王明哲

个人简历

wmzhere@gmail.com

教育背景

2018—2023 博士：清华大学，软件学院，软件系统安全保障小组；导师：姜宇副教授

2014—2018 学士：北京邮电大学，软件学院

研究成果

动态程序分析

背景 动态分析能提供程序执行时的丰富信息，但其应用却受制于额外开销。例如在模糊测试场景下，动态分析：

- 收集慢：占据待测程序 70% 的 CPU 时间
- 处理慢：占据测试工具 85% 的 CPU 时间
- 能力差：无法同时保证插桩灵活性和语义正确性

ATC'21 简化并加速覆盖率收集和处理逻辑

- 将部分桩点逻辑前移到编译期进行，实现单指令插桩，加速覆盖率收集 *23×*
- 抽取逻辑简单的高频处理场景，利用硬件并行执行，加速覆盖率处理 *6×*

效果：方案已实际部署于 *Google* 的 *OSS-Fuzz* 集群

PLDI'22 基于编译器的动态插桩

在编译器插桩的场景下，允许在运行时按需调整桩点：

- 高性能、高语义正确性的插桩方案，实现低至 *3%* 的覆盖率收集开销
- 兼顾代码质量和生成速度的重编译策略，实现低至 *80μs* 的桩点调整时延

系统模糊测试

背景 模糊测试是有效的安全缺陷挖掘方法，但其成熟的应用场景却仅限于库和小型程序。其原因主要包括：

- 粒度粗：只检测“是否崩溃”等简单问题，无法在指令粒度进行分析和控制
- 难适配：需要裁切大型系统为测试单元，严重依赖领域专家

CCS'20 控制流完整性方案的安全性测试

利用模糊测试思想，暴力扫描 CFI 方案的实际可跳转目标

- 利用进程地址空间分析，枚举全部的潜在跳转目标
- 利用轻量级调试和进程快照，高效校验防护逻辑

效果：在 *12* 个开源 *CFI* 方案中找到 *10* 类缺陷

ICSE'21 大型系统的端到端模糊测试

解决数据库等大型系统中覆盖率失灵、崩溃不能复现的痛点

- 多进程、多二进制、多 DSO 的复杂场景下，实现源码级全系统插桩
- 多进程系统的在线异常分析、上下文记录

效果：在 *Postgres, GaussDB, Comdb2* 等知名数据库上找到 *79* 处缺陷

项目经历

插桩平台　对 LLVM 全流程进行深度定制，满足各类工程和研究需求

- 内容：开发 9k 行 C++ 程序
  - 互相打通的数据流 [S&P'22]、控制流 [ASE'19]、机器码 [ASE'20] 插桩
  - 良好的工程化：无需修改编译参数，自动对同一程序生成不同插桩的二进制
- 技能：编译器二次开发、运行时设计

测试平台　从零自主研发的模块化模糊测试平台

- 内容：设计系统架构，搭建基础框架；组织 4 人团队，开发 27k 行 Rust 程序
  - 模块化设计：一个平台支撑多类对象、多种策略的测试
  - 高质量代码：包含文档和测试，部署持续集成，实施代码走查
  - 性能调优：无锁、零拷贝的覆盖率分析；批处理任务调度
- 技能：内核和微结构级别的性能调优

研究计划

精细测试　程序分析主导的动态测试

- 现状：模糊测试中暴力变异为主、覆盖率导向为辅；大量变异均为无效执行
- 新思路：符号化程序输入，利用控制流和数据流分析定位测试位点，精细引导测试
- 预期成果：可解释、可终止的动态测试

抽象测试　隔离外部环境的动态测试

- 现状：待测系统的实际执行涉及无关的环境交互，适配复杂且执行缓慢
- 新思路：自动识别风险位点，从而抽象外部环境，直达关键代码
- 预期成果：易适配、高性能、无误报的大型系统动态测试

荣誉与奖项

2022　Google FuzzBench 模糊测试工具评测，覆盖率第一名
2021　清华大学，一等奖学金
2020　完成珠峰挑战（中国第 44 人）：连续骑车，爬升 8848m
2018　北京邮电大学 ACM/ICPC 竞赛，银奖
2017　全国软件测试大赛，一等奖
2016　北京邮电大学信息安全竞赛，第一名；入选“天枢”信息安全战队

代表工作

PLDI'22	Odin: On-Demand Instrumentation with On-the-Fly Recompilation	*CCF-A*, 第一作者
S&P'22	PATA: Fuzzing with Path Aware Taint Analysis	*CCF-A*, 第二作者
ATC'21	RIFF: Reduced Instruction Footprint for Coverage-Guided Fuzzing	*CCF-A*, 第一作者
ICSE'21	Industry Practice of Coverage-Guided Enterprise-Level DBMS Fuzzing	*CCF-A*, 第一作者
CCS'20	Finding Cracks in Shields: On the Security of Control Flow Integrity Mechanisms	*CCF-A*, 共同一作
ASE'20	Zeror: Speed Up Fuzzing with Coverage-sensitive Tracing and Scheduling	*CCF-A*, 第二作者
ASE'19	VisFuzz: Understanding and Intervening Fuzzing with Interactive Visualization	*CCF-A*, 第二作者
ICSE'18	SAFL: Increasing and Accelerating Testing Coverage with Symbolic Execution and Guided Fuzzing	*CCF-A*, 第一作者
专利'22	CN114168469A, 基于数据库管理系统模糊测试的覆盖率分析方法及系统	第一发明人
专利'21	CN112463581B, 一种对分布式系统进行模糊测试的方法及系统	第一发明人