Comprehensive threat analysis and systematic mapping of CVEs to MITRE framework

Stefano Simonetto
University of Twente
s.simonetto@utwente.nl

Peter Bosch
University of Twente
h.g.p.bosch@utwente.nl

Abstract

This research addresses the significance of threat intelligence by presenting a practical approach to generate a labeled dataset for mapping CVEs to MITRE. By linking Common Vulnerabilities and Exposures (CVEs) with the MITRE ATT&CK framework, the paper outlines a scheme that integrates the extensive CVE database with the techniques and tactics of the ATT&CK knowledge base.

The core contribution lies in a detailed methodology designed to map CVEs onto corresponding ATT&CK techniques and, in turn, to tactics through a data-driven perspective, centering specifically on the labeling provided by NIST. This procedure enhances our understanding of cybersecurity threats and yields a structured, labeled dataset essential for practical threat analysis. It facilitates and improves the recognition and categorization of cybersecurity threats. Furthermore, the paper analyses the dataset in the context of cyber-threat intelligence. It highlights how vulnerability understanding and awareness have improved over the years through the continuous effort to place vulnerabilities in the context of an attack by linking it to abstract techniques.

The dataset allows for a comprehensive cyber attack stage and kill-chain analysis. It serves as a training resource for algorithm development in various use cases, such as threat detection and large language model fine-tuning.

1 Introduction

Over 25 years, from 1999 to 2023, the National Vulnerability Database (NVD) [1] maintained by the National Institute of Standards and Technology (NIST), has been a critical repository for cybersecurity information. During this extended period, the NVD has played a key role in documenting data on vulnerabilities across various systems, software, and technologies. A new CVE is generated each time a security flaw is identified in software or hardware and subsequently reported to the organization.

Despite their importance to the cybersecurity community, CVEs often lack specific guidance on countering identified vulnerabilities. This information gap becomes particularly crucial when considering the role of vulnerabilities in unlocking particular attack patterns. As pointed out by Sadlek et al. (2022), the timely identification of relevant threats before the attackers exploit is fundamental for proactive defense approaches. Sequences of adversarial actions that may evolve into attacks can be identified through multi-step attacks, which can be modeled using the kill-chain concept. This vision consists of ordered phases describing the attacker's progress in achieving objectives (Hutchins et al., 2011).

Natural language processing (NLP) and artificial intelligence (AI) can clarify the relationships between entities and events mentioned in text data. By contextualizing this information, these technologies help build a more comprehensive view of cyber threats and the actors behind them (Arazzi et al., 2023). Recently, indications of generative AI in cyber-threat intelligence have emerged (Ferrag et al., 2023). However, these applications require high-quality and substantial data for effective training to build their knowledge base.

The paper aims to establish a reliable foundation for correlating vulnerabilities with techniques and tactics by implementing a well-defined and structured pipeline. The main contributions of this paper are:

- The creation of a comprehensive dataset, employing a systematic conservative approach to map from CVEs to MITRE techniques and tactics;
- An in-depth examination of vulnerabilities, clarifying their associations with CWEs and the subsequent link to the MITRE framework.

[1] https://nvd.nist.gov/

The resulting dataset extends to threat intelligence, where it aids analysts in identifying potential risks, while also enabling better comprehension of kill chains and the identification of techniques used by adversaries to more effectively defend against attacks.

2 Background and taxonomy

Understanding and addressing vulnerabilities is essential to strengthen applications effectively. Threat identification uses multiple risk factors to prioritize threats according to their severity by using the multiple risk factors and calculating the threat prioritization value, which represents the severity level of the threat (Ma et al., 2009). However, protecting digital assets from potential threats and attacks is a constant challenge that demands expertise and a comprehensive understanding of the company's environment.

As shown in Hemberg et al. (2020), it is possible to go from a CVE to the related techniques and tactics following the path of CVE-CWE-CAPEC-ATT&CK. Before explaining this framework in more detail, we describe each pipeline component.

2.1 CVE

The Common Vulnerabilities and Exposures (CVEs) are unique identifiers assigned to publicly known cybersecurity vulnerabilities. These identifiers help security professionals and organizations communicate about specific weaknesses, ensuring that everyone refers to the same vulnerability with a common name. CVEs are essential for knowledge-sharing, enabling researchers and vendors to collaborate and develop appropriate patches or mitigations to protect systems from potential exploitations. Unfortunately, vulnerabilities can be complex, involving intricate technical details such as specific products and versions.

2.2 CWE

Common Weakness Enumeration (CWE)[2] is a community-developed list of common software weaknesses and security flaws. Unlike CVEs, which identify specific vulnerabilities, CWEs categorize broader classes of weaknesses, embracing various instances of similar vulnerabilities. This classification aids in understanding the root causes of vulnerabilities, facilitating more comprehensive security measures during software development and system deployment. CWEs explain how (conditions and procedures), why a vulnerability can be exploited (cause), and explain the consequences (impact) (Aghaei et al., 2020).

2.3 CAPEC

The Common Attack Pattern Enumeration and Classification (CAPEC)[3] provides a publicly available catalog of common attack patterns that helps users understand how adversaries exploit weaknesses in applications and other cyber-enabled capabilities. CAPEC defines "Attack Patterns" as descriptions of adversaries' common attributes and approaches to exploit known weaknesses in cyber-enabled capabilities. Each attack pattern captures knowledge about how specific parts of an attack are designed and executed and provides guidance on mitigating the attack's effectiveness.

2.4 ATT&CK framework

MITRE ATT&CK is a curated knowledge base and model for cyber adversary behavior, reflecting the various phases of an adversary's attack lifecycle and the platforms they are known to target (MITRE, 2023). It originated from a project to document and categorize post-compromise adversary tactics, techniques, and procedures (TTPs) against Microsoft Windows systems to improve the detection of malicious behavior (Strom et al., 2018). Currently, the framework has been extended to a broad spectrum of environments. At its core, ATT&CK is a behavioral model comprising tactics that denote short-term adversary goals, techniques delineating how these goals are achieved, sub-techniques offering more specific methods at a lower level, and documented adversary usage encompassing procedures and metadata.

The MITRE ATT&CK framework can be used for cyber-threat intelligence enrichment, SOC assessment, defensive gap assessment, behavioral analytics development, red teaming, and adversary emulation.

3 Dataset creation

The main contribution of this paper is the creation of a dataset that links CVEs to MITRE techniques and tactics. The knowledge deriving from CVEs, CWEs, CAPEC and ATT&CK is fragmented, and the available data are disconnected. It seems that

[2] https://cwe.mitre.org/

[3] https://capec.mitre.org/

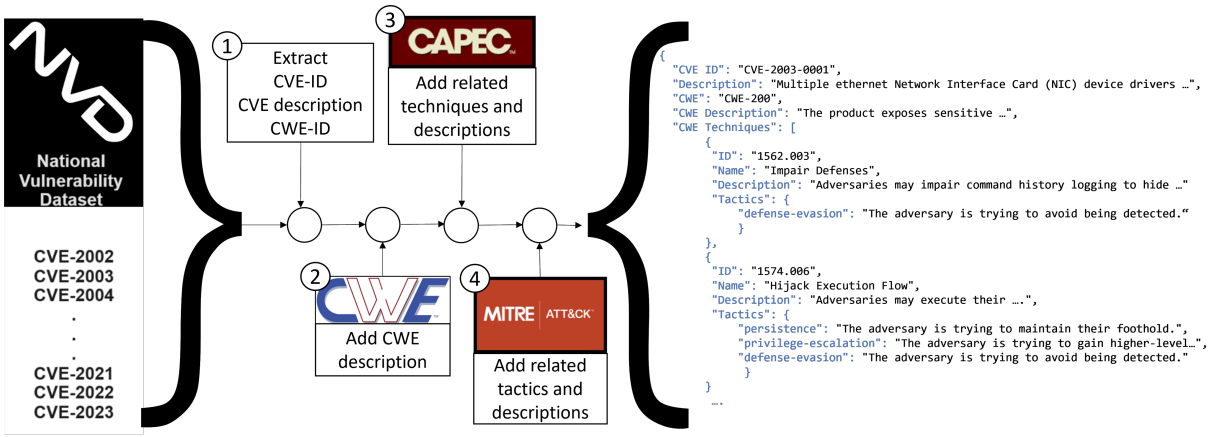

Figure 1: Pipeline of dataset formation

all these organizations are working in sealed environments, missing the bigger picture, to which a vulnerability can be useful to trace back a specific step in the cyber kill chain.

For example, CVEs serve as unique identifiers for publicly recognized cybersecurity vulnerabilities, whereas CWEs aim to abstract and categorize CVEs. Although both frameworks have distinct objectives, combining their knowledge allows us to comprehend the wider context.

To this purpose, we designed a pipeline to systematically retrieve the tactics and techniques associated with any known CVE. This leads to the largest dataset where CVEs are linked with tactics. Other works are proposing ways to achieve the same task as described in section 5, but our approach poses some constraints over the linking from CVEs to tactics to avoid an exploding surface:

1. We adopted only the NIST labeling from CVEs to CWEs: since NIST has to manually label CVEs coming from CNAs (CVE Numbering Authorities), we decided that adopting their labeling was the most neutral approach. If NIST did not provide any labeling, we adopted the labeling from the CNA. NIST is mapping CVEs to CWEs according to "Weaknesses for Simplified Mapping of Published Vulnerabilities." This subset of CWEs was selected through coordination between the NVD and the CWE teams.

2. We avoided linking CWEs between each other: to prevent an exploding attack surface, we chose the strictest approach, avoiding interlinking between CWEs. This decision is rooted in the observation that the relationship from CWEs to techniques, and subsequently from techniques to tactics, is typically not one-to-one but one-to-many. In the realm of threat intelligence, false negatives are dangerous, but also false positives have to be considered.

We got a ground truth dataset that can be used as a baseline for multiple purposes by relying on entities, e.g., MITRE, NIST, etc. The final dataset is available online (Simonetto). The implementation stages are depicted in Fig. 1 and can be summarized in the following subsections.

3.1 Retrieving CVE information

We downloaded data from NVD repository (nvd) and parsed it to extract only CVE ID, CVE description, and CWE ID. To do so, we discarded information that could not be used for the mapping, e.g., Common Platform Enumeration (CPE), impact, CVSS, references, assigners, and others. CVEs that have been assigned a CVE ID but subsequently rejected for any reason are not considered. An example is shown in Listing 3.1, based on data retrieved on 23-1-2024.

```
"CVE ID": "CVE-2023-0001",
"Description": "An information exposure
                vulnerability in...",
"CWE": "CWE-319"
```

3.2 Adding CWE descriptions

Enhance the CWEs by integrating corresponding descriptions that are neither deprecated nor overly general. The CWEs within the dataset span various levels of abstraction, from Pillars, which represent the highest level of abstraction, to more specific classifications, such as classes, bases, and variants, each offering a finer-grained description of

the CWE. An example illustrating the raw format of the data is present at Listing 3.2.

```
"CWE-ID":"319",
"Name":"Cleartext Transmission
       of Sensitive Information",
"Weakness abstraction":"Base",
"Description":"The product transmits
           sensitive or security-critical.."
```

3.3 Bridging CWEs to techniques

CAPEC provides a comprehensive list of attack patterns, each associated with a name, an ID, a description, and other pertinent details that facilitate a deeper understanding of the attack type. These details include the associated CWEs and techniques for each attack pattern. In this study, CAPEC is used to link CWEs to techniques. An example of raw CAPEC data used in this context (Listing 3.3).

```
"CAPEC-ID":"383",
"Name": "Harvesting Information via API
       Event Monitoring",
"Abstraction": "An adversary hosts an
            event within an application..",
"Related weaknesses":"311, 319, 419, 602",
"Technique-ID": "1056.004",
"Technique": "Credential API Hooking"
```

3.4 Linking the related tactics

To complete the analysis, the final step involves establishing connections between techniques and their corresponding MITRE tactic(s). This linkage is crucial for understanding how specific techniques contribute to broader strategic objectives in cybersecurity. By mapping techniques to relevant MITRE tactics, we gain insights into the strategic context in which these techniques are deployed (Listing 3.4).

```
"ID": "1056.004",
"Name": "Input Capture",
"Description": "Adversaries may hook ...",
"Tactics":
    "collection": "The adversary is
                trying to gather data ...",
    "credential-access": "The adversary
                        is trying to steal..."
```

It is important to acknowledge that the NVD database's organizational structure spans from 1999 to the present. We exclusively use CWE descriptions that are neither category nor deprecated, following the specifications provided by CWE-MITRE: "Category is simply a collection of similar weaknesses that do not all share the same combination of the dimensions, so a category should not be used for mapping". For instance, since CWE-388 is categorical, it should not be used for mapping, so the related CWE description is set to unknown. A big gap in the mapping from CVEs to CWE is related to the CVEs that, according to NIST, do not have enough information about the issue to classify it; details are unknown or unspecified. Only during the last year were more than 15% of all CVEs labeled as no-info from NIST. The final observation concerns the absence of connections between CWEs and techniques. CAPEC does not define links for all the CWEs listed in the CWE database, resulting in a significant gap. This gap is noteworthy when considering the overall count of CVEs (without the rejected ones), as only 19,40% have a comprehensive mapping to the corresponding technique(s) and, consequently, to the associated tactic(s) as shown in Eq. 1.

$$\text{CVE to TTPs} = \frac{\text{Entries with technique}}{\text{Total entries}} \times 100 \tag{1}$$

In the development of a threat intelligence dataset, our primary objective is to establish correlations between vulnerabilities and MITRE tactics. Unlike traditional approaches that heavily rely on expert viewpoints, our methodology prioritizes integrating information from four key sources: CVE, CWE, CAPEC and MITRE.

4 Dataset analysis

In addition to making the dataset accessible, we perform an analysis utilizing the insights gathered. Initially, we visualize the yearly distribution of disclosed vulnerabilities, as depicted in Fig. 2, totaling 236,071 CVEs across all years. The visual representation shows an upward trajectory in the annual number of vulnerabilities. This escalating trend implies a continual growth in the overall

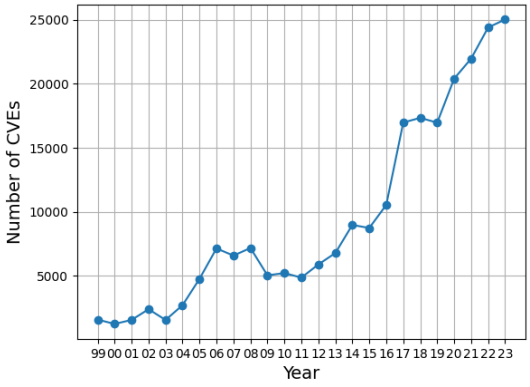

Figure 2: Number of vulnerabilities per year

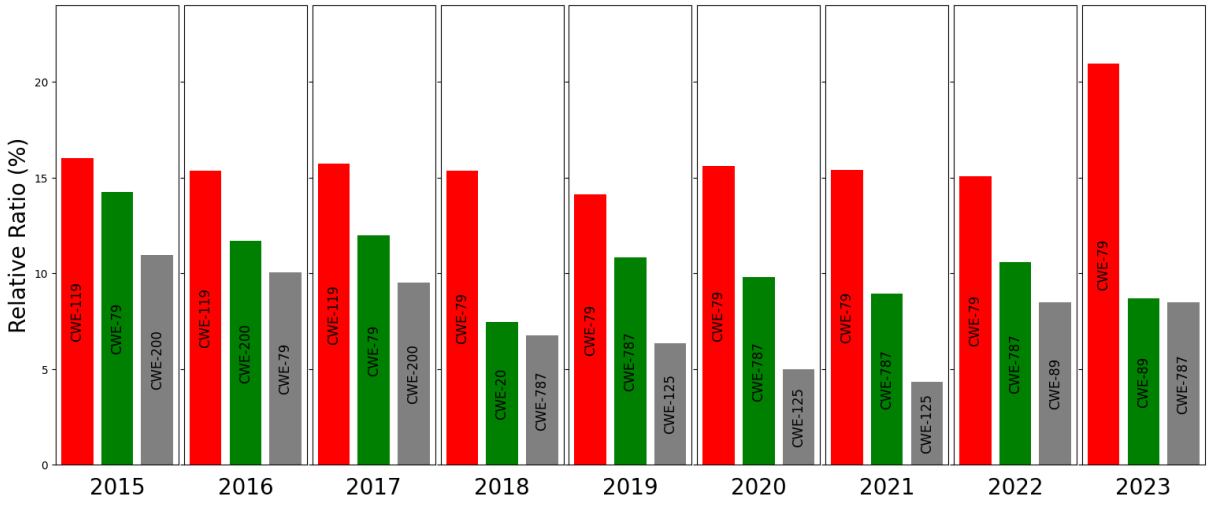

Figure 3: Most common CWEs per year

number of security vulnerabilities discovered and reported over the years. As vulnerabilities proliferate, they contribute to broadening the attack surface, highlighting an increasing array of potential points through which adversaries can exploit weaknesses in systems or applications. This expanding attack surface poses challenges for cybersecurity measures, requiring organizations to adapt and enhance their defenses to address the evolving threat landscape effectively.

Furthermore, we conduct an in-depth analysis of the mapping from CVEs to CWEs. We visualize the data by plotting the top three CWEs identified for each year, as shown in Fig. 3. The findings paint a familiar picture, portraying recurrent patterns in the prevalent vulnerabilities. This observation reveals that certain weaknesses consistently appear to be the leading contributors to security concerns across different time frames. Such insights into the recurring CWEs aid in understanding persistent challenges, guiding efforts toward targeted mitigation strategies, and reinforcing cybersecurity measures against well-established vulnerabilities.

Observing the graph, it is clear that CWE-79 (Improper Neutralization of Input During Web Page Generation) has consistently maintained its status as the most prevalent vulnerability over the last six years. This vulnerability manifests when the application fails to neutralize or incorrectly neutralizes user-controllable input before incorporating it into output, which subsequently serves as a web page to other users. An example of CWE-79 is presented as follows:

```
<body>
 <h1>Welcome <?php echo $_GET['name'];?>
 </h1>
</body>
```

The web application takes a user-supplied input parameter name from the query string and directly echoes it back into the HTML response without validation or sanitization. This creates a vulnerability because if an attacker crafts a URL such as: `http://example.com/welcome.php?name=<script>alert('XSS')</script>`, the script tag will be executed when the page is loaded in a victim's browser, leading to a Cross-Site Scripting attack.

The other relevant during the timeframe taken into account are:

1. CWE-89: "Improper Neutralization of Special Elements used in an SQL Command ('SQL Injection')";
2. CWE-787: "Out-of-bounds Write";
3. CWE-125: "Out-of-bounds Read";
4. CWE-20: "Improper Input Validation";
5. CWE-200: "Exposure of Sensitive Information to an Unauthorized Actor".

Digging deeper into our analysis, we extended our investigation by establishing a mapping between the CWEs and the corresponding techniques documented in the CAPEC mapping. This complex mapping allowed us to connect vulnerabilities

with specific attack techniques, providing a more comprehensive understanding of the potential exploits associated with each weakness. By bridging the gap between CWEs and techniques, we gained valuable insights into how adversaries may leverage identified weaknesses to carry out sophisticated cyber attacks.

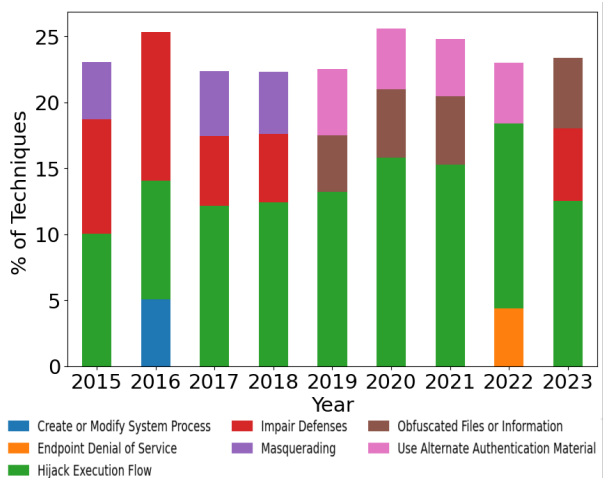

Figure 4: Techniques unlocked by CVEs

After analyzing attack techniques (Fig. 4), we consistently find that "Hijack Execution Flow" emerges as the most frequent technique that attackers can employ. Additionally, in the connection between techniques and tactics, where "Defense evasion" notably stands out as the most prominent tactic that malicious actors may utilize (Fig. 5).

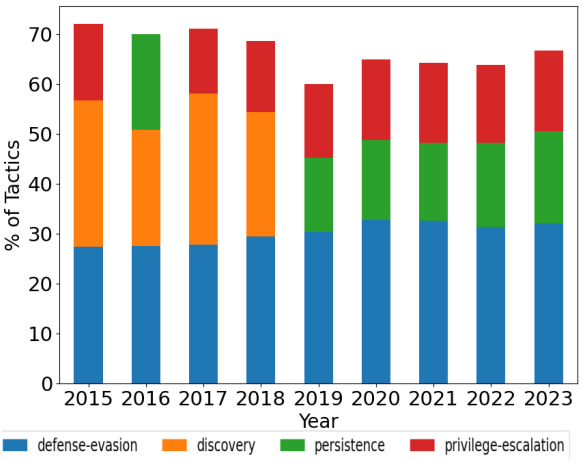

Figure 5: Tactics unlocked by CVEs

We want to emphasize that these findings do not represent what malicious actors employ daily to perform attacks. Instead, they reveal the potential exploits enabled by vulnerabilities that adversaries may leverage. For visualization purposes, we constrained the timeframe from 2015 to 2023 inclusive.

5 Related work

Comprehensive threat intelligence datasets, especially those focused on vulnerabilities, are crucial for cybersecurity research. Understanding and analyzing vulnerabilities are key for fortifying digital systems. Quality data is essential for training machine learning models, enabling them to capture intricate patterns in real-world cybersecurity scenarios (Ferrag et al., 2023). Our dataset establishes a foundation that does not depend on external experts for mapping CVEs to MITRE, unlike previous approaches such as the one proposed by Grigorescu et al. (2022). This aims to provide a more unbiased and objective basis for threat intelligence analysis. This section provides an overview of existing research on threat intelligence datasets.

Vulnerabilities have been thoroughly examined in previous research, Ozment (2007) conducted an in-depth study and analysis of the National Vulnerability Database (NVD), highlighting various limitations. More recently also, Glyder et al. (2021) focuses on a basic analysis of vulnerabilities and scores from the NVD. Data sources about vulnerabilities are widespread, and the most used for threat identification mostly come from two datasets, one from ENISA (2019) and the other that can be extracted from BRON (Hemberg et al., 2020).

5.1 ENISA dataset

In December 2019, the European Union Agency for Cybersecurity (ENISA) released a report titled "State of Vulnerabilities 2018/2019". This report sought insights into the opportunities and constraints within the vulnerability ecosystem. A comprehensive collection of 27,471 pieces of vulnerability information, spanning from January 1, 2018, to September 30, 2019, was compiled from diverse data sources. While analyzing this data, the authors correlated CVEs with MITRE ATT&CK techniques by utilizing shared information from the CAPEC found in both the National Vulnerability Database and ATT&CK. Within the ENISA report dataset, there were 8,077 CVEs identified, corresponding to 52 distinct MITRE ATT&CK techniques or, in this context, labeled instances (Katos et al., 2019). Articles such as (Mendsaikhan et al., 2020), (Lakhdhar and Rekhis, 2021), and (Mendsaikhan et al., 2021) are adopting the ENISA dataset. Mendsaikhan et al. (2020) describes a

method to automatically map software vulnerability using a multi-label classification approach. The authors took the vector representation of the vulnerability description and classified it with various multi-label classification methods to evaluate it in different measures. They found the LabelPowerset method with Multilayer Perceptron. Lakhdhar and Rekhis (2021) provides a multilabel classification approach to automatically map a detected vulnerability to the MITRE tactics that the attacker could use. The authors evaluate machine-learning algorithms (BinaryRelevance, LabelPowerset, ClassifierChains, MLKNN, BRKNN, RAkELd, NLSP, and Neural Networks).

5.2 BRON

In February 2021, Hemberg et al. (2020) published BRON set the standard for the systematic mapping from CVEs to MITRE tactics. BRON is a relational graph that depicts entries from various information sources as distinct types of nodes, and their interconnections are illustrated as edges. Unidirectional links in the sources are identified and portrayed as bidirectional connections within BRON's graph. By leveraging BRON, Abdeen et al. (2023) present a tool that automatically maps CVE entries to ATT&CK techniques based on their textual similarity. SMET achieves this mapping by leveraging ATT&CK BERT, a model that the authors trained using a siamese network architecture as described by SBERT (Reimers and Gurevych, 2019). This works by taking two sentences as input, extracting each sentence embedding using BERT, and then optimising the network weights to maximise the similarity of the two embeddings if the sentences are semantically similar. Another approach, such as the one proposed by Ampel et al. (2021), uses only a subset of the entire dataset made available by BRON. They leveraged a dataset of 24,863 CVEs into 10 of the 14 ATT&CK tactics.

5.3 Runtime comparison

One of the strengths of BRON's approach is bidirectionality because data retrieval from CVEs is possible through tactics and vice versa. This complexity comes to the cost of time-retrieval. Furthermore, the connection between CWEs that are related together leads to an exploding surface of applicable techniques. Considering these factors, our approach significantly enhances the speed of retrieving TTPs related to BRON, focusing only on TTPs relevant to the actual CWE. Our approach's retrieval time is noteworthy for its efficiency, enabling quick and straightforward access to techniques and tactics. To quantify this, we conducted 10 runs and calculated the average time required to retrieve a technique for a selected CVE ('CVE-2023-0001'). Our approach demonstrated a significantly faster performance, with an average retrieval time of only 0.46 seconds per technique, compared to an average of 53.45 seconds per technique with BRON. Additionally, for the same CVE, our approach retrieves only the two techniques strictly related to the CWE, whereas BRON retrieves 84 different techniques.

5.4 Other approaches

Mendsaikhan et al. (2021) describe a method to map the cyber-threat information using a multi-label classification approach. The authors conducted four experiments using three publicly available datasets to train and test seven multi-label classification methods and one pre-trained language model in six evaluation measures. Other than the already cited ENISA dataset, this approach uses two other datasets:

1. TRAM: Threat Report ATT&CK Mapping (TRAM) is a tool developed by MITRE to aid the analyst in mapping finished reports to ATT&CK. TRAM uses a Logistic Regression model to predict the mapping of the ATT&CK technique for a given report. MITRE released the source code and the corresponding dataset used to train the model (for Threat-Informed Defense, 2024). The dataset contains example sentences or phrases representing specific techniques and maps them to one or more techniques. The TRAM dataset represents the short threat information in sentences or phrases. It has 3,005 example sentences mapped to 188 unique MITRE ATT&CK techniques.

2. rcATT: Legoy et al. (2020) implemented a tool called rcATT, a system that predicts tactics and techniques related to given cyber-threat reports. They collected the threat reports referenced in the original MITRE ATT&CK framework per each technique to train the tool. They made their source code and the parsed threat reports publicly available. The rcATT represents the long descriptive information in the form of threat reports. It has 1,490 ex-

ample reports mapped to 227 unique MITRE ATT&CK techniques.

5.5 Unsupervised learning

Researchers have expanded their investigations into vulnerability analysis to incorporate advanced techniques, with a significant emphasis on unsupervised machine learning. Kuppa et al. (2021) proposed a multi-head joint embedding neural network model to automatically map CVEs to ATT&CK techniques. They address the problem of the lack of labels for this task using a novel, unsupervised labeling technique. For the labeling process to be successful, they had to measure the similarity/dissimilarity of ATT&CK technique candidate vectors and CVE description representations. They manually label randomly sampled 200 CVEs found in threat reports with their corresponding ATT&CK techniques and, extract the context phrases, and create candidate vectors.

6 Conclusion

As highlighted by Aota et al. (2020), the labeling process of reports with vulnerability identifiers has thus far been performed manually and has, therefore, suffered from scalability issues due to the shortage of security experts. The versatility of the proposed dataset makes it invaluable for a wide range of applications, showcasing its adaptability and utility across various domains. Its applicability extends to threat intelligence, where analysts can leverage the data to enhance their understanding of potential risks and vulnerabilities. The dataset's rich content and diverse sources provide a comprehensive view of the threat landscape, aiding in the identification and mitigation of potential cyber threats.

Moreover, the dataset is well-suited for kill-chain concatenation, enabling the mapping and analysis of different stages in a cyber attack. This facilitates a more holistic approach to cybersecurity, allowing practitioners to identify patterns, vulnerabilities, and attack vectors throughout the entire kill chain. This insight is crucial for developing effective defense strategies and proactive measures against evolving cyber threats. As highlighted by Kuppa et al. (2021), understanding the attacker's choice of vulnerability for a particular attack stage is a hard problem.

In machine learning and artificial intelligence, the dataset is a valuable resource for training models. Its extensive nature allows for the development of robust machine-learning algorithms capable of recognizing and predicting patterns within complex data. Researchers and developers can refine and enhance the models' language understanding capabilities by exposing language models to a broad range of scenarios and contexts present in the dataset.

7 Limitations

Challenges often arise when dealing with vulnerabilities and weaknesses. The NVD-CWE-noinfo category reflects situations where issues lack adequate details for classification, leaving key information unknown or unspecified. Similarly, the NVD-CWE-Other classification marks that the NIST employs only a specific subset of CWEs for mapping, omitting certain weakness types not covered by this subset. Furthermore, some CAPEC to ATT&CK mappings are absent due to unprovided information from the source. Recognizing the need for advancement, NIST has announced plans to retire all legacy data feeds by 2024, emphasizing a transition to updated application programming interfaces (APIs) to enhance the accuracy and comprehensiveness of vulnerability data. By design choice, we avoid mapping to deprecated or category CWEs, as MITRE suggested. Deprecated CWEs were originally used but introduced unnecessary complexity and depth, while category CWEs are not weaknesses but rather a view that provides a comprehensive categorization and, therefore, inappropriate to describe the root causes of vulnerabilities. The main limitation of this paper is the absence of connections between CWEs and techniques, as highlighted in Section 3. CAPEC does not define links for all the CWEs listed in the CWE database, resulting in a significant gap. This gap is substantial when considering the overall count of CVEs, as only 19,40% have a comprehensive mapping to the corresponding technique(s).

Ethics statement

As the creators of this dataset, we have mapped CVEs to ATT&CK tactics, showing which step an attacker can potentially take. We believe that the benefits of open-source collaboration outweigh the risk of possible misuse by individuals with malicious intent. It enables cybersecurity professionals and researchers to enhance defense strategies and improve overall security posture. We are committed to fostering responsible usage of this dataset within the cybersecurity community, promoting

transparency and ethical practices to maximize its positive impact while minimizing potential harm.

Acknowledgment

We thank the Twente University Centre for Cybersecurity Research (TUCCR) for their essential support and resources, which greatly contributed to this research. Their collaborative environment and guidance were crucial to the completion of this work.

References

Nvd data feeds. https://nvd.nist.gov/vuln/data-feeds. Accessed on: 23-1-2024.

Basel Abdeen, Ehab Al-Shaer, Anoop Singhal, Latifur Khan, and Kevin Hamlen. 2023. Smet: Semantic mapping of cve to att&ck and its application to cybersecurity. In *IFIP Annual Conference on Data and Applications Security and Privacy*, pages 243–260. Springer.

Ehsan Aghaei, Waseem Shadid, and Ehab Al-Shaer. 2020. Threatzoom: Hierarchical neural network for cves to cwes classification. In *International Conference on Security and Privacy in Communication Systems*, pages 23–41. Springer.

Benjamin Ampel, Sagar Samtani, Steven Ullman, and Hsinchun Chen. 2021. Linking common vulnerabilities and exposures to the mitre att&ck framework: A self-distillation approach. *arXiv preprint arXiv:2108.01696*.

Masaki Aota, Hideaki Kanehara, Masaki Kubo, Noboru Murata, Bo Sun, and Takeshi Takahashi. 2020. Automation of vulnerability classification from its description using machine learning. In *2020 IEEE Symposium on Computers and Communications (ISCC)*, pages 1–7. IEEE.

Marco Arazzi, Dincy R. Arikkat, Serena Nicolazzo, Antonino Nocera, Rafidha Rehiman K. A., Vinod P., and Mauro Conti. 2023. Nlp-based techniques for cyber threat intelligence.

ENISA. 2019. Enisa's state of vulnerabilities 2018/2019 report. (Accessed January 19, 2024).

Mohamed Amine Ferrag, Merouane Debbah, and Muna Al-Hawawreh. 2023. Generative ai for cyber threat-hunting in 6g-enabled iot networks. In *2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing Workshops (CCGridW)*, pages 16–25.

Center for Threat-Informed Defense. 2024. The center for threat-informed defense. https://github.com/center-for-threat-informed-defense/tram. Accessed: 2024-06-19.

Jillian Glyder, Andrew Kyle Threatt, Randy Franks, Lance Adams, and Geoff Stoker. 2021. Some analysis of common vulnerabilities and exposures (cve) data from the national vulnerability database (nvd). In *Proceedings of the Conference on Information Systems Applied Research ISSN*, volume 2167, page 1508.

Octavian Grigorescu, Andreea Nica, Mihai Dascalu, and Razvan Rughinis. 2022. Cve2att&ck: Bert-based mapping of cves to mitre att&ck techniques. *Algorithms*, 15(9):314.

Erik Hemberg, Jonathan Kelly, Michal Shlapentokh-Rothman, Bryn Reinstadler, Katherine Xu, Nick Rutar, and Una-May O'Reilly. 2020. Linking threat tactics, techniques, and patterns with defensive weaknesses, vulnerabilities and affected platform configurations for cyber hunting. *arXiv preprint arXiv:2010.00533*.

Eric M Hutchins, Michael J Cloppert, Rohan M Amin, et al. 2011. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *Leading Issues in Information Warfare & Security Research*, 1(1):80.

V Katos, S Rostami, P Bellonias, N Davies, A Kleszcz, S Faily, A Spyros, A Papanikolaou, C Ilioudis, and K Rantos. 2019. State of vulnerabilities 2018/2019. *European Union Agency for Cybersecurity (ENISA), Technical Report*.

Aditya Kuppa, Lamine Aouad, and Nhien-An Le-Khac. 2021. Linking cve's to mitre att&ck techniques. In *Proceedings of the 16th International Conference on Availability, Reliability and Security*, pages 1–12.

Yosra Lakhdhar and Slim Rekhis. 2021. Machine learning based approach for the automated mapping of discovered vulnerabilities to adversial tactics. In *2021 IEEE Security and Privacy Workshops (SPW)*, pages 309–317. IEEE.

Valentine Legoy, Marco Caselli, Christin Seifert, and Andreas Peter. 2020. Automated retrieval of att&ck tactics and techniques for cyber threat reports. *arXiv preprint arXiv:2004.14322*.

Jie Ma, Zhi-tang Li, and Hong-wu Zhang. 2009. A fusion model for network threat identification and risk assessment. In *2009 International Conference on Artificial Intelligence and Computational Intelligence*, volume 1, pages 314–318. IEEE.

Otgonpurev Mendsaikhan, Hirokazu Hasegawa, Yukiko Yamaguchi, and Hajime Shimada. 2020. Automatic mapping of vulnerability information to adversary techniques. In *The Fourteenth International Conference on Emerging Security Information, Systems and Technologies SECUREWARE2020*.

Otgonpurev Mendsaikhan, Hirokazu Hasegawa, Yukiko Yamaguchi, and Hajime Shimada. 2021. Automatic mapping of threat information to adversary techniques using different datasets. *International Journal*

on Advances in Security Volume 14, Number 1 & 2, 2021.

MITRE. 2023. ATT&CK. https://attack.mitre.org/.

James Andrew Ozment. 2007. *Vulnerability discovery & software security*. Ph.D. thesis, University of Cambridge.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Lukáš Sadlek, Pavel Čeleda, and Daniel Tovarňák. 2022. Identification of attack paths using kill chain and attack graphs. In *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium*, pages 1–6. IEEE.

Stefano Simonetto. CVE to MITRE Dataset. https://github.com/stefanosimonetto/data_CVE_MITRE, year = 2024,.

Blake E Strom, Andy Applebaum, Doug P Miller, Kathryn C Nickels, Adam G Pennington, and Cody B Thomas. 2018. Mitre att&ck: Design and philosophy. In *Technical report*. The MITRE Corporation.