Exposé der Bachelorarbeit von Yan Wittmann

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

Hintergrund

Die automatisierte Identifikation von öffentlich bekannten Schwachstellen in Softwareprodukten ist heutzutage ein wichtiger Bestandteil von IT- Sicherheitsmaßnahmen. Die metaeffekt GmbH entwickelt unter einer offenen Lizenz einen Schwachstellen-Scanner, der öffentliche Datenquellen wie die NVD (National Vulnerability Database, NIST), OSV-Datenbanken (Open Source Vulnerabilities, Google) und maschinenlesbare Hersteller-Ratgeber analysiert, um potenzielle Schwachstellen zu erkennen.

Dabei stellen die Heterogenität der von Herstellern eingesetzen Produktidentifikationsstandards (z.B. CPE (Common Platform Enumeration), PURLs (Package URLs)) und die uneinheitlichen Eingabedaten aus unserer vorhergehenden Software-Kompositionsanalyse Herausforderungen dar. Diese Herausforderungen mit dem automatisierten Matching haben einen manuellen Korrekturschritt namens "Produkt-Korrelation" (Product Correlation) nötig gemacht, in dem durch ein Korrelationsteam manuelle Änderungen am automatisierten Prozess in YAML-Dateien vorgenommen werden.

Forschungsthema

Das in der Einleitung genannte manuell gepflegte Korrelationsformat stößt jedoch mittlerweile zunehmend auf Skalierungsprobleme, da die dutzende YAML-Dateien mit teils mehreren tausend Zeilen nicht nur unübersichtlich, sondern auch durch redundante Identifikationseinträge 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, aber auch für ein verbessertes automatisiertes Matching durch besseres Verständnis der Datenlage.

Auf Basis der dargestellten Herausforderungen im Hintergrund ergibt sich folgende Zielsetzung für meine Bachelorarbeit in Kooperation mit der metaeffekt:

- Analyse des Ökosystems der Produktidentifikationsstandards.
- Analyse 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.
- Anpassung des Software-Scanners, um geeignetere Produktdarstellungen für die Schwachstellenanalyse als Eingabe zu erhalten.
- Konzeption und Implementierung eines skalierbaren und zukunftssicheren Korrelationsformats.
- Die bestehende automatisierte Produktidentifikation mit den gewonnenen Erkenntnissen verbessern und weiterentwickeln.

Motivation und Vorarbeiten

Als Werkstudent bei der metaeffekt GmbH trage ich die Hauptverantwortung für die Entwicklung des Schwachstellenscanners und des Produktkorrelationssystems. Die tägliche Arbeit mit diesen Systemen und die damit verbundenen Herausforderungen motivieren mich dazu, durch neue Ansätze ihre Skalierbarkeit und Wartbarkeit zu verbessern.

Literatur und Quellen

Im Rahmen dieser Arbeit wurden unter anderem bereits einige relevanten Quellen und andere Projekte untersucht, Gespräche mit Personen aus der Branche geführt und weitere Recherchen durchgeführt.

Forschungsarbeiten, Talks

- Software Identification Ecosystem Option Analysis
- Graph-Based CPE Matching for Identification of Vulnerable Asset Configurations
- Vortrag: Universal Software Product Identity (Thomas Schmidt, FIRSTCON23)

Tools

- DependencyCheck
- OpenSSF Scorecard

Spezifikationsdokumente

PURL, CPE 2.2/2.3, OSV, CSAF, vers, etc.

Zitate

Durch ein Gespräch mit Thomas Schmidt, BSI:

> Product Identification presents one of the greatest challenges in the field of vulnerability management. If it is not solved, complete automation of vulnerability matching is not possible.
>
> CSAF 3.0 will not be published until it is solved.

Naming Things is Hard:

> There are only two hard things in Computer Science: cache invalidation and naming things.