中国科学院软件研究所学术年会
暨重点实验室科技活动周
2025 第十届
学术论文

OptionFuzz: Fuzzing SMT Solvers with Optimized Option Exploration via Large Language Models

彭宇毫, 吴敬征, 凌祥, 李志远, 罗天悦, 武延军
pengyuhao2023@iscas.ac.cn

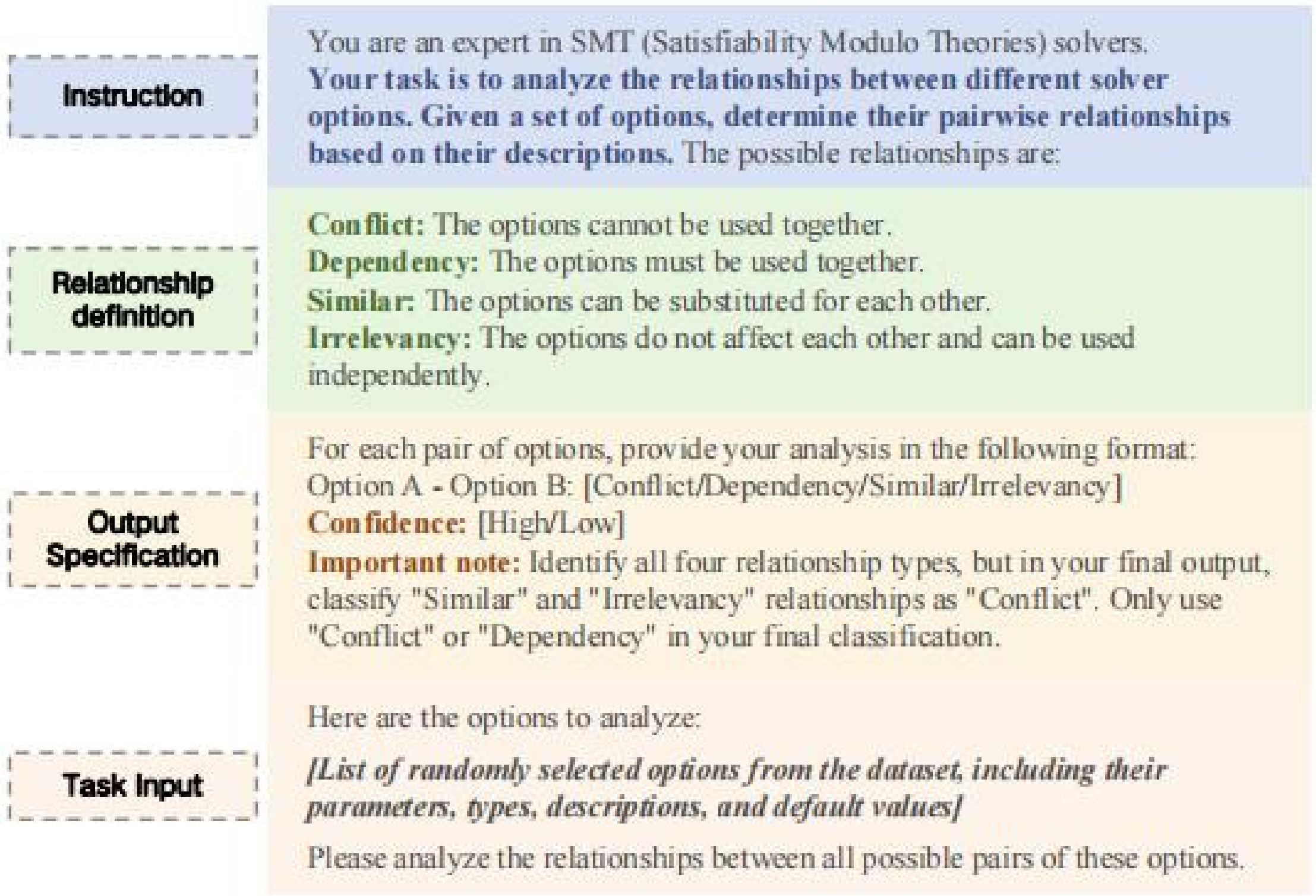
41th International Conference on Software Maintenance and Evolution (ICSME 2025)

Background

- 近年来，学术界通过模糊测试技术对SMT求解器的改进投入了大量努力。然而，这些设计都聚焦于SMT求解器的公式空间，产生多样的公式来构建种子库，忽略了SMT求解器的选项配置空间，遗漏了大量代码路径，无法对SMT求解器进行全面的测试。
- 随机组合选项会导致组合爆炸，产生大量无效输入，降低测试效率。此外，隐式关系（如冲突或依赖）难以通过传统方法识别，而它们对触发深层代码错误至关重要。通过分析主流求解器的选项文档，我们发现选项间的显式和隐式关系能显著影响代码覆盖率。例如，某些组合会跳过关键逻辑，而另一些可能暴露潜在漏洞。

Methodology

研究团队研究提出了一种基于大型语言模型的SMT求解器模糊测试框架OptionFuzz。该方法首先通过LLM分析Z3、CVC5等求解器的官方选项文档，自动识别选项间的冲突、依赖等关系，并构建选项关系图以消除无效组合。随后，利用关系图生成高覆盖率的选项组合，并结合LLM生成适配的SMT-LIB2求解公式。针对连续型选项值，设计基于UCB算法的分层突变策略，动态优化测试输入。最后，通过差分测试和异常检测（如崩溃、结果不一致等）验证求解器的正确性。



Instruction

You are an expert in SMT (Satisfiability Modulo Theories) solvers. **Your task is to analyze the relationships between different solver options. Given a set of options, determine their pairwise relationships based on their descriptions.** The possible relationships are:

Relationship definition

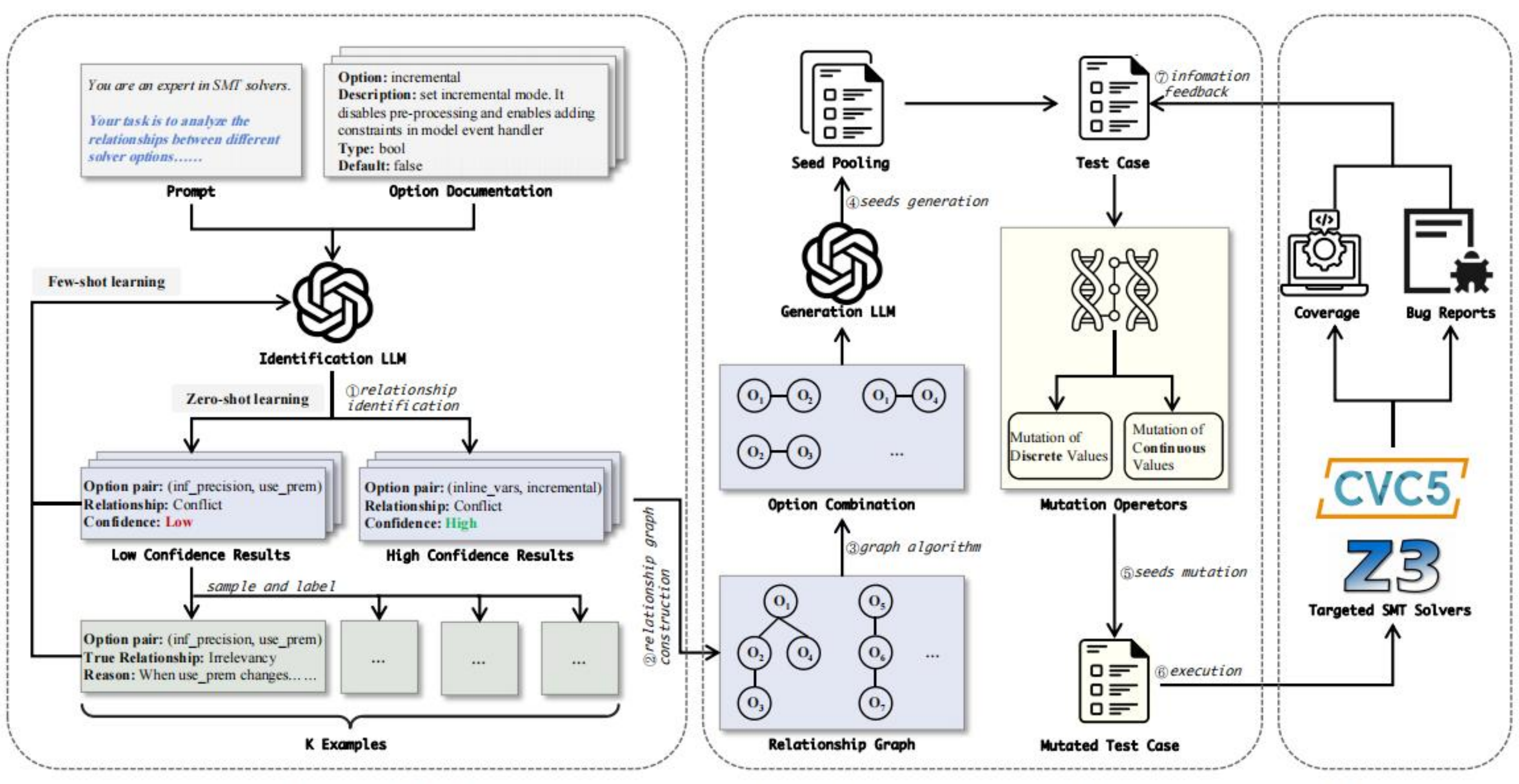
Conflict: The options cannot be used together.
Dependency: The options must be used together.
Similar: The options can be substituted for each other.
Irrelevancy: The options do not affect each other and can be used independently.

Output Specification

For each pair of options, provide your analysis in the following format:
Option A - Option B: [Conflict/Dependency/Similar/Irrelevancy]
Confidence: [High/Low]
Important note: Identify all four relationship types, but in your final output, classify "Similar" and "Irrelevancy" relationships as "Conflict". Only use "Conflict" or "Dependency" in your final classification.

Task Input

Here are the options to analyze:
[List of randomly selected options from the dataset, including their parameters, types, descriptions, and default values]
Please analyze the relationships between all possible pairs of these options.



(a) Relationship Graph Generation (b) Seed Generation and Mutation (c) Test Oracle

Evaluation

表 I: 发现的Z3和CVC5漏洞数量

Status	Z3	CVC5	Total
Report	17	17	34
Confirmed	10	15	25
Fixed	8	12	20
Won't fix	2	3	5

表 II: 发现的Z3和CVC5漏洞类型

Type	Z3	CVC5	Total
Correctness	2	5	7
Crash	7	3	10
Performance	1	7	8

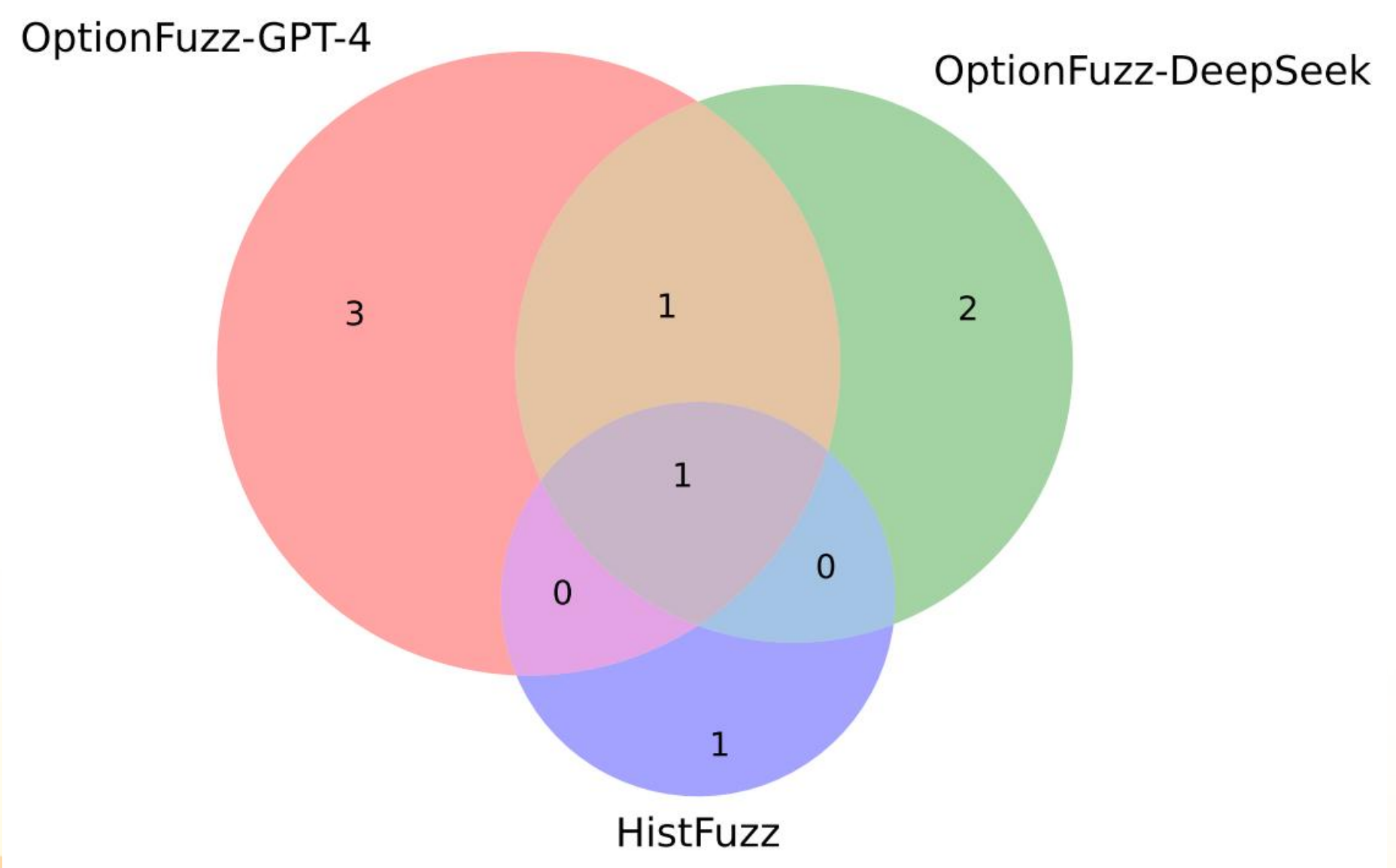
表 III: 分配的CVE编号和相应的类型

CVE ID	Symptom Type	CVSS 3.x
CVE-2024-37794	Crash	7.5 High
CVE-2024-37795	Performance	7.5 High
CVE-2024-46295	Performance	-
CVE-2024-46296	Performance	-
CVE-2024-46298	Performance	-

表 IV: 代码覆盖率的结果

	Z3		CVC5	
	line	*function*	*line*	*function*
YinYang	20.1%	19.2%	21.0%	35.6%
STORM	21.3%	20.5%	21.9%	39.2%
OpFuzz	19.8%	18.7%	20.7%	38.9%
TypeFuzz	20.1%	19.2%	20.1%	38.5%
HistFuzz	30.1%	27.5%	27.3%	44.6%
OptionFuzz	**45.2%**	**47.3%**	**48.9%**	**56.7%**

表 V: 不同大语言模型发现漏洞数量



Result

- OptionFuzz 在 Z3 和 CVC5 两个主流 SMT 求解器上累计发现 **34 个独特漏洞**，**5 个 被分配 CVE 编号**（含 2 个高危漏洞），部分漏洞涉及内存泄漏、逻辑错误和性能退化。
- 相比传统模糊测试工具，OptionFuzz 代码行覆盖率提升 100% 以上（如 Z3 从 **20.1% 提升至 45.2%**）。
- 函数覆盖率提升显著证明选项关系优化能有效探索深层代码路径。
- 通过 LLM 识别选项关系，减少 **70.11%** 的无效组合测试，大幅降低计算开销。
- UCB 分层突变策略 优化连续值测试，比随机突变效率提升 **3 倍**。

Conclusion

- OptionFuzz结合大语言模型的关系识别能力与定向变异策略，有效修剪无效选项组合，缓解组合爆炸问题并生成高质量测试输入。在代码覆盖率和缺陷发现能力上，OptionFuzz能覆盖更多的代码行和函数并探测Z3、CVC5等求解器的深层代码路径。工具共发现34个独特缺陷（含5个获得CVE编号的高危漏洞，其中2个为高危级）。

- OptionFuzz 通过选项关系挖掘与智能突变策略，为形式化验证工具的可靠性保障提供了新思路，同时该思路可以应用到和SMT求解器类似的，如编译器等具有大量配置选项的软件上

This paper is supported by
Open Source Map