- Report on a research paper

(The Open-Source Trap: Unraveling Open-Source Threats in the Software Supply Chain)

Name: K.Sony Reddy

Program Name: Cyber Security

Roll No: 160123737073

Github Repository: https://github.com/KSonyReddy/cs-open_source_trap-

Report

Paper Overview

The paper **"The Open-Source Trap: Unraveling Open-Source Threats in the Software Supply Chain"** by Clayton W. Boozell (SANS Institute, 2024) highlights the growing security risks in open-source ecosystems. It explains how attackers exploit public repositories through techniques like **typo-squatting**, **dependency confusion**, and **revive-jacking**, compromising software supply chains. The research, conducted mainly in a **lab-based PyPI environment**, demonstrates how easily malicious packages can infiltrate trusted systems. While the paper provides valuable awareness and recommendations such as package signing and developer vigilance, it lacks **automation, large-scale data analysis, and cross-ecosystem validation**.

Security Features

Uses SequenceMatcher algorithm with an 85% similarity threshold to identify suspicious package names that closely resemble popular legitimate packages like "requests", "numpy", "pandas", "react", and "express", helping prevent developers from accidentally installing malicious lookalike packages. Implements a weighted scoring system (typo-squat=2 points, abandoned=1 point, revived=3 points) that calculates cumulative risk scores and categorizes packages into Low/Medium/High risk levels, enabling prioritized security review and automated decision-making for package approval.

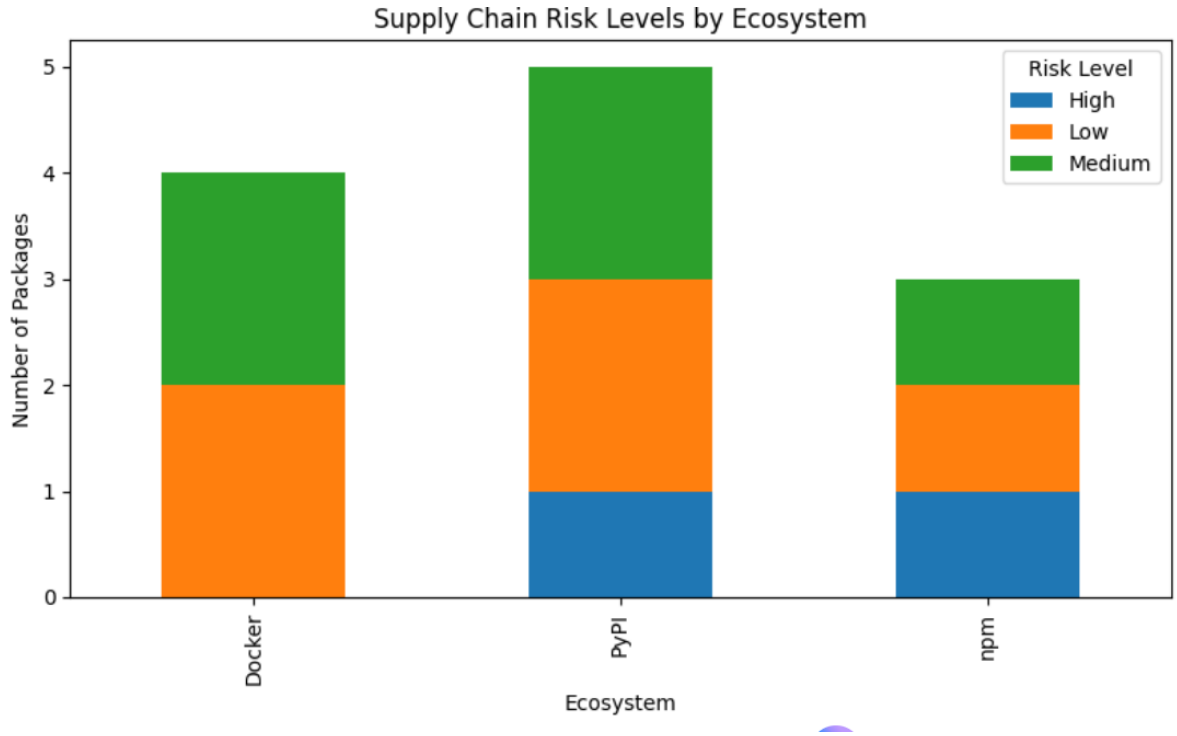
Folder Structure

```
improvisation/
├──code.py                  # Contains code
├── README.md               # Project documentation
```

Screenshots

```
--- SAMPLE OUTPUT ---
         name ecosystem  typo_suspect  abandoned  revived_risk  risk_score  \
0    requests      PyPI         False      False         False           0
1       numpy       npm         False      False         False           0
2      pandas    Docker         False      False         False           0
3       react    Docker         False      False         False           0
4     express      PyPI         False      False         False           0
5   requeests    Docker          True       True         False           3
6       numpi       npm         False       True         False           1
7     pandass      PyPI          True       True         False           3
8      reaact    Docker          True       True         False           3
9      expres      PyPI          True       True         False           3
10     oldlib      PyPI         False       True          True           4
11  unusedpkg       npm         False       True          True           4

   risk_level
0         Low
1         Low
2         Low
3         Low
4         Low
5      Medium
6      Medium
7      Medium
8      Medium
9      Medium
10       High
11       High
```



Learning Outcomes

Demonstrates automated detection of supply chain threats using three core techniques. Typo-squatting detection employs SequenceMatcher with an 85% similarity threshold to identify malicious packages mimicking popular ones like "requests" or "numpy". The weighted risk scoring system assigns points to each threat type and categorizes packages into Low, Medium, or High risk levels, enabling prioritized security review and visualization of threat distribution across ecosystems.

Conclusion

This code provides a basic automated tool to scan packages and identify potential threats before installation.The system outputs results in two formats: a CSV file for detailed review and a bar chart showing which ecosystems have the most risk. While this helps catch obvious threats automatically, it only handles detection.