Automating Cybersecurity TTP Classification Based on Unstructured Attack Descriptions

Felipe Castaño
Digital security and Cybersecurity,
Vicomtech
University of León
Bilbao, Spain
fcastano@vicomtech.org

Amaia Gil-Lerchundi
Digital security and Cybersecurity,
Vicomtech
San Sebastian, Spain
agil@vicomtech.org

Raul Orduna-Urrutia
Digital security and Cybersecurity,
Vicomtech
San Sebastian, Spain
rorduna@vicomtech.org

Eduardo Fidalgo Fernandez
Department of Electrical, Systems and Automation.
University of León.
León, Spain
eduardo.fidalgo@unileon.es

Rocío Alaiz-Rodríguez
Department of Electrical, Systems and Automation.
University of León.
León, Spain
rocio.alaiz@unileon.es

***Abstract*—CTI sources help SOCs to share important information about incidents and attacks. Unstructured text processing gains importance, considering that incident-related information is present in a wide range of sources. The datasets in the literature contain insufficiently lengthy text or a limited number of samples per class. Therefore, we proposed a method to build a semi-automatic dataset using the CTI sources. As a result, we have presented a new dataset of unstructured CTI descriptions called Weakness, Attack, Vulnerabilities, and Events 27k (WAVE-27K). WAVE-27K includes information on 27 different MITRE techniques and 7 tactics, containing 22539 samples associated with a single technique and 5262 samples related to two or more techniques. WAVE-27K is the largest dataset compared to those in the literature. We trained a BERT-based model using WAVE-27K, obtaining a 97.00% micro F1-score, which could validate that the information included on WAVE-27-K has quality sufficient for training machine learning models.**

***Index Terms*—Cybersecurity, Threat Intelligence, Matrix MITRE, Dataset, NLP, BERT**

Type of contribution: *Research in progress*

I. Introduction

Traditional security measures like antivirus software or endpoint detection and response (EDR) solutions have a certain level of effectiveness but struggle to identify emerging threats like zero-day attacks[1]. In response to this challenge, cyber threat hunting emerges as a cybersecurity method that systematically scans networks, systems, and devices to uncover anomalies and potential cyber threats[2].

Cyber Threat Intelligence (CTI) activities start collecting information from different sources, and it ends up extracting knowledge that helps in the decision-making process for proactive defense, like cyber hunting methods [3]. CTI sources are increasing due to better capabilities from the current platforms to process data faster than before. The above facilitates the process of sharing threat information with the cybersecurity community [4]. Incident-related information can be found on open-source intelligence (OSINT) sources, cybersecurity analyst forums, or the broader Internet.

For that reason, it is crucial to automatically process unstructured texts to extract information such as tactics, techniques, and procedures (TTPs) from different free-text sources [5]. Furthermore, it is possible to use unstructured text processing methods as a base to detect Dark web forums or other sources where attacks are explained. The above allows the detection of the attacks as well as detecting networks where such information is shared and the groups behind them. Although there are several available CTI datasets in the state of the art, they often fall short in both sample quantity and data quality. These datasets contain insufficiently lengthy text or a limited number of samples per class.

Our method uses the available information from CTI sources to support the automation of incident classification, reducing costs and ad hoc studies with limited data. That is possible through the utilization of unstructured text processing tools. In this paper, we explore the creation of a dataset employing this approach and its subsequent validation. As a result, we present WAVE-27K, a dataset that contains 27801 CTI descriptions related to 27 different MITRE techniques and 7 tactics. WAVE-27 is the largest dataset compared to those in the state of the art, and it contains the largest number of samples per class as well.

This paper is organized as follows: Section II provides a related work review, offering context for the research. Section III details our methodology, including dataset dataset-building process. Section IV describes the experimental setup, defining details regarding the models and the metrics used for model evaluation. Finally, Section V presents the results, and Section VI contains our findings and future research.

II. Background

There are two main groups in the CTI pattern extraction literature according to their goal. The first group focuses on extracting data from unstructured sources and presenting it as structured information, where it is possible to extract knowledge. The second group includes classification techniques that address CTI unstructured data as a classification problem. The main objective of this group is to relate the text with one

or more known cyberattack techniques. In this section, we present the more significant results of both groups.

A. Information Extraction

Noor et al. implemented three phases to extract information from unstructured data [6]. The first phase consists of data collection from CTI sources. The second phase is the data analysis, where a semantic search method identifies observables, techniques, and procedures. Finally, the last phase is a model that predicts the class of the cyber threat group actor using the information extracted in the previous phase. They collected 327 unstructured reports from 2012 to 2018 related to 36 threat groups, following the steps earlier described. Then, they evaluated different models on the dataset and reported the DLNN model as the more effective with 94% accuracy.

Later, Jo et al. presented a BERT model to extract entities from CTI unstructured data [7]. They combined BERT and BiLSTM layers for this task, focusing on identifying ransomware and related information. Finally, they built a dataset manually annotated by five graduate students, which contains 6791 entities and 4323 relations, where BERT achieved an F1-score of 97.2% for the entity recognition task.

Recently, Siracusano et al. proposed a method using a gpt-3.5-turbo[1] prompt to identify entities and relations [8]. The information is represented as a Structured Threat Information Expression (STIX)[2] bundle, allowing comparison with other works. They focused on identifying malware and building a dataset that contained 204 reports.

B. Classification Techniques

In 2020, Legoy et al. addressed CTI information as a classification problem where the main objective is identifying MITRE ATT&CK[3] tactics and techniques [9]. For this, the authors evaluated the performance of TF-IDF weighting factors [10] against a Word2Vec model in the pre-processing step. In the classification process, they evaluated binary relevance [11] and multi-label approaches. The dataset used in this work contains 1490 reports and MITRE attack and tactic labels. Finally, they reported that models using Word2Vec under-performed the TF-IDF weighting factors. The results showed that the AdaBoost Decision Tree model achieved 61.30% F0.5-score for the multi-label approach and Gradient T Boosting with 65.04% F0.5-score for the binary relevance approach.

Later, Mendsaikhan et al. [12] evaluated the capability of identifying MITRE attacks using a multi-label approach on models such as a fine-tuned BERT model [13], Multi-label k-Nearest Neighbors (MlkNN), and LabelPowerset. They used three public datasets for the training process, Threat Report ATT&CK [4] (TRAM) dataset, ENISA dataset [14], and RCATT dataset [9]. The results showed that BERT achieved the highest performance with 78.01% F1-score, followed by the LabelPowerset method with Multilayer Perceptron (MLP) in the second place with 74.70% F1-score.

Obinate et al. evaluated several deep learning techniques on the classification task as well [15]; for this purpose, they built a dataset using the information available on MITRE ATT&CK and the Attack Pattern Enumerations and Classifications (CAPEC) sources. They collected the description of several threat actors and their malware campaigns, resulting in a dataset[5] that contains 12945 samples. They also use the TRAM dataset. Next, the authors evaluated models such as Linear Regression (LR), Support Vector Machine (SVM), and SecureBERT[16] on both datasets. The results showed SecureBERT achieved the highest F1-score value with 72.50% on their dataset, and SVM achieved the highest F1-Score with 60.90 on the TRAM dataset.

In 2022, Alves et al. [17] implemented 11 different combinations of hyperparameters on Transformers such as BERT [13], RoBERTa[18], SecBERT, and SecRoBERTa. The dataset used in this work contains 9909 sentences related to 253 techniques; the authors used procedure examples from the MITRE ATT&ACK source to collect the data. They used an accuracy metric to evaluate the performance, reporting RoBERTa as the model with the highest performance with an accuracy of 82.64% on the testing dataset.

On the one hand, the information extraction group generates structured information from unstructured sources, and it is helpful in daily cyber-threat intelligence tasks. However, the dataset construction can be complex due to the need to extract structured processes. On the other hand, the classification technique group adds label standardization using the MITRE matrix, allowing the comparison between different implementations and facilitating the integration of the public datasets into the training process. For those reasons, we have decided to focus on the classification techniques group in this work.

C. Datasets

There are several public datasets for the classification of techniques groups detailed in Section II, they contain CTI descriptions to the MITRE matrix using technique labels. It is also important to highlight that since we are focused on the classification techniques group, we do not use datasets from the information extraction approaches. Orbinato et al. [15] presented CTI-to-MITRE with NLP, a dataset that describes a technique with one or multiple sentences by sample, those sentences are related to 14 tactics and 188 different techniques. The authors presented the data in October 2023 from the MITRE ATT&CK framework using the general description in natural language and added extra information using the CAPEC taxonomy. CTI-to-MITRE contains 12945 samples, where around 8000 are related to 37 techniques, and it presents the following information: tactic, technique, sub-technique, technical name, and the event description in a sentence. Despite this dataset containing 12945 samples.

The second dataset is TRAM[6] and is obtained from an open-source platform designed to integrate MITRE ATT&CK matrix across the CTI community by mapping techniques from unstructured text. This platform includes a dataset called TRAM that contains 1482 samples with a description of the event and an MITRE technique related to it. Concerning the

[1] platform.openai.com/docs/models/gpt-3-5
[2] oasis-open.github.io/cti-documentation/stix/intro
[3] attack.mitre.org
[4] github.com/center-for-threat-informed-defense/tram
[5] github.com/dessertlab/cti-to-mitre-with-nlp
[6] github.com/center-for-threat-informed-defense/tram

dataset distribution, the 46% of the techniques contains less than 10 samples. Similarly to CTI-to-MITRE, most of the samples are concentrated on a group of techniques, with 1211 samples related to 30 techniques.

Finally, the European Union Agency for Cybersecurity (ENISA) released a report titled State of Vulnerabilities [14] in 2019. They collected information about 27471 vulnerabilities and mapped the CVEs[7] to the MITRE ATT&CK technique, obtaining 7642 samples related to 50 techniques and nine tactics[8]. Contrary to the other datasets, this dataset can contain more than one MITRE technique associated with a single sample, which means that the results of a model in this scenario will be a multi-class, multi-output classification. Considering that a sample can be related to more than one technique, the average of samples per technique is 1366 samples. However, the dataset contains outliers, such as techniques with less than 100 samples and techniques with more than 2500 samples.

III. Methodology

After revising the literature presented, we consider the need for a larger unstructured CTI text dataset. We hypothesize that enhancing the quality and quantity of available data will improve the performance of state-of-the-art algorithms. Therefore, we have built the Weakness, Attack, Vulnerabilities, and Events 27K dataset (WAVE-27K) using the available OSINT information. We used a model on the available datasets and our own to validate that the building method generate a dataset of sufficient quality.

A. WAVE-27K Dataset Building Method

We used four of the primary CTI sources to build the WAVE-27K. First, the MITRE ATT&CK framework provides base knowledge of techniques and tactics. This source contains a matrix with information that related tactics, and techniques with campaigns and groups that carry out those techniques. Besides, it provides information about software or tools used in an attack and its possible mitigation. We retrieved the MITRE matrix in September 2023, getting information about 14 tactics, 738 techniques, 43 mitigation, 554 software vulnerabilities, 82 tools, 20 campaigns, and 141 groups.

The second source is a list of attack patterns called CAPEC, retrieved in September 2023, that helps understand how adversaries exploit software weaknesses. The list contains columns that provide information such as the name of the attack pattern, description, likelihood, severity, execution flow, and related weakness, among other data. Following the information provided by CAPEC about software weakness, the third source is Common Weakness Enumeration (CWE), which contains a community-developed list of software and hardware weaknesses. That list describes the weakness, background details, technology affected, consequences, architecture affected, and observable examples. We retrieved 958 samples from the CWE source in the collection process.

Finally, Common Vulnerabilities and Exposures (CVE) is the fourth source used and contains information about known vulnerabilities. The CVE source provides a vulnerability description, the vulnerability complexity, and the impact on confidentiality, integrity, and availability of the software. We retrieved around 22400 samples from this source at the moment of the collection process.

Several works have studied the integration process of the previous sources to create a more robust CTI dataset[19], [20], [4]. Following those works, we have proposed an approach to merge the previous sources by linking them using external references. The process involves checking the collected samples of each source and looking for any reference to another source. Later, those external references are analyzed, identifying the target source, extracting the URL, and selecting only the external references that point to the selected sources. After extracting the identifier from the URL, the next step is to check the samples in the target source and match the identifier with the actual sample. Finally, a new relationship is created if the identifier is found in an external reference and a sample in the target source. This method intends to improve data comprehensiveness by integrating incident-related details from various sources. To be precise, we utilized the subsequent categories: MITRE ATT&CK external references for linking with CAPEC IDs, CAPEC external references for associating with CWE IDs, and CVE vulnerabilities for synchronization with CWE IDs. CWE assumes a pivotal role in this procedure, as shown in Figure 1.

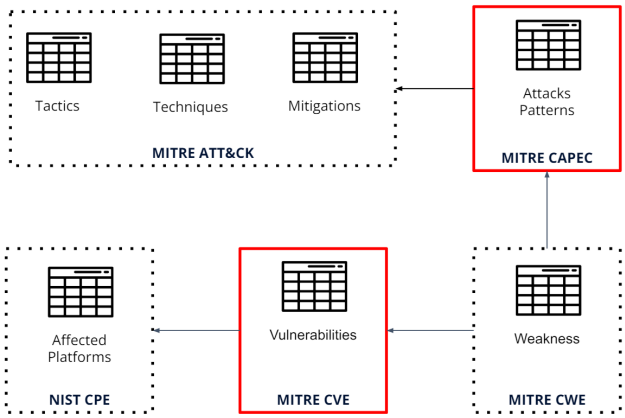

Figure 1. Sources Integration process. Red highlighting represents the sources that have provided external links to relate information with other sources

Once the relations between the sources are extracted, the next step is to create WAVE-27K with CTI descriptions, tactics, and techniques related to them. This step is based on the proposed State of Vulnerabilities report[14]. Therefore, we extracted the vulnerability descriptions reported on the CVE source, added them as the main component of WAVE-27K, and used the extracted relations to link the CVE description with the MITRE tactics and techniques. Following this method it is possible not only to build a new dataset but constantly update the samples using new reports on the sources.

WAVE-27K contains 27801 samples, where 22539 of them are related to one technique, and the other 5262 samples are related to two or more techniques. WAVE-27K includes descriptions of 27 different MITRE techniques and 7 tactics related to these techniques. The number of tactics and techniques included on WAVE-27K correspond with the reported

[7] cve.mitre.org/
[8] https://github.com/enisaeu/vuln-report

on the CTI sources. That indicates that the included techniques and tactics represent the most common incidents found in Security Operations Centers (SOCs), and the number of tactics and techniques included can increase while new reports are generated. A potential limitation is the coverage of the dataset since there is limited control over the class balance within the dataset since they are collected from samples reported on OSINT sources. Moreover, this approach suggests that the samples collected represent the prevailing trends and patterns observed in cyber attacks, providing valuable insights into real-world threat scenarios.

As a result, we present the largest dataset compared to those in the state of the art, which also contains the largest number of samples per class, as shown in Table I. Besides providing a larger number of samples per technique, WAVE-27K contains a more detailed description of the CTI event.

Table I
DATASETS DESCRIPTION AND DISTRIBUTION, COMPARISON BETWEEN PUBLIC DATASETS AND WAVE-27K. WHERE S/T MEANS SAMPLES PER TECHNIQUES AND W/I MEANS WORDS PER DESCRIPTION.

Dataset	Samples	Tactics	Techniques	AVG S/T	AVG W/I
CTI_NLP [15]	12945	14	188	68	15
TRAM [15]	1482	14	80	18	28
ENISA[12]	7642	9	50	1465	45
WAVE-27K	**27801**	7	27	1830	45

IV. EXPERIMENTAL SETUP

Given the results reported by Mendsaikhan et al. [12], Obinate et al. [15] and Alves et al. [17], we used a deep learning model for our experiments due to the performance they achieved. Specifically, we implement SecBERT, a transformer model trained on a corpus focused on cybersecurity. In our experiments, we employed an 80/20 dataset split, where 80% of the data was used for training and the remaining 20% for validation purposes.

In the evaluation process, we address a crucial challenge in the comparative analysis of the four datasets. While the MITRE matrix serves as the common set of labels, we have verified that the subset employed by each dataset differs, taking WAVE-27K as the reference point, we noticed that each CTI, TRAM, and ENISA incorporates only 12, 9, and 8 common labels, respectively. Another challenge in the comparative analysis derives from the methodologies employed in datasets building process. Some of them were generated by extracting information from the MITRE matrix using NLP algorithms, while others contain CVE descriptions and are manually annotated. The different building processes limited the data to specific types of information and introduced complexities in the data direct comparison. This discrepancy precludes a direct comparison. To facilitate comparison, we exclusively employed the SecBERT model for the evaluated process. The first experiment consists of comparing the results between the SecBert trained in each dataset. Then, using the models already trained in the first experiment, we utilized them to evaluate the common classes between WAVE-27K and the other datasets. All of this is to provide a representative measure of the quality of the data regarding the current datasets in the literature.

A. Metrics

F1-score is a widely used metric for binary and multi-class classification tasks. The F1-score strikes a balance between precision and recall, offering an overview of the capacity of the models to classify both positive and negative instances. The F1-score is adapted for multi-label datasets by calculating it for each class independently and then averaging the scores. In this particular scenario, we used the Micro-average F1-score since it is useful when dealing with unbalanced datasets where some classes may have significantly fewer instances than others, considering the above described, we have selected the Micro F1-score as the main metric to measure the performance of the models.

Another metric used for the evaluation process is accuracy (ACC). We used ACC because several authors in the literature report the performance of their models using it, enabling comparisons with these prior works. Finally, we used the Matthews Correlation Coefficient (MCC), a statistical metric that evaluates classification models, especially on unbalanced datasets[21]. MCC considers true and false positives and negatives to provide a single score representing the quality of the classification.

V. RESULTS

The results show that the model trained with the WAVE-27K dataset achieved an F1 Micro Average score of 97.00% in the complete test set, outperforming the results on the other datasets, we used the F1-score micro average as the main metric since it assigns the same importance to each class no matter how many classes it contains, which is helpful for the unbalanced dataset as the presents in this evaluation. This model came second place in ACC and MCC after the model using the ENISA Dataset, as shown in Table II. In the context of the comparison between intersected classes of WAVE-27K and the public datasets, the results demonstrate that the model trained with WAVE-27K outstands the performance of the models trained with CTI, TRAM, and ENISA dataset with Micro F1-scores of 96.46%, 95.50% and 92.15% respectively.

Table II
COMPARISON OF AVAILABLE DATASETS WITH WAVE-27K, THE FIRST PART OF THE TABLE SHOWS THE RESULTS RELATED TO CLASSIFICATION OF ALL CLASSES USING A MODEL TRAINED IN THE SAME DATASPACE, AND THE SECOND PART THE RESULTS USING ONLY THE COMMON CLASSES BY EACH PUBLIC DATASET AND WAVE-27K.

Experiment	Dataset name	Classes	N. Test Samples	ACC	MCC	F1 Micro
Complete Test Set	CTI	188	1942	70.90	70.45	70.90
	TRAM	80	221	52.49	50.82	52.49
	ENISA	50	1147	**88.97**	**78.60**	93.79
	WAVE-27K	27	4171	87.03	74.53	**97.00**
CTI - WAVE-27K	CTI	12	266	74.43	**72.15**	74.43
	WAVE-27K	12	3837	**91.25**	64.70	**96.46**
TRAM -WAVE-27K	TRAM	9	37	59.46	49.46	19.04
	WAVE-27K	9	4171	**91.61**	**65.21**	**96.50**
ENISA - WAVE-27K	ENISA	8	431	**80.22**	38.25	83.48
	WAVE-27K	8	1177	79.86	**49.56**	**92.15**

These metrics provide quantitative insight into the proficiency of the model across diverse datasets, highlighting its robust performance in extracting and classifying cybersecurity related information. However, as we detailed in Section IV, the discrepancy of labels between the datasets precludes a direct comparison. Therefore, we use the results obtained by the model to validate that the data collected in WAVE-27K

contains relevant information for the classification of incidents and provides a level of quality sufficient for training a machine learning model.

VI. Conclusions and Future Work

In this paper, we study how to develop a dataset semi-automatically to enable the linkage of free text descriptions and MITRE TTP labels with a level of quality sufficient for training a machine learning model and achieving comparable results. One of the main advantages of the proposed building method is that it uses information available in CTI sources, which indicates that the techniques and tactics included represent the most common incidents encountered in SOC, and the number of tactics and techniques included may increase as new reports are generated and the dataset is updated.

The studies focused on classifying MITRE TTP incidents from free text descriptions are important since we can link a description of a CTI event, specifically a vulnerability, with an MITRE technique, and thereby derive potential mitigations. Given the importance of robust and well-classified datasets acquired at a low cost, we proposed a new data-gathering technology aimed to collect and easily update a CTI dataset. It can be applied both to the description of incidents encountered in OSINT and to internal repositories generated in the day-to-day operations of a SOC. The results show that the proposed method effectively built a dataset with a level of quality sufficient to train a machine-learning model. We trained and tested a SecBert model in each dataset, achieving a micro F1-score of 97.00% with WAVE-27K in the complete test set, getting the highest value. We also outperform the other model in the common class comparison with an F1-score up to 96.50%.

In future work, we will evaluate other machine learning models we do not use in this work. Besides, we will study the possibility of training specialized models per class to check the effectiveness of the classification task in such a scenario once the number of samples allows it. Furthermore, we will study a cascading classification method, where first we will classify tactics followed by a technique classification using a stacking approach. This way, we would assess if the hierarchical classification improves overall performance.

Another aspect of future work involves facing longer unstructured text and linking them using appropriate tags. Integrating state-of-the-art models into this scenario will expand the scope of attack classification systems, enabling the automatic generation of alerts from free and unstructured text.

Acknowledgment

This work has been partially supported by the European Union's Horizon Europe Framework under the project ATLANTIS (Grant Agreement No. 01073909).

References

[1] P. Kumar, M. Wazid, D. P. Singh, J. Singh, A. K. Das, Y. Park, and J. J. P. C. Rodrigues, "Explainable artificial intelligence envisioned security mechanism for cyber threat hunting," *Security and Privacy*, no. March, pp. 1–14, 2023.

[2] Malwarebytes, "What is cyber threat hunting." [Online]. Available: https://www.malwarebytes.com/cybersecurity/business/what-is-cyber-threat-hunting

[3] N. Sun, M. Ding, J. Jiang, W. Xu, X. Mo, Y. Tai, and J. Zhang, "Cyber Threat Intelligence Mining for Proactive Cybersecurity Defense: A Survey and New Perspectives," *IEEE Communications Surveys and Tutorials*, vol. 25, no. 3, pp. 1748–1774, 2023.

[4] K. Rantos, A. Spyros, A. Papanikolaou, A. Kritsas, C. Ilioudis, and V. Katos, "Interoperability challenges in the cybersecurity information sharing ecosystem," *Computers*, vol. 9, 3 2020.

[5] S. Fujii, N. Kawaguchi, T. Shigemoto, and T. Yamauchi, "Cyner: Information extraction from unstructured text of cti sources with noncontextual iocs," in *International Workshop on Security*. Springer, 2022, pp. 85–104.

[6] U. Noor, Z. Anwar, T. Amjad, and K. K. R. Choo, "A machine learning-based fintech cyber threat attribution framework using high-level indicators of compromise," *Future Generation Computer Systems*, vol. 96, pp. 227–242, 7 2019.

[7] H. Jo, Y. Lee, and S. Shin, "Vulcan: Automatic extraction and analysis of cyber threat intelligence from unstructured text," *Computers & Security*, vol. 120, p. 102763, 9 2022.

[8] G. Siracusano, D. Sanvito, R. Gonzalez, M. Srinivasan, S. Kamatchi, W. Takahashi, M. Kawakita, T. Kakumaru, and R. Bifulco, "Time for action: Automated analysis of cyber threat intelligence in the wild," *arXiv preprint arXiv:2307.10214*, 7 2023. [Online]. Available: http://arxiv.org/abs/2307.10214

[9] V. Legoy, M. Caselli, C. Seifert, and A. Peter, "Automated retrieval of att&ck tactics and techniques for cyber threat reports," *arXiv preprint arXiv:2004.14322*, 2020.

[10] D. Christopher, P. Raghavan, H. Schütze *et al.*, "Scoring term weighting and the vector space model," *Introduction to information retrieval*, vol. 100, pp. 2–4, 2008.

[11] O. Luaces, J. Díez, J. Barranquero, J. J. del Coz, and A. Bahamonde, "Binary relevance efficacy for multilabel classification," *Progress in Artificial Intelligence*, vol. 1, pp. 303–313, 2012.

[12] O. Mendsaikhan, H. Hasegawa, Y. Yamaguchi, and H. Shimada, "Automatic mapping of threat information to adversary techniques using different datasets," *International Journal on Advances in Security Volume 14, Number 1 & 2, 2021*, 2021.

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

[14] V. Katos, S. Rostami, P. Bellonias, N. Davies, A. Kleszcz, S. Faily *et al.*, "State of vulnerabilities 2018/2019: analysis of events in the life of vulnerabilities," *Report/Study*, 2019.

[15] V. Orbinato, M. Barbaraci, R. Natella, and D. Cotroneo, "Automatic mapping of unstructured cyber threat intelligence: An experimental study:(practical experience report)," in *2022 IEEE 33rd International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 8 2022, pp. 181–192.

[16] E. Aghaei, X. Niu, W. Shadid, and E. Al-Shaer, "Language model for text analytic in cybersecurity," *arXiv preprint arXiv:2204.02685*, 2022.

[17] P. M. Alves, G. P. Filho, and V. P. Goncalves, "Leveraging bert's power to classify ttp from unstructured text," *2022 Workshop on Communication Networks and Power Systems, WCNPS 2022*, 2022.

[18] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[19] E. Hemberg, A. Srinivasan, N. Rutar, and U.-M. O'Reilly, "Sourcing language models and text information for inferring cyber threat, vulnerability and mitigation relationships," in *AI4Cyber: AI-enabled Cybersecurity Analytics and Deployable Defense workshop*, 2022.

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

[21] B. W. Matthews, "Comparison of the predicted and observed secondary structure of t4 phage lysozyme," *Biochimica et Biophysica Acta (BBA)-Protein Structure*, vol. 405, no. 2, pp. 442–451, 1975.