Investigating Removals in the National Vulnerability Database

Şevval Şimşek
Boston University
Boston, MA
sevvals@bu.edu

Zhenpeng Shi
Boston University
Boston, MA
zpshi@bu.edu

David Starobinski
Boston University
Boston, MA
staro@bu.edu

David Sastre Medina
Red Hat
Madrid, Spain
asastrem@redhat.com

***Abstract*—The National Vulnerability Database (NVD) enables automation of vulnerability management, security measurement, and compliance. This database is quite dynamic and updates to CVE entries happen for many reasons. We analyze this subject to uncover the reasons, frequency, and nature of these updates. On top of this analysis, we build upon a knowledge graph to predict removed mappings ahead of time.**

***Index Terms*—NVD, Knowledge graphs, CVE, CWE, CPE.**

The National Vulnerability Database (NVD) is one of the largest and most comprehensive vulnerability databases that provides information about relationships between Common Vulnerabilities and Exposures (CVE) entries, related Common Weakness Enumeration (CWE), affected Common Platform Enumeration (CPE) and more. Software Composition Analysis (SCA) and threat modeling tools widely rely on the NVD either to detect or predict potential vulnerabilities in a system.

Some of the NVD mappings between CVE, related CWE and CPE can be incorrect or outdated, leading to removal of these mappings. Yet, so far no work has investigated the frequency and nature of these removals. The first contribution of this work is to perform a quantitative study of these removals. Moreover, the process of removing such mappings is a process that often takes time (i.e., several weeks or months). Hence, being able to automatically predict which mappings should be removed can help speed up this process. The second contribution of this work is to employ and refine techniques based on *knowledge graphs* [1] to perform such predictions.

For our first contribution, to understand the nature of removals from the NVD, we use the CVE history API [2] to obtain the update history for the CVE records that are removed in the November 2022 version of the database. To provide a baseline, we also randomly choose a set of CVE IDs from the database, and fetch update history for these as well. Our preliminary findings show that the number of events (the type of change that is made for the CVE record) for the two sets serve as an indicator for removal. Especially for the CPE-CVE mappings, the number of CPE Deprecation Remap events, that refer to a CPE version being removed due to deprecation, is dramatically larger than for the random set. Similarly, for CWE-CVE mappings, the removed mappings undergo Reanalysis, CWE Remap or Modified Analysis (which all result in CWE records being deleted or updated) considerably more often than mappings in the random set. Focusing on CWE-CVE mapping removals, we find that only 10% of the mappings are complete removals where a new CWE mapping is not added, while 5% of the removals result in a more specific remapping (e.g. CVE now mapped to a CWE item which is a child, member or peer of the old CWE), and 85% of removals result in a new CWE mapping where the new value is not related to the removed CWE value.

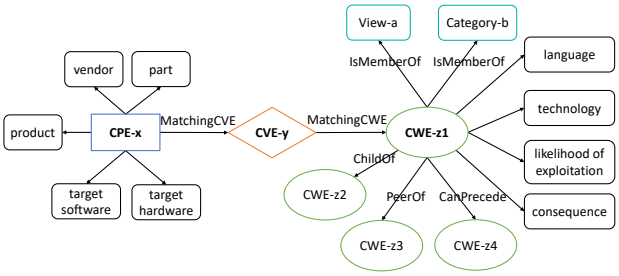

Fig. 1: Structure of the threat knowledge graph [1].

Next, for our second contribution, to predict removals of mappings, we train a knowledge graph using August 2021 data. A *knowledge graph* is a network of entities - i.e. objects, concepts and relationships. In our work we focus on CVE to CWE mappings and CVE to CPE mappings. Using the November 2022 version, we detect which mappings have been removed and use this set as our test set. Using several knowledge graph embeddings, we obtain results for the *Mean Reciprocal Rank (MRR)* metric. The MRR is a number between 0 and 1, where a higher MRR indicates better prediction performance. In our results, we see that the set of removed mappings have substantially lower MRR (≈ 0.2) than the set of random mappings (≈ 0.45). This means that the embedding models recognize weak links in the knowledge graph since they assign them a lower MRR. Using our preliminary results, we therefore argue that knowledge graphs can be employed to flag mappings that rank lower than a determined threshold, and tools relying on the NVD database can implement this methodology to improve their vulnerability detection process.

ACKNOWLEDGMENT

This work was supported in part by the Red Hat Collaboratory at Boston University, under grant #2023-01-RH17.

REFERENCES

[1] Z. Shi, N. Matyunin, K. Graffi, and D. Starobinski, "Uncovering product vulnerabilities with threat knowledge graphs," in *2022 IEEE Secure Development Conference (SecDev)*. IEEE, 2022, pp. 84–90.

[2] NVD, "Vulnerability apis," 2023. [Online]. Available: 'https://nvd.nist.gov/developers/vulnerabilities'