Cyber Threat Analysis Using Natural Language Processing for a Secure Healthcare System

Shareeful Islam
School of Computing and Information Science
Anglia Ruskin University
Cambridge, United Kingdom
Focal Point
Waterloo, Belgium
shareeful.islam@aru.ac.uk

Spyridon Papastergiou
Department of Informatics
University of Piraeus
Piraeus, Greece
Focal Point
Waterloo, Belgium
paps@unipi.gr

Stefano Silvestri
Institute of High Performance Computing and Networking (ICAR)
National Research Council of Italy (CNR)
Naples, Italy
stefano.silvestri@icar.cnr.it

2022 IEEE Symposium on Computers and Communications (ISCC) | 978-1-6654-9792-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISCC55528.2022.9912768

***Abstract*—Cyber threats in the healthcare sector have increased significantly in recent years. Attackers are now using sophisticated techniques to launch multi-phase cyber attacks to compromise the system and leak patient healthcare data. Healthcare organisations need to protect IT infrastructures and understand the threats and possible attack surface for a secure healthcare service delivery. Hence, threat analysis is one of the key activities for tackling the potential risks and ensuring security of a system context. This work presents a threat analysis approach that allows to identify and assess the possible threats within healthcare information infrastructure. The approach considers the existing threat data from widely used repositories and uses Natural Language Processing to identify threats among cyber security news, also evaluating their corresponding level. The preliminary experimental assessment shows promising results, providing a realistic manner to assess the threats, allowing to adopt the proposed approach in real-world contexts.**

***Index Terms*—Healthcare Ecosystem, Cyber Threat, Deep Learning, Natural Language Processing, Healthcare Information Infrastructure**

I. Introduction

The healthcare sector is continuously adopting new technologies from connected healthcare and wearables devices and Internet of Thing (IoT) to medical applications and patient portals for improving overall patient experiences and service delivery. This massive technological transformation not only provides benefits but also increases attack surface where threat actors can exploit possible threats for any potential risk within the Health Care Information Infrastructure (HCII). There are a number of successful cyber attacks in recent years in the healthcare sector, notably ransomware attack in the Ireland's Department of Health and Health Service Executive in 2021 and NHS 2017 [1]. Additionally, there are inherent vulnerabilities in the medical devices, such as flaws in Braun's infusion pump or Medtronic insulin pump, that could pose potential threat to the patient health [2]. Hence, nearly 90% of healthcare organisations have experienced a data breach in 2018 [3]. There is a need to protect the interconnected cyber systems and infrastructures from any potential threats for a secure healthcare service delivery [4].

A large amount of unstructured Natural Language (NL) Cyber Security (CS) data related to the healthcare domain is available on the Internet. This textual data contains often crucial and updated information related to the assets of the HCII and of the healthcare supply chain, such as threats, vulnerabilities, attacks, and other important CS information, which could be very useful to improve the protection of the HCIIs. On the other hand, it is often difficult to identify and extract the relevant information from such kind of sources, which are usually available on blog posts, CS news websites, social media and others, in order to leverage them for the development of CS systems. In particular, the complexity of the NL, which can present polysemy, irony, complex and long sentences and others, in addition to the peculiarities the CS domain, such as a large presence of non-standard abbreviations or acronyms [5], make it difficult to automatically extract the required information buried under the text.

Some of these issues have been recently addressed in literature by specifically tailored Deep Learning (DL) for Natural Language Processing (NLP), which implement Named Entity Recognition (NER) for CS [6]–[9], also integrating domain-specific Knowledge Bases (KBs) and catalogues, allowing to integrate NLP in CS frameworks. NER is a task of Information Extraction that identifies and classifies the named entities mentioned in NL texts, which are words of multi-word expressions belonging to a specific domain. Examples of potential named entities from the CS domain could be attack types (e.g., *Denial of Service*, *fishing*, etc.), assets (*MySQL*, *Apache Tomcat*, etc.), threats (*ransomware*, etc.), vulnerabilities (*Broken Authentication*, *injection*, etc.), and others. Mining and identifying the most updated CS threats from the huge amount of information available in NL documents in the Internet can support the establishment of situational awareness proactively monitoring and preventing CS issues [10], but specifically tailored approaches are required [11].

The research presented in this paper aims to ensure cyber security of the HCIIs by identifying and analysing the

This research was funded by the European Commission, grant number 883273, AI4HEALTHSEC - A Dynamic and Self-Organized Artificial Swarm Intelligence Solution for Security and Privacy Threats in Healthcare ICT Infrastructures.

978-1-6654-9792-3/22/$31.00 ©2022 IEEE

threats, leveraging NLP NER to identify them and also to evaluate their level. In particular, we adopt relevant catalogues to improve the identification of threats within the HCIIs, i.e., Common Attack Pattern Enumeration and Classification (CAPEC), Common Platform Enumeration (CPE), in conjunction with NLP NER approaches, allowing for the extraction and classification of the threats and the corresponding assets from updated data from Internet and, moreover, also allowing for the identification of the level of the threats.

Our work makes three important contributions. Firstly, it focuses on the holistic understanding of the threats taking into account the existing healthcare context. The approach considers two main components to understand the healthcare context and the analysis the threats within this context. In particular, healthcare context is decomposed into *services* and *assets*. Secondly, as mentioned above, the proposed approach adopts the NLP not only to identify the possible threats and assets related to HCIIs among blog, news and social media posts, but also to determine the level of the risk associated to a specific threat. For this purpose, we leverage a CS domain-specific BERT neural language model [12] fine tuned for the Named Entity Recognition (NER) task, using data extracted from a CS news website. The NER model is trained on a dataset annotated using Distant Supervision [13], exploiting the information available in the considered catalogues. Our experimental assessment shows promising results, which allow to test the proposed approach in real-world HCCIs.

The paper, after presenting the related works in Section II, describes the details of the proposed approach in the Section III and the experimental assessment in the Section IV, also including the details of the used datasets and resources and a discussion of the preliminary results. Finally, in Section V, the conclusions and the future work are reported.

II. Related Works

There are several works that focus on threat modelling and assessment. Notably, PASTA and Attack Tree are well known threat modelling methods [14]. PASTA is a risk-centric approach that identifies security flaws and possible impact so that appropriate controls can be determined for the mitigation. The model advocates analyst-business collaboration with the intent to assess, document, and propose countermeasures relative to the likelihood of an attack. Attack Tree follows a tree-based hierarchical structure to describe security of a system. The root node considers goal, while the lower level nodes consider the possible attack to the system. It provides potential attack patterns for specific targets while describing threats aimed at a system and the possible counterattack approaches to realise them. The Centre for Internet Security (CIS) reveals that a number of attacks such as ransomware, data breaches, DDos, and inside threats are commonly used by the attacker in the healthcare sector [15]. A recent study showed that at least 20% of the medical device manufacturers experienced ransomware or malware attacks in the last 20 months [16]. Cyber attacks can target medical devices, such as infusion pump or healthcare services, such as medicine delivery of the healthcare system [17]. The works in literature emphasises on the control like patch management and incident manage to improve security of a hospital. A cyber supply chain threat analysis integrates Random Forest and GBoost algorithms for the threat prediction [18]. The work consider threat intelligence and predicts the TTP deployed for a cyber attack and Gboost provides higher accuracy. A novel threat analysis framework named SHChecker was proposed by [19], combining machine learning and formal analysis capabilities for the Smart healthcare systems (SHSs). The paper considers Internet of Medical Things (IoMT) and adopts a number of ML algorithms including Decision Tree, Artificial Neural Network, K-means Algorithm, etc. The result shows that NN-based algorithms provides less accuracy than DT based algorithms.

Recently, also the Natural Language research community started to propose in literature some methodologies and techniques able to leverage NLP for the definition of innovative CS approaches. In [9], a data and knowledge-driven CS Named Entity Recognition (NER) method is presented, which exploits a BiLSTM-CRF with multi-head self-attention neural network with word embeddings trained on CS closed-domain texts to improve their effectiveness [20], in conjunction with external dictionary knowledge, for the recognition of the details of the assets (application, vendor, version, etc.) involved in CS issues. The results showed an improvement over the baseline. A DL-based architecture for the identification of relevant CS information, such as vulnerability exploitations, attack discoveries and advanced persistent threats, was also presented by [6]. The architecture is formed by a word-embedding layer, a bidirectional LSTM layer, and a Conditional Random Field (CRF) layer, concatenated with a further bidirectional LSTM output, capable to show improvements with respect to the baselines.

Also the BERT-based architectures [12] have been recently exploited for CS NER. An example is the CyBERT model, presented by [21], able to implement a semi-automated CS vetting for industrial control systems (ICS). This model was trained on a specifically-created corpus of labelled sequences from ICS device documentation, collected across a wide range of vendors and devices, improving the obtained results compared to models trained on generic domain. Also in [8] the author proposed a BERT-based model fine tuned for the CS NER task, improving the obtained results using domain dictionaries. In [7], a CS NER approach was implemented using a model that integrates BERT and a BiLSTM-CRF DL architecture. The authors of [22] proposed a semantic schema to describe CS events, leveraging a DL-based information extraction (IE) pipeline able to support the automatic extraction of structured information about data breaches, ransomware and phishing attacks, the discovery and the patches of vulnerabilities, from articles about CS. The authors of [23] presented a method to analyse the severity of CS threats based on the used language through a DL approach, exploiting a corpus of 6,000 tweets describing software vulnerabilities, annotated with authors' opinions toward their severity, also presenting a method for linking software vulnerabilities reported in tweets to CVEs and

NVD databases. Their results showed that an high precision in forecasting high-severity vulnerabilities, also highlighting that reports of severe vulnerabilities online are predictive of real-world exploits.

In summary, the threat modelling approaches provides a guideline to identify and analyse the threats. Several works describe various cyber attacks in healthcare sectors and recently some work adopts NLP DL-models for threat analysis. Our work mainly differs from these contributions specifically we focus on Natural Language Processing not only to extract relevant treats information from the texts, but also to determine the threat level for the healthcare sector.

III. Proposed Approach

The work aims to identify and asses the cyber-threats for securing the healthcare system. It considers security from the context of healthcare ecosystem and other related components.

The proposed approach consists of two main components:

- Healthcare Ecosystem Context Component;
- Threat Assessment Component.

These components are linked upon each other and includes a number of steps to perform specific functionalities. A detailed overview of the components is given below.

A. Healthcare Ecosystem Context Component

Healthcare ecosystem is a complex system that consists of heterogeneous set of actors, entities, and systems (such as hospitals and social service organisations, medical equipment suppliers, pharmacies, health care research labs, devices developers, etc.) who are involved in the healthcare process and service delivery, including patient treatment, appointment, surgery and many others. This ecosystem is huge and includes a widely distributed network, including an interconnected set of healthcare entities (organisations, such as hospital agencies or clinics or individuals, like doctors) that implement healthcare services which provision relies upon interdependent HCIIs (e.g., IT and Operational Technology (OT) systems) comprising interconnected sets of assets (e.g., implants, sensors, healthcare software, such as patients' health records, pathology scanners and servers, medical x-rays equipment).

Within the last decades, there are a significant digital advancements within the whole ecosystem to support the healthcare service delivery and increase the interdependencies between physical and cyber levels. This composite and dynamic nature of digital interconnectivity has altered the threat landscape posing new cyber threats attracting the attention of adversaries to develop new security and privacy challenges committing sophisticated coordinated cyber-attacks that could cause a dramatic impact to the healthcare ecosystem. For instance, a cyber-attack on insecure imaging servers and unprotected data storages supporting medical x-rays can lead to the web exposure of sensitive information of patients, such as medical images and scans; or a comptonisation of a remote monitoring software of defibrillators could allow adversaries to take advantage of the system damaging the hospital equipment or amending of medical device configuration [4]. Therefore, it is necessary to identify and analyse the threats that could pose any potential risk within the ecosystem.

This component investigates the overall healthcare ecosystem context based on the possible services and assets related with the services. Therefore, it includes service and asset inventory of the healthcare information infrastructure. A healthcare entity delivers various services and some of them are critical relating to patient treatment. It is necessary to generate a comprehensive list of services, e.g., patient appointment, remote consultation, surgery schedule, medical report, patient registration, etc. Service is viewed as a business process, where a collection of activities and tasks form a Business Flow, ensuring the proper operation of the service. Each business process is part of a specific healthcare ecosystem and may depend on external actors.

Once the services are identified, it is necessary to decompose the services and identify the assets which are related with them. Our approach advocates to use the Common Platform Enumeration (CPE)[1] catalogue to map the HCII assets with specific classes of applications, operating systems, and hardware devices. CPE provides a structure naming for the assets. The inventory tools and scanners can also assist to automatically identify the assets. The identified assets are the internal system components that are controlled by the examined healthcare organisation(s). We have considered four distinct healthcare areas as presented in Table I to describe the assets within the HCII. Additionally, assets are also categorised depending on its functionalities, as shown in Table II. This allows to determine the importance of each asset within the ecosystem.

TABLE I
Assets areas

Area	Name
1	User interactions with implants and sensors
2	Medical equipment and IT devices
3	Services and processes
4	Interdependent HCIIs – Ecosystem

TABLE II
Assets categories

Area	Name
Influence	Found in most organisations, distinct
Type	Software, hardware, Operating System (OS), information Sensitivity
Sensitivity	Restricted, unrestricted
Criticality	Essential, required, deferrable

B. Threat Assessment Component

This component has the purpose of identifying and prioritising the threats by following the services and assets. Individual threats can be considered as potential stepping stones to security risks (deliberate or accidental), which may

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

affect those services and assets. The identified threats can be categorised through threat taxonomies and assessed in a qualitative manner using threat scales. It consists of two steps, i.e., **identify threats** and **prioritise threats**.

This initial step is the identification of threats and it focuses on the potential threats for each identified asset and considers threat intelligence data for this purpose. There are several available sources that catalogue known threats along with their characteristics, such as Common Attack Pattern Enumeration and Classification (CAPEC)[2] to identify the threats relevant to the HCII. A set of threat characteristics from the CAPEC is considered to describe the threats. A partial list of the characteristics is given below.

- *Abstraction*: Defines the different abstraction levels that apply to an attack pattern. A Meta level attack pattern provides an abstract characterisation of a specific methodology or technique used for an attack and generalisation of a related group of standard level attack pattern. It is often void specific technology or implementation and provides an understanding of a high-level approach.
- *Status*: Defines the different status values that an entry of the CAPEC catalogue including view, category, attack pattern.
- *Description*: A short description of the threat.
- *Vendor and Item*: Respectively identify the vendor and item (e.g., *Google* and *Chrome*) affected by the CS issue.
- *Likelihood of Attack*: Determines the likelihood and severity of an attack that leverages using the attack pattern and may not be completely accurate for all attacks.
- *Related Attack Patterns*: Refers to other attack patterns and related high-level categories. These relationships give insight to similar items that may exist at higher and lower levels of abstraction.
- *Execution Flow*: It is used to provide a detailed step-by-step flow performed by an adversary for a specific attack pattern. It is applicable to attack patterns with an abstraction level of details.
- *Prerequisites*: Indicates one or more prerequisites conditions necessary for an attack.
- *Skills and Resource Required*: Describe skill level or knowledge and possible resources (e.g., CPU cycles, IP addresses, tools) required by an adversary for an attack.
- *Indicators*: The possible indicators including activities, events, conditions, or behaviours that may indicate an attack which could be imminent, in progress, or has occurred. Each Indicator element provides a textual description of the indicator.
- *Consequences*: The possible consequences associated with an attack pattern. The required Scope element identifies the security property that is violated. The optional Impact element describes the technical impact that arises if an adversary succeeds in their attack.
- *Mitigation*: The suitable counter measure to prevent or mitigate the risk of an attack. The approaches described in each mitigation element should help improve the resiliency of the target system, reduce its attack surface, or reduce the impact of the attack if it is successful.

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

The latter step aims to prioritise the threat, so that healthcare organisations can proactively determine the suitable controls to tackle the identified threats. We have considered the history of reported incidents related to those threats for the threat prioritisation. This step follows the threat profiling from the first step and investigates threat-related information through a series of known sources, ranging from CS news websites, CS blogs and social media, to threat and vulnerability catalogues for references of incidents related to specific CAPEC categories for the threat level calculation. For this purpose, we implemented an automated analysis of the various available sources exploiting Natural Language Processing (NLP) NER approach, which can be applied to perform an analysis on unstructured text data as presented in Figure 1. A set of input natural language sources corresponding to threat reports, articles on various blogs/websites, Twitter data, online publicly available datasets, and/or log-files of the HCIIs can be fed as input into the NLP module.

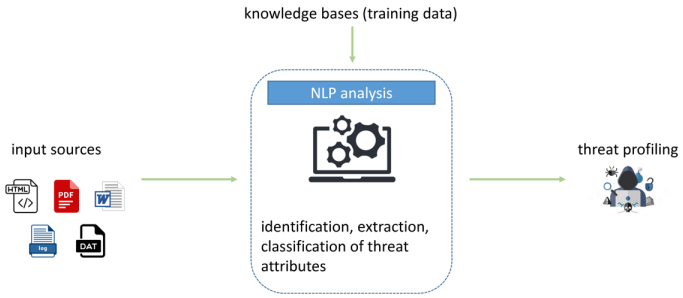

Fig. 1. A conceptual schema of threat profiling and assessment.

The most essential threat attributes can be automatically extracted by the NER module, which has been previously trained on source documents annotated applying DS-based on the available Knowledge Bases (e.g., CAPEC information), and then manually reviewed by domain experts. The NER module is a supervised machine learning method, and thus prior knowledge is needed in order to train the appropriate model. Distant Supervision uses the knowledge extracted from KBs (thesauri, catalogues, dictionaries, etc.), assuming that if a string in text is included in the KB, it can be automatically annotated as a named entity. This approach is automatic, but it suffers of incomplete and noisy annotations [13], which are partial identification of an entity, due to the presence in the thesaurus of only a part of an entity or due to slight differences among the entity listed in the thesaurus and the one in the corpus (e.g., the a synonym in multi-word entity, or a plural version of the same word). For this reason, the data obtained from DS must be manually reviewed, preventing the use of this approach for large datasets. We use the computed list of threats and assets as KBs for DS. This list contains the detected threats per relevant asset and their related categories that operate for

the provision of each identified healthcare service.

We fine-tuned for the NER task a BERT-based model [12] pretrained on a large document collection belonging to the CS domain, using the distantly supervised annotated data. The obtained NER model allows to identify the assets and the threats among a large document collection of news in natural language, which is periodically extracted from the web, including in this way the most updated information. The NER model can address the issues of the DS, such as noisy or incomplete annotation, thanks to the generalisation capabilities of the DL-based method [24], improving the detection of the relevant named entities.

The NER module not only improves the identification of potential threats related to the assets of the HCII, but is also exploited to evaluate the level of the threats, which we correlate to their occurrence in the dataset analysed through the proposed approach. In particular, we calculate the percentage of the occurrence of each identified threat for an asset, increasing the number of the occurrence whenever the same threat and assets are mentioned in the same sentence. Then, we assign a threat level based on this percentage of occurrence, as shown in the next Table III. We assume that if the percentage of occurrence of a specific threat is high in the existing datasets, also its threat level is high. We identified five different levels, from *Very High* to *Very Low*.

TABLE III
THREAT PROFILING AND ASSESSMENT

Threat Level	Percentage of Occurrence Range
Very High	[80-100]
High	[60-80]
Medium	[40-60]
Low	[20-40]
Very Low	[1-20]

IV. EXPERIMENTS

This Section discusses the experiments we conducted for the threat level assessment. It first presents the details of the used datasets and the other adopted tools and resources. Then, it describes the experiments and shows the obtained results. The purpose of the experiments is to assess the usefulness of the proposed methodology, investigating on the effectiveness of the proposed method for the evaluation of the threat level.

A. Datasets

A CS news posts collection has been used for both fine-tuning and test the NER model, as well as for testing the threat level evaluation approach based on the occurrence of the threats and the assets. This corpus has been extracted from The Hacker News website[3], a CS news platform that attracts over 8 million readers monthly, which is daily updated with the latest CS news and provides in-depth reports on current and future CS trends. The website is daily updated and contains tons of documents, describing threats, attacks, vulnerabilities and other CS topics.

For the preliminary experimental assessment presented in this paper, we randomly extracted a limited number of posts from The Hacker News website, equal to 7,410. These posts have been further randomly split into three datasets: a training set (20%) and a test set (10%) for the NER model, and a dataset for testing the threat level evaluation approach (70%, hereinafter, Threat Level (TL) dataset). The features of the datasets, in terms of number of news posts and corresponding word counts, are reported in Table IV.

TABLE IV
DATASETS FEATURES

Dataset	News post count	Word count
NER Training set	1,568	40,142
NER Test set	801	20,247
Threat Level (TL) dataset	5,041	129,553

The text of the news have been extracted through a web crawler and a web scraper specifically implemented using a set of specific Python scripts. It is worth noting that the scripts run once a week, updating the dataset with the latest news, continuously increase the available information for the future real-world applications of the proposed approach.

The CAPEC database used for the DS annotation of the NER datasets is structured as a JSON. It has been preprocessed, extracting the entries labelled as *threat*, their corresponding *product* and *vendor* labels in order to identify the assets, the description of the threats under the *description* label and the content of the *id* label, which include the coding of the corresponding threat (e.g., CVE-2021-37971, CAPEC-103, etc.). The relevant information has been included in a list, used to apply the DS for the annotation of the training set: the assets and their related threats mentioned in each sentence of the blog posts of the training set have been annotated be means of DS, after preprocessing the text by applying lowercasing, tokenization and sentence splitting. As mentioned above, the training set annotated through DS has been manually reviewed. The features of the training set, in terms of number of assets and threats annotated, are reported in Table V.

TABLE V
NER DATASET ANNOTATED ENTITIES

Dataset	Threat count	Asset count
Training set	231	1,198
Test set	115	587

B. Resources and Tools

The NLP NER relies on a BERT model trained on a very large CS document collection named SecBERT[4]. In detail, this model was trained on a corpus formed by: i) APTnotes[5],

[3]https://thehackernews.com

[4]https://github.com/jackaduma/SecBERT

[5]https://github.com/aptnotes/data

a collection of publicly-available papers and blogs (sorted by year) related to malicious campaigns/activity/software that have been associated with vendor-defined APT (Advanced Persistent Threat) groups and/or tool-sets; ii) the text extracted from the website included in Stucco-Data[6], a repository that keeps a list of the data sources that are potentially relevant to cyber security and the source for the web site to make the data sources easy to read (including the texts from CPE, CVE and other databases, as well as blogs, forums, bulletin boards, etc.); iii) a corpus of corpus of 1,000 English news articles from 2017–2019 used for CASIE project [22]; i) the datasets of SemEval 2018 Task 8 SecureNLP [25], a shared task on semantic extraction from CS reports. The model has 12 attention heads, 6 hidden layers and has an hidden size equal to 768. The SecBERT model has been fine-tuned on the NER task using the Huggingface Python library[7], which offers a set of API for training and fine-tuning Transformers-based Neural Language Models.

For the preprocessing of the textual data and the implementation of the DS annotation, we have exploited Spacy[8], a flexible NLP Python library that includes tools for tokenization, sentence splitting and other NLP preprocessing tasks.

C. Preliminary Results and Discussion

The first part of our experiments aimed at verifying the effectiveness of the NER model based on SecBERT, comparing the obtained performances in terms of Precision, Recall and F1-Score [26] with i) the ones obtained using DS and ii) a baseline BERT model (*BERT-base-uncased* [12], pretrained on a large general-domain corpus) fine-tuned on the same training set. The obtained results in terms of Precision, Recall and F1-Score are reported in the next Table VI. As we can see, the metrics confirm that the SecBERT model, pretrained on a large CS closed-domain corpora collection and fine-tuned on the dataset specifically created for our purposes, provides a performance boost with respect to the baseline BERT model and a DS rule-based annotation. Moreover, the performances level of the NER model permits to leverage it to mine the relevant information for the threat level assessment from the larger CS news dataset.

TABLE VI
NER RESULTS

Method	Precision	Recall	F1-Score
DS	0.9569	0.7897	0.8654
BERT	0.9554	0.7859	0.8623
SecBERT	**0.9662**	**0.7995**	**0.8750**

The Table VII shows the number of the assets and threats identified through the NER model among the TL dataset. After the extraction of the relevant entities, the same document collection has been preprocessed, applying sentence splitting, with the purposes of selecting only the sentences where a mention of both an asset and a threat is present, allowing in this way to identify the assets and the corresponding threats. Totally 1,654 sentences containing a mention of both assets and threats have been extracted.

TABLE VII
ENTITIES EXTRACTED IN TL DATASET

Entity Type	Number of Entities
Threat	639
Asset	2,187

A threat occurrence for each asset and the corresponding percentages have been calculated, defining the corresponding level of threat following the ranges of the percentage of occurrence, as shown in the previous Table III. At this point, it was possible to associate the threat level to the assets of the services of the HCIIs (summarised in the previous Table I), identified by the Healthcare Ecosystem Context Component. A mapping among those assets and the couple asset/threat extracted through NLP, with the corresponding threat level, allows for the identification of the threats and the evaluation of their corresponding level. As shown in Table VIII, in our preliminary experiments, it was possible to extract from the CS news 78 threats for 5 different assets. Moreover, each threat has been also characterised by its corresponding level (we identified various threat levels, depending on their occurrence).

TABLE VIII
ENTITIES EXTRACTED IN TL DATASET

Assets	Threats count
Apache Tomcat	33
Adobe Reader	8
Google Chrome	22
Laravel framework	5
Debian Linux	10

The results obtained from these preliminary experiments are promising. Firstly, we were able to create a list of the HCII assets by exploiting the available CS catalogues. Then, although we used a limited number of documents and sources, it was already possible to identify a significant number of threats for a set of assets involved in the HCII, also evaluating their of corresponding level. This confirmed that the proposed approach can be exploited to develop a CS situational awareness framework and to support the monitoring and the prevention of CS incidents in the HCIIs.

V. CONCLUSION AND FUTURE WORK

The paper presents a threat analysis method using Natural Language Processing for securing the healthcare supply chain. The proposed method assesses a specific threat based on the occurrence evaluated through NLP applied to published CS news website. Initially, assets related to the specific HCII are identified and linked with the possible threats that impact on the asset. The approach leverages a NLP NER technique based on a BERT model pretrained on CS closed-domain corpus, also supported by CS catalogues.

[6]http://stucco.github.io/data/
[7]https://huggingface.co
[8]https://spacy.io

We are planning to improve the features of the proposed approach, as well as to test it on larger dataset and to apply it in a real world environment. In detail, the identification of the assets and the vulnerabilities could be refined, by applying Relation Extraction techniques [27] in order to better identify and classify the relation between them. The approach will be also improved, in order to include the identification and the evaluation of the vulnerabilities. The methodology will be also tested on larger datasets. Firstly, the dataset formed by CS news is constantly updated and enlarged, by extracting the more recent news. Moreover, we are including a social media dataset in the threat assessment process, integrating the dataset with a large collection of CS-related tweets. Finally, the method will be tested in real pilot environments, in particular within the pilot studies of the AI4HEALTHSEC[9] EC-funded project.

REFERENCES

[1] D. Rees. (2021) Cyber attacks in healthcare: the position across Europe. [Online]. Available: https://www.pinsentmasons.com/out-law/analysis/cyber-attacks-healthcare-europe

[2] D. McKee and P. Laulheret. (2021) McAfee Enterprise ATR uncovers vulnerabilities in globally used B. Braun infusion pump. [Online]. Available: https://www.mcafee.com/blogs/enterprise/mcafee-enterprise-atr/mcafee-enterprise-atr-uncovers-vulnerabilities-in-globally-used-b-braun-infusion-pump/

[3] Ponemon Institute, "Sixth annual benchmark study on privacy & security of healthcare data," Ponemon Institute, Tech. Rep., 2016.

[4] S. Islam, S. Papastergiou, and H. Mouratidis, "A dynamic cyber security situational awareness framework for healthcare ICT infrastructures," in *PCI 2021: 25th Pan-Hellenic Conference on Informatics*. Volos, Greece: ACM, 2021, pp. 334–339.

[5] M. Tikhomirov, N. V. Loukachevitch, A. Sirotina, and B. V. Dobrov, "Using BERT and augmentation in named entity recognition for cybersecurity domain," in *Natural Language Processing and Information Systems - 25th International Conference on Applications of Natural Language to Information Systems, NLDB 2020*, vol. 12089. Saarbrücken, Germany: Springer, 2020, pp. 16–24.

[6] P. Ma, B. Jiang, Z. Lu, N. Li, and Z. Jiang, "Cybersecurity named entity recognition using bidirectional long short-term memory with conditional random fields," *Tsinghua Science and Technology*, vol. 26, no. 3, pp. 259–265, 2021.

[7] S. Zhou, J. Liu, X. Zhong, and W. Zhao, "Named entity recognition using BERT with whole world masking in cybersecurity domain," in *2021 IEEE 6th International Conference on Big Data Analytics (ICBDA)*. Xiamen, China: IEEE, 2021, pp. 316–320.

[8] Y. Chen, J. Ding, D. Li, and Z. Chen, "Joint BERT model based cybersecurity named entity recognition," in *2021 The 4th International Conference on Software Engineering and Information Management*, ser. ICSIM 2021. Yokohama, Japan: Association for Computing Machinery, 2021, p. 236–242.

[9] C. Gao, X. Zhang, and H. Liu, "Data and knowledge-driven named entity recognition for cyber security," *Cybersecurity*, vol. 4, no. 1, pp. 1–13, 2021.

[10] O. Mendsaikhan, H. Hasegawa, Y. Yamaguchi, and H. Shimada, "Identification of cybersecurity specific content using the Doc2Vec language model," in *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, vol. 1. Milwaukee, WI, USA: IEEE, 2019, pp. 396–401.

[11] M. Ciampi, G. De Pietro, E. Masciari, and S. Silvestri, "Some lessons learned using health data literature for smart information retrieval," in *Proceedings of the 35th Annual ACM Symposium on Applied Computing*. Online: Association for Computing Machinery, 2020, p. 931–934. [Online]. Available: https://doi.org/10.1145/3341105.3374128

[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1*. Minneapolis, Minnesota: ACL, Jun. 2019, pp. 4171–4186.

[13] Y. Yang, W. Chen, Z. Li, Z. He, and M. Zhang, "Distantly supervised NER with partial annotation learning and reinforcement learning," in *Proceedings of the 27th International Conference on Computational Linguistics*. Santa Fe, New Mexico, USA: Association for Computational Linguistics, Aug. 2018, pp. 2159–2169.

[14] N. Shevchenko. (2018) Threat modeling: 12 available methods. [Online]. Available: https://insights.sei.cmu.edu/blog/threat-modeling-12-available-methods/

[15] Center for Internet Security (CIS). (2017) Cyber attacks: In the healthcare sector. [Online]. Available: https://www.cisecurity.org/insights/blog/cyber-attacks-in-the-healthcare-sector

[16] N. Goud. (2017) Malware and ransomware attack on medical devices. [Online]. Available: https://www.cybersecurity-insiders.com/malware-and-ransomware-attack-on-medical-devices/

[17] S. T. Argaw, J. R. Troncoso-Pastoriza, D. Lacey, M. Florin, F. Calcavecchia, D. Anderson, W. P. Burleson, J. Vogel, C. O'Leary, B. Eshaya-Chauvin, and A. Flahault, "Cybersecurity of hospitals: discussing the challenges and working towards mitigating the risks," *BMC Medical Informatics Decis. Mak.*, vol. 20, no. 1, p. 146, 2020.

[18] A. Yeboah-Ofori, H. Mouratidis, U. Ismai, S. Islam, and S. Papastergiou, "Cyber supply chain threat analysis and prediction using machine learning and ontology," in *Artificial Intelligence Applications and Innovations - 17th IFIP WG 12.5 International Conference, AIAI 2021*, vol. 627. Hersonissos, Crete, Greece: Springer, 2021, pp. 518–530.

[19] N. I. Haque, M. A. Rahman, M. H. Shahriar, A. A. Khalil, and A. S. Uluagac, "A novel framework for threat analysis of machine learning-based smart healthcare systems," *CoRR*, vol. abs/2103.03472, 2021.

[20] S. Silvestri, F. Gargiulo, and M. Ciampi, "Improving biomedical information extraction with word embeddings trained on closed-domain corpora," in *2019 IEEE Symposium on Computers and Communications (ISCC)*. IEEE, 2019, pp. 1129–1134.

[21] K. Ameri, M. Hempel, H. Sharif, J. Lopez Jr., and K. Perumalla, "CyBERT: Cybersecurity claim classification by fine-tuning the BERT language model," *Journal of Cybersecurity and Privacy*, vol. 1, no. 4, pp. 615–637, 2021.

[22] T. Satyapanich, F. Ferraro, and T. Finin, "CASIE: extracting cybersecurity event information from text," in *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020*. New York, NY, USA: AAAI Press, 2020, pp. 8749–8757.

[23] S. Zong, A. Ritter, G. Mueller, and E. Wright, "Analyzing the perceived severity of cybersecurity threats reported on social media," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 1380–1390.

[24] J. Fu, P. Liu, and Q. Zhang, "Rethinking generalization of neural models: A named entity recognition case study," in *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020*. New York, NY, USA: AAAI Press, 2020, pp. 7732–7739.

[25] P. Phandi, A. Silva, and W. Lu, "SemEval-2018 task 8: Semantic extraction from CybersecURity REports using natural language processing (SecureNLP)," in *Proceedings of The 12th International Workshop on Semantic Evaluation*. New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 697–706.

[26] F. Gargiulo, S. Silvestri, M. Ciampi, and G. De Pietro, "Deep neural network for hierarchical extreme multi-label text classification," *Applied Soft Computing*, vol. 79, pp. 125 – 138, 2019.

[27] A. Alicante, A. Corazza, F. Isgrò, and S. Silvestri, "Unsupervised entity and relation extraction from clinical records in Italian," *Computers in Biology and Medicine*, vol. 72, pp. 263–275, 2016.

[9]https://www.ai4healthsec.eu