2023 International Conference on Control, Communication and Computing (ICCC) 19-21 May 2023

Extraction of Actionable Threat Intelligence from Dark Web Data

Varsha Varghese
Cyber Forensics and Information Security, Department of Computer Science & Engineering
ER&DCI Institute of Technology
Thiruvananthapuram
varghesevarsha5@gmail.com

Mahalakshmi S
Project Engineer, Cyber Security Group
Centre for Development of Advanced Computing
Thiruvananthapuram
mahalakshmiprabha96@gmail.com

Senthilkumar KB
Scientict_E, Cyber Security Group
Centre for Development of Advanced Computing
Thiruvananthapuram
kbsenthil@cdac.in

2023 International Conference on Control, Communication and Computing (ICCC) | 979-8-3503-3412-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCC57789.2023.10165477

***Abstract*—Darknet has emerged as an excellent platform for cybercriminals to conduct various illicit activities such as operating a fully functional crypto-currency-based marketplace or maintaining a highly anonymous communication channel. However, it also acts as a source of Cyber Threat Intelligence which is the information about the techniques, tactics, or motives of an emerging threat. The life cycle of a Darknet-based Cyber Threat Intelligence solution mainly consists of the Collection, Processing, Analysis, and Production of Data from the Darknet. The Collection and Processing stages deal with accumulating information from various dark web pages and transforming the data into some format using a big data system. The Analysis part is integral to the Threat Intelligence schema since it constitutes the extraction of named entities like Organization, Offensive Activity, Tools, etc., and the determination of relationships between these entities. The production stage is like data mining where some knowledge is extracted that can be relevant to an organization under consideration. Such knowledge can mitigate risks and disrupt any targeted cyber-attack campaigns, making Cyber Threat Intelligence actionable. In this research, an open-source-intelligence toolset is used to scan and collect data from the dark web forums through crawling and scrapping. The collected data is then ingested into a state-of-the-art NLP model that extracts actionable threat intelligence using Named Entity Recognition. The experimental results indicate that the model could identify HackerIds, tools, software, organizations, and other entities in the discussions of dark web forums with better efficiency and accuracy. This could be used to identify the source of a data leak, the release of new malware, evidence of a new exploit, and other offensive activities.**

Keywords—Actionable Threat Intelligence, BERT, Elasticsearch, Hacking Forums, Named Entity Recognition

I. Introduction

The Dark Web can be defined as a part of the Internet that is not indexed by search engines such as Google and requires specialized software to access, such as The Onion Router (TOR). It is different from the deep web which is simply a collection of nonindexed pages that is accessible whenever online, such as our email account, paid content such as magazine subscriptions, streaming services such as Amazon Prime and Netflix (once logged in), or pages that require the exact URL to access. The deep web neither supports criminal activities nor guarantees users' anonymity. On the other hand dark web specifically ensures anonymity and is a hub of unlawful activities. The types of illegal information, products, or services available in darknet marketplaces are astonishingly vast. These range from cloned Mastercard information to stolen social media accounts. Moreover, forged documents, email database dumps, and even malware-as-a-service are available. Also, goods such as illegal drugs, weapons, pornographic content, fraud/counterfeit tutorials, etc. are widely traded in the dark net markets. Besides, there are hacker forums that serve as a suitable platform for sharing hacking tools, zero-day vulnerabilities, exploits, and other illegal products or services. Some of the notorious hacker forums at the time of this writing are nulled, dread, raidforums, 4chan, etc.

Monitoring and analyzing the data from Dark web forums and marketplaces can greatly help in identifying various threat actors, their motives, possible future attacks, and links to other hacker communities. For instance, if confidential data has been extorted from an organization, the threat actors mostly choose their darknet community as the initial place to disclose those details. Subsequently, they would put these data for sale in the darknet marketplace. Similarly, a hacker who penetrated a specific high-security network would try to publicize his/her skills in the hacker community to gain attention. Furthermore, early warnings of zero-day vulnerabilities, their proof of concepts, or exploit codes could be disseminated and discussed in dark web hacker communities. Hence, intelligence from the darknet ecosystem would help in understanding the security posture of an organization and contribute to risk management strategies. Moreover, to deliver a complete view of the threat landscape it is equally important to consider information from dark web sources along with other open sources.

The key challenge in developing a threat intelligence toolset based on the dark web is the data analysis techniques used. Here, the main sources of data are dark web forums and marketplaces. Mostly the content of these will be discussions or communications using real-world language instances. Whilst these can be easily understood by human analysis, they cannot be considered as a solution, especially in the context of actionable automated threat intelligence generation. Therefore, this research aims to identify cyber-security-related entities from crawled dark web forum data for cyber threat intelligence purposes.

II. Related Work

Throughout the literature, there were many studies related to knowledge extraction from big data in the cybersecurity context. Earlier, most of the solutions were either rule-based or used statistical models. In one such approach, a combination of content analysis and social network analysis was used for identifying and ranking hackers from dark web underground forums. The data was collected and pre-processed and a social network interaction graph was generated based on user interaction. Then, content analysis was performed based on users' characteristics and topic

979-8-3503-3412-8/23/$31.00 ©2023 IEEE

preferences. Finally, the Social Network Analysis was done using a PageRank algorithm and the key users(hackers) were identified and ranked [1].

More recent works included a machine learning approach. In one such solution, a combination of various data processing techniques was used to discover cybersecurity information from unstructured hacker forum data [2]. The data collection was done by a Web Crawler to gather posts from hacker forums followed by data cleaning tasks which involved tokenization, stop word removal, tagging, etc. The next stage was word embedding wherein TF-IDF is used to find term relevance. Word embeddings are a type of word representation that allows words with similar meanings to have a similar representation. Both Word2vec and Doc2vec embedding models were used to find which one is better. The analysis module used two stages of clustering. Firstly, theme topic clustering was used to determine the topics related to security context so that non-security-related topics are avoided. Secondly, LDA was used to remove further non-relevant information. The evaluation was done to identify data cleaning efficiency, theme topic clustering performance, and CTI event detection part.

Isuf *et al*. [3] presented a comparative study of Convolutional Neural Network methods and Support Vector Machines (traditional ML techniques), for text classification from underground hacker forums. The two forums used were Hackhound and Nulled.IO. Data pre-processing involved preparation (extracting task-relevant fields from the raw data and storing the result in a useful format) and cleaning (removing parts of the data that act as noise and do not contribute to task performance). They found that SVM performs almost similarly to CNN and for practical implementations SVM can be chosen since CNN is more computationally demanding.

Pavlos et al. [4] proposed a study on using transformer-based models such as BERT base, XLNet, RoBERTa, and ELECTRA, optimized with suitable hyper-parameters for the NLP task called Named entity recognition or NER. NER is an NLP technique that can automatically scan entire articles and detect named entities such as persons, locations organizations, etc and classify them into predefined categories. Fine-tuning was done using a single extra output layer on top depending on the task (For NER, a linear classification layer was used). The dataset used for evaluating the optimized model was DNRTI (Dataset for NER in Threat Intelligence). The results showed that BERT base and XLNet base models excelled in all the metrics (F1- score, Precision, and Recall). When compared with the results of LSTM and BiLSTM methods, they demonstrated that modern transformer-based models have a considerable improvement over previous techniques.

Another BERT-based research for threat intelligence extraction was proposed by Priyanka et al. [5]. They introduced CyBERT which is a BERT-based model that has been fine-tuned using cyber-security domain-specific knowledge. This can be considered as a baseline model for various cybersecurity tasks and thereby they attempt to implement transfer learning. To create the CyBERT model they re-apply MLM(Masked Language Models) on the general BERT model by fine-tuning it with a specially crafted cybersecurity corpus. The performance of CyBERT was analyzed based on tasks such as Vulnerability and exploit search (predicting potential attack vectors given textual descriptions of vulnerabilities) Cybersecurity NER (Identification of named entities) and Cybersecurity Knowledge Graph (CKG) completion (Model to predict relations between entities).

Brian Nafziger proposed an automated open-source toolset[6] that is capable of collecting, processing, and analyzing darknet-specific data for extracting threat intelligence. The toolset is based on the intelligence life cycle and comprises data collection, data processing, data analysis, and data mining (extraction of knowledge). Figure 1 represents the block diagram of the existing toolset wherein the input will be raw data crawled from the dark web and the output will be intelligence or information related to threats

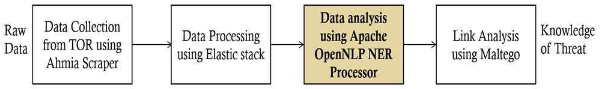

Raw Data
Data Collection from TOR using Ahmia Scraper
Data Processing using Elastic stack
Data analysis using Apache OpenNLP NER Processor
Link Analysis using Maltego
Knowledge of Threat

Fig. 1. Existing System

III. METHODOLOGY

Since the current state-of-the-art models in NLP are pre-trained transformer-based models[8], the proposed solution aims to improve the accuracy of the NLP NER task in the existing system and identify the information with much higher efficiency. The proposed system uses a different model for Named Entity Recognition instead of the existing Apache Open-NLP NER processor. Google's BERT [7] or Bidirectional Encoder Representations from Transformers was the model selected after considering the recent developments in the field of NLP as well as the possibilities of transfer learning.

Some of the profound benefits of using BERT in a domain-specific application such as this is that BERT can understand a word's context. For instance, given two sentences such as "The man was accused of robbing a bank." and "The man went fishing by the bank of the river.", the previous models such as Word2Vec will have the same word embedding for the word "bank" in both sentences. However, for BERT the word embedding for "bank" would be different for each sentence due to its property of bi-directionality. Moreover, BERT comes as a pre-trained model trained on a large corpus and it will have a general language understanding. For performing the downstream tasks the only step we have to do is fine-tune the model. This property of separating the pre-training phase and fine-tuning phase allows the model to adapt to changing requirements. Since the field of cybersecurity is constantly dynamic, having a model that can be agile is a great advantage.

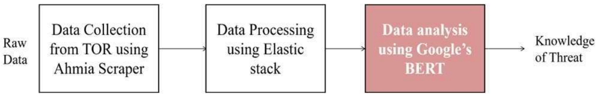

Raw Data
Data Collection from TOR using Ahmia Scraper
Data Processing using Elastic stack
Data analysis using Google's BERT
Knowledge of Threat

Fig. 2. Block diagram of proposed system.

Figure 2 represents the proposed system wherein the Data Analysis module uses Google's BERT instead of the existing Apache Open-NLP Processor. The NLP module will perform the subfunction of data pre-processing tasks. Since the crawled data cannot be processed as it is, data pre-processing steps such as HTML tag removal, stop word removal, etc are necessary. To extract and identify threat-related details, here the NLP task of Named Entity Recognition (NER) will be carried out using the selected model (BERT). NER is the process of identifying entities such as names, organizations, locations, etc and classifying them into a pre-defined group. For making the model perform Named Entity Recognition, fine-tuning should be done followed by testing and evaluation. The process of fine-tuning involves a dataset that is rich in cybersecurity contextual data. Moreover, the vocabulary of the general BERT should be extended to include terms related to cybersecurity.

Figure 3 represents the overall workflow of the NLP module. Since BERT comes as an already pre-trained model it has a basic language understanding. Pre-training typically needs many computational resources and the corpus should be huge. Due to the concept of transfer learning, this task will no longer need to be done again. The most important task that has to be done is the fine-tuning of BERT on a dataset rich in cybersecurity context and could also perform Named Entity recognition. The dataset selected for fine-tuning here is DNRTI or Dataset for Named Entity Recognition in Threat Intelligence [9]. The DNRTI dataset fine-tunes the general BERT model and the resultant model is used to process the incoming crawled data. Since the crawled data as such will not be suitable for ingesting into a fine-tuned BERT it undergoes some pre-processing steps as well. Finally, the BERT performs the NER to identify potential entities associated with threats and attacks which can then be incorporated into a report or analyzed further for more information.

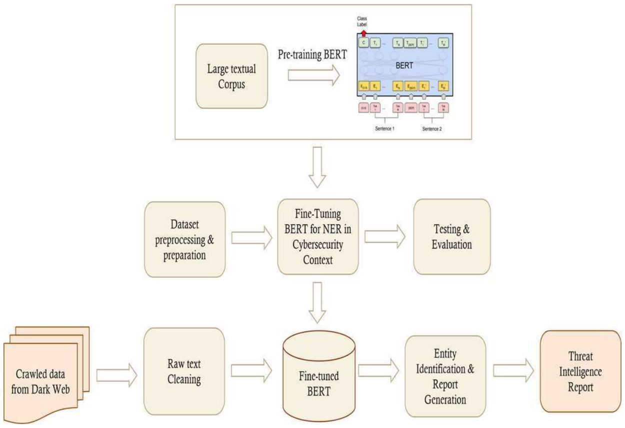
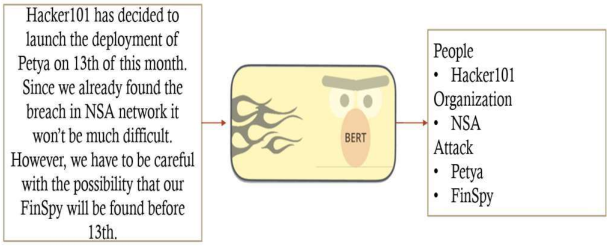

Large textual Corpus
Pre-training BERT
Class Label
BERT
Sentence 1
Sentence 2
Dataset preprocessing & preparation
Fine-Tuning BERT for NER in Cybersecurity Context
Testing & Evaluation
Crawled data from Dark Web
Raw text Cleaning
Fine-tuned BERT
Entity Identification & Report Generation
Threat Intelligence Report

Fig. 3. NLP Module Workflow

Below figure 4 shows a simple example of how the proposed threat intelligence solution will work on some plain textual data.



Hacker101 has decided to launch the deployment of Petya on 13th of this month. Since we already found the breach in NSA network it won't be much difficult. However, we have to be careful with the possibility that our FinSpy will be found before 13th.

People
- Hacker101

Organization
- NSA

Attack
- Petya
- FinSpy

Fig. 4. Entities identification using Natural Language Processing from the sentence to gather Threat Intelligence

As shown in above figure 4, the fine-tuned BERT will be able to correctly identify that entities such as Petya and FinSpy are attacks or threats. For instance, if there is an active discussion going on about a zero-day exploit in a hacker community will at least get the name of the entity being discussed which could greatly help in preparing for the attack or conducting further streamlined searches regarding that attack.

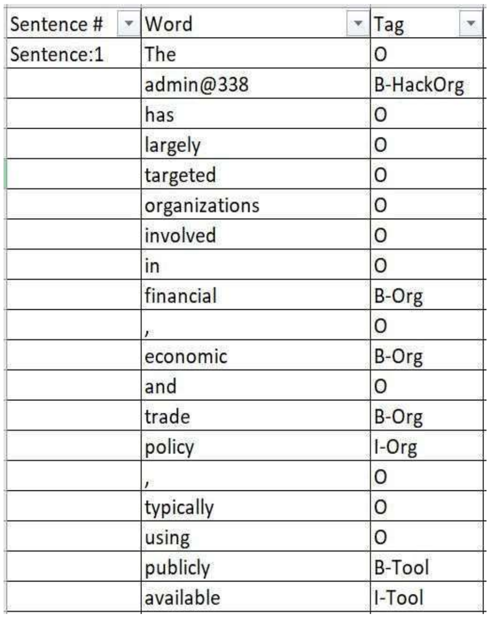
IV. EXPERIMENTAL SETUP

The initial phase of building the darknet threat intelligence toolset constitutes of setting up and configuring Tor Service Over VPN and Ahmia Scrapy crawler to crawl through dark web forums and sites. After crawling, the resulting indexed data was stored in Elasticsearch.

Since the data is stored in JSON format under the index, this can be considered as the data that is ready to be analyzed. Among the crawled dark web forums, there can be a few sites where there is no content present. Such sites can be avoided from the analysis process using the filter in Kibana. Once the data collection and indexing part is over, the next stage is to prepare a suitable high-functioning NLP model based on Google's BERT. Since BERT is already a pre-trained model, it should be fine-tuned so that it will be able to predict and classify entities related to the cybersecurity context. For this fine-tuning process, the DNRTI dataset was prepared and used for training the model.

A. Dataset Selection and Preparation

The dataset selected for this research was the Dataset for Named Entity Recognition in Threat Intelligence (DNRTI)[9]. DNRTI is a selected collection of labeled sentences in the information security domain that are annotated using the BIO format. The BIO labeling is the industry standard labeling form wherein the entities contain either start tags(”B-X”) which indicate the beginning of an entity, or continuation tags(”I-X”) which can be a part of the entity if it has two or more words, or no entity tags(“O”). For instance, the sentence “Joseph belongs to the United States” will be annotated as “B-Person”, ”O”, ”O”, “O”, ”B-Country”, and ”I-Country” (Figure 5). The Sentences in the dataset were collected from various sources such as the websites of security companies, government agencies, threat intelligence reports, and GitHub. The fully annotated DNRTI contains 175220 words. The quality of the dataset was ensured by the authors and the biases were eliminated.



Sentence #	Word	Tag
Sentence:1	The	O
	admin@338	B-HackOrg
	has	O
	largely	O
	targeted	O
	organizations	O
	involved	O
	in	O
	financial	B-Org
	,	O
	economic	B-Org
	and	O
	trade	B-Org
	policy	I-Org
	,	O
	typically	O
	using	O
	publicly	B-Tool
	available	I-Tool

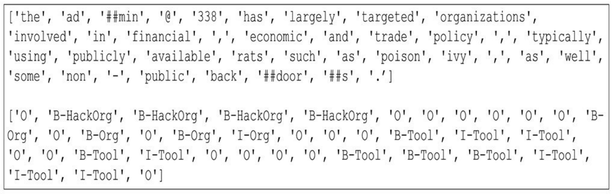
Fig. 5. Labelling in the Dataset.

B. Annotated Labels

The entities in the dataset were categorized into 13 categories which are hacker organization, attack, sample file, security team, tool, time, purpose, area, industry, organization, way, loophole, and features.

The corresponding labels of these 13 categories in the data set are HackOrg, OffAct, SamFile, SecTeam, Tool, Time, Purp, Geo, Idus, Org, Way, Exp, and Features respectively. Among these categories, the "SamFile" category was merged with the "Tool" category, and the "Idus" category was merged with the "Org" category since they contained very similar entities.

Apart from merging the categories, the dataset contained a few errors which were corrected and converted into an excel file. After that, the next important step before training the model was to tokenize the input text. For this BERT tokenizer was used. BERT also leverages the use of WordPiece Tokenization which is breaking a word into multiple tokens as shown in Figure 6.



```
['the', 'ad', '##min', '@', '338', 'has', 'largely', 'targeted', 'organizations',
'involved', 'in', 'financial', ',', 'economic', 'and', 'trade', 'policy', ',', 'typically',
'using', 'publicly', 'available', 'rats', 'such', 'as', 'poison', 'ivy', ',', 'as', 'well',
'some', 'non', '-', 'public', 'back', '##door', '##s', '.']

['O', 'B-HackOrg', 'B-HackOrg', 'B-HackOrg', 'B-HackOrg', 'O', 'O', 'O', 'O', 'O', 'O', 'B-
Org', 'O', 'B-Org', 'O', 'B-Org', 'I-Org', 'O', 'O', 'O', 'B-Tool', 'I-Tool', 'I-Tool',
'O', 'O', 'B-Tool', 'I-Tool', 'O', 'O', 'O', 'O', 'B-Tool', 'B-Tool', 'B-Tool', 'I-Tool',
'I-Tool', 'I-Tool', 'O']
```

Fig. 6. Results of BERT WordPiece tokenization.

C. Training the model and Finetunig the Hyper Parameters

The model "bert-base-uncased" is imported from the huggingface library using the function from_pretrained. There are other released pre-trained BERT models such BERT-Large-Uncased(24-layer, 1024-hidden, 16-heads, 340M parameters), BERT-Base-Cased(12-layer, 768-hidden, 12-heads, 110M parameters), BERT-Large-Cased(24-layer, 1024-hidden, 16-heads, 340M parameters), and BERT-Base-Multilingual Cased (104 languages, 12-layer, 768-hidden, 12-heads, 110M parameters). The reason for selecting an uncased model is that in most of the observed dark web forum posts, the formatting and capitalization of words were given less importance since it is more of an informal platform.

Further, the hyper-parameters were defined such as learning rate, number of epochs, sentence length, max Gradient Norm, Batch size, etc. These parameters were selected based on trial and error for reaching an optimum accuracy.

TABLE I. HYPER PARAMETERS

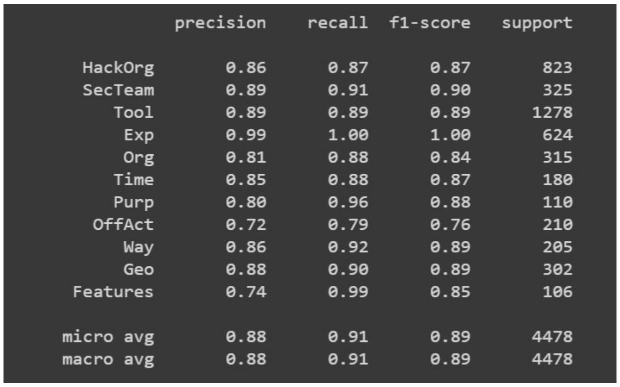
Number of Epochs	20
Batch Size	32
Sentence Length	128
Max Gradient Norm	1.0

D. Evaluation

The performance of BERT based NER model was measured by Precision, Recall, and F1 Score. These values are calculated against tokens and Entity labels. In addition, the Support value is also calculated to identify the true response of data. Besides, these Models tested on unseen sentences to identify the entity and its whereabouts such as B-label (e.g., 'B-HackOrg') if the token constitutes the beginning of a named entity, I-label (e.g., 'I-HackOrg') if it is inside a named entity, but not positioned first, or (iii) O-label ('O') if it is not part of a named entity (i.e., it is outside of it).

V. Results

The fine-tuned BERT model was evaluated based on the F1 score, precision, and recall, for each category and macro and micro scores as shown in Figure 7. Here the results indicate that the BERT model could identify and categorize cyber-threat-related entities with an f1-score close to 0.9 in each category.



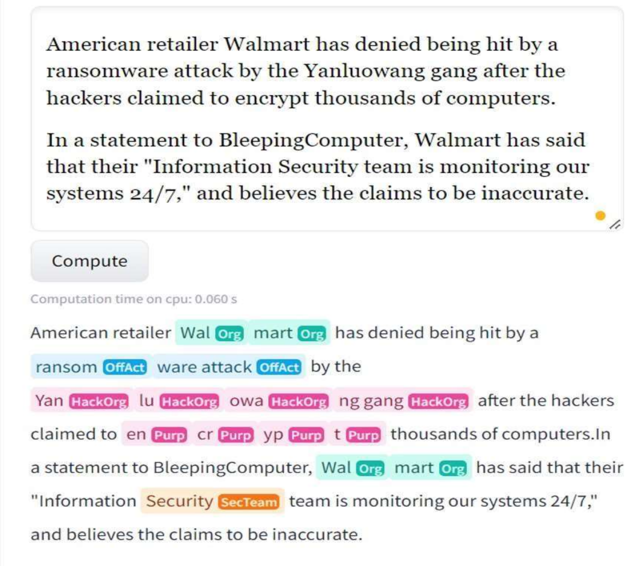
```
              precision    recall  f1-score   support

     HackOrg       0.86      0.87      0.87       823
     SecTeam       0.89      0.91      0.90       325
        Tool       0.89      0.89      0.89      1278
         Exp       0.99      1.00      1.00       624
         Org       0.81      0.88      0.84       315
        Time       0.85      0.88      0.87       180
        Purp       0.80      0.96      0.88       110
      OffAct       0.72      0.79      0.76       210
         Way       0.86      0.92      0.89       205
         Geo       0.88      0.90      0.89       302
    Features       0.74      0.99      0.85       106

   micro avg       0.88      0.91      0.89      4478
   macro avg       0.88      0.91      0.89      4478
```

Fig. 7. Classification Report.

The given figure 8 indicates how the fine-tuned BERT model works on an unseen sentence and how entities are identified and categorized into a specific category.



American retailer Walmart has denied being hit by a ransomware attack by the Yanluowang gang after the hackers claimed to encrypt thousands of computers.

In a statement to BleepingComputer, Walmart has said that their "Information Security team is monitoring our systems 24/7," and believes the claims to be inaccurate.

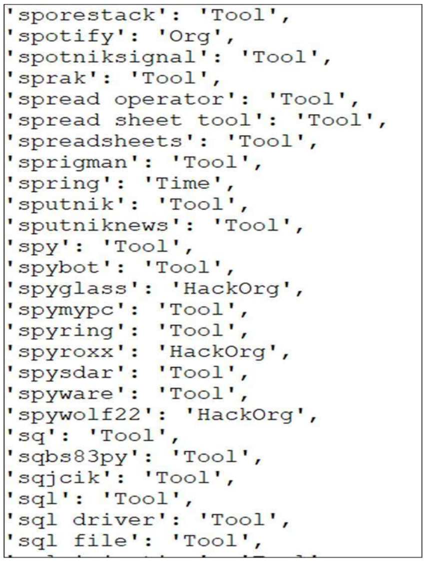
Computation time on cpu: 0.060 s

American retailer Wal Org mart Org has denied being hit by a ransom OffAct ware attack OffAct by the Yan HackOrg lu HackOrg owa HackOrg ng gang HackOrg after the hackers claimed to en Purp cr Purp yp Purp t Purp thousands of computers.In a statement to BleepingComputer, Wal Org mart Org has said that their "Information Security SecTeam team is monitoring our systems 24/7," and believes the claims to be inaccurate.

Fig. 8. Results of entity identification on random sentence

This fine-tuned model was then used to identify entities from the crawled dark web data. Since the model needs textual input, only raw content from the website was taken using an API call to Elasticsearch. The raw content was pre-processed to remove unwanted characters, spaces, stop words, etc. using the NLTK library. Along with that, using regex, URLs and IP addresses were also identified. The raw text was given into the trained model to make predictions and classify the entities into predefined labels as shown in figure 9.

The prediction from the fine-tuned BERT indicates that the model was able to identify HackerIds or Threat Actor names used in the forums although this might not be related to



```
'sporestack': 'Tool',
'spotify': 'Org',
'spotniksignal': 'Tool',
'sprak': 'Tool',
'spread operator': 'Tool',
'spread sheet tool': 'Tool',
'spreadsheets': 'Tool',
'sprigman': 'Tool',
'spring': 'Time',
'sputnik': 'Tool',
'sputniknews': 'Tool',
'spy': 'Tool',
'spybot': 'Tool',
'spyglass': 'HackOrg',
'spymypc': 'Tool',
'spyring': 'Tool',
'spyroxx': 'HackOrg',
'spysdar': 'Tool',
'spyware': 'Tool',
'spywolf22': 'HackOrg',
'sq': 'Tool',
'sqbs83py': 'Tool',
'sqjcik': 'Tool',
'sql': 'Tool',
'sql driver': 'Tool',
'sql file': 'Tool',
```

Fig. 9. Entities identified by the proposed model.

real-life identities. It could also identify software tools, utilities as well as real-world organizations with good accuracy. Hence by employing this model, continuous monitoring and thereby identification of entities in dark web forum discussions can be done. This could help in making a more focused approach to finding new threats, data breaches, or possible exploits.

VI. CONCLUSION

Since the cyber threat landscape is changing rapidly, automated threat Intelligence collection is becoming a necessity rather than a value-add for today's organizations. The Threat Intelligence information extracted from the dark web can be critical in the decision-making process due to various reasons. To extract actionable insights from the raw data there are many open-source tools available for collecting, processing, and analyzing these unstructured data. With the literature survey, it was possible to study and understand the works done in the area of knowledge extraction from textual data using NLP models. Furthermore, an open-source automated threat intelligence toolset based on the Intelligence lifecycle was analyzed. This led to framing a research goal and proposing a deep learning framework called BERT instead of the standard NLP models. The model was fine-tuned using the selected dataset, tested, evaluated, and fed with the raw data crawled from the dark web. The model was able to predict many Software tools used in the industry and identified HackerIds used in the chat forums. Along with this, the feasibility of deploying the model directly from Elasticsearch was studied. Although it supported generic transformer-based models the fine-tuned model was not supported since it contained new categories other than generic ones. It is expected that in the coming Elasticsearch versions, they could include more fine-tuned models with various categories so that deployment will be much easier.

REFERENCES

[1] Huang C, Guo Y, Guo W, Li Y, "HackerRank: Identifying key hackers in underground forums", *International Journal of Distributed Sensor Networks. 2021*, 17(5). doi:10.1177/15501477211015145

[2] Chia-Mei Chen, Dan-Wei Wen, Ya-Hui Ou, Wei-Chih Chao, and Zheng-Xun Cai, "Retrieving Potential Cybersecurity Information from Hacker Forums", *International Journal of Network Security*, vol.23, no.6, pp.1126-1138

[3] I. Deliu, C. Leichter and K. Franke, "Extracting cyber threat intelligence from hacker forums: Support vector machines versus convolutional neural networks", *2017 IEEE International Conference on Big Data (Big Data),* pp. 3648-3656

[4] P. Evangelatos et al., "Named Entity Recognition in Cyber Threat Intelligence Using Transformer-based Models," *2021 IEEE International Conference on Cyber Security and Resilience (CSR),* pp. 348-353

[5] P. Ranade, A. Piplai, A. Joshi and T. Finin, "CyBERT: Contextualized Embeddings for the Cybersecurity Domain," *2021 IEEE International Conference on Big Data (Big Data)*, pp. 3334-3342

[6] Brian Nafziger, "Data Mining in the Dark: Darknet Intelligence Automation", SANS Whitepaper, 2017 . [Online]. Available: https://www.sans.org/white-papers/38175/

[7] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[8] Vaswani, Ashish & Shazeer, Noam & Parmar, Niki & Uszkoreit, Jakob & Jones, Llion & Gomez, Aidan & Kaiser, Lukasz & Polosukhin, Illia, Attention Is All You Need, 2017 . [Online]. Available: https://dl.acm.org/doi/10.5555/3295222.3295349

[9] X. Wang et al., "DNRTI: A Large-Scale Dataset for Named Entity Recognition in Threat Intelligence," *2020 IEEE 19th International Conference on Trust, Security, and Privacy in Computing and Communications (TrustCom)*, pp. 1842-1848