Enriched Cyber Threat Intelligence Through OSINT: A Methodology for Strengthening Cybersecurity Resilience

Carlos Eduardo de Sousa*[1], Robson de Oliveira Albuquerque[1], and João José Costa Gondim[1]

Departamento de Engenharia Elétrica - Universidade de Brasilia, Brazil
kdusousa@gmail.com, robson@redes.unb.br, gondim@unb.br

Abstract. With the rise of cyber threats, Cyber Threat Intelligence has become crucial for organizations to proactively defend themselves. Open Source Intelligence adds valuable information to Cyber Threat Intelligence, offering additional context and deeper insights. This study proposes a methodology to enrich Cyber Threat Intelligence with Open Source Intelligence. The goal is to address the gap between abundant data and the limited capacity to filter and apply it strategically. The research develops methods to verify, validate, and filter Open Source Intelligence data, enhancing the quality and utility of Cyber Threat Intelligence. A case study was conducted to demonstrate this methodology, focusing on the enrichment of an Indicator of Compromise, specifically an IP address. The results show how Open Source Intelligence can provide vital context to enhance Cyber Threat Intelligence and support better decision-making in cybersecurity.

Keywords: Cyber Resilience, Cyber Threat Intelligence, OSINT

1 Introduction

Cyber Threat Intelligence (CTI) is essential for defending against cyberattacks, helping organizations anticipate and counter emerging threats [1]. As threats grow, it's crucial for companies to implement security strategies based on up-to-date data. However, many private-sector cybersecurity teams are hesitant to share CTI due to concerns about corporate reputation [2].

Open Source Intelligence (OSINT) has rapidly expanded, fueled by technological advances and the availability of digital data. OSINT is now crucial in fields like cybersecurity, law enforcement, and military operations. AI has boosted OSINT's efficiency, enabling real-time data processing for geolocation, sentiment analysis, and risk assessment. In conflicts like the current Russo-Ukrainian war, AI-enhanced OSINT has been used for intelligence purposes, highlighting both opportunities and ethical concerns [3]. Effective OSINT relies on tools like

* This work was supported by the FAPDF TECH LEARNING Grant, "Construção de modelos de linguagem natural para processamento de dados em fontes abertas."

geospatial intelligence (GEOINT), social media intelligence (SOCMINT), and human intelligence (HUMINT), which gather data from sources such as social media, satellite images, and public records. Success depends on selecting appropriate tools, cross-referencing for accuracy, and adhering to ethical standards. This approach enriches CTI by providing essential context [4, 5].

For effective CTI enrichment, OSINT must prioritize validation and verification to avoid irrelevant or invalid data [6],[7, 8]. Screening ensures data relevance and accuracy, structuring it to describe threats and key actors effectively [9]. The "5W3H" method is one approach for this representation [10]. This study addresses a cybersecurity gap, hypothesizing that a structured OSINT methodology enhances CTI quality, better preparing organizations for threats. Testing this aims to validate processes, identify best practices, and guide cyber defense optimization.

This study contributes by proposing a comprehensive methodology for CTI enrichment using OSINT, based on the "5W3H" framework to improve data accuracy and strategic use. It also presents a case study on an Indicator of Compromise (IoC) to demonstrate the methodology's application and identifies challenges and best practices for integrating OSINT with CTI to enhance cybersecurity strategies.

Section II reviews **Related Works**, discussing existing CTI and OSINT integration approaches. Section III presents essential **Definitions**, covering key concepts such as TIP, IoC, IoA, OSINT, and Enrichment. Section IV describes the **Proposed Methodology**, outlining the "5W3H-based" method for OSINT-CTI integration. Section V discusses the **Case Study**, demonstrating the methodology's application to an IoC and evaluating results. Section VI concludes with **Conclusions and Future Work**.

2 Related Works

Several approaches to improving CTI enrichment based on OSINT have been proposed in various studies. In this section, we review the approaches described in these works and compare them with our methodology.

The study [7] discusses the Enhanced Threat Intelligence Platform (ETIP), which enhances Threat Intelligence Platforms (TIPs) in handling OSINT data. ETIP integrates cyber threat information with infrastructure data, improving the analysis, visualization, and sharing of information.

In [11], a system automates the extraction and categorization of CTI using OSINT data. It employs convolutional neural networks and syntactic dependency analysis to identify new Indicators of Compromise (IOCs) with 84% accuracy and classify them with 94% efficiency. This approach enriches CTI data, enhances alert generation, and strengthens security measures through detailed threat analysis.

Complementarily, [6] explores the relevance of enriching CTI data through OSINT, emphasizing the integration of public and restricted sources to understand evolving cyber threats. Using advanced data mining and machine learning,

this approach enhances organizations' ability to predict and counter cyberattacks. The study highlights the need for a holistic approach, prioritizing quality and efficient data integration in CTI enrichment.

On the other hand, [4] explores CTI enrichment with OSINT to enhance security policies in critical infrastructures. The integration of OSINT adds context to CTI, improving risk assessment and threat detection. The proposed architecture uses context and risk-based policies, enabling dynamic security adjustments. This model improves integration between intelligence-sharing platforms and access control systems, addressing limitations in access revocation and strengthening security management against emerging threats.

In this article, [5] examines integrating OSINT into CTI practices, emphasizing the need for structured collection and analysis of open source information to improve cybersecurity. Using a semi-systematic review of the literature and expert interviews, the study proposes a conceptual framework to help organizations proactively and effectively address cyber threats.

Based on the studies presented here, the identified gap relates to the need for a more structured and specific methodology for enriching CTI with OSINT, which offers detailed and efficient integration of multiple data sources and practical and dynamic application in real-time security policy adaptation, which has not yet been deeply addressed.

The proposed methodology addresses this gap by structuring a contextual analysis model to enrich CTI based on a clear and comprehensible methodological approach, facilitating the automation of OSINT collection and analysis through guiding questions that orient the identification, extraction, and categorization of relevant data, and integrating shared intelligence into platforms, allowing organizations to adjust their security policies in real time, promoting a more resilient and proactive approach to threat mitigation.

3 Definitions

We present here some necessary definitions.

- **Threat Intelligence Platform (TIP)**: A Threat Intelligence Platform (TIP) centralizes, analyzes, and shares threat data, using sources for real-time insights and proactive risk management. Tools like TIPCE analyzed Threat Intelligence Sharing Platforms (TISPs), extracting 182K Indicators of Compromise (IoCs) to strengthen cybersecurity [12].
- **Observables and Indicators of Compromise (IoCs)**: Observables are network features like IPs or payload hashes that identify security incidents. When linked to a breach, they become Indicators of Compromise (IoCs), such as malicious file hashes or domains, enabling automated actions like IP blacklisting. Though vital for detection, advanced CTI requires manual analysis, using graph theory to develop complex IoCs for detailed threat assessments [13].
- **Indicators of Attack (IoAs)**: Indicators of Attack (IoAs) focus on TTPs to identify ongoing or imminent threats. Unlike IoCs, which reveal past

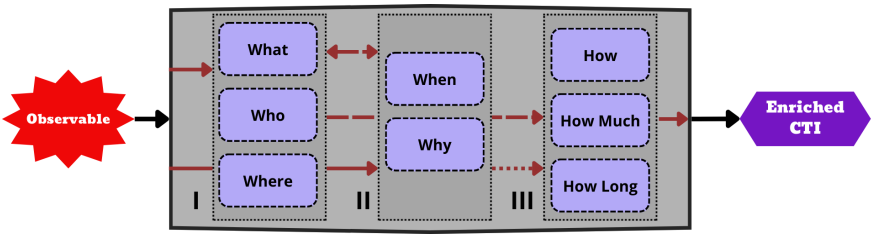

Fig. 1. Enrichment System

breaches, IoAs analyze behavior and context to predict and prevent attacks, enabling real-time threat responses [15].

- **Open Source Intelligence (OSINT)**: Open Source Intelligence (OSINT) collects and analyzes public information to support decision-making. With vast data from the Internet and social media, OSINT relies on core techniques to transform raw data into actionable insights, ensuring effective intelligence gathering despite tool obsolescence [8].
- **Enrichment**: Enrichment adds context and details to raw threat data, enhancing its usefulness for security analysts. It involves integrating socio-organizational data, applying machine learning for classification, or merging sources to improve detection and response. Continuous enrichment ensures accurate CTI and stronger defenses [14].

4 Proposed Methodology

The proposed methodology for enhancing CTI data relies on integrating information from OSINT (Open Source Intelligence) using the 5W3H model as guiding parameters [16]. This approach aims to address critical questions such as: "What" (What is happening, specific threats), "When" (When were the activities detected or occurred), "Where" (Geographic origins or targets), "Who" (involved actors), "Why" (motivations and objectives), "How" (techniques and tools used), "How much" (impact or damage), and "How long" (duration of the threat). This systematic methodology aims to enhance the accuracy and utility of CTI, allowing more effective responses.

These questions are integrated into the CTI enrichment process in a structured, iterative manner, as illustrated in Figure 1. Initially, "What", "Who" and "Where" (I) are prioritized to establish a basic dataset. "When" and "Why" (II) follow, focusing on timeline and motivations. Lastly, "How," "How much," and "How long" (III) emphasize techniques, impact, and duration. This cycle allows continuous validation, ensuring all questions are coherently addressed for a comprehensive threat assessment, enhancing defensive measures.

The "What" focuses on identifying Indicators of Compromise (IoCs), such as malware or phishing attacks, through OSINT tools like forums, security reports, and threat feeds. These sources provide insights into emerging attack

techniques and targeted sectors, helping predict trends and strengthen defenses. The "Where" examines the geographic origin and movement of threats within networks, using IP geolocation and Command and Control (C2) monitoring to trace attacks and map vulnerabilities. Meanwhile, the "Who" delves into threat actors, leveraging OSINT to uncover their identities, affiliations, and motivations, which inform strategies for countering future attacks.

The parameters "When," "Why," "How," "How Much," and "How Long" further enhance threat analysis. "When" tracks the timeline of threats, linking them to external events for predictive insights, while "Why" explores attackers' ideological, political, or economic motivations through OSINT sources like public records and dark web forums. The "How" investigates the methods used by attackers, such as malware and phishing techniques, aiding in the development of targeted defenses. "How Much" quantifies the impact of threats, assessing damage and recovery needs, while "How Long" evaluates their duration and persistence, providing timelines that support effective incident response and risk management strategies.

For the application of the proposed methodology in CTI enrichment using OSINT, relevant and credible open data sources were utilized to provide actionable insights on cyber threats.

The methodology combined various resources for robust data collection and analysis, avoiding over-reliance on specific datasets. Though manual in this case, automating the process through public APIs could enhance scalability in future implementations.

Key sources like VirusTotal and AbuseIPDB provided insights into malware and malicious IP activities, while Censys and the TOR Project highlighted vulnerabilities and anonymization. Tools such as IPInfo, ICANN WHOIS, and Kaspersky's Threat Intelligence Portal added geolocation and threat metadata. Frameworks like MITRE ATT&CK and the Cyber Kill Chain enriched strategic understanding of adversarial tactics.

Unlike automated platforms like ThreatConnect, this methodology emphasizes flexibility and depth over speed, making it accessible to resource-constrained organizations and adaptable to specific threat scenarios.

The proposed methodology offers a structured and flexible approach to enhancing CTI through OSINT, distinguishing itself from automated threat intelligence platforms and OSINT-based tools. Automated platforms like ThreatConnect and IBM X-Force Exchange excel in integrating data sources, automating analysis, and enabling real-time responses, but their high cost and limited customization pose challenges for smaller organizations. In contrast, the proposed methodology emphasizes manual contextual analysis, making it an accessible alternative that provides deeper insights into specific threats, particularly for organizations with constrained budgets.

OSINT tools such as VirusTotal, Shodan, and AbuseIPDB are invaluable for gathering raw data about IoCs, offering cost-effective and diverse sources for analysis. However, these tools often lack contextual frameworks, leading to potential information overload and requiring analysts to invest significant effort

in correlating data manually. By integrating a structured analytical framework, the proposed methodology enriches the raw data provided by these tools, organizing it into actionable intelligence that supports strategic decision-making and enhances the effectiveness of CTI efforts.

5 Case study

In this case study, we apply the proposed methodology to enrich the analysis of an IoC, specifically the IP address X.X.X.X (IP address omitted in compliance with GDPR). This IP was identified in network logs conducting scanning activities, raising concerns about its potential involvement in malicious behavior. While the study begins with a single IoC, the research process inevitably uncovers additional IoCs through related findings. These newly identified IoCs necessitate a retroactive analysis, leveraging the proposed methodology to iteratively expand and refine the enrichment process. By using OSINT, this approach provides deeper insights into the network of IoCs, supporting a more informed and adaptive cybersecurity response.

Table 1 details the application of the methodology to enrich the intelligence surrounding the IP X.X.X.X:

In the proposed methodology, transforming raw data into actionable intelligence requires contextual analysis and effective data filtering to ensure relevance and reliability in cybersecurity decision-making.

Contextual analysis correlated information from OSINT sources like VirusTotal, Shodan, and AlienVault to identify malicious activity linked to the target IP. Tools like GreyNoise and Censys revealed behavioral patterns, such as botnet participation or DDoS involvement, while the MITRE ATT&CK framework and Cyber Kill Chain model mapped the IP to specific attack stages, providing deeper threat insights. Data filtering eliminated irrelevant information by cross-referencing platforms like AbuseIPDB and IPInfo, retaining only corroborated high-risk indicators validated across multiple sources.

The methodology's 5W3H-based approach effectively enriched CTI through recursive searches and iterative analysis. Investigating related files, hashes, and geolocation refined the IP's profile, while connections between domains, motives, and attack methods were established using parameters like "Who," "Why," and "How." This process produced a comprehensive, actionable intelligence profile.

6 Conclusions and Future Work

This study showcased a structured methodology to enhance CTI using OSINT, proving its effectiveness in enriching raw threat data. By applying the 5W3H model, the case study illustrated how OSINT can add valuable context to an IoC, improving decision-making in cybersecurity operations. The methodology was particularly efficient in linking scattered data points, validating sources, and providing actionable intelligence for better threat detection.

Table 1. Threat Intelligence Summary

Aspect	Details
What	**Threat:** Port scanning, Hacking, Exploiting Hosts, Brute-Force and DDoS Attacks activities. **Files:** Unix.Trojan.Mirai-10028259-0, 317477696, nk1, a.bat. **Hashes:** 5e2df75f3b5c4690ca591de9d438a1cbe 2646a9d893b8b6853ddeb928b4cc0b0 **TOR Network:** Exit Node **OS:** Ubuntu Linux 20.04
When	**Created:** 26/01/2011. **First reported:** 13/07/2024. **Last reported:** 16/09/2024*
Where	**Geolocation:** 36.1750,-115.1372. **City - Country:** Las Vegas - EUA. **Associated Address:** 1621 Central Ave, Cheyenne, WY, 82001, United States. **Reported in:** France, Poland, Turkey, USA, and others 10 **Impacted devices:** Netonix Switches. **Target Countries:** Albania, Australia, Austria, Bahrain, Belarus, Belgium, Brazil, Bulgaria, Cambodia, Canada, +50
Who	**Domain:** cs6.top, frantech.ca, meowware.ddns.net. **Hostname:** mails0.lillekarrmaleri.se. **ISP:** Frantech Solutions.
Why	**Motivation:** Economic motivations.
How	**Tools:** Unix.Trojan.Mirai. **TTP:** Enumerates kernel/hardware configuration (Mitre T1082), Reads runtime system information (T1057), Changes its process name (T1036), Reads CPU attributes (T1082), Deletes log files (T1070.001), Enumerates running processes (T1057), Reads hardware information (T1082), Reads network interface configuration (T1016), Deletes Audit logs (T1070.001), Deletes itself (T1070.004), Deletes system logs (T1070.001), Unexpected DNS network traffic destination (T1071.004).
How Much	**Damage:** Nothing found. **Operational disruptions:** Nothing found.
How Long	**Times reported:** 2.655.

Future work will focus on automating parts of the OSINT collection process, incorporating machine learning and large language models (LLMs) to extract insights from feeds and reports, thereby enhancing data categorization and contextualization. Testing the methodology in real-time threat scenarios and expanding its application across different industries will help assess its adaptability, aiming to create a versatile tool for diverse cybersecurity environments. The ultimate goal is to develop a proactive and dynamic intelligence system capable of keeping pace with evolving cyber threats.

References

1. Schlette, D., Caselli, M., Pernul, G.: A Comparative Study on Cyber Threat Intelligence: The Security Incident Response Perspective. IEEE Communications Surveys

and Tutorials. 23(4), 2525–2556 (2021). doi:10.1109/COMST.2021.3117338
2. Security Agency, I.: The President's National Security Telecommunications Advisory Committee Draft NSTAC Report to the President Measuring and Incentivizing the Adoption of Cybersecurity Best Practices. (2021).
3. Ghioni, R., Taddeo, M., Floridi, L.: Open source intelligence and AI: a systematic review of the GELSI literature. AI and Society (2023). doi:10.1007/s00146-023-01628-x
4. Osliak, O., Saracino, A., Martinelli, F., Mori, P.: Cyber threat intelligence for critical infrastructure security. Concurrency and Computation: Practice and Experience. 35(23) (2023). doi:10.1002/cpe.7759
5. Sofie, J., Supervisor, S., Radianti, J.: Unveiling the Potential of Open-Source Intelligence (OSINT) for Enhanced Cybersecurity Posture: A study on OSINT implementation and utilization within organizations and recommendations for increased leverage of OSINT's advantages. (2021).
6. Sun, N., Zhang, Y., Wu, X., Wang, M., Yu, L., Chen, B.: Cyber Threat Intelligence Mining for Proactive Cybersecurity Defense: A Survey and New Perspectives. IEEE Communications Surveys and Tutorials. 25(3), 1748–1774 (2023). doi:10.1109/COMST.2023.3273282
7. González-Granadillo, G., Faiella, M., Medeiros, I., Azevedo, R., González-Zarzosa, S.: ETIP: An Enriched Threat Intelligence Platform for improving OSINT correlation, analysis, visualization and sharing capabilities. Journal of Information Security and Applications. 58, May (2021). doi:10.1016/j.jisa.2020.102715
8. Tanabe, R., de Oliveira Albuquerque, R., da Silva Filho, D., Alves da Silva, D., Costa Gondim, J. J.: OSINT Methods in the Intelligence Cycle. (2021).
9. Tatam, M., Shanmugam, B., Azam, S., Kannoorpatti, K.: A review of threat modelling approaches for APT-style attacks. (2021). doi:10.1016/j.heliyon.2021.e05969
10. da Silva, R. M., Gondim, J. J. C., de O. Albuquerque, R.: Methodology to Improve the Quality of Cyber Threat Intelligence Production Through Open Source Platforms. (2021).
11. Zhao, J., Yan, Q., Li, J., Shao, M., He, Z., Li, B.: TIMiner: Automatically extracting and analyzing categorized cyber threat intelligence from social data. Computers & Security. 95, August (2020). doi:10.1016/j.cose.2020.101867
12. Satvat, K., Gjomemo, R., Venkatakrishnan, V. N.: TIPCE: A Longitudinal Threat Intelligence Platform Comprehensiveness Analysis. In: Proceedings of the Fourteenth ACM Conference on Data and Application Security and Privacy, pp. 349–360. ACM, New York, NY (2024). doi:10.1145/3626232.3653278
13. A Systems Approach to Indicators of Compromise Utilizing Graph Theory. In: 2018 IEEE International Symposium on Technologies for Homeland Security (HST), pp. 1–6. (2018). doi:10.1109/THS.2018.8574187
14. Moriot, C., Lesueur, F., Stouls, N., Valois, F.: How to build socio-organizational information from remote IP addresses to enrich security analysis? In: 2022 IEEE 47th Conference on Local Computer Networks (LCN), pp. 287–290. (2022). doi:10.1109/LCN53696.2022.9843570
15. Hagen, R. A., Helkala, K.: Complexity of Contemporary Indicators of Compromise. (2024).
16. Melo e Silva, A., Gondim, J. J. C., de Oliveira Albuquerque, R., Villalba, L. J. G.: A methodology to evaluate standards and platforms within cyber threat intelligence. Future Internet. 12(6), June (2020). doi:10.3390/fi12060108