Computers & Security 144 (2024) 103972

Contents lists available at ScienceDirect

Computers & Security

journal homepage: www.elsevier.com/locate/cose

Improving quality of indicators of compromise using STIX graphs

Sheng-Shan Chen [a], Ren-Hung Hwang [b], Asad Ali [c], Ying-Dar Lin [d], Yu-Chih Wei [e], Tun-Wen Pai [a,*]

[a] Department of Computer Science and Information Engineering, National Taipei University of Technology, Taipei, Taiwan
[b] College of Artificial Intelligence, National Yang Ming Chiao Tung University, Tainan, Taiwan
[c] National Institute of Cyber Security, Ministry of Digital Affairs, Taipei, Taiwan
[d] Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu, Taiwan
[e] Department of Information and Finance Management, National Taipei University of Technology, Taipei, Taiwan

ARTICLE INFO

Keywords:
Indicators of Compromise (IoC)
Threat Intelligence Platform (TIP)
Cyber Threat Intelligence (CTI)
Structured Threat Information eXpression (STIX)
Open Source INTelligence (OSINT)

ABSTRACT

Cybersecurity relies on Indicators of Compromise (IoCs) to detect and address threats. Although Threat Intelligence Platforms (TIPs) and Open Source Intelligence (OSINT) are common sources for gathering IoCs, their reliability varies. In our study, we enhance the management of IoCs and OSINT by introducing a novel method that reliably assesses IoC's threat severity and confidence scores, focusing on Structured Threat Information eXpression (STIX) for threat associations. Our approach, implemented on OpenCTI, significantly enhances IoC value, as it aggregates threat intelligence from diverse sources utilizing a STIX graph-based approach, which is a unique feature among TIPs. Additionally, our method employs heuristic analysis to optimize IoC scoring. It takes into account factors such as relevance, completeness, timeliness, accuracy, and consistency while emphasizing the confidence of the source. Notably, the proposed method has enhanced the precision of the confidence score, achieving a 25.18% reduction in the average difference of confidence scores compared to the benchmarked platform. The Emotet and Medusa case studies underscore the importance of source credibility in confidence scores, emphasizing our TIP's precision in cybersecurity threat assessment and defense enhancement.

1. Introduction

In recent years, the shift to remote work and a virtualized IT environment, accelerated by COVID-19, has resulted in security breaches and increased vulnerability to cyberattacks in various locations (Mahyoub et al., 2023). The global cost of cybercrime has surged to $8.44 trillion as of 2022 and is projected to reach $23.84 trillion by 2027 (Fleck, 2022). This escalating threat landscape has led to the expansion of the global cyber threat intelligence market. Enterprises are increasingly relying on Cyber Threat Intelligence (CTI) to stay informed about the latest attack trends, enabling them to adapt to evolving threats (Statista, 2023).

CTI encompasses detailed information regarding cyber threats that organizations may encounter (Bandara et al., 2022). It is compiled from a diverse array of sources, including but not limited to Open Source Intelligence (OSINT), internal data, proprietary services, and Information Sharing and Analysis Centers (ISACs). Threat Intelligence Platforms (TIPs) play a crucial role in the aggregation, analysis, and presentation of this information, facilitating the identification of Indicators of Compromise (IoCs) (Azevedo et al., 2019). OSINT involves collecting and processing data from open sources (Hwang et al., 2022), offering advantages such as exploiting available information, low collection costs, and easy data access. However, the credibility of OSINT is questionable, as it can be posted on online resources by anyone or a specific organization. TIPs address this by purchasing commercial intelligence or exploring the dark web for the latest threat intelligence (Connolly et al., 2023). Nevertheless, many TIPs provide threat intelligence that has not been processed or lacks a standardized method for calculating threat indicators, limiting the generation of valuable quality data (Enisa, 2021).

Sharing OSINT is also challenging, as TIPs must use specific standards to expedite the processing and analysis phases of information receipt (Khan and Wallom, 2022). Additionally, a proper quality assessment is necessary to verify the value of collected data for threat intelligence, presenting a significant challenge (Sillaber et al., 2016). Improving the quality of information obtained is crucial, considering the difficulty information security analysts face in sifting through vast amounts of data to find relevant information. Fortunately, the issue of TIP exchange has been addressed with the emergence of Structured Threat Information eXpression (STIX) (OASIS, 2023) and Malware Information Sharing Platform & Threat Sharing (MISP) (Wagner et al., 2016). STIX, a language and serialization format established by OASIS

[*] Corresponding author.
E-mail addresses: rhhwang@nycu.edu.tw (R.-H. Hwang), twp@ntut.edu.tw (T.-W. Pai).

https://doi.org/10.1016/j.cose.2024.103972
Received 9 January 2024; Received in revised form 23 June 2024; Accepted 26 June 2024
Available online 1 July 2024

for exchanging CTI, resolves the problem by classifying each piece of information as a specific object or attribute in its latest version, STIX 2.1. Multiple objects are linked through relationships, enabling quick and complex expression of CTI, including doubts, compromises, and attributions. STIX visually presents all aspects to analysts or other TIPs. While MISP also facilitates threat intelligence exchange, its design primarily supports its specific ecosystem. Given STIX's broader applicability and software-agnostic nature, our focus remains on STIX for its wider utility in threat intelligence exchange.

Given the unclear calculation methods employed by current intelligence platforms in determining IoC threat scores and the varying quality of data sources, there is a significant risk of receiving inaccurate or outdated information. This paper addresses this issue by presenting a method to enhance the IoC quality through a knowledge graph and by establishing a threat score standard. The proposed method enriches intelligence by associating and merging different OSINT sources for a single IoC, following the 18 STIX Domain Objects (SDOs) (OASIS, 2023) dictated by the STIX Heuristic evaluation method. Our study, based on the four dimensions of completeness, accuracy, relevance, and timeliness (Sergio, 2015) and the additional dimension of consistency, allows for an objective evaluation of the IoC threat scores and credibility, thereby improving the quality of threat intelligence. Furthermore, we integrate our method into OpenCTI (Filigran, 2024), an OSINT platform. Our integrated approach enables OpenCTI to collect OSINT and enrich IoC with a docker connector. We evaluated our approach using 232,370 pieces of intelligence, facilitating the creation of enriched IoC and the generation of STIX bundles to identify and calculate scores.

The main contributions as follows:

- We developed a method to improve IoC value using CTI, significantly boosting threat data quality. This enhancement is achieved by integrating a multidimensional analysis framework that incorporates not only the volume and variety of data sources but also the contextual depth and temporal relevance of the information, thereby delivering a more nuanced and comprehensive understanding of cyber threats.
- We introduced the first IoC-specific severity and confidence scoring system, marking a 25.18% average confidence score variation from other platforms.
- We validated our approach within OpenCTI through Emotet and Medusa case studies, proving its effectiveness in creating detailed, actionable IoCs.

2. Background

2.1. Cyber threat intelligence

CTI helps organizations quickly detect and respond to potential cybersecurity threats, offering insights into vulnerabilities, attack methods, malware, threat behaviors, and indicators. This information is valuable for individuals and organizations, aiding in strengthening cybersecurity defenses and responding to threats. The different levels at which CTI operates include tactical, operational, and strategic. Tactical intelligence provides the technical details necessary for defenses against imminent threats, operational intelligence informs the broader context around threat actors and campaigns, and strategic intelligence guides long-term security policies and risk analysis. However, the effectiveness of CTI hinges on four essential qualities: completeness, accuracy, relevance, and timeliness. Completeness ensures CTI covers sufficient threat data to be effectively actionable across potential victims. Accuracy is vital, as the benefits derived from correct threat intelligence must outweigh the costs associated with errors. Relevance ensures the intelligence is directly applicable to the organization's specific threat landscape, allowing for effective countermeasures. Lastly, timeliness dictates that intelligence must be operationalized promptly to outweigh the costs of threat intelligence and mitigate threats efficiently.

However, the quality of threat intelligence varies as different sources contribute to CTI, each with distinct collection, analysis, and sharing methods, resulting in differences in intelligence quality and credibility. Another common issue is the inconsistency in the reliability of sources. Intelligence providers differ in expertise, resources, and technical capabilities, affecting the accuracy and reliability of the intelligence they offer. Accuracy is crucial as intelligence involves vast amounts of information, some of which may need correction, updating, or may be incomplete, impacting the reliability of threat assessment.

To address the above-mentioned issues, the evaluation of the value of threat intelligence is vital. The evaluation involves questions such as the ability to identify attacks, reduction in false alarms, relevance to specific targets, and the time between creating threat events or indicators and recording defensive responses. Establishing a credibility assessment mechanism for different intelligence providers is essential for ensuring reliability. Consistent analysis methods and evaluation standards contribute to the uniformity and comparability of intelligence. Promoting information sharing and collaboration mechanisms facilitates inter-organizational intelligence sharing and timely access to the latest information.

2.2. Indicators of compromise

Indicators of Compromise (IoCs) are crucial for sharing and utilizing threat intelligence, helping identify signs of system or network threats. Specific indicators like IP addresses, domain names, file hash values, and file metadata play a vital role in this process.

- IP addresses: These are often associated with malicious activities, and unusual IP connections may indicate an ongoing or attempted breach.
- Domain names: Cyber-attackers commonly use domain names for phishing, malware distribution, or command and control (C2) communication.
- URLs: Similar to domain names, malicious URLs are frequently employed in phishing attacks or for delivering malware.
- File hash: The hash values of malicious files, such as malware or other harmful payloads, are often shared as IoCs. Comparing file hashes helps organizations identify known threats.

IoCs allow the detection of threat behaviors by analyzing and comparing various types of data, allowing for tracking and identifying attacker activities. These IoCs bolster the security of the entire ecosystem and enhance individual defensive capabilities when they are shared among organizations and security communities. Sharing IoCs enables organizations to quickly gain insights into emerging threats and attacks, supporting the development and implementation of proactive security measures to reduce potential damage and risks.

2.3. Threat intelligence platform

One way to share IoCs is through a Threat Intelligence Platform (TIP). A TIP acts as a central hub where organizations and security communities can exchange, share, and store information about IoCs. These platforms offer a secure and standardized environment for effective sharing and obtaining of the latest threat information. Organizations can upload their IoCs to the TIP to share them with other organizations. Simultaneously, they can also check IoCs provided by other users on the platform to learn about new threats and attacks. This sharing and exchange mechanism helps organizations quickly adapt to the ever-changing threat landscape.

TIPs also come with features and tools to support IoCs management and analysis. These include indexing and searching for threat intelligence, adhering to information exchange standards, and providing visualization and reporting capabilities. These features help organizations better organize, analyze, and apply the IoCs information they receive,

Table 1
Comparative analysis of CTI quality evaluation across different studies.

Studies	Input	Output	Objective		Method	IoC Standard				
			Identification & correlation	Trust establishment		Completeness	Accuracy	Relevance	Timeliness	Consistency
Sillaber et al. (2016)	CTI	Relevance score		✓	Regression & Word embedding			✓		
Schlette et al. (2021)	CTI	Quality score	✓		Data quality	✓	✓			
Zhang et al. (2022)	CTI	Quality assessment	✓		MITRE ATT&CK correlation				✓	
Schaberreiter et al. (2019)	CTI	CTI Source priority		✓	Quantitative params	✓	✓		✓	
Ours	IoC	Severity & Confidence score	✓	✓	Heuristic evaluation	✓	✓	✓	✓	✓

Table 2
Comparative analysis of CTI quality improvement approaches across different studies.

Studies	Input	Output	Objective		Method	IoC standard				
			Accelerated response	Enrichment		Completeness	Accuracy	Relevance	Timeliness	Consistency
Meier et al. (2018)	CTI	Rank	✓		Page rank	✓	✓	✓		
Azevedo et al. (2019)	CTI	Classification	✓		Clustering	✓	✓	✓	✓	
Gao et al. (2020)	CTI	Classification	✓		GCN			✓		
Gonzalez-Granadillo et al. (2021)	CTI	CTI report		✓	Heuristic analysis	✓		✓	✓	
Ours	CTI	STIX graph	✓	✓	Heuristic scoring	✓	✓	✓	✓	✓

enhancing their cybersecurity capabilities and efficiency. Examples of such platforms include MISP, which is widely used to share, store, and correlate Indicators of Compromise of targeted attacks. OpenCTI is an open-source platform that allows organizations to manage their knowledge and observables about cyber threat intelligence (Filigran, 2024). These platforms exemplify the diverse ecosystem of TIPs available to organizations looking to improve their cybersecurity postures through collaborative intelligence sharing.

However, some challenges must be addressed to share IoCs on TIPs successfully. Ensuring the quality and credibility of shared IoC information is one such challenge. OpenCTI, the open-source TIP used in our experimental environment, offers fields for rating the author's reliability and the confidence level of each CTI object to assist with this challenge (OpenCTI, 2024). Despite the availability of these features, our research sought to explore beyond the platform's default mechanisms. We aimed to develop and validate a novel set of criteria and mechanisms for the quality assessment of IoCs, complementing and potentially enhancing the existing capabilities of OpenCTI. Thus, while acknowledging the value of OpenCTI's built-in assessments, our study deliberately did not utilize these fields; instead, it focused on our independent methodological contribution to the field of CTI quality assessment.

3. Related work

In recent years, there has been a significant push to share information about threats, network attacks, and incidents among organizations. OSINT serves as the primary source for this shared data, enabling even small organizations to detect complex attacks without extensive, in-depth investigations. TIPs like MISP (Wagner et al., 2016) and OpenCTI play a key role in collecting, storing, and spreading these data. These platforms use protocols such as STIX (Bandara et al., 2022), TAXII (Connolly et al., 2014), and OpenIoC (Obrst et al., 2012) to enable intelligence sharing between platforms. Despite these improvements, challenges related to outdated or incomplete threat intelligence still persist.

3.1. Evaluating threat intelligence

The recent literature focuses on providing high-quality threat intelligence by addressing these challenges. The literature reveals a dual approach: one aspect focuses on the evaluation of threats, while the other aims to enhance the quality of intelligence. For example, Serrano et al. stress the importance of measurable quality control in TIPs (Serrano et al., 2014). Although they propose a solution, it faces practical implementation difficulties. Schlette et al. suggest a Data Quality (DQ) method to measure and visually represent the quality of threat intelligence (Schlette et al., 2021). However, this method requires expertise since it heavily relies on specific data sources.

Table 1 offers a detailed comparative analysis of various research efforts aimed at assessing the quality of CTI. These studies underline the essential task of scrutinizing network threat intelligence to address its intrinsic challenges effectively. Unlike prior methodologies that might focus on a subset of quality factors, such as completeness, accuracy, relevance, and timeliness. Our approach stands out by equally weighing all these elements, including consistency. This equal consideration ensures a comprehensive evaluation framework that improves upon previous methodologies by providing a nuanced and balanced assessment of CTI quality. Our research contributes to the domain by advocating for a data-driven, dynamically adjustable evaluation strategy that equally values each quality factor. This approach facilitates the development of more accurate and reliable threat intelligence evaluation metrics.

Table 3
Notations.

Category	Symbol	Description
Severity score	SS	Overall of severity score
	SS^A	Accuracy
	SS^R	Relevance
	SS^T	Timeliness
	SS^{CP}	Completeness
	SS^{CS}	Consistency
Confidence score	CS	Overall of confidence score
	CS^{SR}	Source rank
	CS^{SC}	Similarity score

3.2. Enhancing threat intelligence quality

In another study, Zhang et al. introduce a technique for autonomously evaluating sparse threat intelligence nuances (Zhang et al., 2022). They combine this with ATT&CK to identify attack techniques related to IoCs. Schaberreiter et al. expand on this with a parameter-based technique assessing the credibility and overall quality of threat intelligence from the network (Schaberreiter et al., 2019). Their methodology considers aspects like data timeliness, completeness, and reliability, incorporating the SVM machine learning algorithm to discern the confidence score of threat intelligence sources.

In enhancing intelligence credibility, Meier et al. propose FeedRank, a unique ranking system for threat intelligence sources (Meier et al., 2018). This method determines correlations in time and space without relying on baseline facts or operator input. Additionally, a system based on OSINT has been developed to enrich threat intelligence by correlating and amalgamating intelligence through two distinct measures of similarity. Gao et al. integrate various nodes in the infrastructure and their interrelationships using heterogeneous information networks (HinCTI), refining the modeling and analysis of threat intelligence (Gao et al., 2020). Azevedo et al. emphasize the importance of OSINT in improving the security of an organization's network, suggesting a method that uses similarity metrics and associated techniques to consolidate and correlate IoCs (Azevedo et al., 2019). The effectiveness of their method is further validated through experimental evaluations.

In Table 2, we investigate the methodologies used to improve the quality of CTI. This analysis explores various strategies researchers employ to refine the efficacy and reliability of threat intelligence. Each study introduces unique methods for increasing intelligence quality, often focusing on one or two key aspects, such as enrichment or accelerated response mechanisms. Our method, as depicted in our research, uniquely encompasses a holistic view by considering these factors and enhancing the accuracy of an IoC's severity score alongside confidence assessments. Our work includes integrating this comprehensive evaluation and enhancement system with the OpenCTI platform, aiming for practical application and validation.

4. Problem statement

We collect multiple IoCs from various sources to improve their quality and provide an evaluation of this quality, along with the calculation of severity and confidence scores for these IoCs. The problem is stated as given multiple IoCs, we aim to effectively map them into an undirected graph and enhance the accuracy and relevance of threat intelligence score calculations of these IoCs, thus improving the quality of the IoCs. However, our efforts to enhance accuracy and relevance are constrained by the requirements of data integrity and consistency. We divide the problem into three sub-problems. In Table 3, we list the notations used in this paper along with their descriptions. For Severity Score, we consider five factors: Accuracy (SS^A), Relevance (SS^R), Timeliness (SS^T), Completeness (SS^{CP}), and Consistency (SS^{CS}). For Confidence Score, we consider Source Rank (CS^{SR}) and Similarity Score (CS^{SC}).

Data Collection, Extraction, and Normalization

We collect and process IoCs from various OSINT sources to ensure the accuracy and alignment of the processed data with the STIX standard.

- **Inputs:** IoCs from OSINT sources.
- **Outputs:** STIX objects, ensuring precise mapping and standardization of IoC attributes.
- **Objective:** Maximize the accuracy of data extraction and normalization.
- **Constraints:** Ensure data integrity and consistency during extraction and normalization.

Enrichment and Integration

In sub-problem 2, we aim to enhance the initial data, incorporating user-side IoC and ensuring a high correlation between threat intelligence components.

- **Inputs:** IoCs and STIX objects.
- **Outputs:** STIX graph including heuristic components.
- **Objective:** Enhance the correlation and clustering potential of IoC reports.
- **Constraints:** Ensure data relevance and prevent redundancy in the enrichment phase.

Heuristic Severity Score Agent

Our objective in sub-problem 3 is to assign a standardized severity score to each threat, providing stakeholders with a clear and consistent metric to gauge potential threats.

- **Inputs:** STIX graph.
- **Outputs:** Severity Score and Confidence Score.
- **Objective:** Standardize and optimize the quality of the severity score using SS^A, SS^R, SS^T, SS^{CP}, and SS^{CS}.
- **Constraints:** Prevent overfitting or undervaluing any particular metric and ensure uniformity in score calculation.

5. Solution approach

To enhance the quality of IoCs, we need to collect them through OSINT. We use collected IoCs as primary input data and pass them through a system, as shown in Fig. 1, where they undergo multiple phases of collection, analysis, and enrichment. This iterative process aims to enhance the IoC quality, along with the provision of severity and confidence scores, while uncovering hidden relationships and potential attack paths, especially those concealed within intricate relational networks. Here we explain the details of the proposed system architecture.

5.1. System architecture

The main goal of the proposed system is to calculate the severity and confidence score of IoC. Fig. 1 illustrates the overall system architecture, which can be divided into five stages: (1) Data collection, where the challenge of extensive data collection from various OSINT sources is addressed through OpenCTI connectors, laying the foundation for a comprehensive threat intelligence database; (2) Normalization, ensuring a uniform data format for efficient storage and structuring within the OpenCTI graph database, facilitating intricate relationship mapping between cyber entities; (3) Enrichment, where selected IoCs are refined for depth and quality using enrichment connectors, coupled with deduplication and relationship building to weed out redundancies and elucidate the cyber threat landscape; (4) Heuristic scoring, applying a multifaceted scoring system that evaluates IOCs against key indicators:

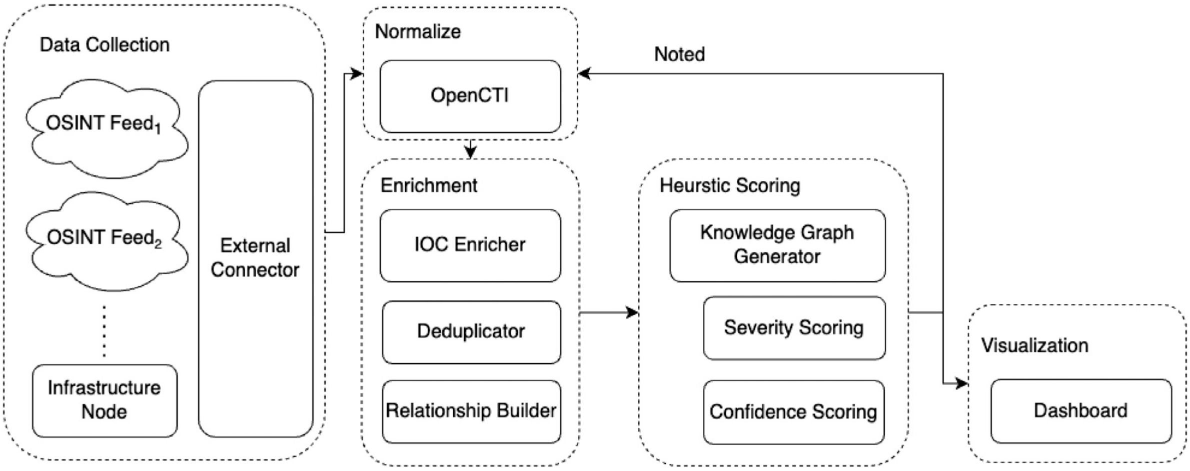


Fig. 1. System architecture.

timeliness, accuracy, completeness, relevance and consistency, providing a nuanced assessment of threat intelligence; and (5) Visualization, the culmination stage that presents the analyzed and scored IoCs in an intuitive visual format, enabling stakeholders to easily understand complex threat relationships and make informed decisions. This section offers an overview of the platform's methodology and the distinctive contributions of each stage of the system architecture, highlighting the integrated approach to enhancing cybersecurity threat assessment.

5.1.1. Data collection

This stage addresses the challenge of collecting and extracting threat intelligence data from various OSINT sources. This stage has a data collection module to collect threat intelligence from multiple cyber threat intelligence providers and categorize them according to function. To promote efficiency and ease of data handling, this module is deployed using Docker and collects data using the following connectors developed by OpenCTI:

- External Import Connectors play a pivotal role in acquiring external intelligence information, which is instrumental for the comprehensive assessment of cybersecurity threats. These connectors coordinate with APIs of other platforms, utilize web crawlers, and employ the Pycti tool developed by OpenCTI to gather a broad spectrum of threat intelligence. Table 4 lists the external connectors, specialized IoC types, and corresponding versions. These settings demonstrate our system's capability to integrate and analyze various forms of threat data, thereby enhancing the overall effectiveness of our threat intelligence platform. Citations corresponding to the sources of IoCs include IPs (AbuseIPDB, 2024), domains, URLs (AlienVault, 2024; VXVault, 2024), file hashes (Abuse.ch, 2024), and CVEs (MITRE, 1999; CISA.gov, 2019).
- Internal Enrichment Connectors are employed to enhance the specificity and reliability of IoC data. These connectors play distinct roles: (1) AbuseIPDB assesses the reputation of IP addresses; (2) VirusTotal conducts comprehensive malware analysis by aggregating data from multiple antivirus engines and databases. Table 5 shows connectors and their functionalities (AbuseIPDB, 2024; VirusTotal, 2024).

5.1.2. Normalization

After collection, the data undergo normalization, via the normalization module, to ensure uniform storage in the OpenCTI database. OpenCTI's graph database storage model, using nodes and edges, is leveraged for its capability to structure, enrich, and represent intricate relationships between cyber entities. The severity scores computed at this phase replace the platform's severity score.

5.1.3. Enrichment

The enrichment stage is composed of three modules: (1) IoC Enricher, (2) Deduplicator, and (3) Relationship Builder.

- IoC Enricher: Due to the vast number of IoCs available, it is impractical to enrich every single IoC. Therefore, our system employs a focused approach, enhancing IoCs actively searched for or queried within the system. Data retrieval processes facilitate this prioritization via APIs, which fetch JSON-formatted data specific to the IoCs under investigation. The enrichment process then utilizes the returned results to augment the IoCs' information, emphasizing the determination of an IoC's maliciousness or the results from antivirus software recognition. These selected IoCs are enriched through connectors, as illustrated in Table 5.
- Deduplicator: To address potential duplication in intelligence collected by OpenCTI, we have referred to the preprocessing steps before handling similarity models (Li et al., 2024), and then instituted the following deduplication steps: (A) Stopword removal, (B) Lemmatization, (C) Weight calculation using Term Frequency–Inverse Document Frequency (TF–IDF), (D) Assessment of data similarity with Cosine Similarity, chosen for its effectiveness in identifying conceptual similarities in content post-preprocessing and its computational efficiency for large-scale datasets. In reviewing the deduplication process for our dataset, a random sample of approximately 9000 records was analyzed, as illustrated in Fig. 2. We have established a similarity threshold of 0.9. This high threshold is chosen to reflect the nature of CTIs that often exhibit only slight variations, such as changes in IP addresses, particularly for those generated from honeypots. The graph indicates a marked increase in the number of similar document pairs identified as duplicates when the threshold is below 0.9, justifying our selection of this threshold to ensure that nearly identical documents are targeted while preserving the granularity of our intelligence data.
- Relationship Builder: Our system employs an automated rule set to parse and process threat intelligence. One such rule is the "Attribution via Attribution" rule which can be understood with an example i.e., If Entity A is attributed to Entity B, and Entity B is attributed to Entity C, then a derived association is established, resulting in Entity A being attributed to Entity C. When these rules are met, the designated actions are automatically triggered, or pertinent events are initiated.

5.1.4. Heuristic scoring

A holistic evaluation of threat intelligence necessitates a comprehensive scoring mechanism. The Heuristic Scoring methodology offers

Table 4
External import connectors and corresponding IoC types.

Connector	IoC type (s)	OpenCTI version
Alienvault	IPs, Domains, URLs, File Hashes	5.3.17
Common vulnerabilities and exposures	CVEs	5.3.17
CISA known exploited vulnerabilities	CVEs	5.3.17
AbuseIPDB IP blacklist	IPs	5.3.17
Abuse.ch SSL blacklist	SSL certificate identifiers	5.3.17
MITRE datasets	Attack techniques, Tactics	5.3.17
MalwareBazaar recent additions	Malware hashes	5.3.17
VX vault URL list	URLs	5.3.17

Table 5
Internal enrichment connectors and their functions.

Connector	Functionality	Version
AbuseIPDB	IP address reputation check	5.3.17
VirusTotal	Multisource malware analysis	5.3.17

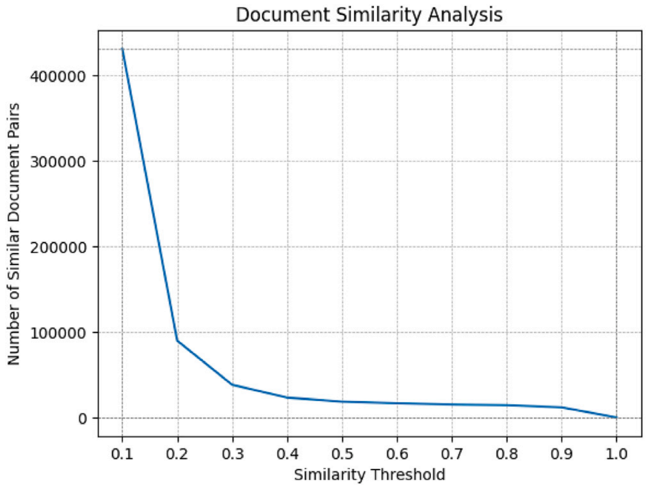


Fig. 2. Similarity comparison of CTI tested by dedupicator.

a well-structured assessment system, utilizing five key indicators: Timeliness, Accuracy, Completeness, Relevance, and Consistency. These indicators cover various aspects of threat intelligence and are vital in understanding their significance in the realm of cybersecurity. Table 7 provides a systematic breakdown of the scoring for each key indicator. By establishing distinct criteria for each score and indicator, this approach offers a multidimensional perspective on the assessment of threat intelligence. Such a scoring system is crucial for ensuring a nuanced and well-informed evaluation. In Section 5.2, we delve into a detailed explanation of this methodology.

5.1.5. Visualization

Visualization is the last stage of our proposed system, designed to provide a detailed visualization of IoCs and their interrelationships and assess their severity. This component makes complex relationships, and varying degrees of threat severity articulated in a structured format for easy comprehension.

5.2. Severity and confidence score calculation

Following are the steps involved in the calculation of severity and confidence scores of IoCs.

Step 1: Establishing standard scoring criteria. The first step is to have a well-defined scoring mechanism. This is depicted in Table 7, where each of the five indicators has a specific scoring system based on predefined criteria.

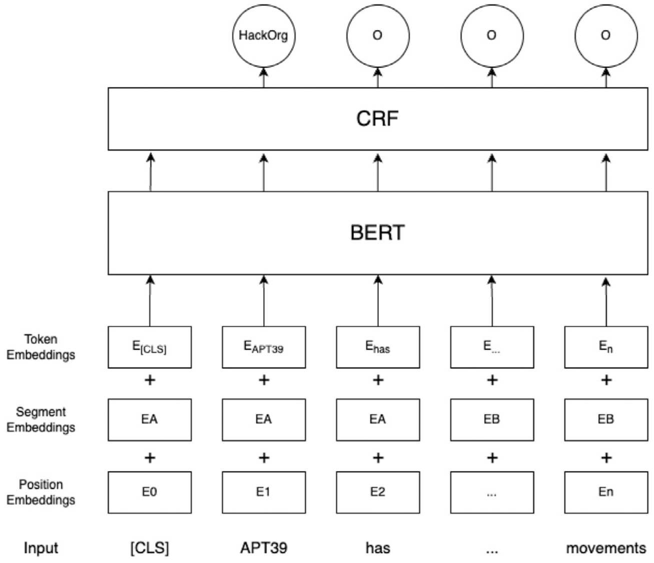


Fig. 3. Architecture of the BERT-CRF model used for relevance standard.

Step 2: In-depth indicator analysis. Each of the five indicators is elaborated. For **Timeliness**, the difference between the “first seen” and “last seen” of each STIX object is analyzed. As mentioned above, this aspect helps various cybersecurity roles differently, emphasizing the need for a personalized approach. **Accuracy** is measured by considering the diversity of external references. The multitude of sources leads to a more comprehensive and reliable threat analysis. For **Completeness**, the STIX 2.1 standard is the primary metric to ensure that the entire trajectory of an IoC is covered, offering a comprehensive view. **Relevance** of threat intelligence is efficiently evaluated using the advanced BERT-CRF model (Chen et al., 2023), which identifies 11 crucial entities such as hacker organizations, offensive actions, sample files, security teams, tools, time and purpose of attack, geographic area, industry, and vulnerabilities. Previous research shows that we use BERT-CRF to achieve good results in threat entities, so we continue to use the previously used model as shown in Fig. 3. BERT (Bidirectional Encoder Representations from Transformers) employs self-supervised learning to generate deep word embeddings from a vast corpus of English text (Vaswani et al., 2017), enhancing word sense disambiguation. It utilizes Masked Language Modeling (MLM) and Next Sentence Prediction (NSP) for pretraining: MLM masks 15% of words in each sentence to predict them based on context, while NSP evaluates the continuity of the sentence sequence. This allows BERT to integrate contextual information from both preceding and following text, improving linguistic accuracy. We used the BERT-base-uncased variant, incorporating 12 transformer blocks and 768 hidden layers with 12 attention heads, totaling 110 million parameters. BERT alone does not suffice for sequence classification; thus, we integrate a Conditional Random Field (CRF) layer post-BERT to model global dependencies and sequence dynamics. The CRF layer learns to maximize the joint probability of label sequences given the

BERT-derived feature vectors H. The CRF function is formalized in Eq. (1). as

$$p(Y \mid H) = \frac{1}{Z(H)} \exp\left(\sum_{i=1}^{n}\sum_{j=1}^{k} \lambda_j f_j \left(y_i, y_{i-1}, H, i\right)\right), \tag{1}$$

where $Z(H)$ normalizes the probabilities to ensure they sum to one, λ_j are feature weights, and f_j are feature functions capturing the patterns in the sequence. This combination ensures precise entity tagging by selecting the label sequence Y that maximizes the probability $p(Y|H)$.

To ensure a nuanced and precise evaluation of each piece's relevance to specific security contexts, the scoring method adopts an averaging approach across these entities, with scores ranging from 0 to 5. This comprehensive strategy guarantees a detailed assessment tailored to the unique demands of diverse security scenarios. Finally, **Consistency** is measured through interactions with external antivirus software interfaces, where their evaluations serve as third-party validations. It is An adaptive, average-based scoring method that accounts for variability in antivirus support and addresses the diversity in IoC types. The importance of these key indicators is explained as follows:

- Timeliness: Threats are evolving rapidly, and knowing a threat at an early stage helps prioritize and deploy resources efficiently. It is the difference between being proactive and reactive in many scenarios.
- Accuracy: A false positive can be as harmful as a missed threat. Hence, verifying threat intelligence against multiple sources ensures higher reliability.
- Completeness: A holistic view of threats is more actionable. With the STIX standard, one can fully understand the threat landscape.
- Relevance: The flood of cyber threats makes distinguishing the crucial from the trivial a paramount task. To maximize the impact of cybersecurity efforts by focusing on threats directly relevant to the organization's specific context.
- Consistency: The reliability of threat intelligence hinges on its consistency, especially given the dynamic nature of cyber threats. Regular validation against multiple antivirus software interfaces provides a solid foundation for this consistency, acting as a crucial third-party check. This ensures that our threat assessments remain steady and reliable across varying conditions and over time, establishing a trusted baseline for cybersecurity defenses.

Step 3: Normalization of the severity score. After scoring each IoC, it is imperative to normalize the results to ensure a unified and intuitive scoring system. The Severity Score quantifies the potential harm an IoC might inflict on an organization's assets. To dynamically optimize model performance, we adopted a method that adjusts the weights of each feature based on their correlation with the target variable, the 'Severity Score'. We calculated the Pearson correlation coefficient for each feature with the severity score, taking absolute values to account for both positive and negative correlations. These coefficients were then normalized so that the sum of all feature weights equals one, thereby ensuring that each feature's influence on the target variable is proportionally adjusted based on its explanatory power. The overall Severity Score (SS) is calculated using the weighted average of the Evaluated Scores in Eq. (2) as

$$\begin{aligned} SS = {} & w_A \cdot SS^A + w_R \cdot SS^R + w_T \cdot SS^T \\ & + w_{CP} \cdot SS^{CP} + w_{CS} \cdot SS^{CS}, \end{aligned} \tag{2}$$

where weights w_A, w_R, w_T, w_{CP}, and w_{CS} are assigned to each threat category based on their respective importance, which are derived from empirical correlation analysis. This refined approach for the calculation of SS enables a comprehensive and equitable evaluation of various threats, ensuring that all dimensions of threat intelligence are appropriately considered in the total SS. The computed feature weights were assigned as follows: Consistency 0.40, indicating its significant impact on the SS; Relevance 0.36, reflecting its substantial influence within the model; Completeness 0.12; Timeliness 0.09 and Accuracy 0.03 in this study.

Table 6
Source ranking and score distribution of CTI source.

CTI Source	Connections	Source rank
FireEye	120 932	100.00
Palo Alto networks	119 699	98.97
ESET	98 748	81.62
Kaspersky	82 608	68.24
Trend micro	76 456	63.15
Symantec	71 560	59.09
Cisco talos	52 854	43.59
McAfee	35 977	29.61
Microsoft	28 161	23.14
US-CERT	26 841	22.04
AlienVault	7615	6.11
Dragos	229	2.46

Step 4: Assessing confidence scores and visualizing IoCs. The confidence score is the trustworthiness of an IoC, encompassing both the accuracy of the CTI and the reliability of its source. To determine this score, we employ two distinct mechanisms: Source Rank and Similarity Score. This holistic approach integrates the authenticity of the source, the corroboration of information between platforms, and the consensus among detection tools. Specifically, our Source Rank method improves the FeedRank mechanism (Meier et al., 2018), to evaluate the reliability of information by examining the reputation of its source. This tailored approach offers a structured methodology to systematically assess each IoC, ensuring a thorough review of its origin, supporting evidence, and consistency across detection tools.

The Source Rank is computed through a well-orchestrated mechanism involving OpenCTI data snapshots, directed graphs based on IoCs, and their timestamps (Eqs. (3), (4), (5)). This detailed approach yields a refined ranking, reflective of the actual threat level. To calculate source rank, first, we create a directed graph and add edges to the directed graph based on IoC and timestamps in the CTI feeds. If feed1 reports the same IoC earlier than feed2 in time, an edge is added between feed1 and feed2, which is represented by Eq. (3) as,

$$A_{ij} = \begin{cases} 1, & \text{if feed}_i \text{ reports an IoC earlier than feed}_j \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where A_{ij} indicates whether there is an edge between nodes i and j in the directed graph. Next, we calculate the PageRank value of each node, which is illustrated in the Eq. (4) as,

$$PR(i) = (1-d) + d \sum_{j \in M(i)} \frac{PR(j)}{L(j)}, \tag{4}$$

where $PR(i)$ is the value of node i, and d is a damping factor, that is generally set to 0.85. $M(i)$ is the set of all nodes pointing to node i and $L(j)$ is the outside degree of node j. Finally, we calculate the source rank value of each node which is shown in Eq. (5) as,

$$CS_i^{SR} = \sum_{j \in M(i)} w_{ij}, \tag{5}$$

where CS_i^{SR} is the source rank value of node i, w_{ij} is the weight of edge (i, j). We set the weight as the number of IoCs in the feed j that are reported later than the feed i. Finally, we sort all CTI feeds according to their source rank values, thus obtaining the final CTI feed ranking results, as shown in Table 6.

The CTIs collected demonstrate that FireEye intelligence has the highest number of connections, indicating that utilizing intelligence from this provider results in the highest Source Rank score. This ranking system is used to identify and prioritize high-quality intelligence sources, reflecting the substantial use and integration of FireEye data in our analysis. The ability to filter and rank intelligence providers based on their connectivity and utilization ensures that our methodology reliably assesses the most influential and credible sources within the cybersecurity landscape.

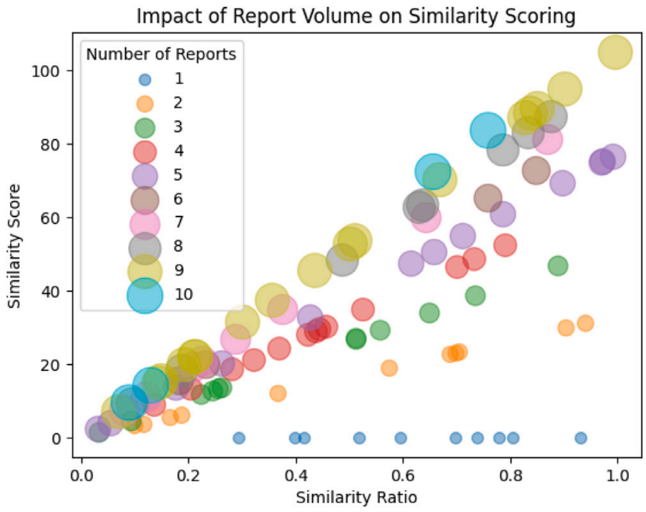


Fig. 4. Score plot of the number of reports per CTI combined with the average similarity.

For the similarity score (CS^{SC}), we devised a dynamic scoring model (Eq. (6)), which maintains an equilibrium in the confidence score, even as the number of reports grows. As shown in Eq. (6), the similarity score of an IoC with respect to its related reports is calculated as

$$CS^{SC} = \frac{\left(\sum_{i=1}^{n} \mathrm{SC}i^2\right) / (n \cdot \log_2(n))}{3} \times 100, \tag{6}$$

where SC represents the similarity score of the ith report, obtained by calculating TF–IDF vectors and cosine similarity to measure the semantic similarity, and n represents the total number of reports. The proposed equation effectively calculates the score for an IoC, considering the similarity scores of associated reports and the total number of these reports. Specifically, the sum of squared similarity scores represents the cumulative similarity between reports. Normalization of the score is achieved by dividing this accumulated value by the total number of reports and subsequently taking the logarithm to base 2, as depicted in Fig. 4. It illustrates the impact of report volume on the calculated CS^{SC}, which is shown on the y-axis, while the x-axis indicates the similarity ratio. This visualization highlights the relationship between the number of reports and the corresponding efficiency scores, demonstrating that as the volume of reports increases, the influence of highly similar reports on the CS^{SC} gradually decreases. This graphical representation aligns with the decreasing function introduced in the equation to maintain a balanced and reliable calculation of scores.

This addresses the influence of the volume of reports and allows each report to have a proportionately increased impact with the increment of total reports. A decreasing function is also introduced to mitigate the potential overemphasis of high similarity on the final score, ensuring a balanced and reliable calculation of scores, so that, with increasing report volume, the influence of highly similar reports gradually decreases. The final score is then scaled by multiplying the result by 100/3 to mitigate the skew. This factor is rooted in the characteristic of our collected intelligence, where three sources, on average, support each IoC. We align the scoring system with the observed average, providing an enhanced confidence measure for reports originating from multiple similar yet diverse sources while circumventing unnecessary distortion effects caused by a high report count. Moreover, this scaling factor is adaptable and can be adjusted to accommodate changes in the average number of sources per IoC as our dataset grows or evolves. The final Confidence Score is integrated with a weighting ratio of 70% for Source Rank and 30% Similarity, providing a nuanced insight into the threat landscape.

6. Implementation

This section introduces the implementation of STIX heuristic scoring system, along with the data sources, tools, and datasets used in the experimentation.

6.1. STIX heuristic scoring system

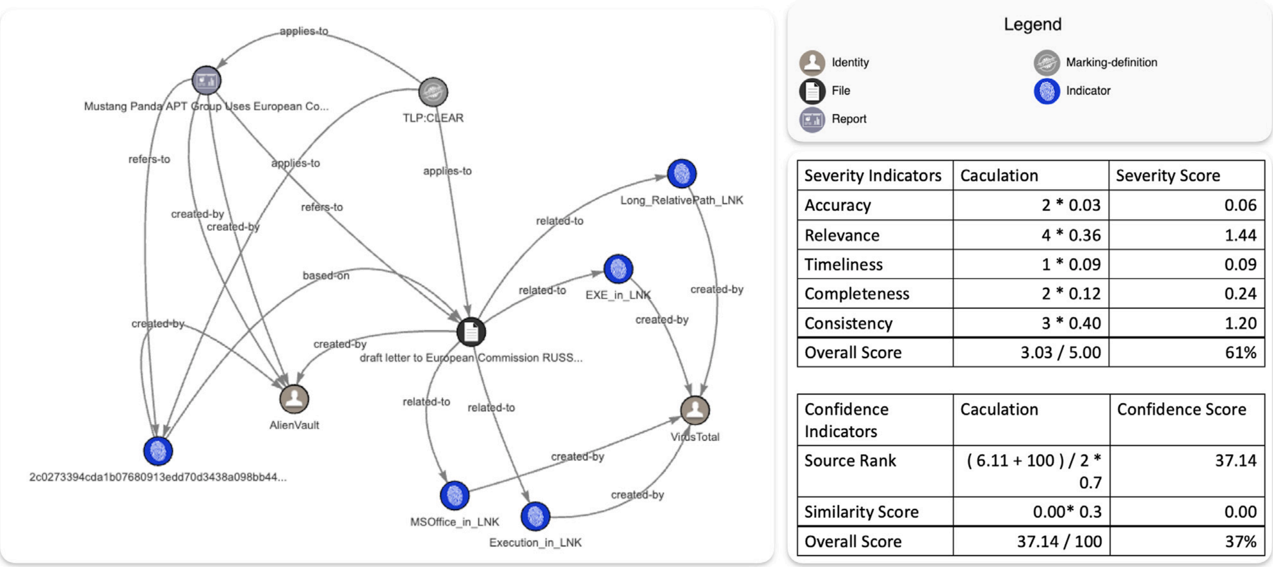
To construct the STIX computing back-end and visual interface for IoC quality enhancement, our system takes advantage of OpenCTI's existing functions and is built upon the pycti library (OpenCTI version 5.3.17), a foundational component of OpenCTI, hence adopting its implementation approach. The system operates in a hardware environment featuring an 11th Gen Intel(R) Core (TM) i7-11700 @ 2.50 GHZ CPU, 48 GB of memory, an Ubuntu 20.04 x64 operating system, and an NVIDIA RTX A5000 graphics card. The system offers two input interfaces: Command-line mode and Web page mode, allowing users a choice based on their needs. Subsequently, according to the process described in the proposal, the system initiates the collection and integration of data related to IoCs, removes extraneous attributes, and proceeds with the scoring computation. Users can view the final heuristic score for each scoring indicator using the Web page interface. An enhancement to visualization functionality has been applied to the oasis-open cti-stix visualization project, with the results shown in Fig. 5. It presents the calculated scores for SS^A, SS^T, and SS^{CP} because these indicators are directly derived from and highly dependent on the structured data within STIX objects. In contrast, SS^R and SS^{CS} require additional contextual analysis that goes beyond the STIX schema, utilizing external APIs for processing, and hence, are not represented in this visual interface.

In Fig. 5, we present an example that elucidates the application of STIX-based scores. The severity score is calculated as 61%, and the confidence score is calculated as 37%. This example delineates the interconnections among entities such as hash files, reports, and identifiers. The severity score, an aggregate metric, quantifies the potential impact of a threat by integrating the assessments across five indicators: Accuracy (0.06), Relevance (1.44), Timeliness (0.09), Completeness (0.24), and Consistency (1.20). The relevance score is notably high because the report covers a substantial amount of useful information. The timeliness score is low as the IoC was identified more than a year ago. Despite this, the Consistency score indicates that the IoC still poses a threat. The confidence score, calculated here as 37%, is influenced by the source rank and the similarity score. The Source Rank is derived from using VirusTotal data, classified as Internal Enrichment Connector, thus scoring 100%. The lower score from AlienVault is averaged, resulting in the Source Rank score presented. However, the confidence score is significantly reduced because the similarity score is 0, as the IoC is supported by only one piece of intelligence. These metrics provide a comprehensive perspective on the threat level posed by an IoC, allowing analysts to rapidly understand the severity and confidence of the presented intelligence.

6.2. Data sources and tools

As shown in Tables 4 and 5, we have chosen to utilize both internal and external docker connectors as our sources of CTI data. We use the STIX library to ensure compliance with the STIX 2.1 specification. Moreover, as previously studied, we have incorporated the BERT-CRF model to calculate the threat score and build the system.

It is important to highlight that the BERT-CRF model is specifically employed for evaluating the relevance of threat intelligence. To achieve a comprehensive and multifaceted assessment, we also integrate heuristic evaluation methods, including analyses based on STIX Graph. This composite approach enables us to delve into a thorough and multidimensional analysis and assessment of threat intelligence, ensuring our system is both precise and comprehensive.



Severity Indicators	Caculation	Severity Score
Accuracy	2 * 0.03	0.06
Relevance	4 * 0.36	1.44
Timeliness	1 * 0.09	0.09
Completeness	2 * 0.12	0.24
Consistency	3 * 0.40	1.20
Overall Score	3.03 / 5.00	61%

Confidence Indicators	Caculation	Confidence Score
Source Rank	(6.11 + 100) / 2 * 0.7	37.14
Similarity Score	0.00* 0.3	0.00
Overall Score	37.14 / 100	37%

Fig. 5. Search for an IoC visualization result.

Table 7
A detailed breakdown of the severity score assessment criteria.

Score	Timeliness	Accuracy	Completeness	Relevance	Consistency
5	Less than 1 week	Corroborated by 5+ sources	Over 80% of the construction graph completed	Recognizes 9–11 entities	Over 80% antivirus detection rate
4	Less than 1 month	Corroborated by 4 sources	60%–79% of the construction graph completed	Recognizes 7–8 entities	60%–79% antivirus detection rate
3	Within 3 months	Corroborated by 3 sources	40%–59% of the construction graph completed	Recognizes 5–6 entities	40%–59% antivirus detection rate
2	Within 1 year	Corroborated by 2 sources	20%–39% of the construction graph completed	Recognizes 3–4 entities	20%–39% antivirus detection rate
1	Indefinite	Corroborated by 1 source	Below 20% of the construction graph completed	Recognizes 1–2 entities	Below 20% antivirus detection rate
0	Unavailable	No corroborating sources	No available data	No recognizable entities	No available data

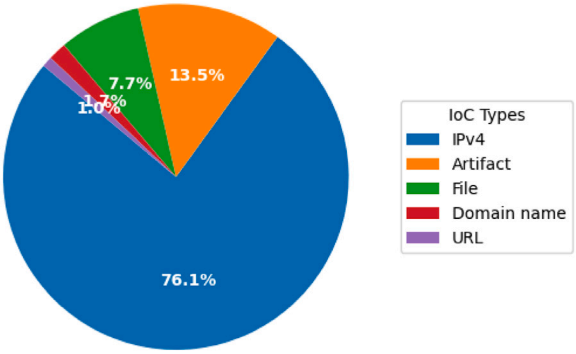


Fig. 6. IoC type distribution.

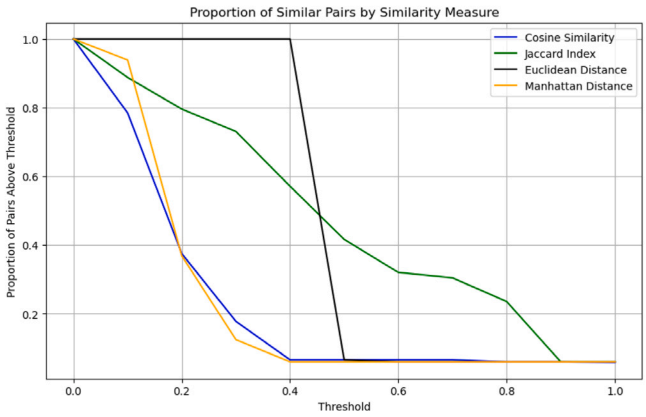


Fig. 7. Proportion of similar pairs at various similarity thresholds for different similarity methods.

6.3. Datasets

Our data set is stored in the local database, where we set up OpenCTI. The number of IoCs is 357,480, the number of CTI reports is 34,183, and the number of observable data is 326,080. We pay special attention to the quality of IoC. The distribution of IoC is shown in Fig. 6.

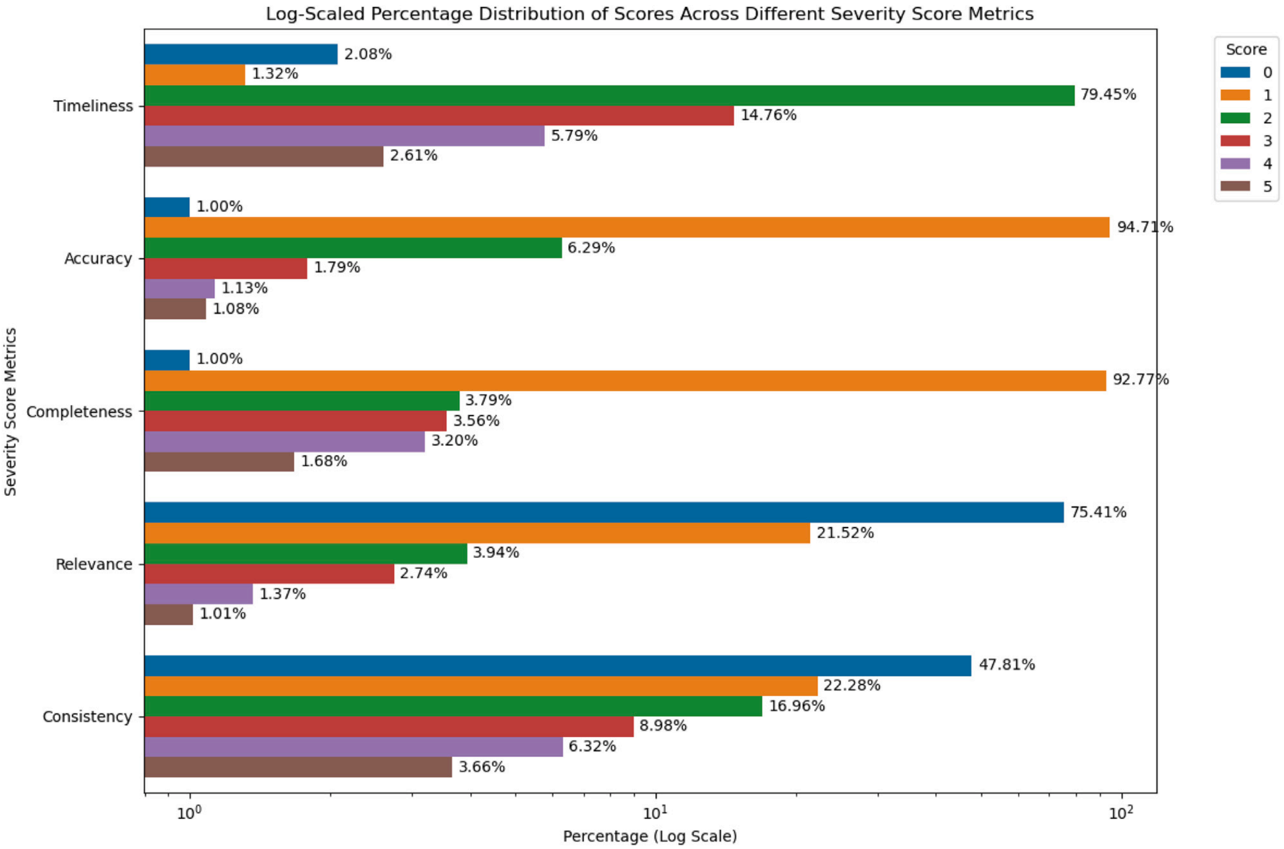


Fig. 8. Log-scaled percentage distribution of scores across different severity score metrics.

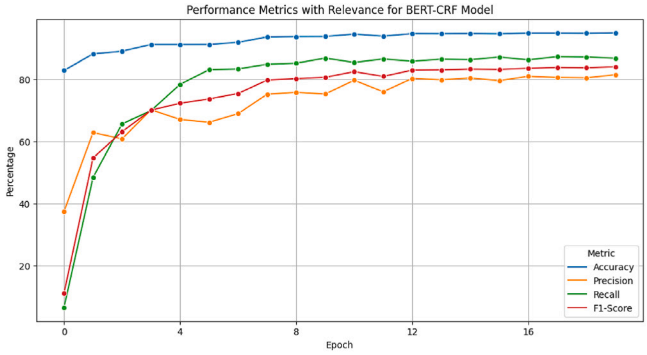


Fig. 9. Perforations of the BERT-CRF model.

7. Results and discussion

In this section, we address two crucial questions about our system:

1. Can the system enhance the value of IoCs?
2. How does our system distinguish itself from other TIPs?

The effectiveness of our system is illustrated through its operation and empirical evaluations using actual IoCs. By aggregating information from various threat intelligence platforms, establishing relationships, and enriching data, our system enhances the quality and availability of IoCs. It outperforms other existing platforms that lack these capabilities. The subsequent subsections delve into case studies of Emotet (Section 7.1) and Medusa (Section 7.2). The Emotet case study indicates how our system improves the quality of IoCs by aggregating and enriching the information effectively. In contrast, the Medusa case study focuses on how our system differentiates itself from other TIPs. It provides a comparative analysis that highlights the challenges of benchmarking in the cybersecurity domain, including the proprietary nature of algorithms and the varied dependence on intelligence data volume across platforms. These exemplify how our system addresses these fundamental questions.

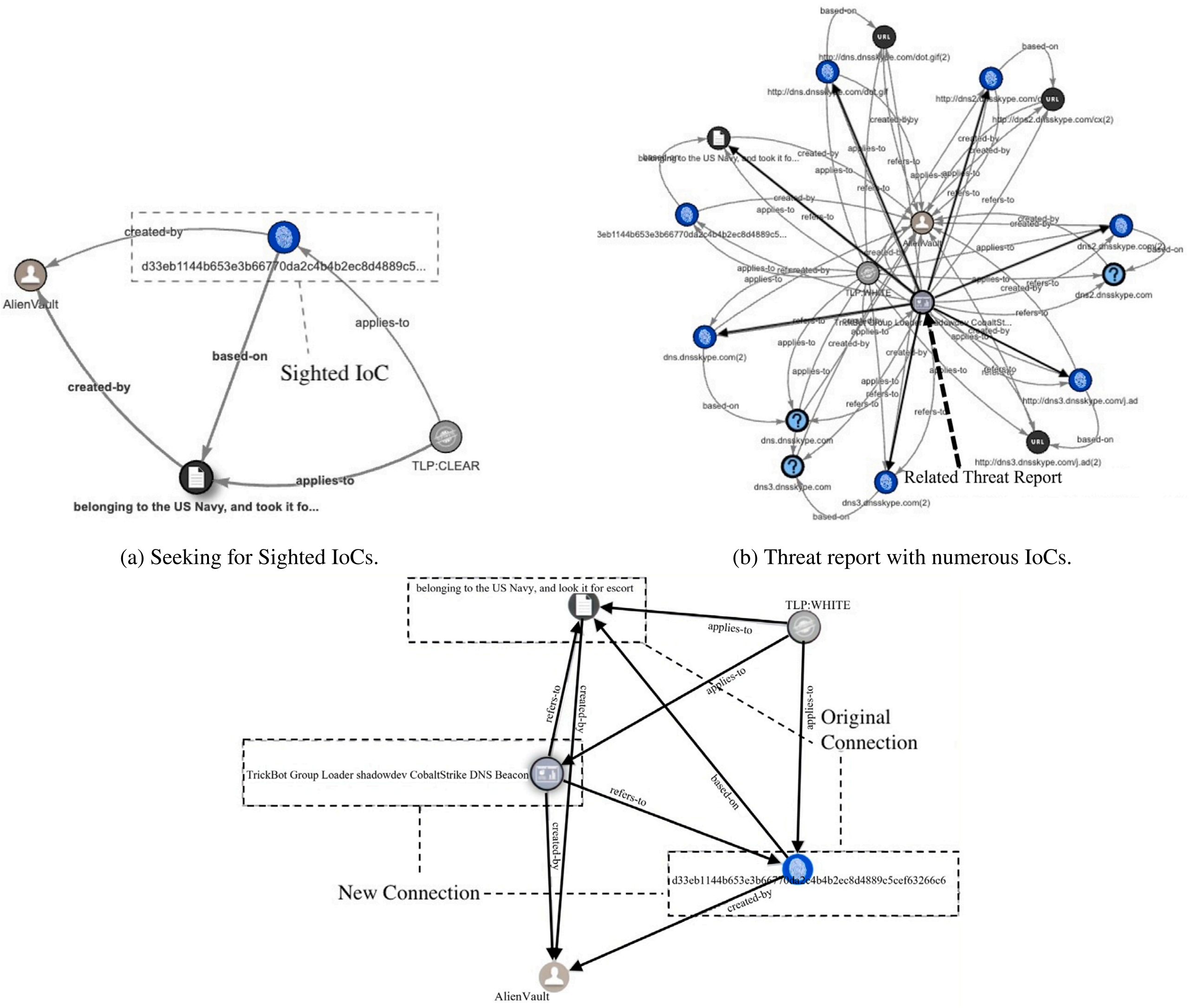
Table 8
Computational efficiency of various similarity algorithms measured in seconds.

Similarity algorithm	Time (s)
Cosine Similarity	0.01
Euclidean Distance	0.01
Manhattan Distance	0.04
Jaccard Index	117.33

7.1. Emotet case study

Emotet is a form of malicious software that is categorized as a keylogger and a banking Trojan. It first surfaced in 2014 and quickly became one of the most damaging malware. Emotet primarily targets the email systems of companies and organizations, infiltrates victims' email accounts, pilfers contact information, and employs social engineering techniques to disseminate malicious emails. It propagates predominantly through spam emails, enticing victims to click on malevolent attachments or links.

To ensure that our quality enhancement measures are based on accurate methods, it is essential to clarify the key technical elements supporting the effectiveness of our system before delving into operational details and specific case studies. A key aspect of our deduplication process involves selecting an appropriate similarity measure to efficiently identify and eliminate duplicate IoCs. In Fig. 7, we evaluated several metrics, including Cosine Similarity, Euclidean Distance, Manhattan


created-by
d33eb1144b653e3b66770da2c4b4b2ec8d4889c5...
AlienVault
applies-to
based-on
Sighted IoC
created-by
TLP:CLEAR
applies-to
belonging to the US Navy, and took it fo...
based-on
http://dns.dnsskype.com/dot.gif(2)
http://dns.dnsskype.com/dot.gif
http://dns2.dnsskype.com/cx(2)
created-by
applies-to
refers-to
AlienVault
TLP:WHITE
3eb1144b653e3b66770da2c4b4b2ec8d4889c5...
dns2.dnsskype.com(2)
dns2.dnsskype.com
dns.dnsskype.com(2)
http://dns3.dnsskype.com/j.ad
http://dns3.dnsskype.com/j.ad(2)
dns.dnsskype.com
dns3.dnsskype.com
dns3.dnsskype.com(2)
Related Threat Report
belonging to the US Navy, and took it for escort
TLP:WHITE
applies-to
applies-to
refers-to
created-by
Original Connection
TrickBot Group Loader shadowdev CobaltStrike DNS Beacon
refers-to
based-on
applies-to
created-by
d33eb1144b653e3b66770da2c4b4b2ec8d4889c5cef63266c6
created-by
New Connection
AlienVault

(a) Seeking for Sighted IoCs.

(b) Threat report with numerous IoCs.

(c) The enriched IoC is easier to comprehend and tidy.

Fig. 10. Emotet hash file case study of IoC type.

Distance, and Jaccard Index, using a representative sample of 500 IoCs extracted from 36,000 CTIs. The y-axis in this Fig. 7 represents the proportion of document pairs that have a similarity score above a given threshold, indicating how consistently each metric can identify similar pairs across different thresholds. Cosine Similarity was chosen for its computational efficiency and consistent performance in detecting conceptual similarities within the preprocessed content. Manhattan sharply declines at a threshold of 0.1, and Euclidean takes a steep drop at 0.4. Table 8 illustrates that, unlike the Jaccard Index, which, despite its ideal performance at lower thresholds, required a significantly longer processing time of 117.33 s, Cosine Similarity maintained consistent and efficient performance across varying thresholds, completing the processing in just 0.01 s. This efficiency is crucial for maintaining the integrity and uniqueness of IoC data in large-scale datasets.

In Fig. 8, it is shown the visual representation provided in the log-scaled percentage distribution graph, we illustrate the proportional distribution of severity scores across different metrics such as timeliness, accuracy, completeness, relevance, and consistency. The graph highlights how scores are distributed, emphasizing the disparities, especially in the lower range scores, which are made more discernible through the use of a log scale due to the extreme values, such as 25,432 counts of 0 scores in Relevance. The experimental data drawn from over 30,000 intelligence entries shows that the distribution of scores in Timeliness largely depends on the age of the intelligence, with a significant portion older than one year, which affects its immediacy and, subsequently, its timeliness score. Most intelligence entries score 1 in accuracy as detected threat indicators inherently possess the attributes that qualify them for this score, indicating that very few entries completely lack relevant attributes. However, both Relevance and Consistency do not have such constraints, resulting in a broader distribution of scores. The results demonstrate that as scores increase, the number of entries at each level decreases, providing evidence that our evaluation criteria effectively quantify and differentiate the quality of IoCs. The decreasing number of entries at higher scores across Accuracy, Completeness, Relevance, and Consistency confirms the robustness of our scoring system, particularly highlighting how our criteria can discern finer details in the threat intelligence data.

Furthermore, for relevance assessment using AI models, additional performance evaluations are required to ensure the usability of the model. The performance of our BERT-CRF model is compared with other advanced models which include instead of SecBERT (jackaduma, 2024) and instead of BERT-BiLSTM (Dai et al., 2019) in Table 9, demonstrating 0.95 accuracy, 0.82 precision, 0.87 recall, and 0.84 F1-scores. This underscores the model's enhanced capability to accurately

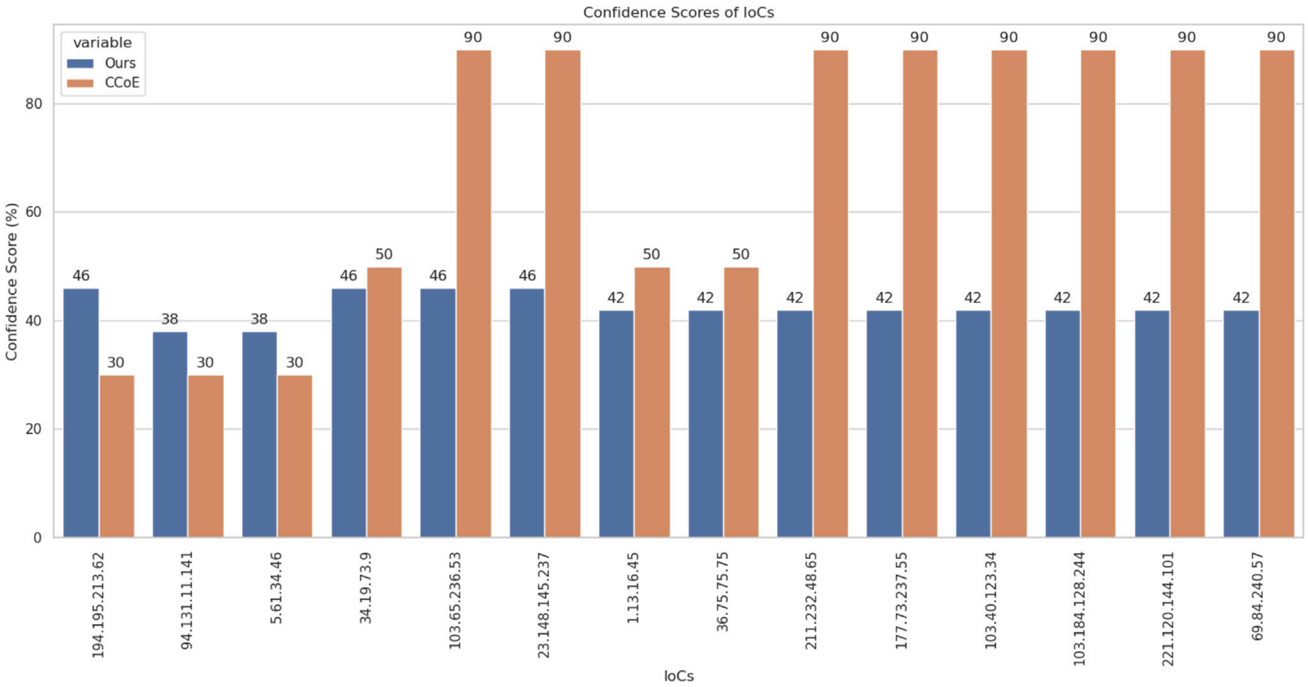


Fig. 11. Comparative analysis of confidence scores for IoCs.

Table 9
Comparison of relevance assessment models.

Model	Accuracy	Precision	Recall	F1-score
SecBERT-BiLSTM-CRF [Ours]	93.41%	77.44%	71.21%	74.20%
BERT-BiLSTM-CRF (Dai et al., 2019)	90.72%	71.46%	52.88%	60.78%
SecBERT-CRF [Ours]	94.82%	82.16%	84.26%	83.20%
BERT-CRF [Ours]	**95.10%**	**81.56%**	**86.88%**	**84.14%**

Table 10
Threat score comparison among different threat intelligence platforms.

TIP	Severity (%)	Confidence (%)
A	11	50
B	10	None
C	90	90
Our	79	24

and efficiently assess the relevance of intelligence entries in various cybersecurity contexts. The hyperparameters for training the model were meticulously set to optimize performance: 20 epochs, an initial learning rate of 5×10^{-5} for BERT and 8×10^{-5} for the CRF fully connected layer, with weight decays of 1×10^{-5} for fine-tuning BERT layers and 5×10^{-6} for the CRF fully connected layer. Fig. 9 shows the performance of the BERT-CRF model. Epoch 3 quickly converged, achieved a good F1-Score of 0.71, and then gradually increased to 0.84.

Fig. 11 detailed comparative analysis, using 17 IoCs related to the Emotet malware as a sample set for average testing, was conducted against benchmarks provided by the Cybersecurity Center of Excellence (CCoE) program under the National Science and Technology Council (NSTC), Taiwan. It demonstrates our system's enhanced precision through a 25.18% reduction in the average difference of confidence scores. The confidence scores for each IoC were calculated using a formula that integrates both the Source Rank and Similarity Score metrics. The Source Rank assesses the reliability of the information based on the reputation and historical accuracy of the source, while the Similarity Score evaluates the consistency of the reported IoC with those found in other credible reports. Specifically, the calculation involves aggregating weighted scores from each contributing source. These scores are then normalized to produce an average confidence score for the set of IoCs.

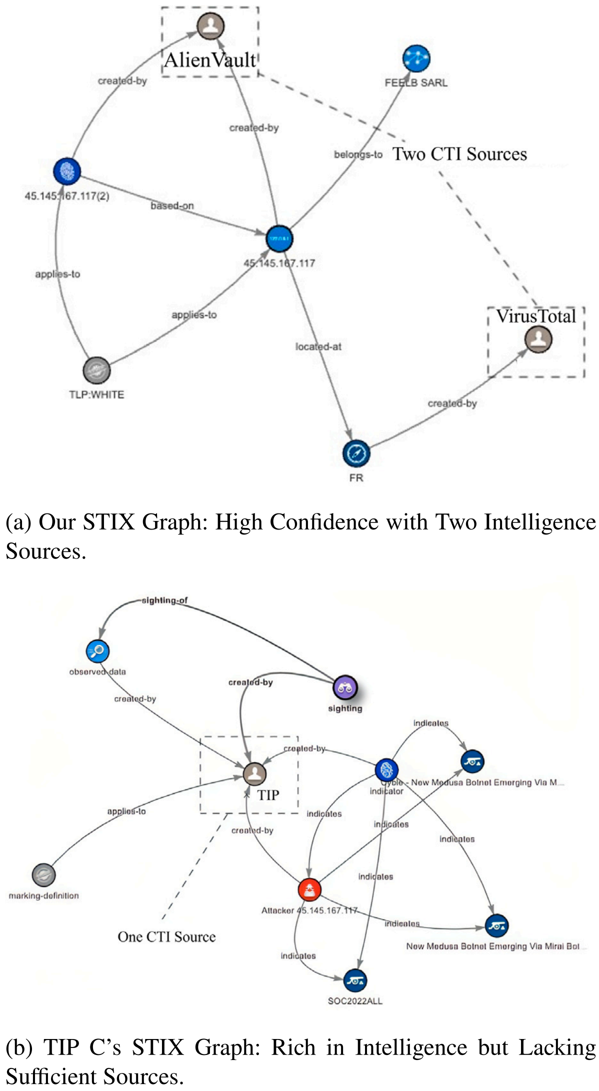
When processing through our system, we initially look for any instances of the input IoC, which is a file hash of an Emotet sample. The results, as depicted in Fig. 10(a), indicate a sighting reported by AlienVault where the IoC was found to be associated with the hash of a file. We then proceed to explore related threat intelligence. The system identified a piece of threat intelligence named "TrickBot Group Loader shadowdev CobaltStrike DNS Beacon", as illustrated in Fig. 10(b), covers additional IoCs, URLs, DNS information, etc. However, TrickBot is a multifunctional Trojan primarily used for financial fraud and malevolent banking activities.

The association between Emotet and TrickBot lies in their user base and propagation methods. Emotet has often been used as an initial infection vector for TrickBot. The emotet is distributed to many users through spam email attacks. Once a system is infected with Emotet, it may download and install TrickBot, thus expanding the infection to a wider range of targets. Currently, we filter out objects unrelated to the IoC, such as IoCs that represent DNS. The next step involves enriching the IoC using the Enricher module and calculating the score. As the final generated graph, depicted in Fig. 10(c), only includes the indicator, report, and other objects, the heuristic component only calculates the score based on these two SDOs.

In other words, a comprehensive attack event should encompass all STIX objects. The relevance model captures entities like HackOrg (TrickBot) and Tool (CobaltStrike). The final scores obtained are Severity Score (61.04%) and Confidence Score (9.54%). Due to the low credibility of the source and the support of a single threat report for this IoC, the confidence score is low. From this experiment, we infer that relying on a single relevant report is inadequate to determine the high credibility of the report itself, leading to a low confidence score.

7.2. Medusa case study

Medusa is a form of malware known as "MedusaHTTP" or "Joker". The backdoor software is intended to infiltrate victims' computer systems and conduct various malicious activities. Medusa's primary objective is to pilfer personal data, sensitive information, and financial



(a) Our STIX Graph: High Confidence with Two Intelligence Sources.

(b) TIP C's STIX Graph: Rich in Intelligence but Lacking Sufficient Sources.

Fig. 12. Comparison of our STIX graph and TIP C's STIX graph: Differences in confidence due to the absence of threat intelligence aids.

credentials. This malware typically penetrates victims' systems through methods such as spam emails, malicious downloads, and exploits vulnerabilities. Once successfully installed, it establishes a backdoor on the victim's system, allowing the attacker to control the infected computer remotely. Given the input of Medusa's C2 IPv4 address, Table 10 lists the threat scores calculated by our system and compares them with other publicly accessible threat intelligence platforms, such as IBM X-Force Exchange (TIP A) (IBM, 2023), CyCraft CyberTotal (TIP B) (Cycraft, 2024), and AlienVault Open Threat Exchange (TIP C) (AlienVault, 2024). This comparison is challenging, as determining the exact calculation methods employed by other platforms is complex and often not transparent.

Hence, we specifically focus on comparing the severity values close to those of TIP C with our STIX graph to demonstrate that our system, though designed based on STIX 2.1, has minor but notable differences compared to TIP C's STIX 2.0.

In Fig. 12(b), the STIX graph generated after analyzing TIP C suggests that the campaign object of TIP C can be identified as an OSINT source. Conversely, in our generated graph Fig. 12(a), the absence of relevant OSINT is evident and contributes to a lower confidence level. Nevertheless, the enrichment module aids in identifying regional STIX objects, which results in a higher threat score. This indicates that our system's confidence calculation is heavily dependent on external processing sources. Moreover, the stark contrast in the severity and confidence scores of TIP A and B, as opposed to TIP C, suggests that the volume and dependability of threat data are crucial factors affecting these metrics. Our analysis is primarily aimed at comparing the severity and confidence scores between TIP C and our system, and not an assessment of the TIPs themselves.

8. Conclusion and future work

In this work, we introduced a system for calculating IoC threat scores and proposed a quality assessment process. The system collected CTI from external sources and network devices of monitored organizations to construct structured information. This system consisted of five main components: data collection, normalization, enrichment, heuristic scoring, and visualization. These components collected and normalized IoCs, then enriched these IoCs, and used heuristic scoring to evaluate IoC threat scores through a series of scores based on STIX and deep learning to measure IoC objectively. This system used weight-heuristic evaluation of enriched IoC based on key indicators such as accuracy, relevance, timeliness, completeness, and consistency to provide IoC's severity score and confidence score.

To navigate these challenges, we adopted a case-study approach. The Emotet case study illustrates the process of IoC enrichment and scoring, highlighting the impact of source credibility on confidence scores. Furthermore, the Medusa case study emphasizes our system's ability to transparently present indicators and perform comparisons with other intelligence platforms. In the future, we plan to incorporate dynamic sandbox technology to analyze the behavioral characteristics of malicious programs, utilizing it as an internal intelligence connector for further comparison with network threat intelligence, thereby generating more valuable intelligence. This incorporation aims to corroborate existing findings and refine the quality of intelligence, ensuring that the IoC threat scores are as accurate and reliable as possible, solidifying the foundation for robust cybersecurity measures.

CRediT authorship contribution statement

Sheng-Shan Chen: Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ren-Hung Hwang:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Asad Ali:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Ying-Dar Lin:** Writing – review & editing, Supervision, Methodology. **Yu-Chih Wei:** Writing – review & editing, Methodology. **Tun-Wen Pai:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Sheng-Shan Chen reports financial support, administrative support, and writing assistance were provided by National Science and Technology Council, Taiwan. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgment

This work was supported in part by the Cybersecurity Center of Excellence (CCoE) program under the National Applied Research Laboratories (NARLabs), Taiwan and National Science and Technology Council (NSTC), Taiwan under 112-2221-E-027-068.

References

Abuse.ch, 2024. Malware sample exchange. URL: https://bazaar.abuse.ch/.

AbuseIPDB, 2024. URL: https://www.abuseipdb.com/.

AlienVault, 2024. AlienVault - open threat exchange. URL: https://otx.alienvault.com. (Accessed 03 January 2024).

Azevedo, R., Medeiros, I., Bessani, A., 2019. PURE: Generating quality threat intelligence by clustering and correlating OSINT. In: 2019 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering. TrustCom/BigDataSE, IEEE, pp. 483–490.

Bandara, E., Shetty, S., Mukkamala, R., Rahaman, A., Liang, X., 2022. LUUNU—Blockchain, MISP, model cards and federated learning enabled cyber threat intelligence sharing platform. In: 2022 Annual Modeling and Simulation Conference. ANNSIM, IEEE, pp. 235–245.

Chen, S.-S., Hwang, R.-H., Sun, C.-Y., Lin, Y.-D., Pai, T.-W., 2023. Enhancing cyber threat intelligence with named entity recognition using BERT-CRF. In: Proceedings of IEEE Global Communication Conference. Kuala Lumpur, Malaysia.

CISA.gov, 2019. Known exploited vulnerabilities catalog: CISA. URL: https://www.cisa.gov/known-exploited-vulnerabilities-catalog.

Connolly, J., Davidson, M., Schmidt, C., 2014. The Trusted Automated Exchange of Indicator Information (Taxii). The MITRE Corporation, pp. 1–20.

Connolly, K., Klempay, A., McCann, M., Brenner, P., 2023. Dark web marketplaces: Data for collaborative threat intelligence. Digit. Threat.: Res. Pract. 4 (4), 1–12.

Cycraft, 2024. CyberTotal. URL: https://cybertotal.cycarrier.com. (Accessed 03 January 2024).

Dai, Z., Wang, X., Ni, P., Li, Y., Li, G., Bai, X., 2019. Named entity recognition using BERT BiLSTM CRF for Chinese electronic health records. In: 2019 12th International Congress on Image and Signal Processing, Biomedical Engineering and Informatics. Cisp-Bmei, IEEE, pp. 1–5.

Enisa, 2021. Exploring the opportunities and limitations of current threat intelligence platforms. URL: https://www.enisa.europa.eu/.

Filigran, 2024. OpenCTI-platform/opencti: Open cyber threat intelligence platform. URL: https://github.com/OpenCTI-Platform/opencti.

Fleck, A., 2022. Infographic: Cybercrime expected to skyrocket in coming years. URL: https://www.statista.com/chart/28878/expected-cost-of-cybercrime-until-2027/.

Gao, Y., Li, X., Peng, H., Fang, B., Philip, S.Y., 2020. Hincti: A cyber threat intelligence modeling and identification system based on heterogeneous information network. IEEE Trans. Knowl. Data Eng. 34 (2), 708–722.

Gonzalez-Granadillo, G., Faiella, M., Medeiros, I., Azevedo, R., Gonzalez-Zarzosa, S., 2021. ETIP: An Enriched Threat Intelligence Platform for improving OSINT correlation, analysis, visualization and sharing capabilities. J. Inf. Secur. Appl. 58, 102715.

Hwang, Y.-W., Lee, I.-Y., Kim, H., Lee, H., Kim, D., et al., 2022. Current status and security trend of osint. Wirel. Commun. Mob. Comput. 2022.

IBM, 2023. IBM X-Force exchange. URL: https://exchange.xforce.ibmcloud.com/. (Accessed 03 January 2024).

jackaduma, 2024. jackaduma/SecBERT. URL: https://huggingface.co/jackaduma/SecBERT.

Khan, S., Wallom, D., 2022. A system for organizing, collecting, and presenting open-source intelligence. J. Data Inf. Manag. 4 (2), 107–117.

Li, L., Huang, C., Chen, J., 2024. Automated discovery and mapping ATT&CK tactics and techniques for unstructured cyber threat intelligence. Comput. Secur. 140, 103815.

Mahyoub, M., Matrawy, A., Isleem, K., Ibitoye, O., 2023. Cybersecurity challenge analysis of work-from-anywhere (WFA) and recommendations based on a user study.

Meier, R., Scherrer, C., Gugelmann, D., Lenders, V., Vanbever, L., 2018. FeedRank: A tamper-resistant method for the ranking of cyber threat intelligence feeds. In: 2018 10th International Conference on Cyber Conflict. CyCon, IEEE, pp. 321–344.

MITRE, 1999. URL: https://cve.mitre.org/.

OASIS, 2023. STIX2.1 introduction. URL: https://docs.oasis-open.org/cti/stix/v2.1/csprd01/stix-v2.1-csprd01.html.

Obrst, L., Chase, P., Markeloff, R., 2012. Developing an ontology of the cyber security domain. In: STIDS. pp. 49–56.

OpenCTI, 2024. Reliability and confidence - OpenCTI Documentation. URL: https://docs.opencti.io/latest/usage/reliability-confidence/.

Schaberreiter, T., Kupfersberger, V., Rantos, K., Spyros, A., Papanikolaou, A., Ilioudis, C., Quirchmayr, G., 2019. A quantitative evaluation of trust in the quality of cyber threat intelligence sources. In: Proceedings of the 14th International Conference on Availability, Reliability and Security. pp. 1–10.

Schlette, D., Böhm, F., Caselli, M., Pernul, G., 2021. Measuring and visualizing cyber threat intelligence quality. Int. J. Inf. Secur. 20, 21–38.

Sergio, C., 2015. The 4 qualities of good threat intelligence. URL: https://www.activeresponse.org/the-4-qualities-of-good-threat-intelligence/.

Serrano, O., Dandurand, L., Brown, S., 2014. On the design of a cyber security data sharing system. In: Proceedings of the 2014 ACM Workshop on Information Sharing & Collaborative Security. pp. 61–69.

Sillaber, C., Sauerwein, C., Mussmann, A., Breu, R., 2016. Data quality challenges and future research directions in threat intelligence sharing practice. In: Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security. pp. 65–70.

Statista, 2023. Cyber threat intelligence market size worldwide. URL: https://www.statista.com/statistics/1230328/cyber-threat-intelligence-market-size-global/.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.

VirusTotal, 2024. VirusTotal. URL: https://www.virustotal.com/gui/.

VXVault, 2024. VX Vault. URL: http://vxvault.net/ViriList.php.

Wagner, C., Dulaunoy, A., Wagener, G., Iklody, A., 2016. Misp: The design and implementation of a collaborative threat intelligence sharing platform. In: Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security. pp. 49–56.

Zhang, S., Chen, P., Bai, G., Wang, S., Zhang, M., Li, S., Zhao, C., 2022. An automatic assessment method of cyber threat intelligence combined with ATT&CK matrix. Wirel. Commun. Mob. Comput. 2022.

Sheng-Shan Chen is pursuing his Ph.D. in Computer Science and Information Engineering at the National Taipei University of Technology (NTUT). He focuses on cyber threat intelligence analysis with a large language model and healthcare security. In addition to his academic pursuits. He has published several notable publications and conferences, including IEEE Global Communications Conference (GLOBECOM), International Conference on Medical and Health Informatics (ICMHI), International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA/AIE), and Hacks in Taiwan Conference (HITCON).

Ren-Hung Hwang (Senior Member, IEEE) received his Ph.D. degree in computer science from the University of Massachusetts, Amherst, Massachusetts, USA, in 1993. He is the Dean of the College of Artificial Intelligence, National Yang Ming Chiao Tung University (NYCU), Taiwan. Before joining NYCU, he was with National Chung Cheng University, Taiwan, from 1993 to 2022. He is currently on the editorial boards of IEEE Communications Surveys and Tutorials and IEICE Transactions on Communications. He received the Best Paper Award from the 6th International Conference on Internet of Vehicles 2019, IEEE Ubi-Media 2018, IEEE SC2 2017, IEEE IUCC 2014, and the IEEE Outstanding Paper Award from IEEE IC/ATC/ICA3PP 2012. He served as the general chair of the International Computer Symposium (ICS), 2016, and International Symposium on Pervasive Systems, Algorithms, and Networks (ISPAN) 2018, International Symposium on Computer, Consumer and Control (IS3C) 2018, IEEE DataCom 2019 (The 5th IEEE International Conference on Big Data Intelligence and Computing). His current research interests include in Deep Learning, Wireless Communications, Network Security, AIoT, and Cloud/Edge/Fog Computing.

Asad Ali is working as a researcher at the National Institute of Cyber Security (NICS), Taiwan. He received his Ph.D. degree in Electrical Engineering and Computer Sciences from the National Yang Ming Chiao Tung University (NYCU), Taiwan in 2022. In 2015, he received his Master degree in Electrical Engineering from National University of Science & Technology (NUST), Pakistan. In 2012, he received his BS degree in electrical engineering from the University of Engineering and Technology, Taxila. His research interests include Cybersecurity, 4G/5G cellular networks, artificial intelligence, network design, and optimization.

Ying-Dar Lin (Fellow, IEEE) is a Chair Professor of computer science at National Yang Ming Chiao Tung University (NYCU), Taiwan. He received his Ph.D. in computer science from the University of California at Los Angeles (UCLA) in 1993. He was a visiting scholar at Cisco Systems in San Jose during 2007–2008, CEO at Telecom Technology Center, Taiwan, during 2010–2011, and Vice President of National Applied Research Labs (NARLabs), Taiwan, during 2017–2018. He was the founder and director of Network Benchmarking Lab (NBL) in 2002–2018, which reviewed network products with real traffic and automated tools, and has been an approved test lab of the Open Networking Foundation (ONF). He also cofounded L7 Networks Inc. in 2002, later acquired by D-Link Corp and O'Prueba Inc. a spin-off from NBL, in 2018. His research interests include network security, wireless communications, network softwarization, and machine learning for communications. His work on multi-hop cellular was the first along this line, and has been cited over 1000 times and standardized into IEEE 802.11s, IEEE 802.15.5, IEEE 802.16j, and 3GPP LTE-Advanced. He is an IEEE Fellow (class of 2013), IEEE Distinguished Lecturer (2014–2017), ONF Research Associate (2014–2018), and received in 2017 Research Excellence Award and K. T. Li Breakthrough Award. He has served or is serving on the editorial boards of several IEEE journals and magazines, including Editor-in-Chief of IEEE Communications Surveys and Tutorials (COMST, 1/2017–12/2020). He published a textbook, Computer Networks: An Open Source Approach, with Ren-Hung Hwang and Fred Baker (McGraw-Hill, 2011).

Yu-Chih Wei is an Associate Professor in the Department of Information and Finance Management at the National Taipei University of Technology. He holds a Ph.D. in Information Management from National Central University, and a B.S. and a M.S. in Information Management from YuanZe University. His research interests include FinTech security, health informatics security, ISRA, SupTech, VANET security, information security management, and business continuity management. Before pursuing an academic career, Dr. Wei was a researcher at the Information & Communication Security Laboratory of Chunghwa Telecom Co., Ltd.

Tun-Wen Pai (Senior Member, IEEE) received his Ph.D. degree in electrical and computer engineering from Duke University, Durham, NC, USA. He works currently as a full professor at National Taipei University of Technology (Taipei Tech), Taipei, Taiwan. He served as the Department Chairman from 2019 to 2022, and served as the Dean of Office of International Affairs from 2022 to 2023 at Taipei Tech. He is also a jointly appointed professor at Department of Computer Science and Engineering, National Taiwan Ocean University where he taught from 1997 to 2018 and served as the Department Chairman from 2002 to 2004. Dr. Pai has received several teaching and research awards from both National Taiwan Ocean University and National Taipei University of Technology. He served as general chairs and regular program committees for several international conferences in the areas of bioinformatics, medical informatics, machine learning, and data engineering, and he also served as a chief guest editor of several special issues from the journals of Biomed Research International, Genes, and Electronics. Dr. Pai is member of ACM and IEEE SMC societies.