Boosting Cyber Security Operations with Knowledge Graphs

David Starobinski

IARPA ReSCIND Proposers' Day

February 28, 2023

BOSTON UNIVERSITY

Security Operation Centers

- Monitor, prevent, detect, investigate, respond, and recover from security incidents
- SoC resources are scarce and highly valuable

General Goals

- Improve the workflow and performance of the SoC
- Automate tasks by leveraging the vast amount of structured and unstructured real-world data available on threats, attacks, and mitigations

Approach

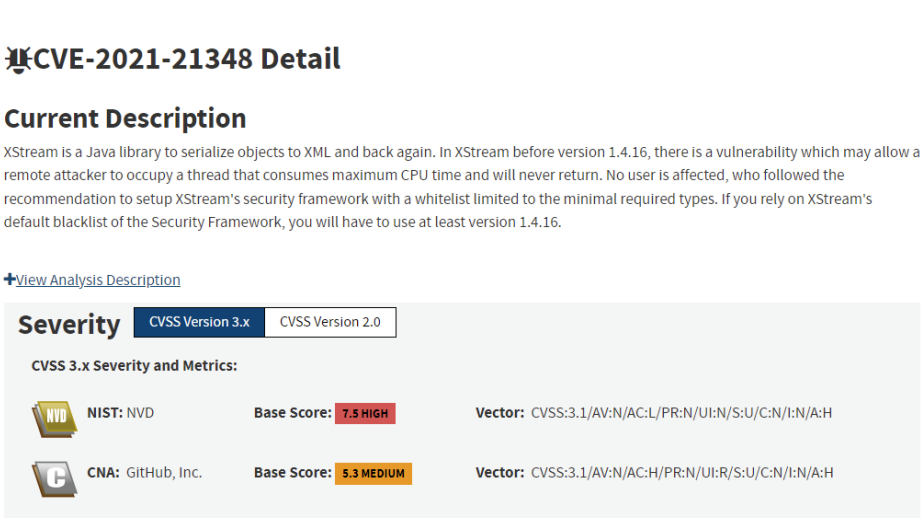
- Develop methods based on **knowledge graphs** to model and derive insights from cyber security data

- Work entails developing **ontologies** to characterize entities, their properties, and relationships between entities

- Use the knowledge graphs for various cyber security activity purposes, such as uncovering hidden relationships, identifying patterns and trends, and querying the data

Case Study: Predicting Unknown Associations between Threat Databases

- National Vulnerability Database (NVD)
  - CVSS scores
  - Associations: CPE <-> CVE, CVE <-> CWE



CVE-2021-21348 Detail

Current Description

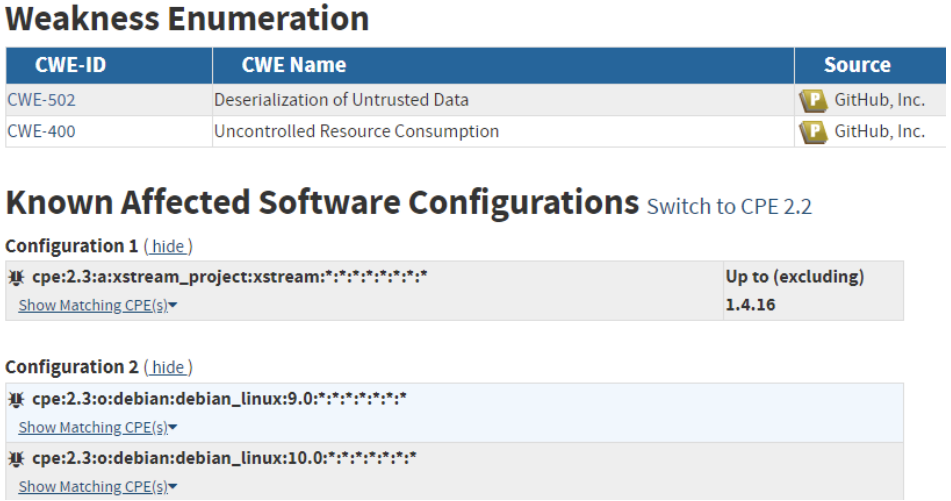
XStream is a Java library to serialize objects to XML and back again. In XStream before version 1.4.16, there is a vulnerability which may allow a remote attacker to occupy a thread that consumes maximum CPU time and will never return. No user is affected, who followed the recommendation to setup XStream's security framework with a whitelist limited to the minimal required types. If you rely on XStream's default blacklist of the Security Framework, you will have to use at least version 1.4.16.

+View Analysis Description

Severity CVSS Version 3.x | CVSS Version 2.0

CVSS 3.x Severity and Metrics:

NIST: NVD	**Base Score:** 7.5 HIGH	**Vector:** CVSS:3.1/AV:N/AC:L/PR:N/UI:N/S:U/C:N/I:N/A:H
CNA: GitHub, Inc.	**Base Score:** 5.3 MEDIUM	**Vector:** CVSS:3.1/AV:N/AC:H/PR:N/UI:R/S:U/C:N/I:N/A:H



Weakness Enumeration

CWE-ID	CWE Name	Source
CWE-502	Deserialization of Untrusted Data	GitHub, Inc.
CWE-400	Uncontrolled Resource Consumption	GitHub, Inc.

Known Affected Software Configurations Switch to CPE 2.2

Configuration 1 (hide)

cpe:2.3:a:xstream_project:xstream:*:*:*:*:*:*:*:* Show Matching CPE(s)	Up to (excluding) 1.4.16

Configuration 2 (hide)

cpe:2.3:o:debian:debian_linux:9.0:*:*:*:*:*:*:* Show Matching CPE(s)

cpe:2.3:o:debian:debian_linux:10.0:*:*:*:*:*:*:* Show Matching CPE(s)

Case Study: Background

Associations by NVD are useful for vulnerability assessment

CVE-ID	CVE-2021-21348
Description	XStream is a Java library to serialize objects to XML and back again. In XStream before version 1.4.16, there is a vulnerability which may allow a remote attacker to occupy a thread that consumes maximum CPU time and will never return. …
Associated CWE	CWE-400 (Uncontrolled Resource Consumption), CWE-502 (Deserialization of Untrusted Data)
Associated CPE by *Aug 4, 2021*	1. cpe:a:xstream project:xstream:*:*, 2. cpe:o:debian:debian linux:*:*

Case Study: Background

However, associations provided by NVD are often incomplete

CVE-ID	CVE-2021-21348
Associated CPE by *Aug 4, 2021*	1. cpe:a:xstream project:xstream:*:*, 2. cpe:o:debian:debian linux:*:*
Associated CPE added from *Aug 4, 2021* to *Mar 29, 2022*	1. cpe:o:fedoraproject:fedora:*:*, 2. cpe:a:oracle:retail_xstore_point_of_service:*:*, 3. cpe:a:oracle:webcenter_portal:*:*, 4. cpe:a:oracle:banking_platform:*:*, 5. cpe:a:oracle:communications_policy_management:*:*, 6. cpe:a:oracle:communications_billing_and_revenue_management_elastic_charging_engine:*:*, 7. cpe:a:oracle:mysql_server:*:*, 8. cpe:a:oracle:business_activity_monitoring:*:*, 9. cpe:a:oracle:communications_unified_inventory_management:*:*, 10. cpe:a:oracle:banking_virtual_account_management:*:*

Case Study: Background

Can we predict some of the unknown associations in advance?

CVE-ID	CVE-2021-21348
Associated CPE by *Aug 4, 2021*	1. cpe:a:xstream project:xstream:*:*, 2. cpe:o:debian:debian linux:*:*
Associated CPE added from *Aug 4, 2021* to *Mar 29, 2022*	1. cpe:o:fedoraproject:fedora:*:*, 2. cpe:a:oracle:retail_xstore_point_of_service:*:*, 3. cpe:a:oracle:webcenter_portal:*:*, 4. cpe:a:oracle:banking_platform:*:*, 5. cpe:a:oracle:communications_policy_management:*:*, 6. cpe:a:oracle:communications_billing_and_revenue_management_elastic_charging_engine:*:*, 7. cpe:a:oracle:mysql_server:*:*, 8. cpe:a:oracle:business_activity_monitoring:*:*, 9. cpe:a:oracle:communications_unified_inventory_management:*:*, 10. cpe:a:oracle:banking_virtual_account_management:*:*

Case Study: Contributions

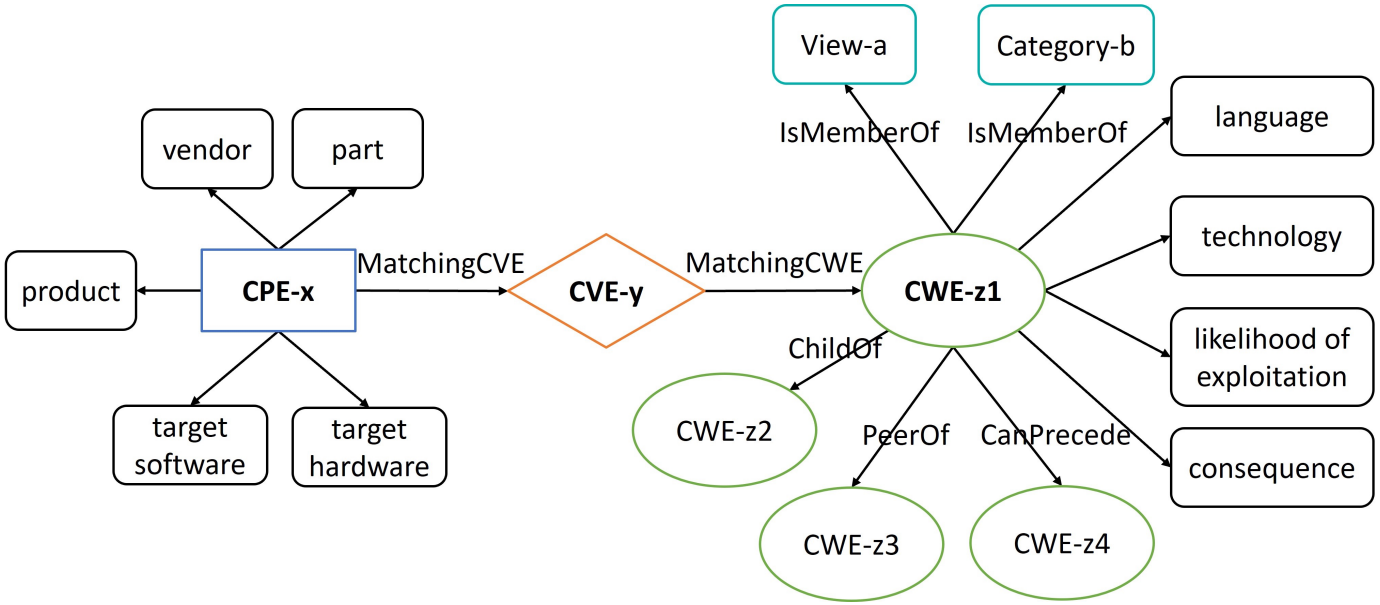
- Propose and implement *threat knowledge graph*
  - Aggregate knowledge from CPE, CVE, and CWE in a graph form
  - Generate triples from threat entries and associations
  - Embed onto a vector space for link prediction

- Complete unknown associations between products and vulnerabilities

Zhenpeng Shi, Nikolay Matyunin, Kalman Graffi, and David Starobinski. "Uncovering Product Vulnerabilities with Threat Knowledge Graphs." In 2022 IEEE Secure Development Conference (SecDev), pp. 84-90.

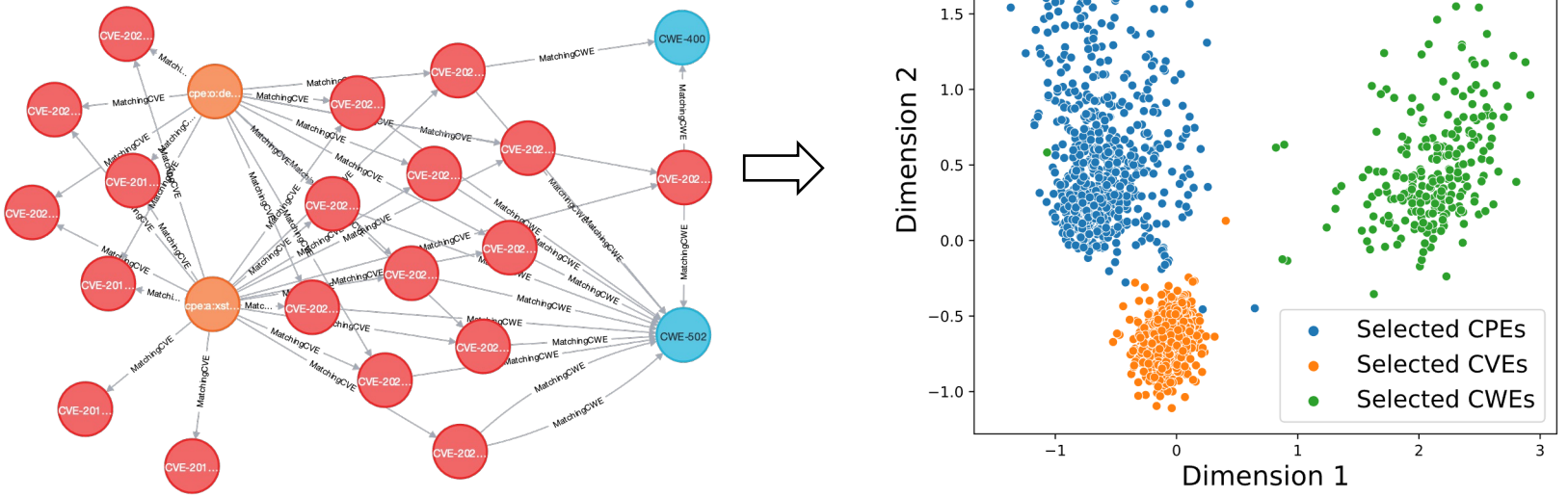
Threat knowledge graph

- *Entities*: use entries of CPE, CVE, and CWE
- *Relations*: use associations by NVD



Knowledge graph embedding

- Translate entities/relations to vectors
- Training



Selected entities as vectors (projected onto 2D space)

Example:
Predicting CPE-CVE associations

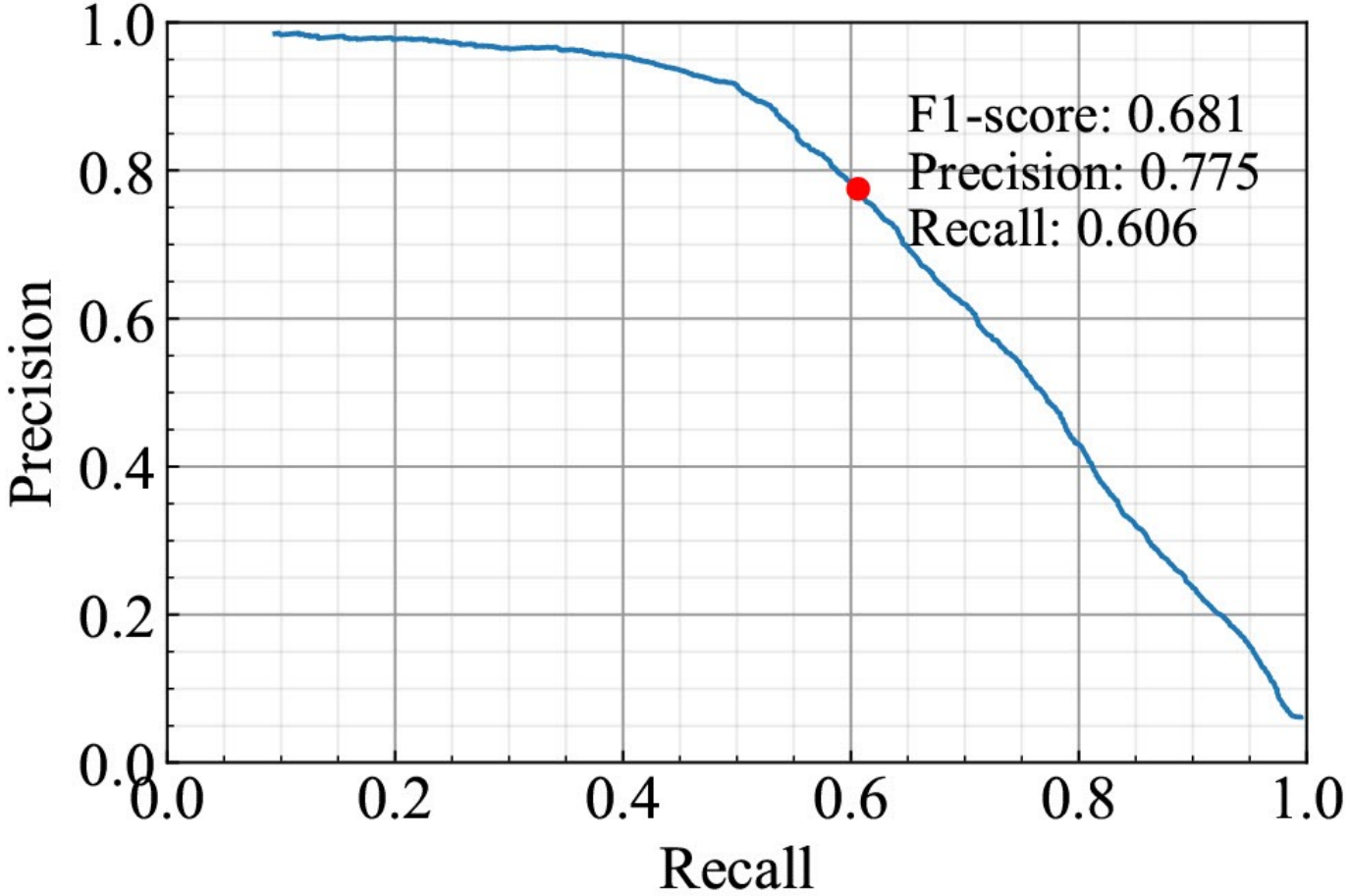
- Threat knowledge graph aggregates data available until Aug 4, 2021
- Predict new triples between Aug 4, 2021 and Mar 29, 2022
- Test set for evaluation
  - Positive triples (ground truth): new CPE-CVE triples
  - Negative triples: replace CPE side of positive triples by random CPEs
  - 50 negative triples for each positive triple
- Evaluation metrics: *precision*, *recall*, *F1-score*

Example: CVE-CPE Prediction

CVE-ID	CVE-2021-21348
Associated CPE by *Aug 4, 2021*	1. cpe:a:xstream project:xstream:*:*, 2. cpe:o:debian:debian linux:*:*
Associated CPE between *Aug 4, 2021* and *Mar 29, 2022*	1. cpe:o:fedoraproject:fedora:*:* 2. cpe:a:oracle:retail_xstore_point_of_service:*:* 3. cpe:a:oracle:webcenter_portal:*:* 4. cpe:a:oracle:banking_platform:*:* 5. cpe:a:oracle:communications_policy_management:*:* 6. cpe:a:oracle:communications_billing_and_revenue_management_elastic_charging_engine:*:* 7. cpe:a:oracle:mysql_server:*:* 8. cpe:a:oracle:business_activity_monitoring:*:* 9. cpe:a:oracle:communications_unified_inventory_management:*:* 10. cpe:a:oracle:banking_virtual_account_management:*:* 11. cpe:a:netapp:clustered data ontap antivirus connector:*:*

- 8 true positives
- 2 false negatives
- 1 false positive

Evaluating CPE-CVE Association Predictions



Future Work

- Augment threat knowledge graph with observed attack behavior

- Use threat knowledge graph to understand, predict, and affect future cyber attack behavior