学校代码：10246
学　　号：19212010035

復旦大學

硕 士 学 位 论 文

（学术学位）

基于多源信息的开源软件漏洞补丁定位方法

Finding Patches for Open Source Software Vulnerabiliies from multiple sources

院　　系：　软件学院
专　　业：　软件工程
姓　　名：　许聪颖
指 导 教 师：　陈碧欢　副教授
完 成 日 期：2021 年 12 月 10 日

目 录

摘 要

开源软件 (Open source software, OSS) 漏洞管理已然成为一个热点问题。开源漏洞数据库为解决漏洞问题提供十分有价值的数据信息，因此，漏洞数据库的数据质量也受到越来越多的关注和研究。具体的问题为：现有漏洞数据库中补丁的质量尚未研究清楚，此外，现有的补丁信息多由人工或基于启发式的识别方法进行收集。这种方法人工成本过高，且过于定制化无法应用于全部的 OSS 漏洞。

empirical study 该如何翻译比价好呢？实证研究？经验性研究？为了解决这些问题，首先，我们进行了实证研究，以了解当前商业旗舰漏洞数据库中开源软件漏洞补丁的质量和特征。我们的研究涵盖五个方面，包括：补丁的覆盖度、一致性、类型、基数和准确性。然后，基于研究的发现，我们提出了第一种名为 TRACER 的自动化方法，用于从多个来源查找开源漏洞的补丁。

实验评估表明：i) 与现有的基于启发式的方法相比，TRACER 能够为多达 273.8% 的 CVE 找到补丁；同时，准确性方面，将 F1 数值提高达 116.8%；ii) 与现有的漏洞数据库相比，TRACER 将召回率（recall）提高达 18.4% ；然而，12.0% 的 CVE 补丁未找到，精度（precision）下降约6.4%。

关键字：关键词 1，关键词 2，关键词 3

中图分类号：TP311

Abstract

Open source software (OSS) vulnerability management has become an open problem. Vulnerability databases provide valuable data that is needed to address OSS vulnerabilities. However, there arises a growing concern about the information quality of vulnerability databases. In particular, it is unclear how the quality of patches in existing vulnerability databases is. Further, existing manual or heuristic-based approaches for patch identification are either too expensive or too specific to be applied to all OSS vulnerabilities.

To address these problems, we first conduct an empirical study to understand the quality and characteristics of patches for OSS vulnerabilities in two state-of-the-art vulnerability databases. Our study is designed to cover five dimensions, i.e., the coverage, consistency, type, cardinality and accuracy of patches. Then, inspired by our study, we propose the first automated approach, named TRACER, to find patches for an OSS vulnerability from multiple sources. Our key idea is that patch commits will be frequently referenced during the reporting, discussion and resolution of an OSS vulnerability.

Our extensive evaluation has indicated that i) TRACER finds patches for up to 273.8% more CVEs than existing heuristic-based approaches while achieving a significantly higher F1-score by up to 116.8%; and ii) TRACER achieves a higher recall by up to 18.4% than state-of-the-art vulnerability databases, but sacrifices up to 12.0% fewer CVEs (whose patches are not found) and 6.4% lower precision. Our evaluation has also demonstrated the generality and usefulness of TRACER.

Keywords: Keyword1, Keyword2, Keyword3
CLC code: TP311

第 1 章 绪论

本章节概述了背景、研究目的与意义[1–2]。

1.1 研究背景

1.2 本文工作概述

1.3 本文篇章结构

如图1-1所示。如表1-1所示。

表 1-1 表格名称

AA \ BB	C1	C2
R1	1	2
R2	3	4

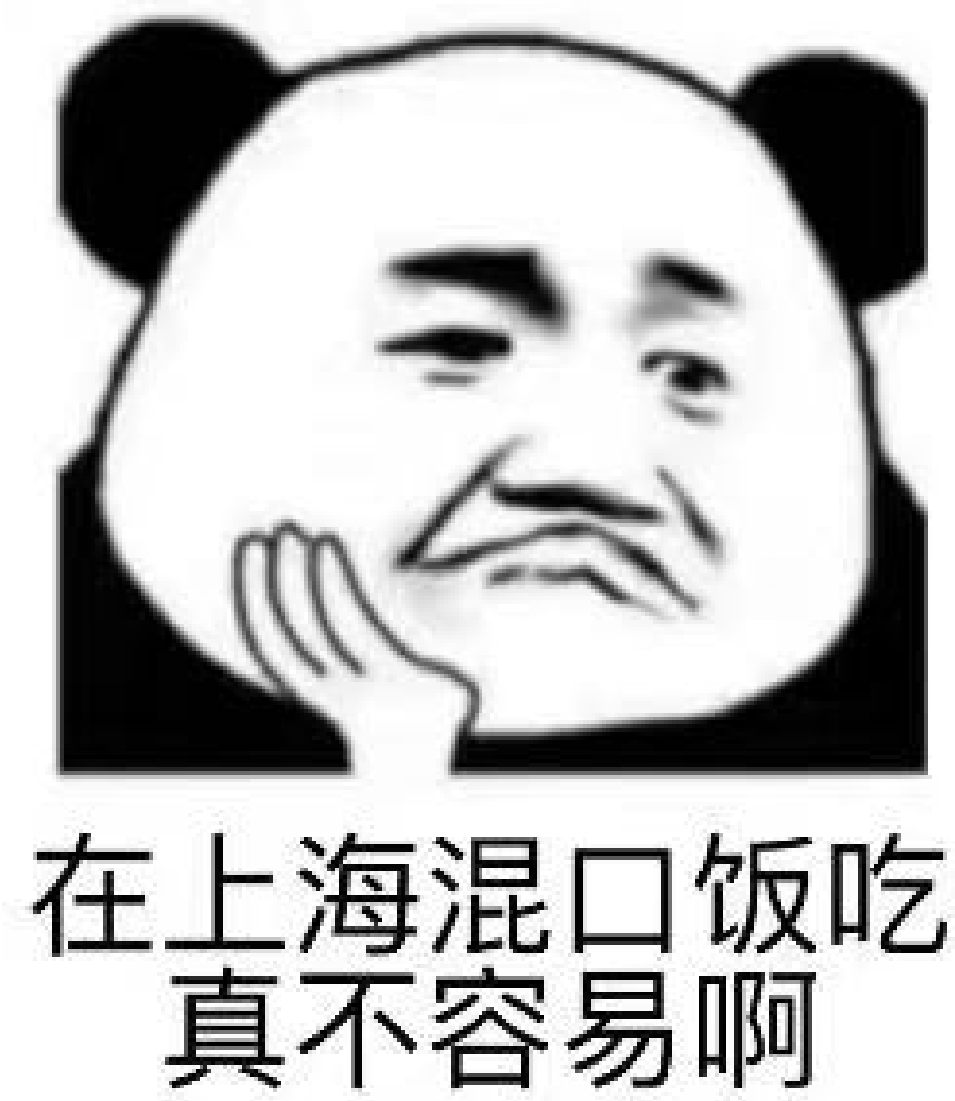

图 1-1 在上海混口饭吃真不容易啊

第 2 章 背景知识

本章节介绍了背景知识。

2.1 CVE–通用漏洞披露

2.2 NVD–(美国) 国家漏洞数据库

2.3 漏洞补丁

第 3 章 实验验证及结果分析

本章节实验评估。

3.1 实验设计

3.2 RQ6：准确性验证

3.3 RQ7：削弱性分析

3.4 RQ8：敏感度分析

3.5 RQ9：通用性分析

3.6 RQ10: 实用性能分析

3.7 讨论

参考文献

[1] 贾培养, 孙鸿宇, 曹婉莹, 等. 开源软件漏洞库综述[J]. 信息安全研究, 2021: 566-574.

[2] MITRE. About cve[EB/OL]. 2021. https://cve.mitre.org/.

致谢

致谢致谢致谢