# Linkage Between CVE and ATT&CK with Public Information

Tomoaki Mimoto[1], Yuta Gempei[1], Kentaro Kita[1], Takamasa Isohara[1],
Shinsaku Kiyomoto[1] and Toshiaki Tanaka[2]
[1] *KDDI Research, Inc., 2–1–15 Ohara, Fujimino-shi, Saitama, 356–8502, Japan*
[2] *University of Hyogo, 7–1–28, Minatojima-minamimachi, Chuo-ku Kobe, Hyogo 650–0047, Japan*

Keywords: CVE, ATT&CK, NVD, CWE, CAPEC, LLM.

Abstract: Establishing rapid and effective cyber threat intelligence collection and analysis methodologies are required to counter the rapidly growing sophistication of cyberattacks. The overview of known vulnerability information and related information can be found in databases such as NVD. However, the relationship between vulnerabilities and TTPs, which are effective CTIs, must be analyzed individually by experts, and many of these relationships are unknown. In this study, we attempt to connect vulnerability information keyed to CVE-IDs with the ATT&CK, which is a knowledge base for TTPs. Specifically, vulnerability information and techniques associated with ATT&CK are each put into an embedding representation with related information, and the similarities between them are evaluated to estimate the techniques related to the CVE-IDs. This study considers the reproducibility problem due to the lack of ground truth in the cybersecurity field by handling only information available from the surface Web.

## 1 INTRODUCTION

Common Vulnerabilities and Exposures (CVE) provide vulnerability information, and each vulnerability is assigned a CVE-ID mainly by MITRE, a non-profit organization in the US. The number of reported vulnerabilities that could be used in cyberattacks has continued to increase rapidly since 2017, and what's more, this is just the tip of the iceberg[1]. The traditional reactive approach to cyberattacks is becoming antiquated, and a change to the proactive countermeasures against potential attacks is being demanded.

Tactics, techniques, and procedures (TTPs) are a concept related to vulnerability threat assessment that focuses on the principles of attacker behavior and attack scenarios. TTPs represent the most fundamental idea of attacers, and analyzing and evaluating TTPs are considered to be the most important for understanding the nature of cyberattacks. Understanding TTPs can facilitate an understanding of the attacker's behavior and help in determining policies for defense. Early identification of potential ways that an attacker can exploit a vulnerability, and knowing where it is in the attack life-cycle, will lead to accurate vulnerability assessments. Therefore, linking CVEs and TTPs to predict attack scenarios that exploit the vulnerabilities, estimate risks, and prioritize responses is expected to be one of the proactive measures against cyber threats.

Prior to the advocacy of the pyramid-of-pain, the ATT&CK framework was created by MITRE, a US non-profit organization, based on the fact that the techniques attackers exploit converge to some extent with the constraints of the target object. ATT&CK is a knowledge base that organizes TTPs and can be used for developing threat models, determining countermeasures, and active threat hunting in cybersecurity. There are several databases related to cyberattacks and vulnerabilities, such as the National Vulnerability Database (NVD) managed by MITRE in addition to ATT&CK, Common Weakness Enumerations (CWE), and Common Attack Pattern Enumeration and Classification (CAPEC) managed by NIST. For reproducibility, this paper constructs data for evaluation from structured databases and discusses the representativeness of vulnerability using LLM. Our experiments showed that aggregating multiple pieces of connectable information improves expressiveness, which suggests the possibility of further enhancing expressiveness through the use of unique and superior information.

[1] https://flashpoint.io/blog/vulndb-uncovers-hidden-vulnerabilities-cve/

Mimoto, T., Gempei, Y., Kita, K., Isohara, T., Kiyomoto, S. and Tanaka, T.
Linkage Between CVE and ATT&CK with Public Information.
DOI: 10.5220/0012722600003767

In *Proceedings of the 21st International Conference on Security and Cryptography (SECRYPT 2024)*, pages 655-660
ISBN: 978-989-758-709-2; ISSN: 2184-7711

## 1.1 Research Topic

There are many challenges to collecting and analyzing cyber threat information to generate Cyber Threat Intelligence (CTI) (Rahman et al., 2023), and we consider two issues here. The first is that the large amount of threat information in its various forms is not well coordinated with each other. Even in structured databases managed by organizations, information may be missing or insufficiently coordinated. For example, connecting vulnerability information with TTPs is considered effective for accurate risk assessment and attack detection, but these connection methods have not been established, and there is little existing research. The second is the scarcity of correct and reproducible available data in the cybersecurity field, for example, there are few tagged corpora. In the previous study (Kuppa et al., 2021), experts manually link CVE-IDs to techniques as the correct answer data, and the results are not disclosed publicly. In addition, it uses multiple non-public information generated by experts and is not reproducible. Rahman et al. also show that the sources of information handled by existing research are often unclear (Rahman et al., 2023). This poses issues of generality and reproducibility, and may hinder the development of the cybersecurity field as a whole.

In this study, we target ATT&CK as information that is not connected clearly to vulnerability information and examine the linkage method between CVE and ATT&CK. We also explain a method to extract some of the linked data between CVE-ID and technique from publicly available information as the correct data, and guarantee the reproducibility of the experiment.

## 1.2 Contribution

This study discusses the possibility of enriching vulnerability information by improving the embedding representation. Vulnerabilities assigned a CVE-ID can be connected to multiple pieces of information by traversing public databases, which can be leveraged to improve embedding representation of vulnerability information. Furthermore, a similar embedding representation can be used to map related information that is not easily associated with the vulnerability to the same space as the vulnerability information, and the similarity of the information enables the linkage between them. In this study, we utilize CWE and CAPEC as information that can be connected to CVE-IDs starting from NVD, and discuss the possibility of connecting CVE-IDs to the ATT&CK technique by evaluating the similarity. In our experiments, we have confirmed that by combining information, it is possible to connect CVE-IDs to TTP chains, which are the clusters of related techniques, with about 87% accuracy, even from only the most basic information sources.

This study also argues for the need to ensure generality and reproducibility through the use of publicly available data. This allows comparison of the methods themselves, independent of the value of the data used, and encourages the development of research in the field of cybersecurity. The problem of lack of ground truth is a challenge in the field of cyber security. In this research target, we construct a dataset by extracting necessary information from public information, and guarantee that all information about the experiment can be reconstructed from public data only.

# 2 DATASET

## 2.1 Public Dataset

In the cyber security field, public databases have been established to share information among organizations in order to combat the vast number of vulnerabilities and attacks that exploit them. In this study, we utilize NVD, CWE, CAPEC, and ATT&CK, which are used by security vendors.

NVD provides an overview of each vulnerability and exposure, including URLs with related information, the organization that registered the CVE, CVSS score, related CWE-ID, affected software version, and update history. In this study, we use information that can be extracted from the NVD regarding vulnerabilities with IDs assigned in 2023 as of 1/23/2024. CWE is a database that systematizes vulnerability types, and in version 4.13, vulnerability types are classified into 934 categories, each of which is assigned a CWE-ID as an identifier. CWE provides an overview of each vulnerability type and related vulnerability types and CAPEC-IDs. CWE is organized from multiple viewpoints such as software development and hardware design, and we use the CWE-1000 dataset, which contains all vulnerability types organized from the perspective of research objectives. CAPEC is a database that systematizes attack patterns, and in version 3.9, there are 559 types, each with a CAPEC-ID as an identifier. CAPEC provides an overview of each attack pattern, the related attack patterns and CWE-ID, the prerequisites for a successful attack, the attack flow, and mitigation measures. CAPEC, as well as CWE, is organized in multiple views, and we use CAPEC-1000, which includes all attack patterns, as our dataset. ATT&CK is a knowledge base that organizes TTPs and can be used for developing threat models, determining countermeasures, and active threat hunting in cyber security. In ATT&CK, as in other structured databases, identifiers are assigned to each of the TTPs and we focus on the techniques of the enterprise field. Technique has sub-techniques for further details, and in v14, they are organized into 201 techniques and 424 sub-techniques. Effective mitigation is linked to each technique in the dataset available from ATT&CK. Moreover, we also utilize group information to construct the TTP chains described below. The groups represent activity clusters known as threat actors, and each group is linked to the techniques they primarily use.

## 2.2 Dataset for Evaluation

Ground truth, i.e., data for evaluation, is necessary to conduct an evaluation experiment. Data for evaluation can be easily constructed when information contained in structured databases is used as the objective variable. However, when making predictions about information with unknown connections, experts often have to tag the information manually. Although this work incurs a significant cost, it can also be used as highly accurate training data, and as a result, a highly accurate model can be expected. However, there are often cases where the correct data generated is not disclosed, and in these cases, the superiority of the model's design method cannot be compared and is not reproducible. In this paper, the correct answers are also constructed from only publicly available data, and the design guarantees generality and reproducibility. Specifically, we utilize AlienVault's Open Threat Exchange (OTX), a crowdsourced threat information sharing platform that is open to anyone with a registered account. OTX provides an SDK[2] to collect threat information called pulse. Pulses include IoCs such as IP addresses and URLs, and may also include CVE-IDs and related techniques from ATT&CK. In this paper, we extracted the pulses from OTX's AlienVault account from 1/1/2023 to 1/11/2023 that contain both CVE-ID and technique, and treated the combination of these pulses as a dataset for evaluation.

## 2.3 Building TTP Chains

Inferring techniques used by attackers for CVE-IDs may allow us to predict the sequence of attack methods, i.e., TTP chains. If a TTP chain can be identified through a technique associated with a CVE-ID, it is possible to predict possible subsequent attack methods and proactively tackle the vulnerability if it could pose a significant risk later on. Therefore, we discuss the linkage between CVE-ID and technique as well as the linkage between CVE-ID and the TTP chain. We apply a method to reproduce TTP chains from techniques (Al-Shaer et al., 2020). We focus on 143 groups in total in the ATT&CK dataset, and represent them as one-hot vectors based on the techniques they use. Since ATT&CK classifies techniques into 201 types, excluding sub-techniques, each attacker group $g_i$ is represented by $g_i \in \{0,1\}^{201}$. Considering a matrix consisting of groups and techniques $M \in \{0,1\}^{143\times 201}$, we obtain a technique $t_j \in \{0,1\}^{143}$ that is represented by a one-hot vector of groups. It is possible to construct highly related technique clusters by evaluating the similarity of these technique vectors and clustering them, and each cluster can be considered as a TTP chain. In our experiment, the same setting as in (Al-Shaer et al., 2020) were used and the final number of clusters was 37. Note that it has been reported that a technique association of about 90% has been found in clusters using this method, and this study treats the results obtained as true (Al-Shaer et al., 2020).

# 3 EMBEDDING VULNERABILITIES

This study discusses the linkage between vulnerabilities assigned a CVE-ID and information that cannot be directly connected to the CVE-ID. As a case study, we attempt to connect CVE-IDs to techniques for ATT&CK or TTPs as information that cannot be directly connected. Specifically, the vulnerability to which the CVE-ID is assigned and the information to be connected, in this case the technique, are each put into an embedding representation, and the technique associated with the CVE-ID is inferred from the similarity between them.

## 3.1 Embedding Representations

Embedding representations of words and sentences have been realized by deep learning models using CNNs and RNNs, but since the transformer-based architecture (Vaswani et al., 2017) was proposed in 2017, various fast and accurate NLP models have been proposed, including BERT (Devlin et al., 2018). This paper uses BERT, one of the major NLP models, to achieve an embedding representation of vulnerabil-

[2] https://github.com/AlienVault-OTX/OTX-Python-SDK

ity information.

The pre-training model used in BERT utilizes BooksCorpus and English Wikipedia as pre-training data, and the model has good comprehension of general terms and sentences, but poor understanding of terms and contexts in the cybersecurity field. Therefore, we use SecBERT (Liberato, 2022), which is a pre-trained model using documents from the cybersecurity field. Using custom heads is generally more accurate than using CLS tokens in the final layer, and in this paper, we obtain the embedding representation of a sentence by the average pooled value of all tokens in the final layer. In the database used in this study, a single description contains multiple sentences, and because these sentences are input together, the number of tokens exceeds 512, which is the upper limit that can be processed by BERT, in some cases. There are various ways to truncate sentences, and here we use 256 tokens each at the beginning and end of a sentence if the number of tokens exceeds 512. With the above heuristic tuning, an embedding representation for a vulnerability assigned a CVE-ID $\boldsymbol{v}_{cve}$ is obtained. Similarly, an embedding representation for a technique in ATT&CK $\boldsymbol{v}_{tec}$ is obtained. Since $\boldsymbol{v}_{cve}$ and $\boldsymbol{v}_{tec}$ are represented by vectors of dimension 768, respectively, it is possible to directly evaluate their similarity.

## 3.2 Use of Multiple Resources

We consider improving the expressiveness of vulnerabilities by adding relevant information. It is possible to link NVD, CWE, and CAPEC with each other using the CWE-ID and CAPEC-ID as keys, so that they can be used as additional information to express the vulnerability to which the CVE-ID is assigned. ATT&CK is also interlinked with tactics, techniques, mitigations, etc., so that, for example, it is possible to check which mitigations are valid for a given technique. Hence, as with CVE-IDs, ATT&CK techniques can also utilize related information such as mitigations and tactics that represents the technique. In this paper, we use $n$ pieces of information related to CVE-IDs and ATT&CK techniques. The final embedding representations of vulnerabilities and techniques are evaluated as a weighted average of the embedding representations of a single information and additional information. Let $\boldsymbol{v_s}$ be the embedding representation of a single information and $\boldsymbol{v_i}$ $(i \in \{1,2,...,n\})$ be the embedding representation of $n$ additional information, the final embedding representation $\boldsymbol{v_m}$ is expressed as follows.

$$\boldsymbol{v_m} = \frac{w_0 \cdot \boldsymbol{v_s} + \sum_{i=1}^{n} w_i \cdot \boldsymbol{v_i}}{w_0 + \sum_{i=1}^{n} w_i} \tag{1}$$

Here, $w_i$ is the weight for each information and they are fixed at $\forall i; w_i = 1$ in the following experiments. By obtaining $\boldsymbol{v_i}$ in the same way as $\boldsymbol{v_s}$, they can be mapped onto the same space, and thus, $\boldsymbol{v_m}$ can be directly compared to each other. There are currently 625 techniques, of which 424 are subtechniques. The embedding representation $\boldsymbol{v}_{tec}$ of each technique is obtained by formula (1) using the embedding representation $\boldsymbol{v_s}$ of the technique and $n$ embedding representations $\boldsymbol{v_i}$ of the sub-techniques. Cosine similarity is used as the similarity measure in this study.

# 4 EXPERIMENT

## 4.1 Estimation of Technique

This paper uses CVE's description, and CWE and CAPEC's description and mitigation as embedding representations of vulnerabilities. The embedding representations of vulnerabilities are represented here by (CVE's description, CWE's description, CWE's mitigation, CAPEC's description, CAPEC's mitigation), where each item is set to 1 if the information is used and 0 if not. For example, if only the CVE's description is used, it is represented as (1, 0, 0, 0, 0). Similarly, the descriptions of techniques and mitigations are used for the embedding representations of techniques, and the embedding representations using both descriptions are represented as (technique, mitigation) = (1,1). We first tested the linkage between CVE-ID and each technique. For evaluation, the top $k$ techniques that are similar for each CVE-ID are selected, and it is considered correct if at least one of the techniques is included in the correct data.

Table 1 shows the results of similarity evaluation of vulnerabilities represented by CVE-IDs and ATT&CK technique. Each raw represents the prediction accuracy at $k \in \{1,2,3,4\}$ for each representation of CVE-IDs and techniques. For each $k$, the top three scores are shown in bold. The accuracy increases gradually as $k$ increases, but it is not linear. On the other hand, as $k$ increases, the error rate also increases. In our experiments, for the estimation of techniques, the error rate was lowest for $k = 1$ in most cases, i.e., the total number of correct techniques relative to the total number of predicted techniques was the largest. For the estimation of technique, surprisingly, we confirmed that (1, 0, 0, 0, 0), i.e., the case in which only the CVE-ID's descriptions are used, is highly accurate. One reason for this may be that the description of a mitigation relates to more than one technique. For example, M1018 is about managing user accounts properly, and there are nearly 100

Table 1: Prediction of techniques related to CVE-ID.

| | *k* | | | |
|---|---|---|---|---|
| CVE, ATT&CK | 1 | 2 | 3 | 4 |
| (1,0,0,0,0), (0,1) | 0.055 | 0.055 | 0.091 | 0.272 |
| (1,0,0,0,0), (1,0) | 0.073 | 0.182 | **0.291** | 0.309 |
| (1,0,0,0,0), (1,1) | **0.109** | **0.291** | **0.382** | **0.491** |
| (1,1,0,1,0), (0,1) | 0.055 | 0.055 | 0.091 | 0.255 |
| (1,1,0,1,0), (1,0) | 0.073 | 0.145 | 0.182 | 0.182 |
| (1,1,0,1,0), (1,1) | **0.164** | **0.273** | **0.291** | 0.327 |
| (1,1,1,1,1), (0,1) | 0.055 | 0.055 | 0.091 | 0.200 |
| (1,1,1,1,1), (1,0) | 0.073 | 0.164 | 0.200 | 0.236 |
| (1,1,1,1,1), (1,1) | **0.127** | **0.255** | **0.309** | **0.345** |
| (0,1,1,0,0), (0,1) | 0.036 | 0.055 | 0.109 | 0.291 |
| (0,1,1,0,0), (1,1) | 0.055 | 0.218 | 0.273 | **0.345** |
| (0,0,0,1,1), (0,1) | 0.036 | 0.055 | 0.109 | 0.200 |
| (0,0,0,1,1), (1,1) | 0.073 | 0.164 | 0.182 | 0.238 |
| (0,0,1,0,1), (0,1) | 0.036 | 0.055 | 0.109 | 0.236 |
| (0,0,1,0,1), (1,1) | 0.073 | 0.091 | 0.127 | 0.182 |

Table 2: Prediction of TTP chains related to CVE-ID (1).

| | *k* | | | |
|---|---|---|---|---|
| CVE, ATT&CK | 1 | 2 | 3 | 4 |
| (1,0,0,0,0), (0,1) | 0.055 | 0.400 | 0.436 | 0.564 |
| (1,0,0,0,0), (1,0) | 0.291 | 0.509 | 0.636 | 0.745 |
| (1,0,0,0,0), (1,1) | 0.345 | 0.600 | 0.727 | 0.800 |
| (1,1,0,1,0), (0,1) | 0.164 | 0.400 | 0.436 | 0.655 |
| (1,1,0,1,0), (1,0) | **0.527** | 0.600 | 0.655 | 0.745 |
| (1,1,0,1,0), (1,1) | 0.509 | **0.709** | **0.764** | **0.873** |
| (1,1,1,1,1), (0,1) | 0.364 | 0.455 | 0.491 | 0.673 |
| (1,1,1,1,1), (1,0) | **0.527** | 0.600 | 0.655 | 0.745 |
| (1,1,1,1,1), (1,1) | **0.545** | 0.655 | **0.800** | 0.800 |
| (0,1,1,0,0), (0,1) | 0.255 | 0.455 | 0.491 | 0.564 |
| (0,1,1,0,0), (1,1) | 0.491 | **0.727** | **0.782** | **0.873** |
| (0,0,0,1,1), (0,1) | 0.164 | 0.400 | 0.436 | 0.636 |
| (0,0,0,1,1), (1,1) | 0.491 | 0.636 | 0.673 | 0.745 |
| (0,0,1,0,1), (0,1) | 0.364 | 0.636 | 0.491 | 0.709 |
| (0,0,1,0,1), (1,1) | **0.564** | **0.709** | 0.709 | **0.818** |

techniques that this approach is effective. Therefore, even if mitigation can be estimated, it does not lead to technique estimation. A result supporting this consideration is that when the representation of technique is (0,1), the CVE-ID is rarely tied to a specific technique. It is considered that in order to connect a CVE-ID to a unique technique, it is necessary to have information that includes a description clearly associated with that technique.

### 4.2 Estimation of TTP Chain

We then tested the linkage between CVE-ID and TTP chains. In the estimation of TTP chains, we first determine the clusters of TTP chains to which the technique connected to the CVE-ID of the correct data belongs. We then select the top *k* techniques that are similar to the embedding representation of the CVE-ID as before, and determine the clusters of their TTP chains. In the experiment, we assume that a vulnerability is correctly predicted when at least one of the predicted clusters is included in the cluster of the correct data. The results of the experiment are shown in table 2. We obtain higher accuracy in estimating TTP chains than in estimating techniques, and even with $k = 1$, the accuracy rate exceeds 56% at maximum. One of the main reasons for the improved accuracy is that the estimation of the TTP chain is a 37-classification task, while the estimation of the technique is a 201-classification task, making it easier to guess. In addition to this, there may be a reason specific to the TTP chain. Some techniques in the same TTP chain are used selectively, and these can be handled with the same mitigation. For example, in our experiment, T1008 and T1104 are included in the same TTP chain. These are techniques that can be used selectively or simultaneously to make it difficult to detect command and control. The mitigations of them are common and characterize the TTP chain. Therefore, unlike the estimation of the technique, the inclusion of the mitigation is considered to contribute to the evaluation of similarity as a cluster. In fact, the estimation of the TTP chain tends to be slightly more accurate when multiple pieces of information are included, especially mitigation, than when only a single piece of information is included. As with the estimation of technique, the accuracy of the TTP chain increases gradually as *k* increases, but the error rate also increases, so it is necessary to determine an appropriate *k* depending on the nature of the task. In our experiments, for the estimation of the TTP chain, the error rate was lowest for $k = 3$ in most cases. The highest accuracy at $k = 3$ is about 80%, which is sufficient when considering that the embedding representation is constructed using only the most basic information (NVD, CWE, CAPEC, and ATT&CK).

The experimental results so far indicate that the NVD's description is the most important sources in terms of representing technique, and mitigation, especially CWE, contributes to the connection between CVE-ID and TTP chain. With the above in mind, the table 3 shows the results when the representation of vulnerabilities is (1,1,1,0,0). The result when the representation of the technique is (1,1) shows almost the highest accuracy in the experiment so far. Especially for ATT&CK, the combination of technique and mitigation improves the accuracy by 5.4 to 16.4%, confirming the effect of combining information. Our experimental results show that when embedding representations of vulnerabilities, it is possible to construct

Table 3: Prediction of TTP chains related to CVE-ID (2).

| | $k$ | | | |
|---|---|---|---|---|
| CVE, ATT&CK | 1 | 2 | 3 | 4 |
| (1,1,1,0,0), (1,0) | 0.400 | 0.673 | 0.691 | 0.764 |
| (1,1,1,0,0), (1,1) | 0.564 | 0.745 | 0.782 | 0.818 |

embedding representations that are more suitable for the purpose by selecting and incorporating sufficient information according to the purpose, even if it is a simple linear combination. On the other hand, it also suggests that the inclusion of unnecessary information reduces the expressive power of the embedding representation.

## 5 RELATED WORK

Here, we introduce some papers related to CVE-IDs and TTPs as related studies. BRON (Hemberg et al., 2020) is an initiative that attempts to connect various types of information starting from tactics and products, and it can be confirmed that the connection between CWE and technique is insufficient. A similar study by MITRE with CVE-ID and technique connection results can be found on Github[3], but it covers only some vulnerabilities up to 2020 and does not allow evaluation for new vulnerabilities. In addition, Kuppa et al. developed a predictive model of the ATT&CK technique associated with CVE-IDs (Kuppa et al., 2021). In the experiment, CVE-IDs were manually linked to techniques in advance, and multiple other information sources were used to suggest the possibility of connecting to unknown techniques, and the model was designed for concept drift. In studies related to TTPs, Ayoade et al. proposed a bias-corrected SVM classifier to classify tactics and techniques in reports from multiple security-related companies (Ayoade et al., 2018), and Li et al. attempted a multi-label classification of TTPs using the semantic similarity of texts using TF-IDF (Li et al., 2019).

## 6 CONCLUSION

This study proposed a method to improve the expressions of vulnerability information using BERT. We evaluated the similarity by applying a weighted average of multiple embedding representations of related information to the vulnerability and the expected connection destinations. This study differs from previous works in that it is highly reproducible because all information is collected from publicly available information. Therefore, it is possible to discuss the superiority of the model construction method itself, independent of the data. As a connection to ATT&CK, we evaluated the linkability of techniques and TTP chains associated with vulnerabilities and confirmed an improvement in accuracy of up to 16.4% with the use of additional information, especially for the TTP chains estimation. Since unnecessary information may be included in the embedding representation, the accuracy of the embedding representation is expected to be further improved by using documents with higher information content and by varying the weights according to the reliability of the information.

[3] https://github.com/center-for-threat-informed-defense/_to_cve

## REFERENCES

Al-Shaer, R., Spring, J. M., and Christou, E. (2020). Learning the associations of mitre att & ck adversarial techniques. In *2020 IEEE Conference on Communications and Network Security (CNS)*, pages 1–9. IEEE.

Ayoade, G., Chandra, S., Khan, L., Hamlen, K., and Thuraisingham, B. (2018). Automated threat report classification over multi-source data. In *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)*, pages 236–245. IEEE.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Hemberg, E., Kelly, J., Shlapentokh-Rothman, M., Reinstadler, B., Xu, K., Rutar, N., and O'Reilly, U.-M. (2020). Linking threat tactics, techniques, and patterns with defensive weaknesses, vulnerabilities and affected platform configurations for cyber hunting. *arXiv preprint arXiv:2010.00533*.

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

Li, M., Zheng, R., Liu, L., and Yang, P. (2019). Extraction of threat actions from threat-related articles using multi-label machine learning classification method. In *2019 2nd International Conference on Safety Produce Informatization (IICSPI)*, pages 428–431. IEEE.

Liberato, M. (2022). Secbert: Analyzing reports using bert-like models. Master's thesis, University of Twente.

Rahman, M. R., Hezaveh, R. M., and Williams, L. (2023). What are the attackers doing now? automating cyberthreat intelligence extraction from text on pace with the changing threat landscape: A survey. *ACM Computing Surveys*, 55(12):1–36.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.