Characterizing License Practices in Maven Central and Their Relationship with CVEs: A CS846 Course Project Proposal

2/27/25

Haonan Zhang, Christina Li, Paul Wooseok Lee

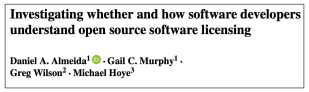
UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Current Studies Primarily Focus on Either License or CVEs

Studies on license:

A Large-Scale Empirical Study of Open Source License Usage: Practices and Challenges

Jiaqi Wu, The State Key Laboratory of Blockchain and Data Security, Zhejiang University, Hangzhou, Zhejiang, China, jiaqiwu@zju.edu.cn
Lingfeng Bao*, The State Key Laboratory of Blockchain and Data Security, Zhejiang University, Hangzhou, Zhejiang, China, lingfengbao@zju.edu.cn
Xiaohu Yang, The State Key Laboratory of Blockchain and Data Security, Zhejiang University, Hangzhou, Zhejiang, China, yangxh@zju.edu.cn
Xin Xia, Huawei, China, xin.xia@acm.org
Xing Hu, The State Key Laboratory of Blockchain and Data Security, Zhejiang University, Hangzhou, Zhejiang, China, xinghu@zju.edu.cn



Investigating whether and how software developers understand open source software licensing

Daniel A. Almeida[1] · Gail C. Murphy[1] · Greg Wilson[2] · Michael Hoye[3]

On the Adoption of Open Source Software Licensing - A Pattern Collection

Petr Pícha, Department of Computer Science and Engineering, University of West Bohemia, Czechia, ppicha@kiv.zcu.cz
Souhaila Serbout, Software Institute (USI), Switzerland, souhaila.serbout@usi.ch

Studies on CVEs:

Understanding the Threats of Upstream Vulnerabilities to Downstream Projects in the Maven Ecosystem

Yuhan Wu[†1], Zeliang Yu[†1], Ming Wen[1*], Qiang Li[1], Deqing Zou[1], Hai Jin[2]
[1]School of Cyber Science and Engineering, Huazhong University of Science and Technology, China
[2]School of Computer Science and Technology, Huazhong University of Science and Technology, China
Email: {shawn, zelianyu, mwenaa, hjzeng, deqingzou, hjin}@hust.edu.cn

Mitigating Persistence of Open-Source Vulnerabilities in Maven Ecosystem

Lyuye Zhang[†*], Chengwei Liu[†], Sen Chen[‡], Zhengzi Xu[†], Lingling Fan[§], Lida Zhao*, Yiran Zhang[¶], Yang Liu[†]
[†]Nanyang Technological University, Singapore
[‡]Continental-NTU Corporate Lab, Nanyang Technological University, Singapore
[§]School of Computer Science and Engineering, Nanyang Technological University, Singapore
[¶]College of Intelligence and Computing, Tianjin University, China
[‖]College of Cyber Science, Nankai University, China

Analyzing the Direct and Transitive Impact of Vulnerabilities onto Different Artifact Repositories

JOHANNES DÜSING and BEN HERMANN, Technical University Dortmund, Germany

What is the Relationship between Licenses and CVE Patterns?

Hypothesis

- Certain license types may influence collaboration and patching processes, potentially correlating with higher or lower vulnerability rates.

Objective

- Systematically analyze license data and CVEs in Maven Central to uncover patterns that can guide more informed artifact and license adoption choices.

Research Questions

RQ1:

- What are the characteristics and trends of license adoption and CVE incidence across Maven Central artifacts?

RQ2:

- Do specific license types correlate with higher or lower vulnerability incidence in Maven Central artifacts?

Dataset and Tools

Goblin Framework

- Neo4j-based Maven Central dependency graph
- Built-in CVE information
- Weaver for on-demand metric computations (Maybe)

Augmenting Goblin with License Data

- Retrieve license metadata through Libraries.io
- Python script with paginated queries to Libraries.io's API
- Merge library and license info into Neo4j nodes

Threats to Validity

Rate Limits and Data Incompleteness:

- Description
  - Libraries.io's API rate limits can cause partial or failed data retrieval.
  - Long delays or incomplete datasets may result if the rate limit is exceeded.
- Mitigation
  - Use incremental and batched queries.
  - Implement retry strategies for rate-limit responses.
  - If these fail, switch to alternative data sources (e.g., public datasets, Sonatype OSS Index), recognizing they might be less comprehensive.

Mapping Between Dependencies, Licenses, and Vulnerabilities

- Description
  - Goblin has Maven dependency + CVE data.
  - External license data may mismatch artifact coordinates or versions.
  - Leads to incomplete or inaccurate mappings.
- Mitigation
  - Strict coordinate matching (groupId, artifactId, version).
  - Omit partial or ambiguous matches to maintain data quality.

Schedule and Milestones

- **Week 1**: Extract Maven dependency data with CVE from Goblin and collect and integrate license data from Libraries.io.
- **Week 2-3**: Analyze CVE incidence trends across Maven artifacts and license adoption trends.
- **Week 4-5**: Investigate the statistical relationship between license types and CVEs.
- **Week 6**: Document methodology, results, and findings as a final report.

Summary

- **Motivation**: Uncover the relationships between the license types and the CVE patterns.
- **Approach**: Goblin framework, Libraries.io.
- **Potential outcome**: Interesting findings to help developers and stakeholders make informed decisions.

UNIVERSITY OF
WATERLOO

FACULTY OF MATHEMATICS