Detecting Unsafe Updates in Software Ecosystems

Yao-Wen Chang

September 27, 2023

Supervised by: Christoph Treude
Co-supervised by: Behnaz Hassanshahi (Oracle Labs, Australia)

BACKGROUND

- What is supply chain?
- **SolarWind's Orion** platform is polluted.
- Malicious action against **Esline-Scope**.

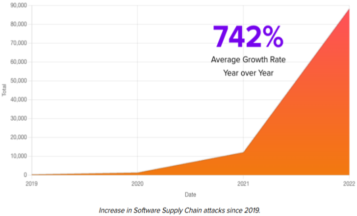


Figure: CI/CD Attack Trend Image source: [McBride(2023)].

OBJECTIVE

- Introduce a new framework.
- Evaluate the effectiveness of the framework.
- Automate the unsafe update explore process.

Research Questions

1. *What is the scope of the impact of the risks that exist within our target Python and Java repositories?*
2. *What are the results if these suspicious updates from contributors in open source projects compromise the target?*
3. *To what extent does this work enhance the security of CI/CD pipelines based on the findings and recommendations from our research?*

OUTLINE

- Related Works
- Research Methods
  - Data Source
  - Framework
  - Pipeline
  - Metric
- Summary
- QA

RELATED WORKS

Static Analysis - Bandit [PyCQA(2023)]

- Parse python source code (AST)
- Pre-defined rules to match the tree node relationship

Machine Learning - Anomaly Detection

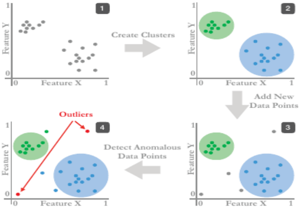


Figure: Anomaly Detection Image source: [Garrett et al.(2019)Garrett, Ferreira, Jia, Sunshine, and Kästner].

RELATED WORKS

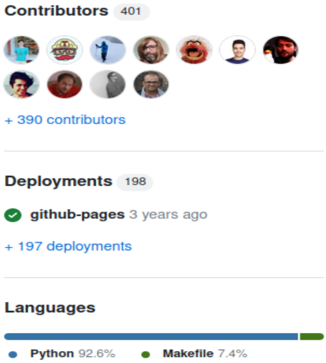
CI / CD Based: in-toto framework [Torres-Arias et al.(2019)Torres-Arias, Afzali, Kuppusamy, Curtmola, and Cappos]

- supply chain layout integrity
- step authentication
- implementation transparency
- graceful degradation of security properties

RESEARCH METHODS I

Data

- Fetch Data
- Clean Data



Contributors 401

+ 390 contributors

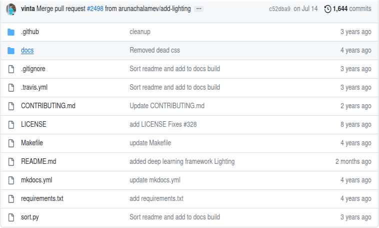
Deployments 198

✓ **github-pages** 3 years ago

+ 197 deployments

Languages

● **Python** 92.6% ● **Makefile** 7.4%



vinta Merge pull request #2498 from arunachalamev/add-lighting ··· c52dba9 on Jul 14 **1,644** commits

.github	cleanup	3 years ago
docs	Removed dead css	4 years ago
.gitignore	Sort readme and add to docs build	3 years ago
.travis.yml	Sort readme and add to docs build	3 years ago
CONTRIBUTING.md	Update CONTRIBUTING.md	2 years ago
LICENSE	add LICENSE Fixes #328	8 years ago
Makefile	update Makefile	4 years ago
README.md	added deep learning framework Lighting	2 months ago
mkdocs.yml	update mkdocs.yml	4 years ago
requirements.txt	add requirements.txt	4 years ago
sort.py	Sort readme and add to docs build	3 years ago

Framework

Introduce papers....

RESEARCH METHODS II

Pipeline

Introduce papers....

RESEARCH METHODS III

Metric

$$SM = (W_p \cdot P) * (W_{tf} \cdot TF) * (\frac{W_{tc}}{TC}) * (W_s \cdot S) \tag{1}$$

$$\begin{aligned}
SM &= \text{Security Scanner Metric} \\
P &= \text{Precision (as a decimal)} \\
W_p &= \text{Weight for Precision} \\
TF &= \text{Total Findings (TP + FP)} \\
W_{tf} &= \text{Weight for Total Findings} \\
TC &= \text{Time Cost} \\
W_{tc} &= \text{Weight for Time Cost} \\
S &= \text{Normalized Severity Score} \\
W_s &= \text{Weight for Severity}
\end{aligned}$$

SUMMARY

This is the first slide of your presentation.

FUTURE WORK

- Data collection.
- Build the system pipeline.
- Contribute to the framework embedded in the system.
- Evaluate the system.

REFERENCES I

Kalil Garrett, Gabriel Ferreira, Limin Jia, Joshua Sunshine, and Christian Kästner.
Detecting suspicious package updates.
In *2019 IEEE/ACM 41st International Conference on Software Engineering: New Ideas and Emerging Results (ICSE-NIER)*, pages 13–16. IEEE, 2019.

Luke McBride.
2023 predictions: What will happen in software supply chain governance?, 2023.
URL `https://blog.sonatype.com/2023-predictions-software-supply-chain-governance`.

PyCQA.
Bandit: a security linter from pycqa, 2023.
URL `https://github.com/PyCQA/bandit`.

REFERENCES II

Santiago Torres-Arias, Hammad Afzali, Trishank Karthik Kuppusamy, Reza Curtmola, and Justin Cappos.
in-toto: Providing farm-to-table guarantees for bits and bytes.
In *28th USENIX Security Symposium (USENIX Security 19)*, pages 1393–1410, 2019.