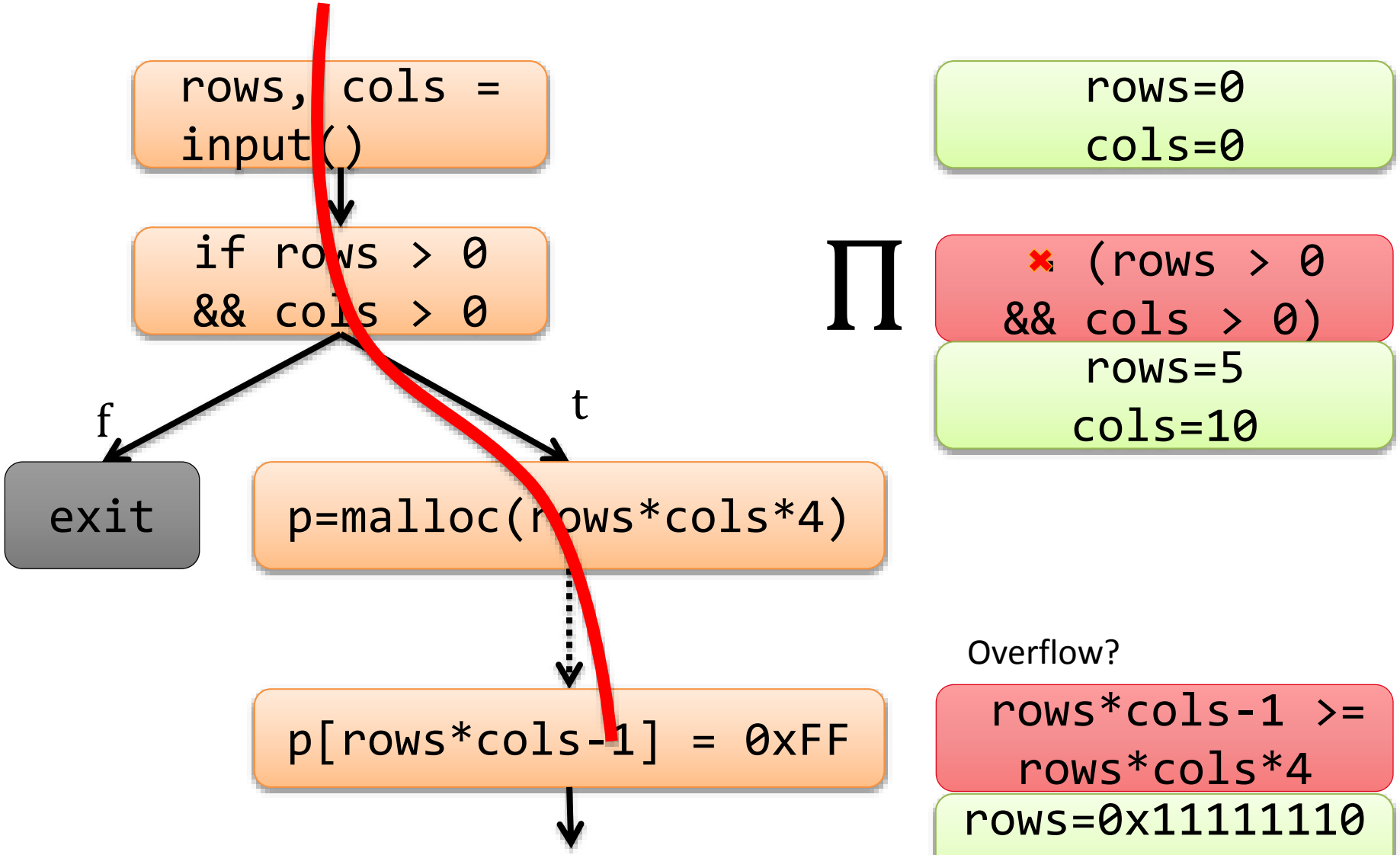
(Updated August 7, 2016)

DARPA officials this morning released partial final, audited results of yesterday's all-day Cyber Grand Challenge (CGC) Final Event—the world's first all-machine cyber hacking tournament—and confirmed that the top-scoring machine was Mayhem, developed by team ForAllSecure of Pittsburgh.

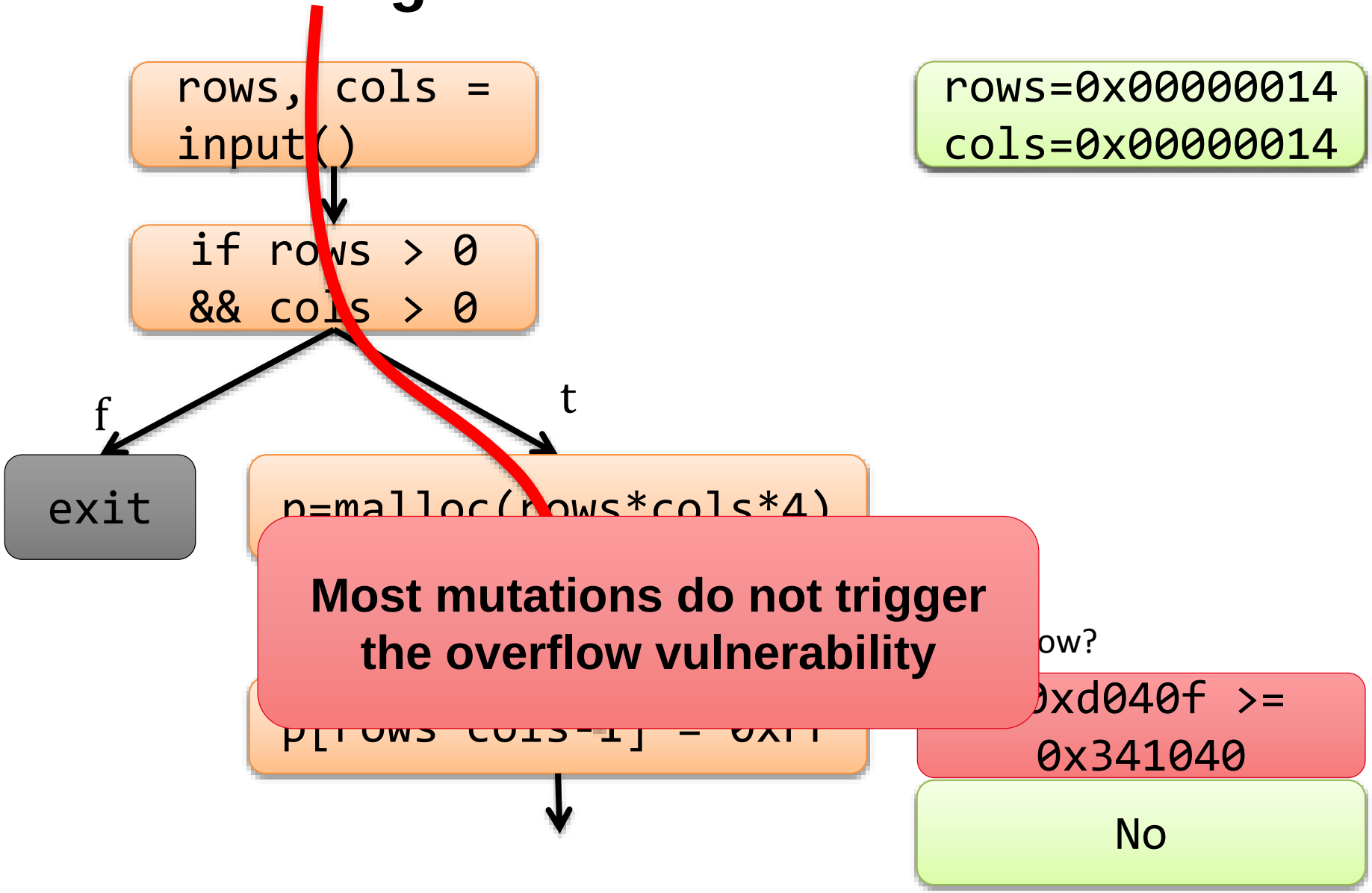
Background: Concolic Execution



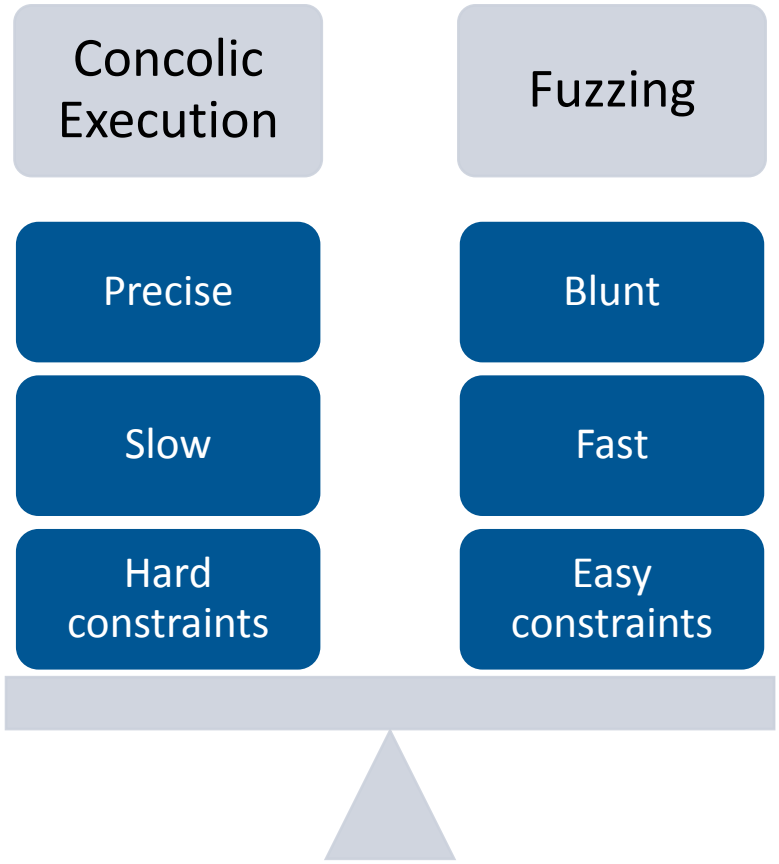
Background: Concolic Execution



Background: Fuzzing



Concolic Execution vs. Fuzzing



SMART

The Synergistic Mayhem AFL Research Tool

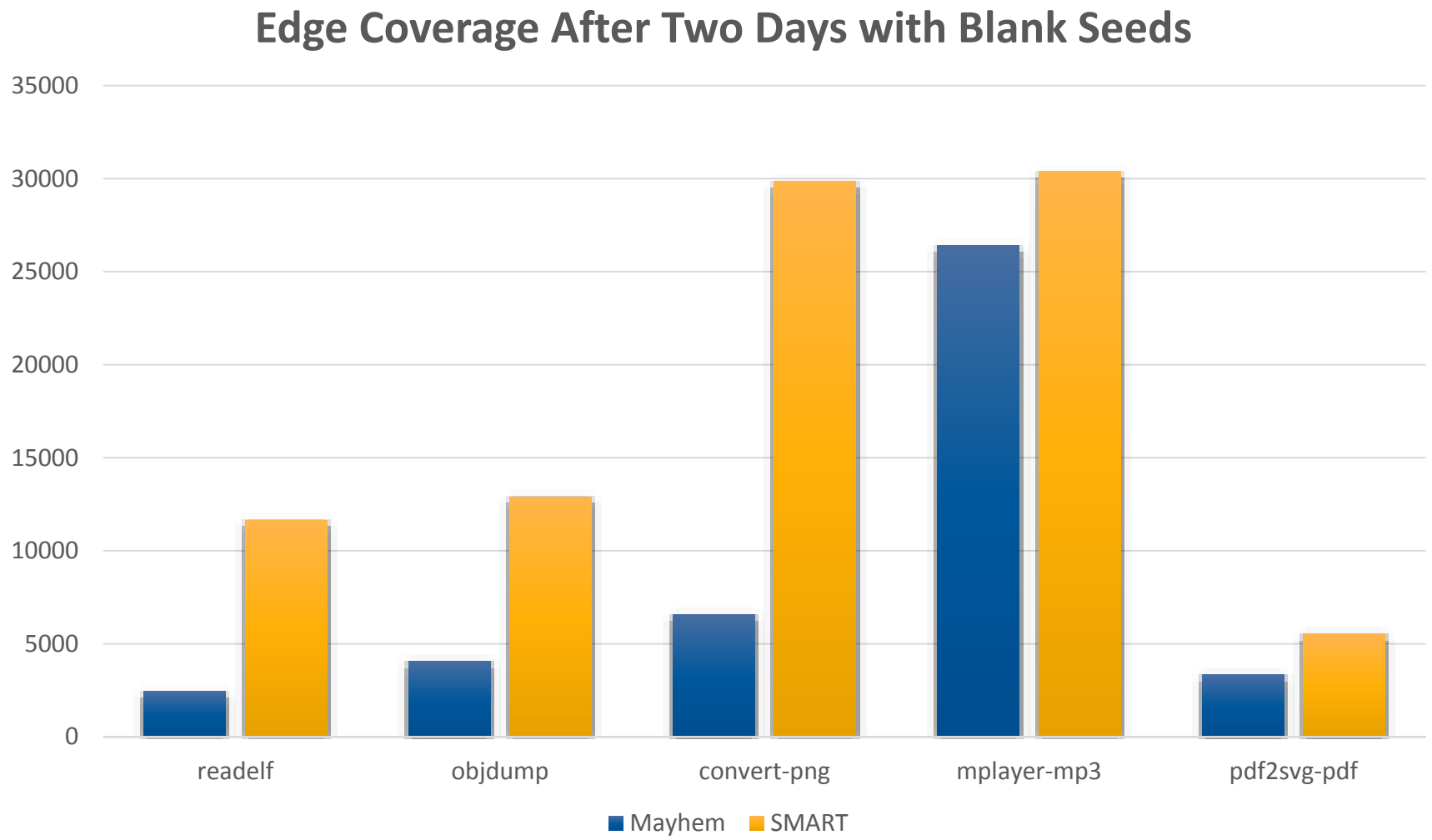
- Concolic execution: Mayhem (ForAllSecure+SEI)
- Fuzzing: AFL
- Periodically synchronize seed files between them

Challenges

How much should we use concolic execution?

- ~10^4 times slower than fuzzing
- Brute force vs. high cost

SMART evaluation



Summary

- Developing new techniques for discovering and mitigating vulnerabilities in the DoD
- Developed vulnerability uniqueness model and used ground truth to explore common fuzzing (mis-)beliefs
- ForAllSecure: Hybrid fuzzing and concolic tester

Team Members

- Edward Schwartz, PhD, CERT
- David Warren, CERT
- Allen Householder, CERT

ForAllSecure Inc.:

- David Brumley, PhD
- Thanassis Avgerinos, PhD
- Tyler Nighswander