ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 07, July 2026

OSINT Based Actionable Cyber Threat Intelligence using Deep Learning

[1] Somnath Tandale, [2] Dhananjay Dakhane, [3] Balaji Venketeshwar

[1] Assistant professor, Department of Computer Engineering, Ramrao Adik Institute of Technology, Nerul, Navi Mumbai, Maharashtra, India

[2] Professor, Department of Computer Engineering, Ramrao Adik Institute of Technology, Nerul, Navi Mumbai, Maharashtra, India

[3] Cyber Defense Researcher, Cybervidyapeeth Foundation, Ranchi, Jharkhand, India

Email: [1] Somnath.tandale@rait.ac.in, [2] dhananjay.dakhane@rait.ac.in, [3] mentor@cybervidyapeeth.in

***Abstract*— The study examines the effectiveness of Open Source Intelligence (OSINT) to produce action able intelligence for cyber threat detection and analysis, specifically investigating the method of collection and analysis of data from sources such as Telegram, Dark web and Twitter, for understanding of evolving cyber threats. Through OSINT research method, this study proposes new framework to identify, track and mitigate emerging cyber threats, such as a framework with integrated of BERT, Heterogeneous Graph Convolutional Network (HGCN) and Variational Autoencoders (VAE) model for enhanced cyber threat detection, BERT model to learn semantically related contexts, HGCN model to capture relationship across multiple platforms, and VAE model for detecting sophisticated anomaly in order to holistically study the heterogeneous data in OSINT. The results and analyses from this study have contributed to the effective use of OSINT based intelligence in identifying potential cyber threats and enhancing cyber threat detection and response process. The approach will use natural language processing (NLP) and machine learning to find relationships among threat indicators, and to produce timely actionable intelligence. This study has made contribution to development of secure and robust cyber threat intelligence system.**

***Index Terms*— Actionable Intelligence, BERT, CTI, Deep learning, HGCN, OSINT, VAE.**

I. INTRODUCTION

Open-Source Intelligence (OSINT) refers to the practice of collecting and analyzing in formation from publicly available sources to generate intelligence. In today's digital age, where vast amounts of data are accessible online, OSINT has emerged as a valuable tool for various purposes, including national security, law enforcement, business intelligence, cybersecurity and research. OSINT encompasses a wide range of sources, such as social media platforms, news websites, blogs, public records, government websites, and online forums. By using these sources, the analysts could discover various aspects including security threats, geopolitical development, market information, competitive intelligence etc. A key benefit of using OSINT is that it is cheap and easy to acquire. Unlike traditional ways of obtaining intelligence, which may require extensive resource and capacity. The information available publicly is sometimes free of charge or cheap to obtain. However, OSINT also carries risks of information overload, validating sources, compliance with ethical constraints, and privacy concerns. However, the vast and constantly growing amount of information in the world coupled with easy availability has made OSINT an essential part of the toolkit of various organizations. The ability of OSINT to bring out useful information in an efficient manner has become the advantage in today's information environment. As technology continues to evolve and the digital landscape expands, the role of OSINT is likely to become even more prominent in shaping strategic decisions and mitigating risks. The amount of time that passes between a new cyber vulnerability being discovered and cybercriminals using it has been decreasing over time. Attackers began searching the internet for susceptible host hours after the exploit made public in order to infect susceptible systems with malware like ransomware and cryptocurrency miners. To maximize the effectiveness of preventive measures, it is therefore crucial for the cybersecurity defense strategy to identify threats and their capabilities as soon as possible [17]. Informal sources like social media platforms, forums, dark webs, and open blogs can indeed provide valuable cyber threat intelligence. This type of information allows for instant publication of threat intelligence in unstructured data formats or natural language on the Internet. Open Source Intelligence (OSINT) refers to the unstructured, publicly accessible threat intelligence gathered from these sources. It's important to consider and analyze information from various sources to enhance cybersecurity measures [19]. Open Source Intelligence (OSINT) related to cybersecurity serves as an early warning system for potential cyber security events, including security vulnerability exploits. By monitoring and analyzing OSINT, organizations can stay informed about emerging threats, security vulnerabilities, and potential exploits, allowing them to take proactive measures to enhance their cybersecurity defenses and protect their systems and data [20]. Monitoring digital traces left behind by activities conducted online, such as on social media, professional blogs, and various web forums, can indeed provide valuable insights into evolving

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 07, July 2026

cyber threats. When the aforementioned traces are considered collectively, organizations may have advanced warnings about impending attacks and threats, enabling them to implement preventive countermeasures to strengthen security systems. For instance, an organization might have advance notice of new exploits appearing on platforms such as Twitter before these are publicly disclosed or on dark web forums before their appearances on social networks like Facebook. Accordingly, organizations might use such sources of information to generate threat intelligence. As such, sources of publicly available data such as forums and social networks could reveal early intelligence on new threats and the nature of the compromise that organizations might utilize to avoid cyber incidents.

A. Twitter(X) as Cyber Intelligence Source

As it is stated in the abstract, Twitter is a very useful OSINT source, because it serves as a natural aggregator for other sources, and has the features of "big data." These features are: big volume of data, diverse set of users, high accessibility, and real-time generation of fresh information. Such features enable Twitter to be a useful information source in terms of cyber threat analysis, security vulnerabilities, and exploits detection. By using Twitter as an OSINT source, organizations are informed in real-time, what the new threat trends are, so their security level can be increased, and they are able to react to upcoming threats with proactive measures [25]. Actually, Twitter's prevalence in the cybersecurity community leads to an environment where it is advantageous for both offense and defense practitioners to have a space for discussions, to report occurrences, and to quickly distribute relevant indicators for vulnerabilities, attacks, malware, and any other cyber event that is of value to security analysts. This communication on Twitter allows professionals in the field to maintain current with latest trends, threats, and trends that benefit the community by enabling knowledge sharing and quick proactive actions when needed. Sharing information and indicators through Twitter may help to boost the community awareness.

B. Darkweb as Cyber Intelligence Source

The Dark web has high value in gathering intelligence related to cyber threats, criminal enterprises, and other illegal activities. The Dark Web is part of the internet that is not indexed by normal search engines, requires a special software or setting in order to access (like Tor) and is often perceived as being anonymously related, encrypted and illegal, as it is used for multiple illegal transactions like: sales of data stolen, weapons, drugs etc. The Dark Web is routinely monitored by security experts and law enforcement to detect potential cyber threats and criminal acts for purposes of strengthening the cyber security postures and fighting cybercrimes. It differs from the surface web, a part of the World Wide Web that can be found using search engines such as Google, by its reliance on an overlay network which obfuscates users' identity and activity.

C. Telegram as a Cyber Threat Intelligence Source

In this paper, we investigate how Telegram can be a useful resource for the collection of cyber threat intelligence information by taking advantage of its huge number of users and its anonymous nature. By gathering and inspecting data from public channels and chats, we uncover communication patterns of threat actors, malware propagation, and discussions on vulnerabilities. Our study reveals an in-depth understanding on TTPs of threat actors, by which we can discover new threats and trends. Based on our study, we propose a frame work that covers data collection, data processing and data analysis using Telegram as an information source to improve cyber threat intelligence. Our results indicate that Telegram is an important OSINT source as it delivers real-time threat intelligence information that can aid in preventing cyber threats. This paper opens up opportunities for the investigation into new methods of collecting threat intelligence and will deepen the knowledge base of the cybersecurity community on the activities of threat actors in Telegram.

II. LITERATURE REVIEW

In the recent growing communication technologies, the organizations such as government, businesses, societies, and healthcare, are significantly dependent on Internet access. While internet access is an essential part of these infrastructure, the risk of cybersecurity threats is also higher. Moreover, security threat intelligence is commonly extracted from formal and informal sources, which shows that the information threats are in a structured format. The Common Vulnerabilities and Exposures (CVE) and National Vulnerability Database (NVD) are simple examples of formal sources as well as dark webs, forums, public blogs, and social media platforms are examples of informal sources [1]. The threats and damage that arise from malicious action in the cyber networks involve hacker attacks, rumors, and frauds are considered the most serious security threats, which need to be effectively identified in their early stage [5]. So, the prevention of these threats is noted as an important research topic, where the cyber threat intelligence approach deals with the Advanced Persistent Threat (APT) attacks that quickly identify the steps of attacks and provide a significant solution for prevention [7] [9]. In addition, the Cybersecurity Knowledge Graph (CKG) is an approach, which can provide easy and timely identification of threats [5]. Several traditional and conventional techniques are outdated due to the emergence of several machine learning (ML) approaches. Natural language processing (NLP) is a technique that actively identifies and monitors threats with higher scalability and false positive management. In addition, the NLP supports efficient feature extraction from unstructured data with the help of Named Entity Extraction (NER) techniques [2]. Moreover, among several NLP techniques the

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 07, July 2026

Bidirectional Encoder Representations from Transformers (BERT) model was a pre- trained transformer-based approach, which showed extraordinary performance in threat identification. The BERT model is utilized by several cybersecurity applications, including spam email identification, malware detection, and intrusion as well as anomaly detection for effective threat identification [6] [10] [11] [12]. In addition, an accurate and lightweight NLP pipeline technique was utilized in the feature extraction process which, extracts structured threat features from the unstructured threats, and some of the NLP techniques involving tagging, lemmatization, tokenization, stop word removal, and filtering are also used in the feature extraction [8]. Even though the objective of the NER technique focused on the classification of name entities related to security threats, the model failed to pay attention to prioritizing and analyzing the threats [3] [14]. Moreover, the implementation of end-to-end ML techniques in threat identification showed some limitations such as higher power consumption, and the need for vast amounts of trustable datasets [2] [15]. In addition, the MITRE Adversarial Tactics, Techniques, and Common Knowledge (ATTCK) was an automatic threat identification framework that showed poor accuracy and scalability due to a minimized false positive rate [1]. Furthermore the Convolution neural network (CNN) model was also utilized in the identification and classification of threats such as phishing, malicious, and spam but the CNN model faced limitations on overfitting and higher computational cost due to the requirements of vast amount of data while training [8]. The extraction related issues were eliminated in the Joint entity and relation extraction model but the model showed limited effectiveness in threat identification due to the implementation of simple edge fusion [5]. The study mentioned by [1] “Automated Emerging Cyber Threat Identification and Profiling Based on Natural Language Processing” introduces a novel framework for automatic identification and profiling of emerging cyber threats using Twitter messages and the MITRE ATT&CK knowledge base. By machine learning algorithms and real-time threat analysis, the system aims to provide timely alerts to cybersecurity teams, enabling proactive mitigation strategies against potential threats. In this system, a sequence of experiments and the implementation in an financial institution in Brazil proves that it is capable to detect threats, profiles their intent and to create alarms and therefore prevent possible action as in the case of PetitPotam. It was confirmed that the relevance of the alert of threat intelligence and also the benefit of using automated approaches in finding the useful intelligence from social media for detecting threats. Extraction of Actionable Threat Intelligence from Dark Web Data [2]. A detailed analysis of deriving actionable threat intel from Dark Web data. Highlights the importance of automated threat intel collection in the ever-changing landscape of the cyber threat world. Demonstrates the capabilities of state-of-the-art transformer-based models (BERT and XLNet) to improve NER, in specific regard to cybersecurity knowledge. This also covers the use of machine learning methods, processing approaches and NLP techniques to retrieve cyber-security information from hacker forums emphasizing data pre-processing and model assessment. The study stresses on the importance of NLP models such as BERT in detecting and classifying cyber-security related entities while also showing a satisfactory result with regards to entity extraction from the Dark Web. Detecting cyber threats from Twitter with Deep Neural Networks [3] is a brand-new methodology using neural networks for cyber threat detection from Twitter, which demonstrates outstanding performance of classifying relevant tweets and extracting useful information relevant to cyber security with a high level of accuracy. The tool uses advanced machine learning techniques (convolutional and bidirectional long short-term memory neural networks) to parse the Twitter data through a 3-stage pipeline architecture allowing it to mine relevant security intelligence information effectively. This ability of the tool to filter the tweets according to certain keywords, classify tweets as relevant or irrelevant, and extract the named entities proves that the tool can increase an organization's cybersecurity monitoring and threat detection at large. Comparing to these traditional methods, deep neural network architectures are proved to perform better, thus they will be useful in the field of cyber threat intelligence. The paper [22], titled "An Expert System for Classifying Harmful Content on the Dark Web", which mainly aims to build a classification system that could automatically categorize harmful contents by category and by threat on the dark web, and it also designs an expert system for this task. Using the keywords identified in forum posts, the system achieves a category identification of 72.5% and threat level identification of 83% based on the white box methodology. The system provides an approach to classify a number of different malicious contents and services besides the malicious software and the hack related information and underlines the importance of transparency and trustworthiness of AI based system in the context of national security and public safety. The limitations that still need to be addressed include supporting more categories and improving the knowledge bases and rule definitions. All in all, this research confirms the ability of expert systems to counter the dark web threat while identifying areas for future work and enhancements. The Author [8] explicitly states the opinions concerning threat intelligence in an automated framework for extracting Cyber threat intelligence from Twitter using novelty classification [8], which could facilitate security analysts in quickly detecting and responding to the emerging cyber-attacks. The model analyzes tweets by security influential accounts on Cyber security, and it classifies tweets as normal or anomalous w.r.t. Cyber threats. By analyzing tweets from influential Cyber security-related accounts, the model classifies tweets as either normal or anomalous in relation to Cyber threats. The framework is evaluated using a

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 07, July 2026

portion of the tweets available for the year 2018 that have the highest relevancy towards Cyber threats. In comparison to standard binary classifiers (SVM, MLP, and CNN), the novelty classifier outperforms others in precision and overall performance.

III. RESEARCH GAP

Research Gap 1: OSINT lacks an intelligent analysis mechanism- OSINT analysis does not currently use intelligent mechanisms. The tools employed are populating the gathered data and its clear and direct connections. Incorporating advanced Machine learning models would reduce reliance on humans. Research Gap 2: Need fully end-to-end neural network One significant area is the exploration of multi-task learning architectures to enhance the processing pipeline into a fully end-to-end neural network, which could improve efficiency and performance. Addition ally, there is a need for research on adapting models over time to remain robust against evolving cyber threats and changes in the language used in tweets. The challenges of real-time processing and scalability when applied to larger datasets or extensive Twitter streams remain unaddressed, indicating a need for optimization in dynamic environments.

Research Gap 3: While numerous studies focus on the technical advancements and methodologies for enhancing detection capabilities, there is a lack of comprehensive exploration into the privacy concerns and regulatory frameworks that govern the use of AI in monitoring clandestine activities. Moreover, the scalability issues and real-time implementation problems for these systems in a dynamic dark web context are often not highlighted in the current literature. It is evident that a multidimensional approach needs to be adopted, which does not just further the technical aspect, but also questionably analyzes the ethical and legal constraints needed for the application of AI in cyber security measures against the dark web.

Research Gap 4: Automation in the detection of misleading information or processes should be adopted rather than manual checking.

IV. METHODOLOGY

1. Input Data: CybAttT Dataset: The CybAttT (Cyber Attack Tweets and Texts) Dataset is a corpus of unstructured Open-Source Intelligence (OSINT) created for the purpose of cyber threat intelligence. In order to account for threat actors' communication and dissemination patterns and to collect real world threats, attacker dialogue, and exploit releases across varied platforms (such as Twitter, Dark Web forums, Telegram channels and cybercrime marketplaces), this dataset aggregates heterogenous OSINT from varied sources in the digital realm. We endeavor to provide a corpus that mitigates many of OSINT's inherent shortcomings- namely the noise present within it, the multilingual nature, and the contextual ambiguities.

2. Preprocessing: Our pipeline takes in raw, messy data and begins the task of creating and refining a normalized, useful corpus. First, we filter out common language words (like 'the' or 'and') and retain words pertinent to our domain (like 'ransomware' or 'exploit') by employing a domain specific stop-word list. After that we use NER filtering. In this step we identify and keep cyber security entities (such as malware "WannaCry" or vulnerabilities "CVE-2021-44228") with the use of spaCy or BERT based NER. We also fix the typos in the text (e.g. "phising" -> "phishing") using Levenshtein distance and cyber security dictionaries. We also filter non-dictionary words (slang, code-like entities) by matching them against our cyber security dictionary. To reduce words to their root forms, we employ stemming and lemmatization techniques using Porter Stemmer and WordNet lemmatizer. Finally, we tokenize the text into individual words and punctuation using the BERT tokenizer, ensuring compatibility with downstream NLP tasks.

3. Feature Extraction: Our feature extraction process involves several techniques to transform the preprocessed text data into meaningful numerical representations. We start by converting tokens to contextual embeddings using BERT-base, which captures semantic relationships between words, such as the distinction between “malware” and “ransomware”. To enable scalable processing, we employ hashing vectorization, transforming text into fixed-length numerical vectors via feature hashing. Additionally, we apply Non-Negative Matrix Factorization (NMF) to reduce dimensionality by decomposing term-document matrices into latent topics, such as “phishing campaigns” and “zero-day exploits”. To visualize threat clusters and anomalies, we use t-SNE to project high-dimensional embeddings into 2D/3D space. Finally, we combine multiple similarity-based features, including semantic similarity using BERT embeddings and cosine similarity, lexical similarity using the Jaccard index on hashtags and keywords, and graph similarity based on user-post interactions from Dark web forums, to create a comprehensive set of hybrid features.

4. Algorithm: HGCN with VAE: We propose a novel architecture that integrates a Heterogeneous Graph Convolutional Network (HGCN) with a Variational Autoencoder (VAE) to model complex relationships in Open-Source Intelligence (OSINT) data. The HGCN is formulated to learn structural relationships between different kinds of nodes- users, posts, keywords, and entities, which are connected with different types of edges-user-post, keyword co-occurrence, and entity relations. With message passing mechanisms along heterogeneous edges, the HGCN can model interactions of various orders, for example user-post-keyword. For transfer learning purposes, we pre-train and initialize the HGCN layers with similar weights from other OSINT datasets. The VAE compresses the node embeddings produced by the HGCN into latent variables and then reconstructs them in order to perform anomalies, such as an outlier post, and also to generate synthetic threat scenarios.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 07, July 2026

By multi-tasking the HGCN and VAE, we can then make use of structural and probabilistic generative information.

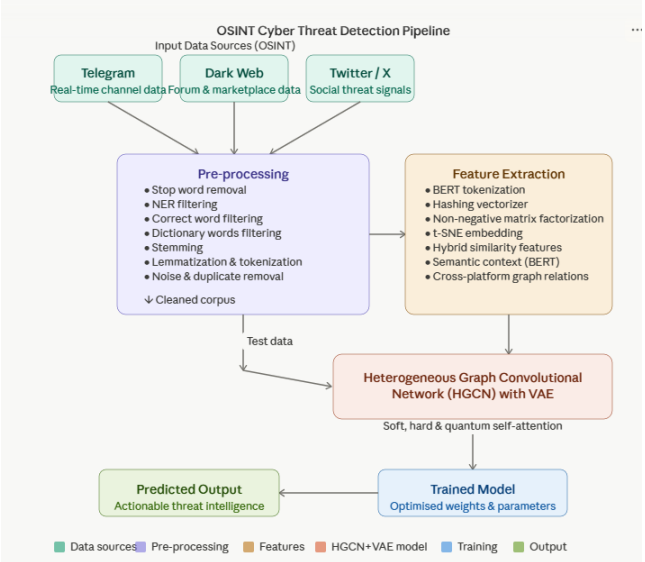

Figure 1. OSINT Cyber Threat Detection Framework

Table 1: Performance Metrics

Metrics/Model	BERT	HGCN+ VAE	BERT+HGCN+VAE
Accuracy (%)	85	88	92
Precision (%)	82	86	90
Recall (%)	89	91	94
F1 Score (%)	83	87	91

V. CONCLUSION

We propose a new framework that leverages BERT, HGCN and VAE for OSINT based cyber threat detection. This system may represent the first such framework that can take OSINT data of heterogeneous nature, make them learnable and detect the threats by leveraging their underlying features. Through the convergence of such state-of-the-art technology, we strive to enable security researchers to achieve better efficiency and effectiveness in identifying and investigating malicious activities in the world and take adequate measures, and set a new paradigm in the field of cybersecurity threat detection.

REFERENCES

[1] Holanda, R. M. (2023). Automated emerging cyber threat identification and profiling based on natural language processing. *IEEE Access*, *11*, 58915–58936. https://doi.org/10.1109/ACCESS.2023.XXXXXXX

[2] Silvestri, S., Islam, S., Papastergiou, S., Tzagkarakis, C., & Ciampi, M. (2023). A machine learning approach for the NLP-based analysis of cyber threats and vulnerabilities of the healthcare ecosystem. *Sensors*, *23*(2), 651. https://doi.org/10.3390/s23020651

[3] Silvestri, S., Islam, S., Amelin, D., Weiler, G., Papastergiou, S., & Ciampi, M. (2024). Cyber threat assessment and management for securing healthcare ecosystems using natural language processing. *International Journal of Information Security*, *23*(1), 31–50. https://doi.org/10.1007/s10207-023-00718-7

[4] [4]. Irshad, E., & Siddiqui, A. B. (2023). Cyber threat attribution using unstructured reports in cyber threat intelligence. *Egyptian Informatics Journal*, *24*(1), 43–59. https://doi.org/10.1016/j.eij.2022.11.002

[5] [5]. Guo, Y., Liu, Z., Huang, C., Wang, N., Min, H., Guo, W., & Liu, J. (2023). A framework for threat intelligence extraction and fusion. *Computers & Security*, *132*, 103371. https://doi.org/10.1016/j.cose.2023.103371

[6] Ferrag, M. A., Ndhlovu, M., Tihanyi, N., Cordeiro, L. C., Debbah, M., Lestable, T., & Thandi, N. S. (2024). Revolutionizing cyber threat detection with large language models: A privacy-preserving BERT-based lightweight model for IoT/IIoT devices. *IEEE Access*, *12*, 1–17.https://doi.org/10.1109/ACCESS.2024.3363469

[7] Coyac-Torres, J. E., Sidorov, G., Aguirre-Anaya, E., & Hernandez-Oregon, G. (2023). Cyberattack detection in social network messages based on convolutional neural networks and NLP techniques. *Machine Learning and Knowledge Extraction*, *5*(3), 1132–1148. https://doi.org/10.3390/make5030060

[8] Wagner, T. D., Mahbub, K., Palomar, E., & Abdallah, A. E. (2019). Cyber threat intelligence sharing: Survey and research directions. *Computers & Security*, *87*, Article 101589. https://doi.org/10.1016/j.cose.2019.101589

[9] Thapa, C., Jang, S. I., Ahmed, M. E., Camtepe, S., Pieprzyk, J., & Nepal, S. (2022). Transformer-based language models for software vulnerability detection. In *Proceedings of the 38th Annual Computer Security Applications Conference (ACSAC 2022)* (pp. 481–496). Association for Computing Machinery. https://doi.org/10.1145/3564625.3567985

[10] Seyyar, Y. E., Yavuz, A. G., & Ünver, H. M. (2022). An attack detection framework based on BERT and deep learning. *IEEE Access*, *10*, 68633–68644. https://doi.org/10.1109/ACCESS.2022.3186484

[11] Seyyar, Y. E., Yavuz, A. G., & Ünver, H. M. (2022). An attack detection framework based on BERT and deep learning. *IEEE Access*, *10*, 68633–68644. https://doi.org/10.1109/ACCESS.2022.3186484

[12] Li, Z. X., Li, Y. J., Liu, Y. W., Liu, C., & Zhou, N. X. (2023). K-CTIAA: Automatic analysis of cyber threat intelligence based on a knowledge graph. *Symmetry*, *15*(2), 337. https://doi.org/10.3390/sym15020337

[13] Piplai, A., Mittal, S., Joshi, A., Finin, T., Holt, J., & Zak, R. (2020). Creating cybersecurity knowledge graphs from malware after action reports. *IEEE Access*, *8*, 211691–211703. https://doi.org/10.1109/ACCESS.2020.3039234

[14] Ma, P., Jiang, B., Lu, Z., Li, N., & Jiang, Z. (2020). Cybersecurity named entity recognition using bidirectional long short-term memory with conditional random fields. *Tsinghua Science and Technology*, *26*(3), 259–265. https://doi.org/10.26599/TST.2019.9010065

[15] Dixit, P., & Silakari, S. (2021). Deep learning algorithms for cybersecurity applications: A technological and status review. *Computer Science Review*, *39*, Article 100317. https://doi.org/10.1016/j.cosrev.2020.100317

[16] Lughbi, H. (2025). *CybAttT dataset* [Data set]. GitHub. https://github.com/HudaLughbi/CybAttT

[17] Govardhan, D., Krishna, G., & Sai, C. (2023). Key challenges and limitations of the OSINT framework in the context of cybersecurity. In *Proceedings of the 2023 International Conference on Electronics, Communication, and Aerospace Applications (ICECAA 2023)*. IEEE. https://doi.org/10.1109/ACCESS.2020.2965257

[18] Dale, D., McClanahan, K., & Li, Q. (2023). AI-based cyber event OSINT via Twitter data. In *Proceedings of the 2023 International Conference on Computing, Networking and Communications (ICNC 2023)*. IEEE. https://doi.org/10.1109/ICNC57223.2023.10074187

[19] Pai, Y., & Prasad, U. K. (2021). Open source intelligence and its applications in next generation cyber security: A literature review. *International Journal of Applied Engineering and Management Letters (IJAEML)*, *5*(2), 1–25. https://doi.org/10.5281/zenodo.5171580

[20] Zhang, H., Liu, B., & Wu, H. (2021). Smart grid cyber-physical attack and defense: A review. *IEEE Access*, *9*, 29641–29659. https://doi.org/10.1109/ACCESS.2021.3058628

[21] Yi-Ting Huang, Chi Yu Lin, Ying-Ren Guo, Kai-Chieh Lo, Yeali S. Sun, and Meng Chang Chen. Open Source Intelligence for Malicious Behavior Discovery and Interpretation.IEEE Tranction, Volume 19, DOI: 10.1109/TDSC.2021.3119008

[22] Rafailă, C., Gurzău, F., Grumăzescu, C., & Bica, I. (2023). MTAFinder: Unified OSINT platform for efficient data gathering. In *Proceedings of the 2023 International Conference on Electronics, Computers and Artificial Intelligence (ECAI 2023)*. IEEE. https://doi.org/10.1109/ECAI58194.2023.10194004

[23] Suryotrisongko, H., Musashi, Y., Tsuneda, A., & Sugitani, K. (2022). Robust botnet DGA detection: Blending XAI and OSINT for cyber threat intelligence sharing. *IEEE Access*, *10*, 43662–43676. https://doi.org/10.1109/ACCESS.2022.3162588

[24] Aleroud, A., Alhussien, N., & Albert, C. (2022). From theory to practice: Towards an OSINT framework to mitigate Arabic social cyberattacks. In *Proceedings of the 2022 International Conference on Intelligent Data Science Technologies and Applications (IDSTA 2022)*. IEEE. https://doi.org/10.1109/IDSTA55301.2022.9923049

[25] Rosa, L., Freitas, M., Mazo, S., Monteiro, E., Cruz, T., & Simões, P. (2019). A comprehensive security analysis of a SCADA protocol: From OSINT to mitigation. *IEEE Access*, *7*, 42156–42176. https://doi.org/10.1109/ACCESS.2019.2906926

[26] Gong, S., Cho, J., & Lee, C. (2018). A reliability comparison method for OSINT validity analysis. *IEEE Transactions on Industrial Informatics*, *14*(12), 5428–5435. https://doi.org/10.1109/TII.2018.2857213Marinho, R., & Holanda, R. (2023). Automated emerging cyber threat identification and profiling based on natural language processing. *IEEE Access*, *11*, 58915–58936. https://doi.org/10.1109/ACCESS.2023.3260020