NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY
POLITEHNICA BUCHAREST

Faculty of Automatic Control and Computers
Computer Science and Engineering Department

Ph.D. Summary Thesis

Vulnerability Prioritization based on Early Vulnerabilities Exposure, Emergent Vulnerability Trends and Contextual Attack Scenarios

Octavian Grigorescu

Thesis advisor:

Prof. dr. ing. Răzvan Rughiniș

BUCHAREST
2023

CONTENTS

ABSTRACT

The discipline of cybersecurity has become progressively critical in protecting societal processes within the interconnected global community of today. With the proliferation of internet users, devices, and services, there is an increasing imperative to effectively manage and protect against cyber threats. This thesis utilizes the capabilities of Artificial Intelligence (AI) to fundamentally transform cybersecurity methodologies, also touching the subject of the Internet of Things (IoT). Although AI does increase the likelihood of cybersecurity threats, it also presents unprecedented prospects for proactive defense strategies. By utilizing cutting-edge artificial intelligence applications, this study aims to gain an advantage over adversaries by promptly detecting vulnerabilities, evaluating risks across all relevant aspects, and devising sophisticated cybersecurity approaches.

Each of the three fundamental research questions addressed in the thesis motivates the investigation and development of its three main contributions. The initial inquiry examines the potential for improving the timely identification of cybersecurity vulnerabilities through the implementation of advanced data analysis and machine learning (ML) methods. The subsequent section examines efficacious approaches and instruments for evaluating and controlling cybersecurity vulnerabilities, emphasizing the distinction between diverse cyber threats. The third inquiry examines the process by which proactive cybersecurity defense strategies, such as risk assessment systems and continuous security management, are developed and integrated into the operations of an organization.

The preliminary contribution investigates empirical research concerning early exposure to vulnerability. The proposed approach encompasses several components: the development of a Security News Aggregator, the application of machine learning (ML) models to identify vulnerabilities in news websites and Twitter, the utilization of textual descriptions to forecast the severity of software vulnerabilities, and the integration of Natural Language Processing (NLP) techniques with pre-existing ontologies to extract exploits and attack vectors from cybersecurity news.

The second contribution focuses on empirical research in the field of risk assessment. It utilizes novel methodologies such as honeytokens that monitor attack execution, the implementation of web application honeypots, language model analysis of current cybersecurity trends, and an examination of emerging vulnerability trends reported in cybersecurity news.

The third contribution tackles proactive cyber defense through empirical research. The project entails several key components: the creation of the CODA footprint, which facilitates ongoing security management; evaluations of risk assessment frameworks and methods; exploration of security breaches in the Internet of Things; implementation of probability and attack graph models within a contextual risk scoring system; and utilization of a contextual priority scoring system to prioritize vulnerability patching.

By implementing sophisticated AI methods, this thesis provides substantial contributions across the domain of cybersecurity. The thesis conducts a critical analysis of the proposed

methodologies, emphasizing their advantages, difficulties faced during implementation, and possible usages. These studies not only introduce innovative approaches to cybersecurity but also establish a standard for subsequent investigations in this continuously developing field.

Keywords: Language Models, Natural Language Processing, Common Vulnerability Scoring System, MITRE Adversarial Tactics, Techniques, and Common Knowledge, Contextual Risk Scoring, Remediation Effort Prioritization

1 INTRODUCTION

1.1 Context

The rapid transitions occurring in the realm of cybersecurity are driven by the increasing complexity of digital environments and the unrelenting growth of cyber threats. The complexity of hostile actors has created significant problems for current cybersecurity vulnerability detection and management approaches. Conventional methods are frequently reactive and find it difficult to adequately address new cybernetic dangers.

Furthermore, a thorough grasp of the nature and traits of cyber threats is necessary due to their immense quantity and diversity. It is now crucially important to identify efficient processes and instruments for evaluating and managing cybersecurity risks, particularly when it comes to differentiating between various cyber threats. As businesses and governmental structures look to strengthen their defenses, they are depending more and more on cutting-edge technology like artificial intelligence (AI), natural language processing (NLP), and machine learning (ML) or deep learning (DL) models to understand and address cyber threats.

While already common knowledge in the cybersecurity community, the use of AI as a tool to deal with the most common cyber-attacks (ransomware attacks, IoT attacks, cloud attacks, and blockchain attacks) is also highlighted by Mijwil et al. (2023) [376]. Not only that the authors describe these attacks as being common, but also as evolving, in that attackers keep on finding new and more sophisticated approaches to finding and exploiting cyber vulnerabilities.

The focus on cutting-edge technology as a response to the complexity of the cyber landscape is also illustrated by recent activity in the cybersecurity academic community. For instance, Nobrega & Rutkowski (2022) [377] argue that Artificial Intelligence enables the identification of attack patterns and the automation of certain cybersecurity procedures. Zhang et al. (2023) [378] mention specific procedures that can be automated through AI. Baharadiya (2023) [379] discusses the application of ML in cybersecurity. Ahsan et al. (2023) [4] provide a comprehensive review of existing databases, ML techniques, and DL techniques used for cybersecurity.

Notably, there is a growing interest in the application of DL models for cybersecurity purposes. Dixit & Silakari (2021) [380] argue that deep learning techniques improve the performance of cybersecurity systems, while Sewak et. al (2022) [381] provide an extensive review of existing solutions using deep learning (DL) for cybersecurity purposes. Lin et. al (2020) [382] distinguish between various types of DL approaches.

In this context, the research outlined in this thesis holds relevant significance within the current

cybersecurity landscape. On the one hand, it addresses the need for turning academic research focused on cutting-edge techniques for cyber threat mitigation into viable proactive solutions that can be applied in real-world scenarios. On the other hand, the current research also expands theoretical knowledge on cybersecurity issues using empirical data. More precisely, the research questions and objectives of the current work aim to contribute with practical and theoretical advancements in early detection, risk assessment, and proactive cybersecurity defense strategies.

1.2 Problem statement

The problem addressed in the current research originates from two sources: the insufficient level of development of existing cybersecurity systems relying on conventional methods for cyber threat detection and mitigation and the gap between theoretical advancements in cybersecurity and their practical implementation within organizational contexts. Each of the two issues is detailed in the following lines.

Conventional cybersecurity systems face significant challenges in the rapidly changing landscape of cyber threats. By relying on traditional methods of cyber threat mitigation, these systems often struggle to adapt to the intricate and diverse nature of contemporary cyber threats. Consequently, organizations using such systems become more susceptible to exploitation.

Moreover, existing methods for assessing and managing cybersecurity risks may lack the necessary granularity to distinguish between different categories of cyber threats. The inadequacy in discerning between nuances of cyber-attacks hinders the ability to effectively prioritize and address the most critical threats. Without comprehensive risk assessment frameworks, organizations may allocate resources less efficiently, making them more vulnerable to sophisticated and targeted cyberattacks. This adds to the focus on reactive measures that are prevalent within the cybersecurity strategies of organizations, potentially hindering the development and implementation of proactive defense plans. Consequently, organizations may be exposed to new and unforeseen cyber threats.

On the other hand, there exists a gap in the integration of cutting-edge technologies into practical cybersecurity systems and solutions. Failure to effectively leverage technology such as NLP and AI, especially ML and DL, may impede the development of viable alternatives to conventional cybersecurity systems. Without practical and accessible solutions, organizations may struggle to apply state-of-the-art cybersecurity practices, leaving them no choice but to fall back on using conventional cybersecurity systems and methods, despite the drawbacks presented in the previous lines.

1.3 Objectives

Given the context and the problem statement presented in the previous sections, the following research questions were advanced:

- **Research Question 1** How can the early detection of cybersecurity vulnerabilities be enhanced using advanced data analysis and machine learning techniques?
- **Research Question 2** What are the effective methodologies and tools for assessing and managing cybersecurity risks, particularly in distinguishing between different types of cyber threats?
- **Research Question 3** In what ways can proactive cybersecurity defense strategies, including continuous security management and risk scoring systems, be developed, and integrated into organizational practices?

The research questions provided the basis for identifying seven specific research objectives of the current research:

1. **Development of Advanced Cybersecurity Detection and Management Systems:** To design and implement innovative systems and tools that enhance the detection, analysis, and management of cybersecurity threats. This includes the creation of platforms like Yggdrasil and CODA Footprint, which provide real-time insights and continuous security management capabilities. These platforms have been implemented and are now operational.
2. **Utilization of Cutting-Edge Technologies in Cybersecurity:** To leverage state-of-the-art technologies such as machine learning and deep learning models, natural language processing, and other artificial intelligence techniques for improving the efficiency and effectiveness of cybersecurity practices. This involves using advanced language models like BERT and RoBERTa for tasks like vulnerability detection, severity prediction, and trend analysis.
3. **In-depth Analysis of Cyber Threats and Vulnerability Trends:** To conduct comprehensive studies that analyze and categorize various types of cyber threats and vulnerabilities. This includes understanding the behavior and techniques of attackers, assessing emergent trends in cybersecurity, and identifying the most critical threats that need immediate attention.
4. **Enhancing Risk Assessment and Prioritization in Cybersecurity:** To develop and refine methods for evaluating and prioritizing cybersecurity risks. This objective encompasses the analysis of risk evaluation frameworks, the development of test-driven security approaches for IoT, and the implementation of probability and attack graph models for contextual risk scoring.
5. **Exploring Proactive Approaches in Cyber Defense:** To shift the focus from reactive to proactive strategies in cybersecurity. This involves the development of systems and

methodologies that allow for early detection and preemptive action against cyber threats, thereby reducing the potential impact of such threats.

6. **Integrating Practical Solutions into Cybersecurity Practices:** To provide actionable tools and frameworks that can be integrated into everyday cybersecurity operations. This includes offering solutions for continuous security management, test-driven security in IoT, and risk assessment, which can be directly applied in various organizational contexts.
7. **Contributing to the Theoretical and Practical Knowledge in Cybersecurity:** To enrich the academic and practical understanding of cybersecurity, providing insights and findings that can be leveraged by researchers, practitioners, and policymakers in the field.

1.4 Thesis structure

The seven chapters of the current work are organized into two main parts, each addressing specific aspects of the research: the state-of-the-art and empirical studies. As such, each chapter is presented briefly in the current section.

Chapter 1 represents the introduction of the thesis and comprises stating the research context along with the challenges that need to be addressed, followed by formulating three research questions and seven objectives that guided our work, and presenting the structure of the thesis.

Chapter 2 provides a comprehensive survey of the current state of the art in cybersecurity. It explores advanced techniques in early vulnerability detection and management (Section 2.1), evolving strategies in cybersecurity risk assessment and management (Section 2.2), and the shift towards a proactive paradigm in cybersecurity with a focus on managing and prioritizing vulnerabilities (Section 2.3).

Chapter 3 delves into empirical studies related to early vulnerability exposure and proposes original contributions to enhance technologies in this field. The chapter begins with an introduction followed by the development of a Security News Aggregator (Section 3.2). Next, ML models are employed for the early detection of vulnerabilities from news websites (Section 3.3), and Twitter (Section 3.4). The severity prediction of software vulnerabilities based on their text description is explored (Section 3.5), and exploits and attack vectors are extracted from cybersecurity news using NLP techniques (Section 3.6) and mapped to existing ontologies (Section 3.7).

Chapter 4 consists of empirical studies that advance the state of the art on trends regarding vulnerabilities and attack tactics. It involves the use of honeytokens to understand and influence the execution of an attack (Section 4.2), the publication of a web application honeypot in the wild (Section 4.3), a case study grounded in language models to analyze the latest cybersecurity trends (Section 4.4), and an analysis of emergent vulnerability trends in

cybersecurity news (Section 4.5).

Chapter 5 advances, through empirical studies, the state of the art related to proactive cyber defense. It includes the description of the CODA footprint continuous security management platform (Section 5.2), an analysis of risk evaluation frameworks and risk assessment methods (Section 5.3), an investigation into why IoT security is failing and the need for a test-driven security approach (Section 5.4), the application of probability and attack graph models in a contextual risk scoring system (Section 5.5), and the contextual prioritization of vulnerabilities and remediation efforts (Section 5.6).

Chapter 6 provides a critical analysis of the proposed approaches, highlighting their advantages (Section 6.1) and the problems faced during their development (Section 6.2), as well as envisioning potential applications (Section 6.3).

The final chapter (**Chapter 7**) summarizes the original contributions of the research (Section 7.1) and outlines potential directions for further research (Section 7.2).

2 EARLY VULNERABILITY EXPOSURE

The initial contribution of the empirical studies present in this thesis examines novel systems and approaches that have been developed to improve the detection and management of cyber threats in the domain of cyber security. This segment of the thesis provides an overview of six discrete yet interconnected research endeavors, all of which have made substantial contributions to the domain of cybersecurity.

The creation of the "Security News Aggregator" signifies an innovative endeavor in the management of the overwhelming volume of cybersecurity-related content. The primary objective of this platform is to effectively prioritize and filter critical security news, placing particular emphasis on emergent vulnerabilities and zero-day attacks. The high costs of cyberattacks serve as evidence of the critical importance of effectively overseeing the enormous and continuously expanding quantity of cybersecurity data in order to ensure timely and efficient distribution of urgent upgrades and critical updates.

Subsequently, the "Early Detection of Vulnerabilities from News Websites Using Machine Learning Models" research paper makes significant progress by analyzing news articles for indications of emergent cyber threats using sophisticated machine learning techniques. By capitalizing on the functionalities of Support Vector Machines, Multinomial Naïve Bayes classifiers, and a fine-tuned BERT model, this model demonstrates exceptional precision in vulnerability detection. This further emphasizes the effectiveness of Natural Language Processing (NLP) in the timely identification of cyber threats.

The article "Yggdrasil – A CSCL System for the Early Detection of Cybernetic Vulnerabilities" presents a novel automated system that employs tweets as a source of data in order to identify potential threats. By utilizing the BERT language model to examine tweets that contain links to cybersecurity articles, this methodology showcases the capacity of transfer learning to augment collaborative learning and timely threat detection within the realm of cybersecurity.

Furthermore, "Severity Prediction of Software Vulnerabilities based on their Text Description" implements a deep learning methodology to assess the severity of software vulnerabilities. The present investigation employs a Multi-Task Learning framework in conjunction with a pre-trained BERT model to forecast the severity levels of vulnerabilities exclusively on the basis of their textual descriptions. This novel methodology exemplifies the capacity of deep learning to deliver prompt and precise evaluations of the severity of vulnerabilities.

An efficient approach is proposed in "Extracting Exploits and Attack Vectors from Cybersecurity News Using NLP" for the automated labeling of articles pertaining to cybersecurity. By utilizing Named Entity Recognition, this methodology effectively retrieves and classifies

crucial data pertaining to emerging vulnerabilities and avenues of attack, thereby augmenting comprehension and adaptability in the face of cyber threats.

Finally, the paper "CVE2ATT&CK: BERT-Based Mapping of CVEs to MITRE ATT&CK Techniques" tackles the crucial issue of associating particular attack techniques with Common Vulnerabilities and Exposures (CVEs). The objective of this study is to autonomously establish these critical connections by annotating CVEs using techniques derived from the MITRE ATT&CK framework and developing models, including language models based on BERT. The accomplishment of this undertaking makes a substantial contribution to the cybersecurity community's capacity to comprehend and efficiently mitigate cyber threats.

Collectively, these six research contributions constitute an all-encompassing and multifaceted investigation into state-of-the-art methodologies and techniques within the realm of cybersecurity. The significance of early detection, advanced data analysis, and the incorporation of machine learning and natural language processing (NLP) is emphasized in the continuous struggle against cyber threats.

2.1 Takeaway

Within the realm of cybersecurity, the initial contribution of the empirical studies segment unveils novel strategies and systems that have been developed to detect and manage cyber threats at an early stage. This represents the culmination of research efforts in the field. This segment of the thesis integrates the findings and suggestions for further research from six separate research studies, all of which have made contributions to the comprehension and progression of cybersecurity practices.

The "Security News Aggregator" research has effectively suggested the creation of an all-encompassing platform that consolidates a wide array of cyber-security news, including corrections and vulnerabilities, security breaches and CVEs. By procuring its information from a diverse range of reputable sources, this platform serves as a beacon for current events, potentially assisting in the detection of zero-day threats and emerging susceptibilities. Subsequent efforts shall be directed towards augmenting the functionalities of the platform by means of automating the identification of pertinent Twitter users, fortifying the article evaluation system, and investigating dark web exploit kits. The necessity for dynamic and responsive cybersecurity information systems is highlighted by this ongoing effort.

In the section entitled "Early Detection of Vulnerabilities from News Websites Using Machine Learning Models", a prototype system designed to identify emergent cyber threats in news articles is presented. Notwithstanding the encouraging outcomes attained by the BERT model, prospective improvements encompass broadening the dataset and conducting trials with alternative language models. Furthermore, further development will involve the integration of this prototype into a more comprehensive system that employs diverse Open Source Intelligence (OSINT) sources to automatically identify early cyber threats. The objective of this suggested

expansion is twofold: to enhance the existing model and to establish a more all-encompassing system that can detect threats at an early stage.

"Yggdrasil - A CSCL System for the Early Detection of Cybernetic Vulnerabilities" is an article that concentrates on task, integration, and focus criteria to develop an advanced typology for CSCL systems in cybersecurity. Both experts and non-experts are assisted in navigating community-generated knowledge on emergent cyber vulnerabilities by the Yggdrasil system. Following an experiment involving transfer learning from a previously developed model [130], this study validates the predictive capabilities of the BERT model and data fusion methods with respect to pertinent information. Subsequent pursuits encompass augmenting the dataset and delving deeper into the feasibility of transfer learning as a methodology for devising CSCL systems in the domain of cybersecurity. This study delineates a path towards the development of automated tools that augment collaborative learning and the exchange of knowledge within the cybersecurity community.

The subsequent segment of the initial contribution within the empirical studies section of the thesis further investigates novel approaches and systems in the field of cybersecurity. It specifically emphasizes the utilization of natural language processing (NLP) to extract cyberattack information, link CVEs to MITRE ATT&CK techniques, and predict vulnerability severity.

The deep learning model proposed in the study "Severity Prediction of Software Vulnerabilities based on their Text Description" streamlines the vulnerability assessment procedure. By employing a multi-task architecture, this model enables system administrators to efficiently assess the severity of threats posed by recently disclosed vulnerabilities. Subsequent enhancements will comprise refining the hyperparameters of the model and integrating it into an independent application that will deliver instantaneous notifications concerning high-risk and critical threats. Furthermore, an exploration of more expansive architectures, such as RoBERTa [181], could potentially lead to a decrease in inaccuracies in the predicted metrics. This research represents a significant advancement in devising accessible and streamlined methods for non-experts to assess cybersecurity threats.

The study titled "Extracting Exploits and Attack Vectors from Cybersecurity News using NLP" presented a novel approach for the automated classification of articles pertaining to intrusions and vulnerabilities. Serra et al. [179] employed sophisticated natural language processing (NLP) methods, such as bidirectional word-level LSTM and custom models incorporating Bloom embeddings and residual CNNs [180]. Further research is warranted to incorporate dark web exploit kits and other cybersecurity-related content into the data sources. Additionally, a user-friendly interface will be designed to facilitate the filtration of news that is pertinent to individual users' requirements. The objective of this approach is to improve the usability and relevance of cybersecurity information for end-users by optimizing its accessibility and applicability.

Finally, the research project titled "CVE2ATT&CK: BERT-Based Mapping of CVEs to MITRE

ATT&CK Techniques" resolved the crucial issue of how to link CVEs to attack techniques automatically. By employing a multi-label task methodology and pioneering the use of BERT-based architectures, this research achieved notable advancements in the precise labeling of critical vulnerability exploits (CVEs). Further research will be devoted to improving the annotated CVE corpus. This will involve investigating approaches such as Few-Shot Learning [220] and Semi-Supervised Learning [221]. These methods will be utilized to address the drawbacks of inadequate training data and label imbalance. Future plans also include the incorporation of supplementary information sources [222], with the objective of rectifying the incongruities present in CVE descriptions.

The extensive research delineated in these studies highlights the ever-changing and swiftly progressing characteristics of the cybersecurity field. At the core of these developments lies the incorporation of cutting-edge methodologies and technologies, including collaborative learning systems, natural language processing, and deep learning. By implementing these comprehensive strategies, organizations can effectively monitor and address the constantly changing cyber threats, thereby optimizing threat detection and management in a way that prioritizes user needs.

The forthcoming research outlined in these studies not only aim to improve and strengthen the functionalities of the current systems but also intends to generate significant advancements in the wider domain of cybersecurity. Continuous innovation plays a pivotal role in the development of user-friendly and intuitive systems that surpass precision and responsiveness, while also catering to the needs of a wide range of users. The specific focus on machine learning models indicates a strategic transition towards implementing more proactive and well-informed approaches to defending against cyber threats.

Fundamentally, the accumulated knowledge from these research endeavors demonstrates an ever-changing cybersecurity environment, characterized by a dedication to ingenuity and adjustment. By incorporating cutting-edge technologies and methodologies, these studies establish the foundation for a cybersecurity infrastructure that is more resilient and adaptable, enabling it to efficiently confront the intricacies and difficulties posed by contemporary cyber threats.

3 VULNERABILITY AND ATTACK TACTICS TRENDS

Within the dynamic realm of cybersecurity, safeguarding web applications represents an imperative frontier. These technological applications, which facilitate user-database interaction via the internet, offer substantial advantages but also present substantial security vulnerabilities. Critical vulnerabilities may arise due to improper programming or misconfigurations, thereby enabling unauthorized individuals to compromise entire systems and gain access to sensitive data. The stakes are extremely high, as malicious actors are constantly seeking to monetize compromised data in addition to causing reputational harm. The second contribution I have made explores this field by means of four separate studies, each providing distinct perspectives on cybersecurity developments, vulnerabilities, and novel defensive strategies.

This chapter proceeds with "HUNT: Using Honeytokens to Understand and Influence the Execution of an Attack". The present study presents an advanced intrusion detection system that makes use of honeypots and honeytokens. It has been purposefully engineered to discern extensive, aimless assaults from more concentrated, specific threats. The system employs a strategic approach to traps that bear resemblance to alluring resources in order to classify assaults, decipher the motivations of attackers, gather forensic evidence, and ultimately eliminate threats. By utilizing interconnected honeytokens, which each pose unique exploitation challenges, a comprehensive scenario can be created in which the abilities and motivations of attackers can be evaluated.

Following this, "Web Application Honeypot Published in the Wild" is devoted to learning penetration techniques and detecting cyberattacks. The research entails the implementation of honeypots within a cybernetic infrastructure. To entice attackers, these honeypots are outfitted with "Capture the Flag"-style challenges. The knowledge acquired over a two months deployment on the internet consists of an examination of human and automated interactions with these honeypots. This analysis offers significant insights into the tactics and behaviors of threats.

"What are the Latest Cybersecurity Trends? A Case Study Grounded in Language Models" is a research based on Natural Language Processing that utilizes the RoBERTa language model. This study examines 2264 news articles pertaining to security. Text embeddings, dimensionality reduction, and topic clustering are employed in order to classify these articles into pertinent categories. By employing this methodology, a thorough assessment of the development and importance of diverse cybersecurity trends is possible, providing insight into the present condition of cyber threats and defenses.

"Analysis of Emergent Vulnerability Trends in Cybersecurity News" concludes our discussion. This study facilitates the prioritization of software patching through the examination of vul-

nerability trends. It introduces a substantial dataset of manually annotated cybersecurity news articles by utilizing Transformer architectures. The articles are classified as pertinent or inconsequential by the refined models, which also utilize clustering techniques to identify recurring patterns, particularly with regard to vendor exposure. The aforementioned system functions as an indispensable asset for cybersecurity analysts, augmenting their routine investigative and research proficiencies.

Collectively, these studies offer a comprehensive perspective on the cybersecurity environment, tackling the intricate difficulties associated with protecting web applications and the information they process. Every research study provides distinct perspectives and approaches, which enhance our comprehension and capacity to combat cyber threats.

3.1 Takeaway

The subsequent contribution within the empirical studies section of the thesis provides an in-depth analysis of the strategic implementation and inventive utilization of honeypots within the realm of cybersecurity. This extensive research endeavor encompasses an examination of honeypot solutions as well as an investigation of the most recent cybersecurity trends via sophisticated language models.

We have examined the entire continuum of honeypot solutions in "HUNT: Using Honeytokens to Understand and Influence the Execution of an Attack", from academic research initiatives to fully-fledged commercial deployments. Our inquiry uncovered the advantages and disadvantages of high-interaction and low-interaction honeypots. It was observed that high-interaction honeypots (HIH) exhibited notable efficacy in preventing advanced persistent threats and targeted attacks. As a consequence of these discoveries, we devised the HUNT framework, which is a decentralized honeynet composed of web application devices and effectively managed through a centralized console. The traps, which operate as microservices in the cloud, have been carefully designed to symbolize distinct vulnerabilities. Each trap has a different level of difficulty, accommodating attackers with a wide spectrum of abilities. A hacker who successfully compromises one HUNTER Task is directed to an additional, more arduous task, thereby ensnarement in this intricate web and notification of response teams. This configuration guarantees that the production environment that it emulates is not affected in any way. Every individual trap independently transmits attack data to the central console, facilitating the compilation and examination of attacker profiles, identities, as well as tactics, techniques, and procedures (TTPs). Subsequently, for evaluating its efficacy, it is imperative to implement the HUNT framework in real-world production settings, simulating an authentic application. Consistent with sophisticated security systems, ongoing adaptation is critical for maintaining a competitive edge against contemporary threats. To maintain its efficacy against a diverse range of cyber threats, it is imperative to extend the protocol coverage of HUNT beyond HTTP.

Subsequently, we investigated the incorporation of honeypot systems into web applications as part of the exhaustive "Web Application Honeypot Published in the Wild" study, with the objectives of detecting and classifying attackers, redirecting them to simulated environments, and analyzing their actions. By virtue of its pioneering methodology, this endeavor has exhibited efficacy in the detection of attack patterns and tools employed against web applications. The surge in assaults and exploitation of vulnerabilities in widely used web applications—many of which still have unresolved security issues—emphasizes the growing necessity for such solutions. The implementation of honeypots poses distinctive difficulties due to the limitations and the necessity of concealing architectural details, which serves to thwart adversaries' attempts to bypass these systems. Developed as a supplementary security measure to conventional systems such as intrusion detection and firewalls, our honeypot has demonstrated its efficacy by furnishing insightful information regarding the tools and activities of malicious actors. By integrating secure versions of applications devoid of known vulnerabilities into the honeypot configuration, the probability of discovering zero-day exploits, which are not yet publicly known but are being actively employed in attacks, is increased. The configuration of our honeypot, which is inspired by a Capture the Flag game featuring multiple tiers of difficulty, is intentionally crafted to evaluate the competence of attackers. This guarantees that the honeypot is both easily observable and sufficiently difficult to distinguish automated tools from human intervention in the event of an attack. The statistics and information extracted from these captured attacks have proven to be of immense value. Predictively, we intend to improve the honeypot in order to enable a greater number of human interactions, thus accumulating a more comprehensive dataset for analysis. By basing the honeypot on a widely used web application that is consistently targeted by malicious actors, we guarantee a substantial level of user involvement. Furthermore, it is our intention to incorporate a number of enhancements into the system. To begin, we shall present a variety of exploration paths that leverage information on vulnerabilities derived from the Top 10 Web Application Security Risks. Our objective is to reveal unprecedented hacking techniques. Secondly, our objective is to establish pathways that provide enhanced interaction opportunities, such as shell access, for potential attackers. By implementing this configuration, we will have the capacity to utilize specialized surveillance tools in order to closely observe their activities, thereby gaining more profound understanding of their methodologies. These improvements are vital in advancing our comprehension of the actions of attackers and will provide valuable insights for the creation of subsequent generations of sophisticated honeypots.

The third study, "What are the Latest Cybersecurity Trends? A Case Study Grounded in Language Models", presents a novel automated pipeline that utilizes the most recent advancements in natural language processing (NLP) technologies to process cybersecurity news articles. By employing RoBERTa, a Transformer-based language model, this pipeline generates embeddings of the articles that are contextualized. Before they are clustered, these embeddings are reduced in dimension. The most effective clustering method was determined to be the combination of Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) and UMAP (Uniform Manifold Approximation and Projection). This methodol-

ogy not only facilitates the efficient clustering of articles but also permits the extraction and comprehensive analysis of the most significant keywords from each cluster. The present stage of the undertaking lays the foundation for a number of ambitious improvements in the future. The development of a system to generate notifications regarding potential cybersecurity threats comes first. The topics extracted by the model will inform the development of these notifications, which will be customized specifically for software applications running on user devices. Additionally, there are intentions to develop a specialized interface. By delivering daily updates on pertinent subjects and showcasing the most recent developments in cybersecurity, this interface will serve as a very useful tool for users seeking to remain well-informed. Finally, there is a goal to augment the corpus through the integration of supplementary sources. By adopting this approach, the clustering process can be considerably improved; by incorporating a more diverse array of topics into the analysis, the quantity of outliers can be diminished, resulting in more precise and all-encompassing observations of the cybersecurity environment.

The study titled "Analysis of Emergent Vulnerability Trends in Cybersecurity News" ultimately presents a groundbreaking system that automates the surveillance of cybersecurity news in search of emergent threats. This innovation signifies a substantial progression within the domain of cybersecurity analysis. By utilizing advanced Transformer-based models as its foundation, this system effectively reduces the laborious process of manually reviewing cybersecurity news. The main purpose of this system is to improve the prioritization of software vulnerability patching through the identification of emerging and current threats. In order to accomplish this, an extensive dataset was compiled and applied to refine these Transformer-based models through the exploration of diverse configurations and architectures. The system's integrated infrastructure demonstrates proficiency in retrieving articles, eliminating extraneous content, extracting relevant data, and subsequently constructing clusters centered on detected software vulnerabilities. Following this, security analysts are presented with these clusters, which substantially streamlines the procedure of cybersecurity news analysis. The accuracy score of the classification model that has been implemented within the system is quite remarkable at 0.91. Notwithstanding these developments, certain aspects have been identified that require further enhancement. An example of such an area is the enhancement of the article clustering procedure, which is at present rated poorly by Silhouette. This suggests that the clusters lack sufficient separation and have a tendency to converge, thereby reducing the clustering algorithm's effectiveness. Furthermore, it is imperative to improve the process of extracting comprehensive information from articles that are considered pertinent by the classification model. A way for surmounting this obstacle is through the construction of a sequence labeling model. The aforementioned model possesses the ability to extricate precise and comprehensive data pertaining to cybersecurity vulnerabilities, including but not limited to CVEs (Common Vulnerabilities and Exposures), attack vectors, and various categories of cyber vulnerabilities. By incorporating this model into the system, the profundity and accuracy of information extraction could be substantially enhanced, thereby bolstering the system's capacity to detect and assess cybersecurity threats.

By virtue of their compilation, these studies not only substantiate the efficacy of novel method-

ologies in cybersecurity but also establish the foundation for subsequent progressions. By incorporating honeypots, natural language processing (NLP) methods, and automated pipelines, an organization can adopt a strategic stance in response to the constantly changing cybersecurity challenges.

4 PROACTIVE CYBER DEFENSE - VULNERABILITIES MANAGEMENT AND REMEDIATION EFFORT PRIORITIZATION

This component of the PhD study has as its central objective the development of a vulnerability and remediation quantification system that can be used in prioritizing cybersecurity team tasks. Furthermore, the development has had a profound impact on issues pertaining to security and privacy [297]. In order to reduce costs and accommodate physical operating conditions, the IoT is subject to inherent technological and market constraints, such as limited storage and processing capabilities in devices like sensors, despite its rapid expansion. Due to the reliance on batteries, these constraints, in conjunction with the critical need for real-time processing and energy efficiency, contribute to the insufficiency of security measures in numerous IoT systems. This has resulted in a decline in consumer and business confidence, which is exacerbated by the industry's inability to convincingly resolve these security concerns. Empirical investigations into the security challenges of the Internet of Things underscore the critical nature of developing novel solutions in this rapidly progressing domain.

"CODA Footprint Continuous Security Management Platform" is the title of the initial article. This platform functions as an all-encompassing resolution for the immediate examination and evaluation of the critical services of an organization. In pursuit of guaranteeing the uninterrupted functioning and protection of critical services, it tackles the intricacy introduced by the proliferation of cloud services, the integration of various devices, and bring-your-own-device (BYOD) policies. The platform represents a substantial advancement in the management of the complex cybersecurity requirements of contemporary organizations, particularly with regard to the Internet of Things.

"Analyzing Risk Evaluation Frameworks and Risk Assessment Methods" is the subject of the following article. In light of the escalating prevalence of cyber-attacks and vulnerabilities in the digital realm, this article emphasizes the escalating significance of cyber insurance. This paper addresses the difficulties encountered by security engineers when attempting to quantify the monetary consequences of intrusions and evaluate the level of risk exposure that an organization is exposed to. The research emphasizes the necessity for enhanced risk assessment frameworks in various sectors such as the Internet of Things (IoT), urging for improvements in both academic and corporate protocols.

"Why IoT Security Is Failing – The Need for a Test-Driven Security Approach" is the title of our third article, which provides a more comprehensive analysis of cybersecurity. This study provides a in-depth investigation of the security susceptibilities inherent in the Internet of Things (IoT) ecosystem. A test-driven security framework is suggested as a means to oversee and conduct security testing on Internet of Things (IoT) applications at every stage of their

development. This methodology is regarded as an essential instrument in addressing nascent cyber risks that are specific to the Internet of Things, underscoring the need for ongoing and stringent security evaluations in IoT settings.

Furthermore, the article "Probability and Attack Graph Models in Contextual Risk Scoring System" presents an all-encompassing software solution that employs attack graph models and probability-based methodologies to manage and quantify risk in computer networks. By generating a network score, which serves as a quantitative indicator of network risk exposure, it demonstrates a dependable and effective approach to evaluating the security of computer networks. This method is essential in today's interconnected digital world, as it provides a flexible approach to cybersecurity in a variety of environments. Its importance is emphasized by the increasing intricacy of networks and the worldwide obstacle of safeguarding digital assets against advanced threats.

Finally, the paper "Contextual Remediations Prioritization System Designed to Implement Theoretical Principles of CVSS v4" proposes a novel way to prioritize cyber vulnerabilities in an era of rapid technological advancement and complex device interconnectivity. The rapid expansion of the digital ecosystem often leads to poor or insufficient settings, creating cyber infrastructure vulnerabilities. The essential issue of vulnerabilities prioritizing in Vulnerability Risk Management (VRM) remains a concern. This study provides a novel solution that aligns with CVSS v4's new features, the present method utilizing dynamic scoring and contextual vulnerability information to prioritize remediations more effectively. In an increasingly susceptible digital economy, this is essential for cyber infrastructure security and resilience.

In brief, the aforementioned four articles collectively tackle the complex issues surrounding security, in vulnerability prioritization as well as in IoT ecosystems, introducing novel approaches and frameworks that bolster the resilience and security of IoT environments. By implementing test-driven security approaches, contextual risk scoring, and continuous security management and risk evaluation, these contributions provide vital insights into the ever-expanding IoT landscape's security.

4.1 Takeaway

The third contribution within the empirical studies section of the thesis provides a thorough examination of novel methodologies employed in the field of cybersecurity. It specifically emphasizes frameworks for risk assessment, contextual risk scoring systems, IoT security, and ongoing security management.

The research paper titled "CODA Footprint Continuous Security Management Platform" presents an innovative centralized platform designed to continuously monitor devices connected to a network. This development signifies a substantial advancement in the field of cybersecurity for organizations. The fundamental aim of this endeavor is to design and implement a distributed software framework that effectively gathers all pertinent configuration

information from a wide range of network-connected devices, including but not limited to virtual machines (VMs), servers, routers, switches, and firewalls. Following transmission to a central intelligence hub, this information is processed and incorporated into a virtual footprint to facilitate continuous monitoring. The overarching goal of this research is to substantially improve the functionalities of the CODA Footprint security platform in the near future. The proposed improvements entail the consolidation of three significant modules. The establishment of agent-based local auditing systems constitutes the initial step. These systems will conduct an extensive investigation of the internal network, generating an automated topology that comprehensively depicts the network environment. The objective of the second module is to enhance the network analysis functionalities of the platform by providing a more detailed and all-encompassing perspective of the network's configuration and functioning. Finally, the third module, which presents the most innovative features, is self-reactive, performing intelligent decisions. By employing a combination of real-time and historical data, this module will possess the ability to independently respond to events and arrive at well-informed decisions. It is anticipated that the incorporation of these modules will significantly enhance the platform's efficacy in safeguarding organizational infrastructures from threats, thereby offering a more resilient and adaptable methodology to cybersecurity administration.

The research article titled "Analyzing Risk Evaluation Frameworks and Risk Assessment Methods" provides a comprehensive analysis of several cybersecurity frameworks, including STIX, XCCDF, OVAL, SCAP, and MAEC. The research highlights the importance of standardization in these frameworks, casting doubt on the purpose of multiple formats if a universally understandable language for security specialists and machines is the ultimate objective. Additionally, the research investigates methods of forecasting which vulnerabilities will be exploited as opposed to those that will remain unexploited. It suggests that a more comprehensive comprehension of the progressive inclinations and methodologies of hackers could be obtained through the examination of historical data, specifically by investigating the provenance of significant assaults that were hitherto undisclosed. To identify patterns, this methodology would entail conducting an exhaustive examination of Twitter feeds, historical news, and timelines depicting renowned assaults. The study also emphasizes the significance of understanding the manner in which hazards spread throughout a network. In light of the multitude of devices that utilize the identical infrastructure and, consequently, the identical vulnerabilities, the research inquiry concerns the most appropriate formula for converting risk into a numeric value. Understanding the correlation between vulnerabilities and the financial repercussions for businesses is an additional financial aspect of this issue, given that businesses are frequently reluctant to disclose information regarding intrusions. Historical data on breaches and losses, particularly those involving renowned brands and significant assaults, may offer valuable insights, according to the study, which capitalizes on the media's propensity to reveal information, sometimes against the will of the entities involved. An aspect that is frequently disregarded, according to the research, is the human element in cybersecurity. Subsequent research is recommended to monitor a variety of news channels, Twitter feeds, and most significantly, the dark web. Emerging vulnerabilities and zero-day exploits, including those that are traded on the illicit

market, may be discovered through this monitoring. The objective of this research is to devise countermeasures and strategies to potential cyber activities through the acquisition of knowledge regarding the resources and technologies they employ. The importance of a multifaceted comprehension of cybersecurity, which takes into account the technical and human factors that influence the dynamics of cyber threats, is emphasized by this exhaustive approach.

Following this, the article titled "Why IoT Security is Failing: The Need for a Test Driven Security Approach" undertakes a critical analysis of the present condition of security in the realm of Internet of Things (IoT) and puts forth an innovative resolution to bolster safeguards in this swiftly progressing sector. The primary objective of the Dynamic IoT-System Security Testing (DISST) model is to mitigate the manifold risks and susceptibilities that are intrinsic to a wide range of IoT environments. The importance of a continuous monitoring system that can evaluate the entire infrastructure of IoT applications on a periodic basis throughout their development lifecycle is emphasized in this model. The study emphasizes the unavoidability of security concerns and vulnerabilities in the Internet of Things (IoT) as a result of the vast quantity of interconnected devices, the intricacy of systems, and the variety of devices, applications, services, and protocols. It highlights the challenge associated with incident detection and proposes approaches such as monitoring network communications, analyzing activity records, conducting penetration tests, and engaging in ethical hacking to uncover weaknesses and effectively address incidents. The research emphasizes the critical obligation of IoT providers to create more secure devices and uphold their security. It argues that "Security by Design" is a crucial concept for mitigating multiple security challenges in the Internet of Things. The DISST model is proposed as an auxiliary instrument to ongoing research and development endeavors, facilitating the identification of security vulnerabilities and weaknesses throughout the entire development cycle, without impeding the implementation or technological strategy. Furthermore, it is possible to incorporate this model with frameworks such as Anastasia in order to train the capabilities of Intrusion Detection Systems (IDS) and Extended Security Information and Event Management (XL-SIEM) in identifying and addressing particular attack patterns, or to optimize policies. In anticipation of future developments, the research highlights the necessity of augmenting the DISST model to encompass a wide range of protocols spanning multiple network stack tiers, as well as to incorporate diverse industries and numerous categories of IoT devices. Furthermore, considerable emphasis is placed on allocating resources towards machine learning initiatives in order to develop a more sophisticated system that can generate security test cases autonomously. The incorporation and growth of sophisticated technologies are essential for the development of IoT security, guaranteeing that it remains abreast of the escalating intricacy and magnitude of IoT infrastructures as well as the ever-changing realm of cybersecurity risks.

Furthermore, the research article titled "Probability and Attack Graph Models in Contextual Risk Scoring System" introduces a novel approach that combines modeling, prioritization, data collection, selection, and aggregation in order to calculate risk scores for network security. This endeavor implements an all-encompassing strategy for assessing risk by combining probability-based and attack graph models. The term "risk score", as it is defined in this

research, can be interpreted in various ways, including the probability of system attacks, the severity of particular vulnerabilities, and the percentage of noncompliance. These calculations are carried out utilizing a combination of experimental and mathematical models that are substantiated by established theoretical foundations. Notwithstanding the successful attainment of its principal aims, the study recognizes the obstacles encountered in ascertaining the effectiveness and precision of the suggested resolution. The aforementioned challenge emerges as a result of the cybersecurity domain's swift evolution and the predominantly empirical construction of assessment models. Despite the standardization efforts of the Security Content Automation Protocol (SCAP) and other community initiatives, the accurate assessment of system risks continues to be a source of ambiguity. Nevertheless, the research furnishes dependable data and outcomes, presenting a versatile framework capable of accommodating present and forthcoming network security demands by integrating cutting-edge concepts and perpetually refining APIs. The initiative distinguishes itself from current solutions through its all-encompassing methodology for aggregating and evaluating data in order to conduct risk analysis. In addition to utilizing a probabilistic model to evaluate diverse categories of data in context, it emphasizes data standardization for uniform risk evaluation metrics. In addition, the study underscores the significance of process prioritization and business type consideration when developing a risk calculation framework that can be modified to accommodate distinct organizational activities and security objectives. This research emphasizes the intricacy of network evaluation, recognizing that the utilization of experimental data prevents the inclusion of all facets by a single method. Nevertheless, it posits that an approach that integrates concrete elements, mathematical concepts, statistical methodologies, and community security benchmarks may produce approximative outcomes. The aforementioned findings play a crucial role in averting system failures, financial losses, and data corruption, thereby augmenting the overall efficacy of network security approaches. Further investigation is warranted to enhance the risk calculation model in order to more accurately represent the ever-changing landscape of network security. This enhancement will integrate statistical methodologies and conform to community security benchmarks in order to formulate prevention strategies that are more accurate and proactive.

Finally, the main goal of the paper entitled "Contextual Remediations Prioritization System Designed to Implement Theoretical Principles of CVSS v4" is to find a solution for prioritizing vulnerabilities' remediation designed for implementation of certain missing fundamentals of CVSS V4. The goal was achieved by implementing a solution grounded in combining multiple types of knowledge in modeling a risk calculation system. To achieve this, we drilled down into CVSS solutions and leveraged data collected from an intelligent agent system and from the Yggdrasil Threat Intelligence Service. Based on these data, we succeeded in creating three indicators used in the vulnerability's score computation: static, dynamic, and contextual. As Jung et al. [366] concluded, there are also some limitations because it is hard to evaluate a new proposal as we need a structure and certain standards to compare different approaches and extract future improvements. What sets this project apart from other established solutions is the incorporation of numerous data types that were purposefully selected for risk assessment,

in addition to evaluating them using a probabilistic model within the given context. An additional benefit of our solution pertains to the standardization of data, which enables the integration of said data through the utilization of consistent risk evaluation parameters. In this paper, we have presented a fundamental risk calculation model. Nevertheless, the ranking of security aspects may be modified in conformity with the specific activities of the organization, according to a structure with distinct levels of importance. This project highlights the fact that evaluating a network is a complex activity that cannot be fully captured by a single method, as it is dependent on a multitude of results obtained from empirical data. However, by utilizing a methodology that combines concrete elements and mathematical principles with statistical approaches and industry security norms, it is possible to obtain approximative outcomes that assist in mitigating the risks associated with data corruption, monetary depletion, or system shutdowns.

As a whole, these studies emphasize the significance of ongoing innovation in the field of cybersecurity, highlighting the necessity for comprehensive and adaptable strategies to tackle the ever-changing and intricate characteristics of cyber threats. By incorporating sophisticated monitoring systems, frameworks for risk assessment, and test-driven security methodologies into these initiatives, a more robust and secure digital infrastructure is established.

5 CONCLUSIONS

5.1 Personal Contributions

In the constantly changing and dynamic field of cybersecurity, the identification of vulnerabilities in a timely and efficient manner is critical. The initial contribution of this study, entitled "Early Vulnerability Exposure", explores novel approaches and systems that seek to transform the process of detecting and controlling cybersecurity vulnerabilities right from their incipient stages. This section of the research comprises a collection of sophisticated studies that, by combining the most recent developments in machine learning and natural language processing (NLP), explore uncharted territories in vulnerability detection. Every individual study included in this contribution not only tackles the issue of effectively managing the ever-growing quantity of cybersecurity threats, but also presents distinct viewpoints and methodologies to improve the effectiveness and precision of vulnerability assessment. The aforementioned introduction establishes the foundation for an exhaustive examination of these studies, emphasizing their collective and individual importance within the wider framework of cybersecurity.

- The "Security News Aggregator" introduces a substantial advancement towards the establishment of a framework that can effectively sift and rank critical security news. This application is designed to handle the administration of a substantial amount of daily cybersecurity data, with an emphasis on zero-day threats, newly discovered vulnerabilities, and critical patches.
- "Early Detection of Vulnerabilities from News Websites Using Machine Learning Models" establishes a model to identify emergent cybernetic vulnerabilities in news articles, thereby extending this trajectory. The accuracy of this model, which employs Support Vector Machines, Multinomial Naïve Bayes classifiers, and a fine-tuned BERT model, is exceptionally high. This result validates the effectiveness of natural language processing (NLP) in the early detection of vulnerabilities.
- "Yggdrasil – A CSCL System for the Early Detection of Cybernetic Vulnerabilities" presents a collaborative learning in a cybersecurity-enhanced automated system for identifying threats from tweets (messages posted on the Twitter platform). By utilizing the BERT language model to analyze tweets that are connected to cybersecurity articles, this system demonstrates remarkable accuracy, thereby highlighting the potential of transfer learning in this field.
- The approach referred to as "Severity Prediction of Software Vulnerabilities based on their Text Description" redirects attention towards the anticipation of vulnerability severity. By employing a deep learning methodology and a Multi-Task Learning framework incorporating a pre-trained BERT model, this research successfully forecasts severity

scores based solely on textual descriptions of vulnerabilities with remarkable accuracy.

- A novel methodology is proposed in the article "Extracting Exploits and Attack Vectors from Cybersecurity News using NLP" to label articles pertaining to vulnerabilities and cyberattacks automatically. By employing Named Entity Recognition, this research efficiently extracts and classifies critical data pertaining to newly discovered vulnerabilities, thereby augmenting comprehension of attack vectors and exploits.
- The paper "CVE2ATT&CK: BERT-Based Mapping of CVEs to MITRE ATT&CK Techniques" tackles the issue of correlating CVEs with attack methodologies. By utilizing MITRE ATT&CK techniques to annotate critical vulnerabilities (CVEs) and developing models, including language models based on BERT, this effort has successfully automated the establishment of these vital connections.

In the diverse realm of cybersecurity, the comprehension and evaluation of risks are very important in order to protect vital IT infrastructures. The subsequent contribution, entitled "Vulnerability and Attack Tactics Trends", investigates novel methodologies for identifying and controlling cybersecurity threats. This segment encompasses a collection of research articles that thoroughly examine the intricacies of cyber-attacks, the application of honeypots and honeytokens, and the evaluation of cybersecurity trends. The primary objective of every study is to augment comprehension regarding attack methodologies and to devise sophisticated systems that can identify and alleviate cyber threats. Through the implementation of advanced technologies and analytical approaches, these research endeavors make substantial contributions to the domain of cybersecurity by presenting novel perspectives on the actions of malicious actors and the progression of cyber hazards. This introductory section provides a comprehensive outline of the aforementioned studies, establishing the framework for an in-depth examination of their approaches, results, and ramifications within the wider domain of cybersecurity risk assessment.

- The research paper titled "HUNT: Using Honeytokens to Understand and Influence the Execution of an Attack" presents a novel intrusion detection system that makes use of honeypots and honeytokens. This system demonstrates proficiency in differentiating between broad-scale, aimless assaults, and more specific, targeted perils. The research endeavours to classify assaults, comprehend the motivations of attackers, collect forensic evidence, and ultimately eradicate the menace by constructing enticement-laden traps. By deploying interconnected honeytokens, each possessing a unique level of exploitation difficulty, an exhaustive scenario can be created to assess the capabilities of an attacker.
- The objective of "Web Application Honeypot Published in the Wild" is to develop an intelligent system capable of identifying cyber-attacks and gaining knowledge of penetration techniques. As part of this cybernetic infrastructure-integrated system, honeypots resembling "Capture the Flag" challenges are constructed. Insights into the behavior of attackers are provided by the study's findings, which are based on observations of both automated and human interactions with the honeypot during its two-month deployment on the internet.

- "What Are the Latest Cybersecurity Trends? A Case Study Grounded in Language Models" identifies major trends in security news through the use of Natural Language Processing, more specifically the RoBERTa language model. The research groups articles into clusters by analyzing 2264 articles using text embeddings, dimensionality reduction, and topic clustering. This approach facilitates an assessment of the progression and significance of diverse cybersecurity trends.
- The original contribution of "Analysis of Emergent Vulnerability Trends in Cybersecurity News" is to provide assistance in the prioritization of software upgrading through the examination of vulnerability trends. This investigation presents an extensive collection of cybersecurity news articles that have been painstakingly annotated, followed by employing Transformer architectures in their processing. Utilizing clustering techniques, the refined models categorize articles as pertinent or inconsequential and uncover patterns pertaining to software vendor exposure. The integration of this system facilitates the daily research and investigation efforts of cybersecurity analysts.

In an age characterized by the proliferation and escalating sophistication of cybersecurity threats, proactive defense strategies are critical. "Proactive Cyber Defense - Vulnerabilities Management and Remediation Effort Prioritization", the third contribution of this research, centers on the advancement and deployment of sophisticated systems and methodologies designed to efficiently oversee and alleviate cybersecurity vulnerabilities. This segment comprises research that examines the complexities of contemporary data infrastructures, the obstacles presented by the Internet of Things (IoT), and the nuances of risk evaluation in the field of cybersecurity. These studies seek to improve the capacity of cybersecurity professionals to proactively detect and resolve vulnerabilities through the development of novel frameworks and models. As a result, the security posture of organizations is strengthened. The aforementioned introduction establishes the groundwork for a comprehensive examination of these studies, emphasizing their contributions to a more proactive and knowledgeable strategy concerning cybersecurity.

- The "CODA Footprint Continuous Security Management Platform" presents an all-encompassing resolution for the real-time auditing and analysis of a company's critical services. This platform is intended to guarantee that the essential services of the organization are sufficiently safeguarded and that the information security defenses remain operational continuously. It tackles the difficulties presented by the proliferation of cloud services, the incorporation of diverse device types, and policies such as "bring your own device" (BYOD), all of which have substantially amplified the intricacy of cybersecurity administration.
- The article "Analyzing Risk Evaluation Frameworks and Risk Assessment Methods" discusses the escalating significance of cyber insurance as the number of cyber-attacks, vulnerabilities, and data exposures continues to rise. The research emphasizes the obstacles that security engineers encounter when attempting to quantify the financial impact of cyberattacks and the complexities they encounter when evaluating the risk exposure

of their organizations. This study highlights the necessity for enhanced methodologies to be implemented in the business and academic spheres.

- "Why IoT Security Is Failing. The Need for a Test Driven Security Approach" explores the security risks and complexities that are inherent in the Internet of Things (IoT) in its swift evolution. The study presents a framework for monitoring and security testing Internet of Things (IoT) applications. It supports the use of a test-driven methodology to assess security throughout the entire development lifecycle. Using this strategy, we hope to combat the novel forms of cyberattacks that are emerging in the IoT ecosystem.
- The software solution proposed in the "Probability and Attack Graph Models in Contextual Risk Scoring System" study is an all-encompassing approach to risk management and quantification in computer networks. This solution comprises novel approaches for collecting, processing, and evaluating data from network devices utilizing attack graph models and probability-based models. A network score is generated as a quantitative assessor of the network's risk exposure; it provides a reliable and efficient method for evaluating the security of computer networks.
- The research presented in "Contextual Remediations Prioritization System Designed to Implement Theoretical Principles of CVSS v4" introduces a novel approach for prioritizing vulnerabilities in cyber infrastructures, developed to address the challenges of the new CVSS v4 version. This method leverages dynamic scoring and contextual information, aiming for a more effective prioritization strategy. It responds to the inadequacies of current Vulnerability Risk Management solutions, promising enhanced security by accurately identifying and addressing vulnerabilities.

The first contribution, titled "Early Vulnerability Exposure", encompasses a comprehensive and multifaceted methodology towards cybersecurity, with a particular focus on the strategic management and timely identification of vulnerabilities. By means of a sequence of pioneering investigations, this study not only augments the existing comprehension of vulnerability detection and management but also establishes a foundation for forthcoming progressions in the domain. The incorporation of sophisticated machine learning methodologies, particularly NLP and BERT models, represents a substantial progression in the effective management of cybersecurity risks. As a group, these studies provide cybersecurity practitioners with invaluable insights and tools that facilitate a more informed and proactive approach to the ever more complex and frequent cybersecurity challenges. The results and approaches outlined in this contribution are positioned to generate a consistent improvement leading to a more secure digital space.

The second contribution, titled "Vulnerability and Attack Tactics Trends", provides a holistic framework for comprehending and controlling cybersecurity risks. By conducting an extensive body of research, this study contributes to the advancement of knowledge regarding the methodologies employed in cyber-attacks and presents novel approaches to identify and mitigate risks. The utilization of cutting-edge technologies, including honeytokens, honeypots, and sophisticated language models, proves the capability of these instruments to differentiate among a multitude of cyber threats. The insights derived from these studies not only

augment the body of theoretical knowledge in the domain but also furnish cybersecurity professionals with practical tools and systems. The aforementioned discoveries play a crucial role in the formulation of risk assessment methodologies, facilitating a more knowledgeable and efficient reaction to cybersecurity obstacles. In general, the contribution makes a substantial stride forward in the comprehension of cybersecurity risks and the creation of instruments that efficiently mitigate these risks.

The third contribution, titled "Proactive Cyber Defense - Vulnerabilities Management and Remediation Effort Prioritization", signifies a substantial advancement in the domain of cybersecurity. This statement embodies a proactive methodology for overseeing and reducing cybersecurity risks, focusing specifically on emerging obstacles like IoT security and the intricacies of contemporary IT infrastructures. The research articles contained in this contribution present novel frameworks and solutions that aid cybersecurity practitioners in the identification, evaluation, and prioritization of vulnerabilities. The incorporation of these methodologies represents a significant progression in cybersecurity protocols, redirecting attention from reactive to proactive approaches. This contribution not only enhances the existing theoretical comprehension of vulnerability management but also provides organizations with practical tools and methodologies to enhance their overall cybersecurity posture. The findings and frameworks established through this investigation play a crucial role in directing the trajectory of cybersecurity in the future, underscoring the need for a comprehensive and progressive strategy in the management of digital threats.

5.2 Directions for Future Research

As of the present moment, there is an indisputable fascination surrounding language models in the domain of cybersecurity, which signifies an emerging area of research ripe for exploration. As we reflect on the course of forthcoming investigations, our vision comprises a sequence of ambitious yet achievable goals that seek to enhance our comprehension and capacities in this pivotal field.

Regarding our initial contribution, we suggest a novel methodology that entails performing fresh experiments utilizing the most recent language models to surface in the field. This entails training cutting-edge models for predicting severity, with a particular emphasis on the Common Vulnerabilities and Exposures that have been publicly disclosed since 2022. Also, the importance of staying informed of cybersecurity developments cannot be overstated, especially those that emerge from the shadowy domains of the Dark Web. In pursuit of our objective, we intend to conduct a series of experiments utilizing the LLM and T5 architectures on a recently assembled CVE-Mitre ATT&CK mappings dataset that contains more than 10,000 entries. The enhancement of our present dataset will require the careful inclusion of relevant details for every entry, including comprehensive articles that explain the properties of each Common Vulnerability Exposure and the Common Weakness Enumeration. This will result in a substantial improvement in the training effectiveness of the model.

The subsequent contribution emphasizes the pragmatic implementation of our research in tangible, real-life contexts. This involves the implementation of a honeypot, which is carefully crafted to simulate an authentic application, thus providing significant visibility into possible cybersecurity risks. We propose that in our pursuit of thorough coverage of cybersecurity advancements, we should incorporate information from a variety of news sources. This will enhance our knowledge base and cultivate a more sophisticated comprehension of emergent threats. Key to our proposed approach is the utilization of named entity recognition methods to extract pertinent entities from the data, including the versions of the software that have been compromised, the vulnerability severity, and the secure versions. The knowledge acquired from this procedure will be essential in improving the correspondence between secure and vulnerable software versions within the CODA Footprint agent system, thereby bolstering the security posture as a whole.

Our third contribution aims to fundamentally transform how cybersecurity hazards are evaluated and ranked. Through the utilization of the Contextual Risk Scoring System, we intend to implement a probability-driven methodology that will enhance our ability to rank attack scenarios in order of importance. Further enhancing this undertaking will be the integration of scan outcomes generated by recently developed agents tailored for the Linux and MacOS operating systems. The remediation endeavor component will require the Remediations Workflow System to undergo an ongoing cycle of testing and analysis. Additionally, by capitalizing on the extensive collection of more than three million identified applications, we are dedicated to build an NLP model that can forecast the Common Platform Enumeration (CPE) by analyzing the varied application formats. The prioritization system will be enhanced by incorporating data from the CVE - Mitre ATT&CK mappings into this initiative, thus revealing potentially hazardous attack scenarios that require immediate attention. Finally, a comprehensive examination of kill chain assaults, integrating insights from Adversarial Tactics, Techniques, and Common Knowledge as well as the Contextual Risk Scoring System, will establish a fundamental basis for a cybersecurity framework that is more robust and adaptable.

Employing these concerted efforts, our dual objective is to broaden the boundaries of cybersecurity investigation and establish a more robust digital landscape.

LIST OF PUBLICATIONS

Journals

- **O. Grigorescu**, A. Nica, M. Dascalu, R. Rughinis. CVE2ATT&CK: BERT-Based Mapping of CVEs to MITRE ATT&CK Techniques. Algorithms. 2022; 15(9):314.
  https://doi.org/10.3390/a15090314
  WOS:000858120500001; IF(2022)=2.3
- **O. Grigorescu**, L. Botezatu, A. Mutu, D. Turcanu. Contextual Remediations Prioritization System Designed to Implement Theoretical Principles of CVSS v4
  (Accepted for publication)
- I Branescu, **O. Grigorescu**, M. Dascalu. Mapping Common Vulnerabilities and Exposures to MITRE ATT&CK Tactics. Advances in Cybersecurity and Reliability. 2024
  (Accepted for publication)
- C. Săndescu, A. Dinisor, C-V. Vladescu, **O. Grigorescu**, D. Corlatescu , M. Dascalu, R. Rughinis, Extracting Exploits and Attack Vectors from Cybersecurity News using NLP, Buletin UPB 2022
  WOS:000805648400006
- **O. Grigorescu**, V. Vitan, D. Iorga, M. Dascalu, and R. Rughinis. Analysis of Emergent Vulnerability Trends in Cybersecurity News
  (In the process of publication and indexing)

Conferences

- **O. Grigorescu**, C. Săndescu and R. Rughiniș, "CODA footprint continuous security management platform", 2016 15th RoEduNet Conference: Networking in Education and Research, Bucharest, 2016, pp. 1-5, doi: 10.1109/RoEduNet.2016.7753223.
  WOS:000390713800024
- Sandescu, C., Rughinis, R., **Grigorescu, O.** (2017). HUNT: Using Honeytokens To Understand and Influence The Execution Of An Attack. In Proc. eLSe 2017 – The International Scientific Conference eLearning and Software for Education, Vol. 1, p. 511, Bucharest, "Carol I" National Defence University
- C. Săndescu, **O. Grigorescu**, R. Rughiniș, R. Deaconescu and M. Calin, "Why IoT security is failing. The Need of a Test Driven Security Approach", 2018 17th RoEduNet Conference: Networking in Education and Research (RoEduNet), Cluj-Napoca, 2018, pp. 1-6, doi: 10.1109/ROEDUNET.2018.8514135.
  WOS:000517570500013

- R. E. Radu, **O. Grigorescu** and R. V. Rughiniș, "Security News Aggregator", 2019 18th RoEduNet Conference: Networking in Education and Research (RoEduNet), Galati, Romania, 2019, pp. 1-8, doi: 10.1109/ROEDUNET.2019.8909609.
  WOS:000520513500024
- D. Iorga, D.Corlătescu, **O. Grigorescu**, C. Săndescu, M. Dascălu, R. Rughiniș "Early Detection of Vulnerabilities from News Websites using Machine Learning Models", 2020 19th RoEduNet Conference: Networking in Education and Research (RoEduNet), Romania, 2020
  WOS:000654265900006
- R. Radu, C. Săndescu, **O. Grigorescu**, R. Rughiniș "Analyzing Risk Evaluation Frameworks and Risk Assessment Methods", 2020 19th RoEduNet Conference: Networking in Education and Research (RoEduNet), Romania, 2020
  WOS:000654265900028
- **O. Grigorescu**, C. Săndescu, A. Caba "Web Application Honeypot Published in the Wild", 2020 19th RoEduNet Conference: Networking in Education and Research (RoEduNet), Romania, 2020
  WOS:000654265900020
- D. Iorga, D.Corlătescu, **O. Grigorescu**, C. Săndescu, M. Dascălu, R. Rughiniș "Yggdrasil—early detection of cybernetic vulnerabilities from Twitter", 2021 23rd International Conference on Control Systems and Computer Science (CSCS23) Romania, 2021
- I. Babalau, D. Corlatescu, **O. Grigorescu**, C. Sandescu and M. Dascalu, "Severity Prediction of Software Vulnerabilities based on their Text Description", 2021 23rd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 2021, pp. 171-177, doi: 10.1109/SYNASC54541.2021.00037.
  WOS:000786477000026
- C. Vladescu, M. -A. Dinisor, **O. Grigorescu**, D. Corlatescu, C. Sandescu and M. Dascalu, "What are the Latest Cybersecurity Trends? A Case Study Grounded in Language Models", 2021 23rd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 2021, pp. 140-146, doi: 10.1109/SYNASC54541.2021.00033.
  WOS:000786477000022
- **O. Grigorescu**, A. Minea, T. Dumitru and R. Rughiniș, "Probability and Attack Graph models in Contextual Risk Scoring System", 2022 21st RoEduNet Conference: Networking in Education and Research (RoEduNet), 2022, pp. 1-9, doi: 10.1109/RoEduNet57163.2022.9921100.

REFERENCES

[1] S. N. Narayanan, A. Ganesan, K. Joshi, T. Oates, A. Joshi, and T. Finin, "Early detection of cybersecurity threats using collaborative cognition," in *2018 IEEE 4th international conference on collaboration and internet computing (CIC)*. IEEE, 2018, pp. 354–363.

[2] S. Narayanan, Ashwinkumar Ganesan, K. Joshi, T. Oates, A. Joshi, and Timothy W. Finin, "Cognitive Techniques for Early Detection of Cybersecurity Events," *arXiv.org*, 2018.

[3] Eugene Fink, Mehrbod Sharifi, and J. Carbonell, "Application of Machine Learning and Crowdsourcing to Detection of Cybersecurity Threats," 2011.

[4] M. Ahsan, K. E. Nygard, R. Gomes, M. M. Chowdhury, N. Rifat, and J. F. Connolly, "Cybersecurity Threats and Their Mitigation Approaches Using Machine Learning – A Review," *Journal of Cybersecurity and Privacy*, vol. 2, no. 3, pp. 527–555, jul 10 2022.

[5] A. Handa, A. Sharma, and S. K. Shukla, "Machine learning in cybersecurity: A review," *WIREs Data Mining and Knowledge Discovery*, vol. 9, no. 4, feb 17 2019.

[6] S. A. Salloum, M. Alshurideh, A. Elnagar, and K. Shaalan, *Machine Learning and Deep Learning Techniques for Cybersecurity: A Review*. Springer International Publishing, 2020, pp. 50–57.

[7] A. Salih, S. T. Zeebaree, S. Ameen, A. Alkhyyat, and H. M. Shukur, "A Survey on the Role of Artificial Intelligence, Machine Learning and Deep Learning for Cybersecurity Attack Detection," in *2021 7th International Engineering Conference "Research & Innovation amid Global Pandemic" (IEC)*. IEEE, feb 24 2021.

[8] S. B. Son, S. Park, H. Lee, Y. Kim, D. Kim, and J. Kim, "Introduction to MITRE ATT&CK: Concepts and use cases," in *2023 International Conference on Information Networking (ICOIN)*, 2023, pp. 158–161.

[9] M. Ahmed, S. Panda, C. Xenakis, and E. Panaousis, "MITRE ATT&CK-driven cyber risk assessment," in *Proceedings of the 17th International Conference on Availability, Reliability and Security*, ser. ARES '22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: https://doi.org/10.1145/3538969.3544420

[10] T. Wang, S. Qin, and K. P. Chow, "Towards vulnerability types classification using pure self-attention: A common weakness enumeration based approach," in *2021 IEEE 24th*

International Conference on Computational Science and Engineering (CSE), 2021, pp. 146–153.

[11] F. Alenezi and C. P. Tsokos, "Machine learning approach to predict computer operating systems vulnerabilities," in *2020 3rd International Conference on Computer Applications & Information Security (ICCAIS)*, 2020, pp. 1–6.

[12] F. N. Alenezi and T. Mehmood, "Data-driven predictive model of windows 10's vulnerabilities," in *2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 2022, pp. 1–5.

[13] C. Elbaz, L. Rilling, and C. Morin, "Fighting n-day vulnerabilities with automated cvss vector prediction at disclosure," in *Proceedings of the 15th International Conference on Availability, Reliability and Security*, ser. ARES '20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3407023.3407038

[14] ——, "Towards automated risk analysis of "one-day" vulnerabilities," in *RESSI 2019-Rendez-vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information*, 2019, pp. 1–3.

[15] A. K. S. Ruchi Sharma and H. Pham, "Software security evaluation using multilevel vulnerability discovery modeling," *Quality Engineering*, vol. 35, no. 2, pp. 341–352, 2023. [Online]. Available: https://doi.org/10.1080/08982112.2022.2132404

[16] P. Kuehn, D. N. Relke, and C. Reuter, "Common vulnerability scoring system prediction based on open source intelligence information sources," 2022.

[17] A. Bonandir and S. Yussof, "An analysis of common vulnerability and exposure (cve) of software products in the year 2016," *International Journal of Advanced Science and Technology*, vol. 112, pp. 157–166, 2018.

[18] M. Vanamala, X. Yuan, and K. Roy, "Topic modeling and classification of common vulnerabilities and exposures database," in *2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, 2020, pp. 1–5.

[19] C.-H. Han and C. Han, "Semi-quantitative cybersecurity risk assessment by blockade and defense level analysis," *Process Safety and Environmental Protection*, vol. 155, pp. 306–316, 11 2021.

[20] Z. Amin, "A practical road map for assessing cyber risk," *Journal of Risk Research*, vol. 22, no. 1, pp. 32–43, 2019. [Online]. Available: https://doi.org/10.1080/13669877.2017.1351467

[21] R. Leszczyna, "Review of cybersecurity assessment methods: Applicability perspective," *Computers & Security*, vol. 108, p. 102376, 9 2021.

[22] J. Crotty and E. Daniel, "Cyber threat: its origins and consequence and the use of qualitative and quantitative methods in cyber risk assessment," *Applied Computing and Informatics*, dec 26 2022.

[23] A. A. Ganin, P. Quach, M. Panwar, Z. A. Collier, J. M. Keisler, D. Marchese, and I. Linkov, "Multicriteria Decision Framework for Cybersecurity Risk Assessment and Management," *Risk Analysis*, vol. 40, no. 1, pp. 183–199, sep 5 2017.

[24] I. D. Sánchez-García, J. Mejía, and T. San Feliu Gilabert, "Cybersecurity Risk Assessment: A Systematic Mapping Review, Proposal, and Validation," *Applied Sciences*, vol. 13, no. 1, p. 395, dec 28 2022.

[25] S. F. Ahmed and N. A. Hikal, "A Review of Cyber-security Measuring and Assessment Methods for Modern Enterprises," *JOIV : International Journal on Informatics Visualization*, vol. 3, no. 3, aug 10 2019.

[26] O. Giuca, T. M. Popescu, A. M. Popescu, G. Prostean, and D. E. Popescu, *A Survey of Cybersecurity Risk Management Frameworks*. Springer International Publishing, aug 15 2020, pp. 240–272.

[27] M. Campos, E. Gomes, and R. Machado, "Sensors for detection of cyber threats on industrial environment using a high interaction ics/scada honeynet1," in *2022 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0&IoT)*, 2022, pp. 1–5.

[28] K. Chawda and A. D. Patel, "Dynamic & hybrid honeypot model for scalable network monitoring," in *International Conference on Information Communication and Embedded Systems (ICICES2014)*, 2014, pp. 1–5.

[29] Y. Xu, Y. Jiang, L. Yu, and J. Li, "Brief industry paper: Catching iot malware in the wild using honeyiot," in *2021 IEEE 27th Real-Time and Embedded Technology and Applications Symposium (RTAS)*, 2021, pp. 433–436.

[30] J. Danani and J. Jani, "Honeypot-a tool to trap website hackers," *Proceedings Published in International Journal of Computer Applications®(IJCA)*, pp. 8–13, 2012.

[31] A. Iskhakova, R. Meshcheryakov, A. Iskhakov, and S. Timchenko, "Analysis of the vulnerabilities of the embedded information systems of iot-devices through the honeypot network implementation," in *Proceedings of the IV International research conference "Information technologies in Science, Management, Social sphere and Medicine" (ITSMSSM 2017)*. Atlantis Press, 2017/12, pp. 363–367. [Online]. Available: https://doi.org/10.2991/itsmssm-17.2017.75

[32] M. Keramati, "Dynamic risk assessment system for the vulnerability scoring," *International Journal of Information and Communication Technology Research*, vol. 9, no. 4, pp. 57–68, 2017.

[33] S. Neuhaus and T. Zimmermann, "Security trend analysis with cve topic models," in *2010 IEEE 21st International Symposium on Software Reliability Engineering*, 2010, pp. 111–120.

[34] G. Spanos, A. Sioziou, and L. Angelis, "Wivss: A new methodology for scoring information systems vulnerabilities," in *Proceedings of the 17th Panhellenic Conference on Informatics*, ser. PCI '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 83–90. [Online]. Available: https://doi.org/10.1145/2491845.2491871

[35] U. Bartwal, S. Mukhopadhyay, R. Negi, and S. Shukla, "Security orchestration, automation, and response engine for deployment of behavioural honeypots," in *2022 IEEE Conference on Dependable and Secure Computing (DSC)*. IEEE, 2022, pp. 1–8.

[36] F. Mayorga, J. Vargas, E. Álvarez, and H. D. Martinez, "Honeypot network configuration through cyberattack patterns," in *2019 International Conference on Information Systems and Computer Science (INCISCOS)*, 2019, pp. 150–155.

[37] R. Colbaugh and K. Glass, "Proactive defense for evolving cyber threats," in *Proceedings of 2011 IEEE International Conference on Intelligence and Security Informatics*. IEEE, 7 2011.

[38] A. Marotta and M. McShane, "Integrating a Proactive Technique Into a Holistic Cyber Risk Management Approach," *Risk Management and Insurance Review*, vol. 21, no. 3, pp. 435–452, 12 2018.

[39] P. Katsumata, J. Hemenway, and W. Gavins, "Cybersecurity risk management," in *2010 - MILCOM 2010 MILITARY COMMUNICATIONS CONFERENCE*. IEEE, 10 2010.

[40] K.-J. Huang and K.-H. Chiang, "Toward a Self-Adaptive Cyberdefense Framework in Organization," *SAGE Open*, vol. 11, no. 1, p. 215824402098885, 1 2021.

[41] A. S. Makaryan and M. M. Putyato, "Conceptual Approach to the Implementation of the Proactive Defense Subsystem of the Operational Cybersecurity Center," in *2021 XXIV International Conference on Soft Computing and Measurements (SCM)*. IEEE, may 26 2021.

[42] K. G. Crowther, Y. Y. Haimes, and M. E. Johnson, "Principles for Better Information Security through More Accurate, Transparent Risk Scoring," *Journal of Homeland Security and Emergency Management*, vol. 7, no. 1, jan 11 2010.

[43] Humza Naseer, Atif Ahmad, S. Maynard, and G. Shanks, "Cybersecurity Risk Management Using Analytics: A Dynamic Capabilities Approach," *International Conference on Interaction Sciences*, 2018.

[44] Y. Badr, F. Biennier, and S. Tata, "The Integration of Corporate Security Strategies in Collaborative Business Processes," *IEEE Transactions on Services Computing*, vol. 4, no. 3, pp. 243–254, 7 2011.

[45] S. Cho, I. Han, H. Jeong, J. Kim, S. Koo, H. Oh, and M. Park, "Cyber kill chain based threat taxonomy and its application on cyber common operational picture," in *2018 International Conference On Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA)*. IEEE, 2018, pp. 1–8.

[46] L. Sadlek, P. Čeleda, and D. Tovarňák, "Identification of attack paths using kill chain and attack graphs," in *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium*. IEEE, 2022, pp. 1–6.

[47] S. Yang, Y. Shi, and F. Guo, "Risk assessment of industrial internet system by using game-attack graphs," in *2019 IEEE 5th International Conference on Computer and Communications (ICCC)*. IEEE, 2019, pp. 1660–1663.

[48] T. W. Purboyo and Kuspriyanto, "A review of network security metrics," 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:212463847

[49] A. Kundu, N. Ghosh, I. Chokshi, and S. K. Ghosh, "Analysis of attack graph-based metrics for quantification of network security," in *2012 Annual IEEE India Conference (INDICON)*. IEEE, 2012, pp. 530–535.

[50] H. Wang, Z. Chen, J. Zhao, X. Di, and D. Liu, "A vulnerability assessment method in industrial internet of things based on attack graph and maximum flow," *Ieee Access*, vol. 6, pp. 8599–8609, 2018.

[51] I. Shrestha and M. Hale, "Detecting dynamic security threats in multi-component iot systems," 2019.

[52] S.-S. Yoon, D.-Y. Kim, K.-K. Kim, and I.-C. Euom, "Vulnerability exploitation risk assessment based on offensive security approach," *Applied Sciences*, vol. 13, no. 22, p. 12180, 2023.

[53] S. Mishra, A. Albarakati, and S. K. Sharma, "Cyber threat intelligence for iot using machine learning," *Processes*, vol. 10, no. 12, p. 2673, 2022.

[54] X. Duan, M. Ge, T. H. M. Le, F. Ullah, S. Gao, X. Lu, and M. A. Babar, "Automated security assessment for the internet of things," in *2021 IEEE 26th Pacific Rim International Symposium on Dependable Computing (PRDC)*. IEEE, 2021, pp. 47–56.

[55] G. George and S. M. Thampi, "A graph-based decision support model for vulnerability analysis in iot networks," in *Security in Computing and Communications: 6th International Symposium, SSCC 2018, Bangalore, India, September 19–22, 2018, Revised Selected Papers 6*. Springer, 2019, pp. 1–23.

[56] P. Griffioen and B. Sinopoli, "Assessing risks and modeling threats in the internet of things," *arXiv preprint arXiv:2110.07771*, 2021.

[57] V. G. Massaro, L. Capacci, and R. Montanari, "Towards context-aware risk assessment scoring system for iot/iiot devices," 2023.

[58] R. Kasprzyk and A. Stachurski, "A concept of standard-based vulnerability management automation for it systems," 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:114936229

[59] D. Waltermire, S. D. Quinn, H. Booth, K. Scarfone, and D. Prisaca, "The technical specification for the security content automation protocol (scap): Scap version 1.3," 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:69690628

[60] U. Tariq, A. O. Aseeri, M. S. Alkatheiri, and Y. Zhuang, "Context-aware autonomous security assertion for industrial iot," *IEEE Access*, vol. 8, pp. 191 785–191 794, 2020.

[61] P. Anand, Y. Singh, A. K. Selwal, P. K. Singh, and K. Z. Ghafoor, "Ivqfiot: An intelligent vulnerability quantification framework for scoring internet of things vulnerabilities," *Expert Systems*, vol. 39, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:239202722

[62] V. Malik and S. Singh, "Security risk management in iot environment," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 22, pp. 697 – 709, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:203320904

[63] S. Rizvi, N. McIntyre, and J. Ryoo, "Computing security scores for iot device vulnerabilities," *2019 International Conference on Software Security and Assurance (ICSSA)*, pp. 52–59, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:231851175

[64] N. A. Khan, A. Awang, and S. A. A. Karim, "Security in internet of things: A review," *IEEE access*, vol. 10, pp. 104 649–104 670, 2022.

[65] J. Zheng, X.-s. Zhang, and X.-h. Pan, "A host deployed vulnerability assessment system based on oval," in *2010 International Conference on Computer Application and System Modeling (ICCASM 2010)*, vol. 2. IEEE, 2010, pp. V2–123.

[66] G. Lee, I.-s. Ko, and T.-h. Kim, "A vulnerability assessment tool based on oval in system block model," in *International Conference on Intelligent Computing*. Springer, 2006, pp. 1115–1120.

[67] K. Papachristou, T.-I. Theodorou, S. Papadopoulos, A. Protogerou, A. Drosou, and D. Tzovaras, "Runtime and routing security policy verification for enhanced quality of service of iot networks," *2019 Global IoT Summit (GIoTS)*, pp. 1–6, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:198145865

[68] M. Dayalan, "Cyber risks, the growing threat," *ResearchGate*, 2017.

[69] Z. M. Smith and E. Lostri, *The hidden costs of cybercrime*. McAfee, 2020.

[70] M. Fichtenkamm, G. F. Burch, and J. Burch, "Cybersecurity in a covid-19 world: Insights on how decisions are made," *ISACA Journal*, vol. 2, no. 1, pp. 1–11, 2022. [Online]. Available: https://www.isaca.org/resources/isaca-journal/issues/2022/volume-2/cybersecurity-in-a-covid-19-world

[71] K. Bissell, R. Lasalle, and P. Dal Cin, "2019 cost of cybercrime study— 9th annual— accenture," *Ninth Annual Cost of Cybercrime Study*, 2019. [Online]. Available: https://www.accenture.com/us-en/insights/security/cost-cybercrimestudy.

[72] L. Ponemon, "Cost of data breach study," *Ponemon Institute*, 2017.

[73] L. Columbus, "Roundup of cybersecurity forecasts and market estimates," Forbes, 2020. [Online]. Available: https://www.forbes.com/sites/louiscolumbus/2020/04/05/2020-roundup-ofcybersecurity-forecasts-and-market-estimates

[74] A. Talalaev, "Website hacking statistics you should know in 2021. patchstack. retrieved february 2, 2022," 2021. [Online]. Available: https://patchstack.com/website-hackingstatistics/

[75] "Faqs - cve," The MITRE Corporation. [Online]. Available: https://cve.mitre.org/about/faqs.html

[76] F. Cremer, B. Sheehan, M. Fortmann, A. N. Kia, M. Mullins, F. Murphy, and S. Materne, "Cyber risk and cybersecurity: a systematic review of data availability," *The Geneva Papers on risk and insurance-Issues and practice*, vol. 47, no. 3, pp. 698–736, 2022.

[77] "Common vulnerability scoring system sig," Forum of Incident Response and Security Teams. [Online]. Available: https://www.first.org/cvss

[78] "Common vulnerabilities and exposures," MITRE, 2023. [Online]. Available: https://cve.mitre.org/

[79] S. FIRST, "Common vulnerability scoring system sig," 2018. [Online]. Available: https://www.first.org/cvss

[80] Y. Li and Q. Liu, "A comprehensive review study of cyber-attacks and cyber security; emerging trends and recent developments," *Energy Reports*, vol. 7, pp. 8176–8186, 2021.

[81] F. Ö. Sönmez, "Classifying common vulnerabilities and exposures database using text mining and graph theoretical analysis," *Machine Intelligence and Big Data Analytics for Cybersecurity Applications*, pp. 313–338, 2021.

[82] 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.

[83] R. Martin, S. Christey, and D. Baker, "A progress report on the cve initiative," in *Proceedings of the 14th Annual Computer Security Incident Handling Conference (FIRST)*, 2002.

[84] B. E. Strom, A. Applebaum, D. P. Miller, K. C. Nickels, A. G. Pennington, and C. B. Thomas, "Mitre att&ck: Design and philosophy," in *Technical report*. The MITRE Corporation, 2018.

[85] "Live cyber attack threat map," ThreatCloud Intelligence – Threatclqud, 2019. [Online]. Available: https://threatmap.checkpoint.com/ThreatPortal/livemap.html

[86] "Cyber threat intelligence," SecurityWizardry, 2019. [Online]. Available: https://www.securitywizardry.com/radar.htm

[87] B. Cui, S. Moskal, H. Du, and S. J. Yang, "Who shall we follow in twitter for cyber vulnerability?" in *Social Computing, Behavioral-Cultural Modeling and Prediction: 6th International Conference, SBP 2013, Washington, DC, USA, April 2-5, 2013. Proceedings 6*. Springer, 2013, pp. 394–402.

[88] S. Trabelsi, H. Plate, A. Abida, M. Aoun, A. Zouaoui, C. Missaoui, S. Gharbi, and A. Ayari, "Monitoring software vulnerabilities through social networs analysis," in *12th International Conference on Security and Cryptography, SECRYPT*, 2015.

[89] S. Mittal, P. K. Das, V. Mulwad, A. Joshi, and T. Finin, "Cybertwitter: Using twitter to generate alerts for cybersecurity threats and vulnerabilities," in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2016, pp. 860–867.

[90] C. Sabottke, O. Suciu, and T. Dumitraș, "Vulnerability disclosure in the age of social media: Exploiting twitter for predicting {Real-World} exploits," in *24th USENIX Security Symposium (USENIX Security 15)*, 2015, pp. 1041–1056.

[91] T. Dumitras, "How to predict which vulnerabilities will be exploited," 2019.

[92] N. Tavabi, P. Goyal, M. Almukaynizi, P. Shakarian, and K. Lerman, "Darkembed: Exploit prediction with neural language models," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[93] O. C. Moholth, R. Juric, and K. M. McClenaghan, "Detecting cyber security vulnerabilities through reactive programming," in *Proceedings of the 52nd Hawaii International Conference on System Sciences, Grand Wailea, Maui, Hawaii, USA*, 2019, pp. 1–10.

[94] "Glossary of security terms — sans institute." [Online]. Available: https://www.sans.org/security-resources/glossary-of-terms/

[95] "Vocabulary — niccs - national initiative for cybersecurity careers and studies," Cybersecurity and Infrastructure Security Agency. [Online]. Available: https://niccs.us-cert.gov/about-niccs/glossary

[96] "Cybersecurity glossary and vocabulary — cybrary." [Online]. Available: https://www.cybrary.it/cybersecurity-glossary

[97] C. Hobbs, M. Moran, and D. Salisbury, *Open source intelligence in the twenty-first century: new approaches and opportunities*. Springer, 2014.

[98] C. Andrew, R. J. Aldrich, and W. K. Wark, *Secret intelligence: A reader*. Routledge, 2009.

[99] D. R. Hayes and F. Cappa, "Open-source intelligence for risk assessment," *Business Horizons*, vol. 61, no. 5, pp. 689–697, 2018.

[100] L. Rosa, M. Freitas, S. Mazo, E. Monteiro, T. Cruz, and P. Simões, "A comprehensive security analysis of a scada protocol: From osint to mitigation," *IEEE Access*, vol. 7, pp. 42 156–42 168, 2019.

[101] H. Chen, R. Liu, N. Park, and V. Subrahmanian, "Using twitter to predict when vulnerabilities will be exploited," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data Mining*, 2019, pp. 3143–3152.

[102] N. Dionísio, F. Alves, P. M. Ferreira, and A. Bessani, "Cyberthreat detection from twitter using deep neural networks," in *2019 international joint conference on neural networks (IJCNN)*. IEEE, 2019, pp. 1–8.

[103] M. S. Abdullah, A. Zainal, M. A. Maarof, and M. N. Kassim, "Cyber-attack features for detecting cyber threat incidents from online news," in *2018 Cyber Resilience Conference (CRC)*. IEEE, 2018, pp. 1–4.

[104] S. Zhou, Z. Long, L. Tan, and H. Guo, "Automatic identification of indicators of compromise using neural-based sequence labelling," *arXiv preprint arXiv:1810.10156*, 2018.

[105] G. Husari, E. Al-Shaer, M. Ahmed, B. Chu, and X. Niu, "Ttpdrill: Automatic and accurate extraction of threat actions from unstructured text of cti sources," in *Proceedings of the 33rd annual computer security applications conference*, 2017, pp. 103–115.

[106] X. Liao, K. Yuan, X. Wang, Z. Li, L. Xing, and R. Beyah, "Acing the ioc game: Toward automatic discovery and analysis of open-source cyber threat intelligence," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 755–766.

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

[108] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 29, no. 1, 2015.

[109] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[110] T. K. Landauer and S. T. Dumais, "A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge." *Psychological review*, vol. 104, no. 2, p. 211, 1997.

[111] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001.

[112] H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to information retrieval*. Cambridge University Press Cambridge, 2008, vol. 39.

[113] Z. S. Harris, "Distributional structure," *Word*, vol. 10, no. 2-3, pp. 146–162, 1954.

[114] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *European conference on machine learning*. Springer, 1998, pp. 137–142.

[115] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, pp. 293–300, 1999.

[116] A. M. Kibriya, E. Frank, B. Pfahringer, and G. Holmes, "Multinomial naive bayes for text categorization revisited," in *AI 2004: Advances in Artificial Intelligence: 17th Australian Joint Conference on Artificial Intelligence, Cairns, Australia, December 4-6, 2004. Proceedings 17*. Springer, 2005, pp. 488–499.

[117] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA*, 2018, pp. 4171—4186.

[118] S. Khan, "Bert, roberta, distilbert, xlnet—which one to use," *Towards Data Science*, 2019. [Online]. Available: https://towardsdatascience.com/bertroberta-distilbert-xlnet-which-one-to-use-3d5ab82ba5f8

[119] "The hacker news — 1 trusted cybersecurity news site." [Online]. Available: https://thehackernews.com/

[120] "Threatpost — the first stop for security news." [Online]. Available: https://threatpost.com/

[121] "Ars technica." [Online]. Available: https://arstechnica.com/

[122] "Security affairs - read, think, share ... security is everyone's responsibility." [Online]. Available: https://securityaffairs.co/wordpress/

[123] Scikit-learn, "Countvectorizer." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html

[124] L. Lipponen, "Exploring foundations for computer-supported collaborative learning," in *Computer support for collaborative learning*. Routledge, 2023, pp. 72–81.

[125] P. Dillenbourg, "What do you mean by collaborative learning?" 1999.

[126] Á. M. De Jesús and I. F. Silveira, "Game-based collaborative learning framework for computational thinking development," *Revista Facultad de Ingeniería Universidad de Antioquia*, no. 99, pp. 113–123, 2021.

[127] N. Wahyuningtyas and I. Idris, "Increasing geographic literacy through the development of computer supported collaborative learning," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 7, pp. 74–85, 2020.

[128] X. Huang, "Improving communicative competence through synchronous communication in computer-supported collaborative learning environments: A systematic review," *Education Sciences*, vol. 8, no. 1, p. 15, 2018.

[129] G. Stahl, "Investigation 2. a theory of group cognition in cscl," in *Theoretical investigations: Philosophical foundations of group cognition*. Springer, 2021, pp. 27–61.

[130] D. Iorga, D. Corlătescu, O. Grigorescu, C. Săndescu, M. Dascălu, and R. Rughiniș, "Early detection of vulnerabilities from news websites using machine learning models," in *2020 19th RoEduNet Conference: Networking in Education and Research (RoEduNet)*. IEEE, 2020, pp. 1–6.

[131] A. Halavais, "Computer-supported collaborative learning," *The International Encyclopedia of Communication Theory and Philosophy*, pp. 1–5, 2016.

[132] P. Dillenbourg, S. Järvelä, and F. Fischer, *The evolution of research on computer-supported collaborative learning: From design to orchestration*. Springer, 2009.

[133] H. Jeong, C. E. Hmelo-Silver, and K. Jo, "Ten years of computer-supported collaborative learning: A meta-analysis of cscl in stem education during 2005–2014," *Educational research review*, vol. 28, p. 100284, 2019.

[134] X. Tian and Z. Li, "Collaborative learning for information security topics: A pilot study." in *AMCIS*, 2020.

[135] X. Yuan, T. Zhang, A. A. Shama, J. Xu, L. Yang, J. Ellis, W. He, and C. Waters, "Teaching cybersecurity using guided inquiry collaborative learning," in *2019 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2019, pp. 1–6.

[136] J.-W. Strijbos, "Assessment of (computer-supported) collaborative learning," *IEEE transactions on learning technologies*, vol. 4, no. 1, pp. 59–73, 2010.

[137] J. Roschelle, "A review of the international handbook of computer-supported collaborative learning 2021," 2020.

[138] V. S. Kumar, "Computer-supported collaborative learning: issues for research," in *Eighth annual graduate symposium on Computer Science, University of Saskatchewan*. Citeseer, 1996.

[139] L. Silva, A. J. Mendes, and A. Gomes, "Computer-supported collaborative learning in programming education: A systematic literature review," in *2020 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, 2020, pp. 1086–1095.

[140] J. Murphy PhD, E. Sihler, M. Ebben PhD, and G. Wilson, "Building a virtual cybersecurity collaborative learning laboratory (vccll)," in *2014 World Congress in Computer Science, Conference Proceedings: Computer Engineering and Applied Computing*, 2014.

[141] Á. Lédeczi, M. MarÓti, H. Zare, B. Yett, N. Hutchins, B. Broll, P. Völgyesi, M. B. Smith, T. Darrah, M. Metelko *et al.*, "Teaching cybersecurity with networked robots," in *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, 2019, pp. 885–891.

[142] G. Stahl, T. D. Koschmann, and D. D. Suthers, *Computer-supported collaborative learning*. Citeseer, 2006.

[143] Z. Chen and C. Demmans, "Csclrec: Personalized recommendation of forum posts to support socio-collaborative learning." *International Educational Data Mining Society*, 2020.

[144] D. Zimbra, A. Abbasi, D. Zeng, and H. Chen, "The state-of-the-art in twitter sentiment analysis: A review and benchmark evaluation," *ACM Transactions on Management Information Systems (TMIS)*, vol. 9, no. 2, pp. 1–29, 2018.

[145] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, "Sentiment strength detection in short informal text," *Journal of the American society for information science and technology*, vol. 61, no. 12, pp. 2544–2558, 2010.

[146] M. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," in *2013 fourth international conference on computing, communications and networking technologies (ICCCNT)*. IEEE, 2013, pp. 1–5.

[147] A. Severyn and A. Moschitti, "Twitter sentiment analysis with deep convolutional neural networks," in *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, 2015, pp. 959–962.

[148] M. Pota, M. Ventura, R. Catelli, and M. Esposito, "An effective bert-based pipeline for twitter sentiment analysis: A case study in italian," *Sensors*, vol. 21, no. 1, p. 133, 2020.

[149] "Sentence transformers, sbert," HuggingFace, 2022. [Online]. Available: https://huggingface.co/sentence-transformers

[150] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[151] F. Chollet *et al.*, "Keras: The python deep learning library," *Astrophysics source code library*, pp. ascl–1806, 2018.

[152] M. Honnibal, "Spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing, sentometrics research," Sentometrics Research. Available at: https://sentometrics-research.com, 2017.

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

[154] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, "Glue: A multi-task benchmark and analysis platform for natural language understanding," *arXiv preprint arXiv:1804.07461*, 2018.

[155] R. Caruana, "Multitask learning: A knowledge-based source of inductive bias1," in *Proceedings of the Tenth International Conference on Machine Learning*. Citeseer, 1993, pp. 41–48.

[156] Z. Han, X. Li, Z. Xing, H. Liu, and Z. Feng, "Learning to predict severity of software vulnerability using only vulnerability description," in *2017 IEEE International conference on software maintenance and evolution (ICSME)*. IEEE, 2017, pp. 125–136.

[157] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[158] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[159] "National vulnerability database nvd - data feeds," National Institute of Standards and Technology. [Online]. Available: https://nvd.nist.gov/vuln/data-feeds

[160] J. X. Morris, E. Lifland, J. Y. Yoo, J. Grigsby, D. Jin, and Y. Qi, "Textattack: A framework for adversarial attacks, data augmentation, and adversarial training in nlp," *arXiv preprint arXiv:2005.05909*, 2020.

[161] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[162] R. Rehurek, "models.word2vec – word2vec embeddings — gensim." [Online]. Available: https://radimrehurek.com/gensim/models/word2vec.html

[163] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," *arXiv preprint arXiv:1508.04025*, 2015.

[164] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[165] "Find pre-trained models — kaggle." [Online]. Available: https://www.kaggle.com/models?tfhub-redirect=true

[166] I. Turc, M.-W. Chang, K. Lee, and K. Toutanova, "Well-read students learn better: On the importance of pre-training compact models," *arXiv preprint arXiv:1908.08962*, 2019.

[167] Y. Zhu, R. Kiros, R. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba, and S. Fidler, "Aligning books and movies: Towards story-like visual explanations by watching movies and reading books," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 19–27.

[168] Barkly, "Study reveals 64% of organizations experienced successful endpoint attack in 2018," *Business Wire*, 2018. [Online]. Available: https://www.businesswire.com/news/home/20181016005758/en/Study-Reveals-64-of-Organizations-Experienced-Successful-EndpointAttack-in-2018

[169] A. L. Queiroz, S. Mckeever, and B. Keegan, "Eavesdropping hackers: Detecting software vulnerability communication on social media using text mining," in *The Fourth International Conference on Cyber-Technologies and Cyber-Systems*, 2019, pp. 41–48.

[170] S. Trabelsi, H. Plate, A. Abida, M. M. B. Aoun, A. Zouaoui, C. Missaoui, S. Gharbi, and A. Ayari, "Monitoring software vulnerabilities through social networks analysis," in *2015 12th International Joint Conference on e-Business and Telecommunications (ICETE)*, vol. 4. IEEE, 2015, pp. 236–242.

[171] J. Wei and K. Zou, "Eda: Easy data augmentation techniques for boosting performance on text classification tasks," in *Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference Natural Language Processing, Hong Kong, China*, 2019, pp. 6381—-6387.

[172] V. Atliha and D. Šešok, "Text augmentation using bert for image captioning," *Applied Sciences*, vol. 10, no. 17, p. 5978, 2020.

[173] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural architectures for named entity recognition," in *The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA*, 2016, p. 260–270.

[174] X. Wu, S. Lv, L. Zang, J. Han, and S. Hu, "Conditional bert contextual augmentation," in *Computational Science–ICCS 2019: 19th International Conference, Faro, Portugal, June 12–14, 2019, Proceedings, Part IV 19*. Springer, 2019, pp. 84–95.

[175] B. Y. Lin, F. F. Xu, Z. Luo, and K. Zhu, "Multi-channel bilstm-crf model for emerging named entity recognition in social media," in *Proceedings of the 3rd Workshop on Noisy User-generated Text*, 2017, pp. 160–165.

[176] Q. Qiu, Z. Xie, L. Wu, L. Tao, and W. Li, "Bilstm-crf for geological named entity recognition from the geoscience literature," *Earth Science Informatics*, vol. 12, pp. 565–579, 2019.

[177] Q. Zhu, X. Li, A. Conesa, and C. Pereira, "Gram-cnn: a deep learning approach with local context for named entity recognition in biomedical text," *Bioinformatics*, vol. 34, no. 9, pp. 1547–1554, 2018.

[178] "spacy · industrial-strength natural language processing in python." [Online]. Available: https://spacy.io/

[179] J. Serrà and A. Karatzoglou, "Getting deep recommenders fit: Bloom embeddings for sparse binary input/output networks," in *Proceedings of the Eleventh ACM Conference on Recommender Systems*, 2017, pp. 279–287.

[180] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE transactions on image processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[181] 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.

[182] L. Ou-Yang, "Newspaper3k: Article scraping & curation—newspaper 0.0. 2 documentation," 2021. [Online]. Available: https://github.com/codelucas/newspaper

[183] "National institute of standards and technology: National vulnerability database (nvd), vulnerability metrics," National Institute of Standards and Technology, US. [Online]. Available: https://nvd.nist.gov/vuln-metrics/cvss#

[184] E. Ma, "Github - makcedward/nlpaug: Data augmentation for nlp." [Online]. Available: https://github.com/makcedward/nlpaug

[185] Y. Qi, "Github - qdata/textattack: Textattack is a python framework for adversarial attacks, data augmentation, and model training in nlp https://textattack.readthedocs.io/en/master/." [Online]. Available: https://github.com/QData/TextAttack

[186] "News and advice on the world's latest innovations — zdnet." [Online]. Available: https://www.zdnet.com/

[187] "National institute of standards and technology: National vulnerability database (nvd), nvd dashboard," National Institute of Standards and Technology, US. [Online]. Available: https://nvd.nist.gov/general/nvd-dashboard

[188] "Mapping MITRE ATT&CK® to cves for impact," The Center for Threat-Informed Defense, Bedford, MA, USA. [Online]. Available: https://mitre-engenuity.org/cybersecurity/center-for-threat-informed-defense/our-work/mapping-attck-to-cve-for-impact/

[189] J. Baker, "CVE + MITRE ATT&CK to understand vulnerability impact," Medium. [Online]. Available: https://medium.com/mitre-engenuity/cve-mitre-att-ck-to-understand-vulnerability-impact-c40165111bf7

[190] S. Roe, "Using MITRE ATT&CK with threat intelligence to improve vulnerability management." [Online]. Available: https://outpost24.com/blog/Using-mitre-attack-with-threat-intelligence-to-improve-vulnerability-management

[191] B. Ampel, S. Samtani, S. Ullman, and H. Chen, "Linking common vulnerabilities and exposures to the mitre att&ck framework: A self-distillation approach," *arXiv preprint arXiv:2108.01696*, 2021.

[192] A. Kuppa, L. Aouad, and N.-A. Le-Khac, "Linking cve's to mitre att&ck techniques," in *Proceedings of the 16th International Conference on Availability, Reliability and Security*, 2021, pp. 1–12.

[193] "Threat report ATT&CK mapping (TRAM)," Github. [Online]. Available: https://github.com/center-for-threat-informed-defense/tram/

[194] S. Yoder, "Automating Mapping to ATT&CK: The Threat Report ATT&CK Mapper (TRAM) Tool," Medium. [Online]. Available: https://medium.com/mitre-attack/automating-mapping-to-attack-tram-1bb1b44bda76

[195] M. T. Ribeiro, S. Singh, and C. Guestrin, "Model-agnostic interpretability of machine learning," *arXiv preprint arXiv:1606.05386*, 2016.

[196] O. Grigorescu, "CVE2ATT&CK dataset," TagTog. [Online]. Available: https://www.tagtog.com/readerbench/MitreMatrix/

[197] ——, "CVE2ATT&CK repository," GitHub. [Online]. Available: https://github.com/readerbench/CVE2ATT-CK

[198] "Vulnerability database." [Online]. Available: https://vuldb.com/

[199] "Exploit database-exploits for penetration testers, researchers, and ethical hackers." [Online]. Available: https://www.exploit-db.com/

[200] TagTog, "Api documentation v1." [Online]. Available: https://github.com/tagtog/tagtog-doc/blob/master/API-projects-v1.md

[201] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," *Intelligent data analysis*, vol. 6, no. 5, pp. 429–449, 2002.

[202] TextAttack, "Documentation webpage." [Online]. Available: https://textattack.readthedocs.io/en/latest/index.html

[203] Q. Yanjun, "Textattack. augmentation recipes." [Online]. Available: https://textattack.readthedocs.io/en/latest/3recipes/augmenter_recipes.html

[204] R. Alazaidah and F. K. Ahmad, "Trending challenges in multi label classification," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 10, pp. 127–131, 2016.

[205] "spacy 101: Everything you need to know," spaCy. [Online]. Available: https://spacy.io/usage/spacy-101

[206] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining multi-label data," *Data mining and knowledge discovery handbook*, pp. 667–685, 2010.

[207] R. Rifkin and A. Klautau, "In defense of one-vs-all classification," *The Journal of Machine Learning Research*, vol. 5, pp. 101–141, 2004.

[208] G. Tsoumakas and I. Vlahavas, "Random k-labelsets: An ensemble method for multilabel classification," in *European conference on machine learning*. Springer, 2007, pp. 406–417.

[209] I. Rish *et al.*, "An empirical study of the naive bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, no. 22, 2001, pp. 41–46.

[210] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020.

[211] "Grid search," Scikit. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

[212] D. A. Forsyth, J. L. Mundy, V. di Gesú, R. Cipolla, Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "Object recognition with gradient-based learning," *Shape, contour and grouping in computer vision*, pp. 319–345, 1999.

[213] W.-t. Yih, X. He, and C. Meek, "Semantic parsing for single-relation question answering," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2014, pp. 643–648.

[214] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," *arXiv preprint arXiv:1404.2188*, 2014.

[215] "Word representation for cyber security vulnerability domain," Github. [Online]. Available: https://github.com/unsw-cse-soc/Vul_Word2Vec

[216] I. Beltagy, K. Lo, and A. Cohan, "Scibert: A pretrained language model for scientific text," *arXiv preprint arXiv:1903.10676*, 2019.

[217] "Secbert model," HuggingFace. [Online]. Available: https://huggingface.co/jackaduma/SecBERT

[218] "Bce with logit loss," Pytorch. [Online]. Available: https://pytorch.org/docs/stable/generated/torch.nn.BCEWithLogitsLoss.html

[219] Y. Dong, W. Guo, Y. Chen, X. Xing, Y. Zhang, and G. Wang, "Towards the detection of inconsistencies in public security vulnerability reports," in *28th USENIX security symposium (USENIX Security 19)*, 2019, pp. 869–885.

[220] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM computing surveys (csur)*, vol. 53, no. 3, pp. 1–34, 2020.

[221] G. Kasieczka, B. Nachman, D. Shih, O. Amram, A. Andreassen, K. Benkendorfer, B. Bortolato, G. Brooijmans, F. Canelli, J. H. Collins *et al.*, "The lhc olympics 2020 a community challenge for anomaly detection in high energy physics," *Reports on progress in physics*, vol. 84, no. 12, p. 124201, 2021.

[222] "Common weakness enumeration webpage," MITRE, 2023. [Online]. Available: https://cwe.mitre.org/

[223] L. Spitzner, "Honeypots: Catching the insider threat," in *19th Annual Computer Security Applications Conference, 2003. Proceedings.* IEEE, 2003, pp. 170–179.

[224] G. H. Kim and E. H. Spafford, "Experiences with tripwire: Using integrity checkers for intrusion detection," 1994.

[225] A. Harper, E. Balas, and H. Gen III, "The birth of roo," *Black Hat Briefings*, 2005.

[226] N. Provos, "Honeyd-a virtual honeypot daemon," in *10th dfn-cert workshop, hamburg, germany*, vol. 2, 2003, p. 4.

[227] S. Kumar, P. Singh, R. Sehgal, and J. Bhatia, "Distributed honeynet system using gen iii virtual honeynet," *International Journal of Computer Theory and Engineering*, vol. 4, no. 4, p. 537, 2012.

[228] J. Levine, R. LaBella, H. Owen, D. Contis, and B. Culver, "The use of honeynets to detect exploited systems across large enterprise networks," in *IEEE Systems, Man and Cybernetics SocietyInformation Assurance Workshop, 2003.* IEEE, 2003, pp. 92–99.

[229] M. Müter, F. Freiling, T. Holz, and J. Matthews, "A generic toolkit for converting web applications into high-interaction honeypots," *University of Mannheim*, vol. 280, pp. 6–1, 2008.

[230] F. De Gaspari, S. Jajodia, L. V. Mancini, and A. Panico, "Ahead: A new architecture for active defense," in *Proceedings of the 2016 ACM workshop on automated decision making for active cyber defense*, 2016, pp. 11–16.

[231] A. Shabtai, M. Bercovitch, L. Rokach, Y. Gal, Y. Elovici, and E. Shmueli, "Behavioral study of users when interacting with active honeytokens," *ACM Transactions on Information and System Security (TISSEC)*, vol. 18, no. 3, pp. 1–21, 2016.

[232] A. Radovici, R. Cristian, and R. ŞERBAN, "A survey of iot security threats and solutions," in *2018 17th RoEduNet conference: networking in education and research (RoEduNet)*. IEEE, 2018, pp. 1–5.

[233] I. Florea, L. C. Ruse, and R. Rughinis, "Challenges in security in internet of things," in *2017 16th RoEduNet Conference: Networking in Education and Research (RoEduNet)*. IEEE, 2017, pp. 1–5.

[234] I. Florea, R. Rughinis, L. Ruse, and D. Dragomir, "Survey of standardized protocols for the internet of things," in *2017 21st International Conference on Control Systems and Computer Science (CSCS)*. IEEE, 2017, pp. 190–196.

[235] "Know your enemy: Defining virtual honeynets," September 2002. [Online]. Available: http://ivanlef0u.fr/repo/madchat/reseau/defense/DefiningVirtualHoneynets.pdf

[236] I. Livshitz, "What's the difference between a high interaction honeypot and a low interaction honeypot," 2020. [Online]. Available: https://www.guardicore.com/2019/1/high-interaction-honeypot-versus-low-interaction-honeypot/

[237] S. Symanovich, "What is a honeypot? how it can lure cyberattack ers," *NortonLifeLock, May*, vol. 26, 2020. [Online]. Available: https://us.norton.com/internetsecurity-iot-what-is-a-honeypot.html

[238] "Low, medium and high interaction honeypot," January 2019. [Online]. Available: https://www.guardicore.com/2019/1/high-interaction-honeypot-versus-low-interaction-honeypot/

[239] N. Provos, "Developments of the honeyd virtual honeypot," *http://honeyd.org*, 2005.

[240] "Using honeyd configurations to build honeypot systems," 2021. [Online]. Available: https://searchsecurity.techtarget.com/Using-HoneyD-configurations-to-build-honeypot-systems

[241] B. Lutkevich, "Lamp (linux, apache, mysql, php)," 2021. [Online]. Available: https://whatis.techtarget.com/definition/LAMP-Linux-Apache-MySQL-PHP

[242] "Usage statistics of apache." [Online]. Available: https://w3techs.com/technologies/details/ws-apache

[243] "Wordpress market share," 2023. [Online]. Available: https://kinsta.com/wordpress-market-share/

[244] "Wordpress: List of security vulnerabilities," 2023. [Online]. Available: https://www.cvedetails.com/vulnerability-list/vendor_id-2337/product_id-4096/

[245] C. Herley, "So long, and no thanks for the externalities: the rational rejection of security advice by users," in *Proceedings of the 2009 workshop on New security paradigms workshop*, 2009, pp. 133–144.

[246] I. Ion, R. Reeder, and S. Consolvo, "{"... No} one can hack my {Mind"}: Comparing expert and {Non-Expert} security practices," in *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, 2015, pp. 327–346.

[247] A. Hern, "Wannacry, petya, notpetya: How ransomware hit the big time in 2017," *The Guardian*, vol. 30, no. 12, p. 2017, 2017. [Online]. Available: https://www.theguardian.com/technology/2017/dec/30/wannacry-petya-notpetya-ransomware

[248] C. Li, H. Wang, Z. Zhang, A. Sun, and Z. Ma, "Topic modeling for short texts with auxiliary word embeddings," in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, 2016, pp. 165–174.

[249] F. Esposito, A. Corazza, F. Cutugno *et al.*, "Topic modelling with word embeddings." in *CLiC-it/EVALITA*, 2016.

[250] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in neural information processing systems*, vol. 26, 2013. [Online]. Available: https://proceedings.neurips.cc/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf

[251] M. Grootendorst, "Bertopic: Leveraging bert and c-tf-idf to create easily interpretable topics," *Zenodo, Version v0*, vol. 9, 2020. [Online]. Available: https://doi.org/10.5281/zenodo.4381785

[252] R. J. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Pacific-Asia conference on knowledge discovery and data mining*. Springer, 2013, pp. 160–172.

[253] F. Hamborg, N. Meuschke, C. Breitinger, and B. Gipp, "news-please: A generic news crawler and extractor," pp. 218–223, 2017.

[254] C. Gormley and Z. Tong, *Elasticsearch: the definitive guide: a distributed real-time search and analytics engine*. "O'Reilly Media, Inc.", 2015.

[255] I. Montani, "spacy/spacy/lang/en/stop_words.py at master - explosion/spacy - github." [Online]. Available: https://github.com/explosion/spaCy/blob/master/spacy/lang/en/stop_words.py

[256] "sentence-transformers/roberta-base-nli-stsb-mean-tokens · hugging face." [Online]. Available: https://huggingface.co/sentence-transformers/roberta-base-nli-stsb-mean-tokens

[257] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," *arXiv preprint arXiv:1908.10084*, 2019.

[258] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," *arXiv preprint arXiv:1802.03426*, 2018.

[259] "Official documentation for umap parameters." [Online]. Available: https://umaplearn.readthedocs.io/en/latest/parameters.html

[260] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, 2016.

[261] A. Mohamed, "An effective dimension reduction algorithm for clustering arabic text," *Egyptian Informatics Journal*, vol. 21, no. 1, pp. 1–5, 2020.

[262] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.

[263] "Official documentation for hdbscan parameters." [Online]. Available: https://hdbscan.readthedocs.io/en/latest/parameterselection.html

[264] "sklearn.cluster.meanshift — scikit-learn 1.4.1 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.cluster.MeanShift.html

[265] "Python scikit-learn module accessible at." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.featureextraction.text.TfidfVectorizer.html

[266] "Python scikit-learn module accessible at." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.featureextraction.text.CountVectorizer.html

[267] [Online]. Available: https://amp-theguardian-com.cdn.ampproject.org/c/s/amp.theguardian.com/world/2020/jul/16/russian-state-sponsored-hackerstarget-covid-19-vaccine-researchers

[268] W. Helen, C. Clive, and F. Henry, "Russia-linked hackers accused of targeting covid-19 vaccine developers — ars technica." [Online]. Available: https://arstechnica.com/information-technology/2020/07/russia-linked-hackers-accused-of-targeting-covid-19-vaccine-developers/

[269] T. Seals, "Nation-state attackers actively target covid-19 vaccine-makers — threatpost." [Online]. Available: https://threatpost.com/russia-north-korea-attackingcovid-19-vaccine-makers/161205/

[270] P. Paganini, "Hackers target covid-19 vaccine supply chain and sell the vaccine in darkweb." [Online]. Available: https://securityaffairs.com/112433/hacking/covid-19-attacks-2.html

[271] T. Seals, "Lazarus group hits covid-19 vaccine-maker in espionage attack — threatpost." [Online]. Available: https://threatpost.com/lazarus-covid-19-vaccine-maker-espionage/162591/

[272] P. Paganini, "Ema: Some of pfizer/biontech covid-19 vaccine data was leaked online." [Online]. Available: https://securityaffairs.co/wordpress/113326/data-breach/ema-data-breach.htm

[273] "ENISA Threat Landscape 2022." [Online]. Available: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2022

[274] "CVE Details." [Online]. Available: https://www.cvedetails.com/browse-by-date.php

[275] "CVSS v3.1 Specification Document." [Online]. Available: https://www.first.org/cvss/v3.1/specification-document

[276] L. Allodi and F. Massacci, "Comparing Vulnerability Severity and Exploits Using Case-Control Studies," *ACM Transactions on Information and System Security*, vol. 17, no. 1, pp. 1:1–1:20, Aug. 2014.

[277] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," May 2019.

[278] P. Frode de la Foret, S. Ruseti, C. Sandescu, M. Dascalu, and S. Travadel, "Interpretable Identification of Cybersecurity Vulnerabilities from News Articles," in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*. Held Online: INCOMA Ltd., Sep. 2021, pp. 428–436.

[279] I. Beltagy, M. E. Peters, and A. Cohan, "Longformer: The Long-Document Transformer," Dec. 2020.

[280] Y. Ming, P. Xu, H. Qu, and L. Ren, "Interpretable and steerable sequence learning via prototypes," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 903–913.

[281] C. Vladescu, M.-A. Dinisor, O. Grigorescu, D. Corlatescu, C. Sandescu, and M. Dascalu, "What are the Latest Cybersecurity Trends? A Case Study Grounded in Language Models," in *2021 23rd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, Dec. 2021, pp. 140–146.

[282] 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," Jul. 2019.

[283] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.

[284] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," *arXiv preprint arXiv:1909.11942*, 2019.

[285] M. Liberato, "Secbert: Analyzing reports using bert-like models," Master's thesis, University of Twente, 2022.

[286] "HuggingFace Transformers." [Online]. Available: https://huggingface.co/docs/transformers/main/en/index

[287] "Scikit-learn-contrib/hdbscan," Jun. 2022. [Online]. Available: scikit-learn-contrib

[288] "PRAW: The Python Reddit API Wrapper — PRAW 7.6.0 documentation." [Online]. Available: https://praw.readthedocs.io/en/stable/

[289] K. McKee, "Kurtmckee/feedparser," Jun. 2022.

[290] "Newspaper3k: Article scraping & curation — newspaper 0.0.2 documentation." [Online]. Available: https://newspaper.readthedocs.io/en/latest/

[291] "PyMongo 4.1.1 Documentation — PyMongo 4.1.1 documentation." [Online]. Available: https://pymongo.readthedocs.io/en/stable/

[292] "MongoDB — Build Faster. Build Smarter." [Online]. Available: https://www.mongodb.com

[293] "FastAPI." [Online]. Available: https://fastapi.tiangolo.com/

[294] "Svelte - Cybernetically enhanced web apps." [Online]. Available: https://svelte.dev/

[295] "Carbon Design System." [Online]. Available: https://carbondesignsystem.com/carbondesignsystem.com

[296] M. Bostock, "D3.js - Data-Driven Documents." [Online]. Available: https://d3js.org/

[297] R. H. Weber, "Internet of things–new security and privacy challenges," *Computer law & security review*, vol. 26, no. 1, pp. 23–30, 2010.

[298] J. Scambray, S. McClure, G. Kurtz, McClure, Scambray, and Kurtz, *Hacking exposed: network security secrets & solutions*. Osborne/McGraw-Hill New York, 2001, vol. 118.

[299] T. Mahmood and U. Afzal, "Security analytics: Big data analytics for cybersecurity: A review of trends, techniques and tools," in *2013 2nd national conference on Information assurance (ncia)*. IEEE, 2013, pp. 129–134.

[300] B. R. Rowe and M. P. Gallaher, "Private sector cyber security investment strategies: An empirical analysis," in *The fifth workshop on the economics of information security (WEIS06)*, 2006.

[301] S. M. Tisdale, "Cybersecurity: Challenges from a systems, complexity, knowledge management and business intelligence perspective." *Issues in Information Systems*, vol. 16, no. 3, 2015.

[302] P. Herzog, "Open source security testing methodology manual (osstmm)," *ISECOM. Available: http://www.isecom.org/research/.[Accessed 2015]. Open Web Application Security Project (OWASP), Attack. Available: https://www.owasp.org/index.php/Category:Attack*, 2010. [Online]. Available: http://www.isecom.org/research/osstmm.html

[303] "Remediating computer security threats using distributed sensor computers." [Online]. Available: http://patents.justia.com/patent/9374385

[304] "Distribution of security rules among sensor computers." [Online]. Available: http://patents.justia.com/patent/9350750

[305] M. Souppaya, K. Scarfone *et al.*, "Guide to enterprise patch management technologies," *NIST Special Publication*, vol. 800, p. 40, 2013. [Online]. Available: http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-40r3.pdf

[306] "Continuous vulnerability assessment & remediation guideline," University of California, Berkeley, 2023, accessed: 2023-12-15. [Online]. Available: https://security.berkeley.edu/continuous-vulnerability-assessment-remediation-guideline

[307] S. G. Kelekar, "Systems and methods for real-time network-based vulnerability assessment," 2012, patent number US8127359. [Online]. Available: http://www.google.com/patents/US8127359

[308] C. M. Stuart, K. George, K. Robin, A. B. Marshall, J. M. Michael, M. P. Christopher, M. C. David, and A. Christopher, "System and method for network vulnerability detection and reporting," 2006, patent number US7152105. [Online]. Available: http://www.google.com/patents/US7152105

[309] OWASP, "OWASP ASVS – Open Web Application Security Project Application Security Verification Standard," 2023, accessed: 2023-12-15. [Online]. Available: https://github.com/OWASP/ASVS

[310] M. C. Ivan Arce, "Automating penetration tests - black hat," 2001, accessed: 2023-12-15. [Online]. Available: https://www.blackhat.com/presentations/bh-usa-01/IvanAcre/bh-usa-01-Ivan-Arce.ppt

[311] [Unknown Author], "The mathematics behind an automated penetration testing framework," 2014, accessed: 2023-12-15. [Online]. Available: https://www.securityforum.at/wpcontent/uploads/2014/05/SF14_Slides_Simos.pdf

[312] ——, "Continuous auditing: Is it fantasy or reality?" *Information Systems Control Journal*, vol. 5, 2002, accessed: 2023-12-15. [Online]. Available: http://www.isaca.org/Groups/Professional-English/continuous-monitoring-auditing/GroupDocuments/ISACA%20Continuous%20Auditing.pdf

[313] H. Joh and Y. K. Malaiya, "A framework for software security risk evaluation using the vulnerability lifecycle and cvss metrics," in *Proc. International Workshop on Risk and Trust in Extended Enterprises*. Citeseer, 2010, pp. 430–434.

[314] U. K. Singh, C. Joshi, and N. Gaud, "Information security assessment by quantifying risk level of network vulnerabilities," *International Journal of Computer Applications*, vol. 156, no. 2, pp. 37–44, 2016.

[315] A. Khazaei, M. Ghasemzadeh, and V. Derhami, "An automatic method for cvss score prediction using vulnerabilities description," *Journal of Intelligent & Fuzzy Systems*, vol. 30, no. 1, pp. 89–96, 2016.

[316] T. Wen, Y. Zhang, Y. Dong, and G. Yang, "A novel automatic severity vulnerability assessment framework." *J. Commun.*, vol. 10, no. 5, pp. 320–329, 2015.

[317] G. Da, M. Xu, J. Zhang, and P. Zhao, "Joint cyber risk assessment of network systems with heterogeneous components," *arXiv preprint arXiv:2006.16092*, 2020.

[318] RSA, "The rsa digital risk index," 2020, accessed: November 1, 2020. [Online]. Available: https://www.rsa.com/en-us/tools/digital-risk-index-form

[319] Tenable, "3 things you need to know about prioritizing vulnerabilities," 2018, accessed: November 1, 2020. [Online]. Available: https://static.tenable.com/marketing/whitepapers/WhitepaperThree_Things_You_Need_to_Know_About_Prioritizing_Vulnerabilities_eBook.pdf

[320] ——, "Tenable lumin," 2020, accessed: November 14, 2020. [Online]. Available: https://www.tenable.com/products/tenable-lumin

[321] Trust Data Solutions, "Cyber security risk services," 2020, accessed: November 14, 2020. [Online]. Available: https://trustsds.com/consulting-services/enterprise-riskand-compliance/cyber-security-risk-assessment/

[322] UpGuard, "A complete third-party risk management platform," 2020, accessed: November 14, 2020. [Online]. Available: https://www.upguard.com/product/vendorrisk

[323] Cisco, "Cyber security and insurance," 2020, accessed: November 14, 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/security/cyberinsurance/index.html

[324] Nationwide, "What is cyber insurance?" 2020, accessed: November 14, 2020. [Online]. Available: https://www.nationwide.com/lc/resources/smallbusiness/articles/what-is-cyber-insurance

[325] "Openscap," 2020, accessed: November 14, 2020. [Online]. Available: https://www.openscap.org/

[326] NIST, "National checklist program repository," 2020, accessed: November 14, 2020. [Online]. Available: https://nvd.nist.gov/ncp/repository

[327] Trusted Computing Group, "Trusted network communications," 2020, accessed: November 14, 2020. [Online]. Available: https://trustedcomputinggroup.org/workgroups/trusted-network-communications/

[328] Center for Internet Security, "Cis benchmarks," 2020, accessed: November 14, 2020. [Online]. Available: https://www.cisecurity.org/cis-benchmarks/

[329] Open Vulnerability and Assessment Language, "Open vulnerability and assessment language," MITRE, 2020, accessed: November 14, 2020. [Online]. Available: https://oval.mitre.org/

[330] C. Nie, J. Li, and S. Wang, "Modeling the effect of spending on cyber security by using surplus process," *Mathematical Problems in Engineering*, vol. 2020, 2020.

[331] Leader Team Global Insurance Broker, "Risk consulting," 2020, accessed: November 14, 2020. [Online]. Available: https://leaderteam.ro/practice/leader-team-riskconsulting/

[332] Corero, "Mirai botnet attack type," 2020, accessed: 2020. [Online]. Available: https://www.corero.com/resources/ddosattack-types/mirai-botnet-ddos-attack.html

[333] A. Gluhak, S. Krco, M. Nati, D. Pfisterer, N. Mitton, and T. Razafindralambo, "A survey on facilities for experimental internet of things research," *IEEE Communications Magazine*, vol. 49, no. 11, pp. 58–67, 2011.

[334] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," *Future generation computer systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[335] OWASP, "Owasp fuzzing," 2020, accessed: 2020. [Online]. Available: https://www.owasp.org/index.php/Fuzzing

[336] Armour Project, "Generic test patterns and test models for iot security testing," 2016, accessed: 2020. [Online]. Available: https://www.armour-project.eu/wpcontent/uploads/2016/08/D21-Generic-test-patterns-and-test-modelsfor-IoT-security-testing.pdf

[337] Anastacia Project, "Attack threats analysis and contingency actions initial report," 2020, version 0.5. [Online]. Available: http://www.anastaciah2020.eu/deliverables/ANASTACIA-WP2-T2.2-CNR-D2.2-AttackThreatsAnalysisAndContingencyActionsInitialReport-v0.5.pdf

[338] ——, "Initial security enforcement manager report," 2020, version 1.0. [Online]. Available: http://www.anastaciah2020.eu/deliverables/ANASTACIA-WP3-T3.1-UMU-D3.1-InitialSecurityEnforcementManagerReport-v1.0.pdf

[339] H. Guo, J. Ren, D. Zhang, Y. Zhang, and J. Hu, "A scalable and manageable iot architecture based on transparent computing," *Journal of Parallel and Distributed Computing*, vol. 118, pp. 5–13, 2018.

[340] J. Mocnej, M. Miškuf, P. Papcun, and I. Zolotová, "Impact of edge computing paradigm on energy consumption in iot," *IFAC-PapersOnLine*, vol. 51, no. 6, pp. 162–167, 2018.

[341] OWASP, "Owasp top 10 application security risks," 2017, accessed: 2020. [Online]. Available: https://www.owasp.org/index.php/Top_10-2017_Top_10

[342] C. Martin, R. Nasr, M. Hoersken, and T. Fuechtler, "Automating information security assessments using intelligent software agents," in *2016 14th Annual Conference on Privacy, Security and Trust (PST)*. IEEE, 2016, pp. 736–744.

[343] resin-os, "Docker balena," 2020, accessed: 2020. [Online]. Available: https://github.com/resin-os/balena

[344] O. Grigorescu, C. Săndescu, and R. Rughiniș, "Coda footprint continuous security management platform," in *2016 15th RoEduNet Conference: Networking in Education and Research*. IEEE, 2016, pp. 1–5.

[345] S. Khan and S. Parkinson, "Review into state of the art of vulnerability assessment using artificial intelligence," *Guide to Vulnerability Analysis for Computer Networks and Systems: An Artificial Intelligence Approach*, pp. 3–32, 2018.

[346] Y. Cherdantseva and J. Hilton, "A reference model of information assurance & security," in *2013 international conference on availability, reliability and security*. IEEE, 2013, pp. 546–555.

[347] K. Kent, S. D. Quinn, and P. Mell, "The security content automation program (scap): Automating compliance checking, vulnerability management, and security measurement," Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology, Report 7343, 2006.

[348] C. Xiaolin, T. Xiaobin, Z. Yong, and X. Hongsheng, "A markov game theory-based risk assessment model for network information system," in *2008 International Conference on Computer Science and Software Engineering*, vol. 3. IEEE, 2008, pp. 1057–1061.

[349] A. Karbowski, K. Malinowski, S. Szwaczyk, and P. Jaskóła, "Critical infrastructure risk assessment using markov chain model," *Journal of Telecommunications and Information Technology*, no. 2, pp. 15–22, 2019.

[350] F. Sun, J. Pi, J. Lv, and T. Cao, "Network security risk assessment system based on attack graph and markov chain," in *Journal of Physics: Conference Series*, vol. 910, no. 1. IOP Publishing, 2017, p. 012005.

[351] J. Shin, H. Son, and G. Heo, "Cyber security risk evaluation of a nuclear i&c using bn and et," *Nuclear Engineering and Technology*, vol. 49, no. 3, pp. 517–524, 2017.

[352] R. Munir, J. P. Disso, I. Awan, and M. R. Mufti, "A quantitative measure of the security risk level of enterprise networks," in *2013 Eighth International Conference on Broadband and Wireless Computing, Communication and Applications*. IEEE, 2013, pp. 437–442.

[353] H. Gao, J. Zhu, and C. Li, "The analysis of uncertainty of network security risk assessment using dempster-shafer theory," in *2008 12th International Conference on Computer Supported Cooperative Work in Design*. IEEE, 2008, pp. 754–759.

[354] Y. Duan, Y. Cai, Z. Wang, and X. Deng, "A novel network security risk assessment approach by combining subjective and objective weights under uncertainty," *Applied Sciences*, vol. 8, no. 3, p. 428, 2018.

[355] H. Owen and B. Byers, "Automation support for cve retrieval," National Institute of Standards and Technology, Official API Documentation, 2021.

[356] V. L. Sujay, "Number of internet of things (iot) connected devices worldwide from 2019 to 2023, with forecasts from 2022 to 2030," Statista, 2023. [Online]. Available: https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/

[357] "Security issues of IoT: Securing your IoT Device in 2023," Device Authority Ltd. [Online]. Available: https://www.deviceauthority.com/blog/security-issues-of-iot-securing-your-iot-device-in-2023/

[358] "Data security: How a proactive c-suite can reduce cyber-risk for the enterprise," The Economist Intelligence Unit, 2016. [Online]. Available: https://impact.economist.com/perspectives/technology-innovation/data-security-how-proactive-c-suite-can-reduce-cyber-risk-enterprise/article/data-security-how-proactive-c-suite-can-reduce-cyber-risk-enterprise

[359] "Common vulnerability scoring system version 4.0: Specification document." [Online]. Available: https://www.first.org/cvss/v4.0/specification-document

[360] "Why organizations struggle with vulnerability management?" [Online]. Available: https://heimdalsecurity.com/blog/vulnerability-management-challenges/

[361] "Forum of incident response and security teams." [Online]. Available: https://www.first.org/

[362] "Common vulnerability scoring system version 4.0: Specification document," Forum of Incident Response and Security Teams. [Online]. Available: https://www.first.org/cvss/v4.0/specification-document

[363] "Cisa known exploited vulnerabilities catalog." [Online]. Available: https://www.cisa.gov/known-exploited-vulnerabilities-catalog

[364] "Exploit prediction scoring system (epss)." [Online]. Available: https://www.first.org/epss/

[365] "Stakeholder-specific vulnerability catego- rization (svcc)." [Online]. Available: https://www.cisa.gov/stakeholder-specific-vulnerability-categorization-ssvc

[366] B. Jung, Y. Li, and T. Bechor, "Cavp: A context-aware vulnerability prioritization model," *Computers Security*, vol. 116, p. 102639, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404822000384

[367] V. Ahmadi Mehri, P. Arlos, and E. Casalicchio, "Automated context-aware vulnerability risk management for patch prioritization," *Electronics*, vol. 11, no. 21, 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/21/3580

[368] "Rudder cve plugin," accessed: 2024. [Online]. Available: https://docs.rudder.io/reference/6.2/plugins/cve.html

[369] C. S. O. Grigorescu and R. Rughiniș, "Coda footprint continuous security management platform," 2016, pp. 1–5.

[370] "Automatic system for early detection of cyber vulnerabilities." [Online]. Available: https://yggdrasil.codaintelligence.com/

[371] "Cis benchmarks list," accessed: 2024. [Online]. Available: https://www.cisecurity.org/cis-benchmarks

[372] "Nist, security content automation protocol," 2022. [Online]. Available: https://csrc.nist.gov/projects/security-content-automation-protocol

[373] "Openscap portal." [Online]. Available: https://www.open-scap.org/

[374] D. Iorga, D.-G. Corlatescu, O. Grigorescu, C. Sandescu, M. Dascalu, and R. Rughinis, "Yggdrasil—early detection of cybernetic vulnerabilities from twitter," in *2021 23rd International Conference on Control Systems and Computer Science (CSCS)*. IEEE, 2021, pp. 463–468.

[375] I. Babalau, D. Corlatescu, O. Grigorescu, C. Sandescu, and M. Dascalu, "Severity prediction of software vulnerabilities based on their text description," in *2021 23rd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*. IEEE, 2021, pp. 171–177.

[376] M. Mijwil, O. J. Unogwu, Y. Filali, I. Bala, and H. Al-Shahwani, "Exploring the top five evolving threats in cybersecurity: An in-depth overview," *Mesopotamian journal of cybersecurity*, vol. 2023, pp. 57–63, 2023.

[377] K. De Nobrega and A.-F. Rutkowski, "The ai family: The information security managers best frenemy?" 2022.

[378] Z. Zhang, H. Ning, F. Shi, F. Farha, Y. Xu, J. Xu, F. Zhang, and K.-K. R. Choo, "Artificial intelligence in cyber security: research advances, challenges, and opportunities," *Artificial Intelligence Review*, pp. 1–25, 2022.

[379] J. Bharadiya, "Machine learning in cybersecurity: Techniques and challenges," *European Journal of Technology*, vol. 7, no. 2, pp. 1–14, 2023.

[380] P. Dixit and S. Silakari, "Deep learning algorithms for cybersecurity applications: A technological and status review," *Computer Science Review*, vol. 39, p. 100317, 2021.

[381] M. Sewak, S. K. Sahay, and H. Rathore, "Deep reinforcement learning in the advanced cybersecurity threat detection and protection," *Information Systems Frontiers*, vol. 25, no. 2, pp. 589–611, 2023.

[382] G. Lin, S. Wen, Q.-L. Han, J. Zhang, and Y. Xiang, "Software vulnerability detection using deep neural networks: a survey," *Proceedings of the IEEE*, vol. 108, no. 10, pp. 1825–1848, 2020.