Bachelorarbeit Yan Wittmann

an der Hochschule Mannheim, Fakultät für Informatik, im SS2025
in Kooperation mit der metaeffekt GmbH.

Hintergrund

Die metaeffekt GmbH entwickelt eine Software zur automatisierten Identifikation von Schwachstellen in Softwareprodukten, die frei unter einer offenen Lizenz verfügbar ist. Dieser Prozess basiert auf den Ergebnissen einer vorhergehenden Software-Kompositionsanalyse. Ein langfristiges Ziel ist die vollständige Automatisierung der Schwachstellenidentifikation.

Für die Schwachstellenzuordnung werden öffentlich zugängliche Datenquellen wie die NVD, OSV-Datenbanken und Security Advisories unterschiedlicher Autoritäten, Institutionen und Herstellern verwendet. Um eine automatisierte Verarbeitung dieser Dokumente zu ermöglichen, geben diese ihre Produktdaten mit verschiedenen Standards wie CPE, PURLs oder sogar proprietäre IDs an. Dies macht individuelle Anpassungen am Code für jede neue Datenquelle erforderlich.

Auch die Eingabedaten, die aus der Software-Kompositionsanalyse stammen, variieren oft, da unterschiedliche Betriebssysteme, Paketmanager oder Projektkonventionen zu uneinheitlichen Darstellungen derselben Software führen.

Derzeit analysiert der Schwachstellen-Scanner diese unterschiedlichen Produktidentifikatoren mithilfe individueller Prüfregeln. Der Erfolg dieses Ansatzes ist jedoch stark formatabhängig und erreicht bei den meisten nur begrenzte Erfolgsraten bei der automatischen Zuordnung. Einige Formate wie PURLs lassen sich leichter verarbeiten, während CPEs oder Microsoft Produkt-IDs oft nur ungenau oder gar nicht zugeordnet werden können.

Ein manueller Prozess namens “Produkt-Korrelation” wurde daher vor einigen Jahren eingeführt, um die Ergebnisse des Automatismus durch ein Korrelations-Team händisch nachpflegbar zu machen. Hierbei werden manuelle Korrekturen in YAML-Dateien dokumentiert, die durch ein einfaches Web-UI unterstützt werden. Jedoch hat dieses System zunehmende Skalierungsprobleme, da die vielen teils tausende Zeilen lange YAML-Dateien unübersichtlich und schwer wartbar geworden sind. Es besteht Bedarf für ein zukunftsfähiges, verbessertes Format und mit einem neuen Konzept für die Modellierung von Produkten und deren Beziehungen.

Ziel der Bachelorarbeit

Ziel der Arbeit mit der metaeffekt ist es, die Qualität der Schwachstellenidentifikation zu verbessern. Hierbei sollen folgende Aspekte behandelt werden:

- Analyse des Ökosystems der Produktidentifikationsstandards.
- Analyse weiterer öffentlicher Tools im selben Kontext hinsichtlich deren Methoden zur Datenverarbeitung und -nutzung.
- Vergleich der internen Eingabedaten mit externen Identifikationsformaten.
- Untersuchung der Schwächen des bestehenden YAML-Korrelationsformats.
- Konzeption und Implementierung eines skalierbaren und zukunftssicheren Korrelationsformats.
- Nutzung des erlangten Wissens, um eine Anpassung der automatisierten Produktidentifikation durchzuführen.