IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 25 July 2025, accepted 28 August 2025, date of publication 5 September 2025, date of current version 15 September 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3606477

RESEARCH ARTICLE

Evaluating Cyber Threat Intelligence: Accuracy, Completeness, Relevance, and Freshness

ASAD ALI[1], **REN-HUNG HWANG**[2], **(Senior Member, IEEE), AND YING-DAR LIN**[3], **(Fellow, IEEE)**
[1]National Institute of Cyber Security, Taipei 100, Taiwan
[2]Institute of Computational Intelligence, National Yang Ming Chiao Tung University, Tainan 711, Taiwan
[3]Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan

Corresponding author: Asad Ali (asad.ali@nics.nat.gov.tw)

This work was supported in part by the National Institute of Cyber Security, Taiwan.

ABSTRACT In the realm of cybersecurity, the extraction of Cyber Threat Intelligence (CTI) is vital for acquiring accurate Tactics, Techniques, and Procedures (TTPs) and Indicators of Compromise (IoC) from reputable sources. The extracted CTI must be evaluated to ensure high quality, and different CTI platforms, such as VirusTotal, AlienVault, and MetaDefender, often yield varying evaluations. A robust method is required that automatically extracts CTI, evaluates it, and provides a verdict on IoCs, determining whether they are malicious, while also evaluating IoC information across multiple platforms. In this work, we propose an automated mechanism that first extracts TTPs and IoCs from reputable threat reports, submits the extracted CTI to multiple platforms, evaluates the platform responses using four key metrics —accuracy, freshness, completeness, and relevance, and provides weighted verdicts for IoCs. We tested 600 IoCs in February 2025, and the weighted verdict matched those of VirusTotal for 79.4%, AlienVault for 87.4%, and MetaDefender for 39.6% of the IoCs. The results also show that VirusTotal provides a consistent evaluation of various types of IoCs in terms of freshness, completeness, and relevance of information, whereas AlienVault shows inconsistencies across all IoC types, and MetaDefender shows inconsistency for some. VirusTotal also outperforms the other two when it comes to providing fresher and more complete intelligence, while AlienVault provides the most relevant information in terms of Structured Threat Information eXpression (STIX) 2.1 objects.

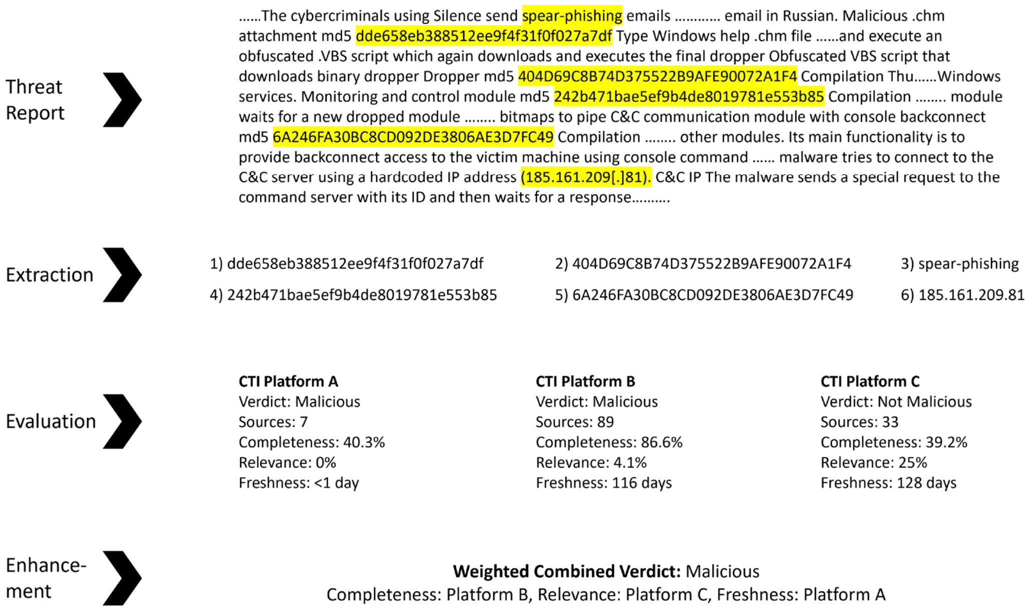
INDEX TERMS Cyber threat intelligence (CTI), indicators of compromise (IoCs), CTI platforms, CTI evaluation, natural language processing.

I. INTRODUCTION

In the realm of cybersecurity, Cyber Threat Intelligence (CTI) refers to the process of collecting, organizing, analyzing, and disseminating intelligence on threats posed to the security of organizations, governments, and industries. This information is useful in mitigating cyberattacks, as it provides stakeholders with valuable intelligence about potential threats and helps them proactively identify these threats and associated threat groups [1]. CTI is gathered by monitoring various open-source intelligence sources, forums, and the dark web, and contains information about Advanced Persistent Threat (APT) groups. The MITRE framework provides a comprehensive list of Tactics, Techniques, and Procedures (TTPs) used by these APT groups, and existing vulnerabilities within organizations, governments, and industries [2].

The APT groups and TTPs outlined by MITRE represent sophisticated threat intelligence [3], [4]. In addition to this high-level intelligence, CTI includes more detailed low-level intelligence, such as Indicators of Compromise (IoCs). These IoCs include hashes, IP addresses, domain names, and URLs. This low-level intelligence is usually available in the form of static blocklists that contain predefined lists of known malicious IoCs, such as IP blocklists or domain blocklists. Despite their utility, static blocklists have limitations due to infrequent updates, which may lead to outdated threat data. Although updating such blocklists is indeed the responsibility of security operations teams, there are challenges to ensuring

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.



......The cybercriminals using Silence send spear-phishing emails .... email in Russian. Malicious .chm attachment md5 dde658eb388512ee9f4f31f0f027a7df Type Windows help .chm file ......and execute an obfuscated .VBS script which again downloads and executes the final dropper Obfuscated VBS script that downloads binary dropper Dropper md5 404D69C8B74D375522B9AFE90072A1F4 Compilation Thu......Windows services. Monitoring and control module md5 242b471bae5ef9b4de8019781e553b85 Compilation .... module waits for a new dropped module .... bitmaps to pipe C&C communication module with console backconnect md5 6A246FA30BC8CD092DE3806AE3D7FC49 Compilation .... other modules. Its main functionality is to provide backconnect access to the victim machine using console command ...... malware tries to connect to the C&C server using a hardcoded IP address (185.161.209[.]81). C&C IP The malware sends a special request to the command server with its ID and then waits for a response.........

1) dde658eb388512ee9f4f31f0f027a7df
2) 404D69C8B74D375522B9AFE90072A1F4
3) spear-phishing
4) 242b471bae5ef9b4de8019781e553b85
5) 6A246FA30BC8CD092DE3806AE3D7FC49
6) 185.161.209.81

CTI Platform A	CTI Platform B	CTI Platform C
Verdict: Malicious	Verdict: Malicious	Verdict: Not Malicious
Sources: 7	Sources: 89	Sources: 33
Completeness: 40.3%	Completeness: 86.6%	Completeness: 39.2%
Relevance: 0%	Relevance: 4.1%	Relevance: 25%
Freshness: <1 day	Freshness: 116 days	Freshness: 128 days

Weighted Combined Verdict: Malicious
Completeness: Platform B, Relevance: Platform C, Freshness: Platform A

FIGURE 1. Example scenario.

that updates occur in a timely manner. For example, organizations may lack the resources to continuously monitor emerging threats or may not prioritize regular updates to blocklists. In addition, maintaining an up-to-date blocklist is also challenging due to the sheer volume of emerging threats. For example, the AV-TEST Institute registers around 450,000 new malicious programs and potentially unwanted applications daily, making it difficult for blocklists to be foolproof or comprehensive [5]. As a result, APT groups can exploit these gaps that are not yet reflected in the blocklists. Here, real-time Cyber Threat Intelligence Platforms (CTIP) come into play, making use of artificial intelligence and machine learning algorithms to continuously monitor data sources to provide up-to-date information [6].

Many organizations that publish TTP reports also disseminate IoC data via real-time feeds or automated threat intelligence platforms. This enables earlier access to critical data, and this form of IoC sharing significantly reduces the delays associated with static blocklists. This highlights the importance of threat intelligence platforms that offer real-time updates and continuous monitoring. Popular real-time threat intelligence platforms include VirusTotal, AlienVault, MetaDefender, AbuseIPDB, and MISP. These platforms perform data collection, processing, and analysis [7]. Upon a search request, they also provide information about various IoCs and keep looking for new ones. While these IoCs assist in updating blocklists, they fail to provide a comprehensive view of the threat as they lack contextual information regarding the APT groups employing these IoCs, their motives, targeted industries, or TTPs used by them. This information is found in CTI reports published by vendors such as CrowdStrike. Therefore, extracting APT group data and their associated TTPs from such reports, in addition to IoCs, is essential so that organizations get complete information about a threat.

A. CTI EXTRACTION AND EVALUATION

CTI extraction provides valuable information from unstructured CTI reports [8]. It enables organizations to obtain insights into the TTPs and IoCs employed by various APT groups. It is important to broaden the input used for CTI extraction and use sources such as CTI reports from various sources, hacker forums, the deep web, and social media. Once APT groups, TTPs, and IoCs are extracted, they must be evaluated, as incorrect or outdated intelligence can compromise an organization's cybersecurity posture. Therefore, it is important to evaluate the extracted intelligence so that high-quality CTI can be provided to help organizations improve their defense mechanisms and make better-informed decisions about their security policies. Although multiple parameters can be used to evaluate the CTI, the most commonly applied criteria are accuracy, timeliness, completeness, and relevance, which are discussed in subsequent sections.

B. EXTRACTION AND EVALUATION ISSUES

Several challenges arise during the extraction and evaluation of CTI. The first challenge involves selecting reliable CTI sources, as there is a lot of fake information available online, and it is of utmost importance to select correct and reliable sources to extract the CTI. Second, the absence of automation makes manual extraction of APT groups, TTPs, and IoCs from unstructured CTI reports a labor-intensive task [9]. A third challenge involves false positives and negatives, which can result in inaccurate CTI. It is important to accurately evaluate an IoC as malicious or non-malicious. Finally, CTI collected from different platforms varies in source count, and CTI must be weighed according to the number of sources involved to increase confidence in the accuracy of the CTI.

C. MOTIVATION

To provide a solution to the identified issues, we propose a combined automated mechanism for CTI extraction, evaluation, and platform assessment that can be explained with the help of Figure 1, where an example scenario is provided. It can be seen that IoCs and TTPs are mentioned in a threat report. They are automatically extracted from threat reports and then fed to open-source CTI platforms A, B, and C. Figure 1 illustrates the evaluation of a sample IoC, revealing that each platform provides a different evaluation, where two platforms mark the IoC as malicious, and one platform marks it as not malicious. In the next step, the automated mechanism provides a weighted combined verdict. Our primary contribution is the proposal of a dynamic weighting mechanism that aggregates IoC data from multiple platforms, gives weights to the obtained information according to the number of sources involved, and provides a combined verdict that is more reliable as compared to the verdict of a single platform. In addition to that, the mechanism also provides information about which platform performs better in terms of providing more complete, relevant, and fresh information.

D. CONTRIBUTIONS

The proposed method has potential applications that extend beyond the quality assessment of IoCs and platforms. The method is helpful for organizations in making informed decisions and implementing adaptive cybersecurity strategies, as it provides a reliable combined verdict and a comprehensive comparison of the completeness, relevance, and freshness of given IoCs across platforms. This allows organizations to pick and choose information that is tailored to their cybersecurity strategy and helps them adapt to the ever-changing threat landscape. The major contributions of this work are as follows.

1) We propose an automated mechanism that extracts CTI from reputable threat reports, inputs the extracted information into CTI platforms, and collects responses from various CTI platforms.
2) We propose a dynamic weighting mechanism that collects information about an IoC from various platforms, assigns dynamic weights to the obtained information based on the number of contributing sources, and provides a combined verdict that is more reliable than the verdict from any single platform.
3) The proposed mechanism provides the CTI evaluation in terms of accuracy, freshness, completeness, and relevance.
4) Our proposed mechanism also provides an evaluation of which platform performs better in providing complete, relevant, and fresh information.

The remainder of this paper is organized as follows. In Section II, we describe the background of CTI extraction and evaluation. Section III provides a review of the related work. We present our problem statement in Section IV. Section V provides the architecture for the proposed solution. In section VI, we present the implementation, results, and their evaluation. Section VII concludes the paper, along with some suggestions for future work.

II. BACKGROUND

In this section, we discuss the extraction and evaluation of CTI, along with related work.

A. CTI EXTRACTION

CTI extraction is the process of collecting, analyzing, and disseminating information about potential cyber threats to organizations, governments, and industries. These threats are typically posed by APT groups that employ various TTPs to compromise security. CTI also includes information about the IoCs used by these groups. Such data is obtained from open-source intelligence feeds, vendor-published CTI reports, and hacker forums.

The CTI extraction process consists of several steps, including data collection, preprocessing, analysis, and dissemination to relevant stakeholders, enabling informed decision-making. This information is typically used to enhance incident response capabilities, strengthen security awareness, and proactively mitigate cyber threats before they cause harm. In recent years, CTI extraction has become an essential component of cybersecurity strategy, as the volume and sophistication of cyber threats continue to grow. Leveraging CTI extraction enables organizations to proactively counter potential threats and safeguard critical assets.

B. CTI EVALUATION

The CTI evaluation involves assessing the quality of the CTI through various parameters, including accuracy, timeliness, relevance, completeness, and reliability. ***Accuracy*** measures the correctness of information. ***Timeliness*** refers to how quickly CTI is disseminated after detection and is often assessed in terms of freshness (i.e., how recently it was updated). ***Completeness*** refers to the extent of information provided about the threat, and it is inherently subjective. ***Relevance*** refers to the extent to which a piece of information is relevant to the domain, infrastructure, and environment of the organization.

The CTI quality evaluation process involves several steps, including identifying key CTI data sources, evaluating the accuracy and reliability of the data, and analyzing the data to identify potential gaps or inconsistencies. In addition, the CTI quality evaluation process involves collaboration with key stakeholders, including security analysts and incident

TABLE 1. CTI extraction related work-I.

Purpose	Mining Sources	Studies	Approach
IoC Extraction	Threat Reports	[10]	CTI miner for IoC parsing
		[11]	NER
		[12]	Sequence labelling
		[13]	Collection, aggregation, clustering
		[14]	NER, Sequence labelling
	Forums	[15]	Terms identification, sequence labelling
		[16]	Topic classification
	Blogs	[17]	Topic classification, term and relation extraction
TTP Extraction	Threat Reports	[18]	Semantic extraction
		[19]	NER, Terms identification
		[20]	Terms identification
		[21]	Terms identification, Semantic extraction
		[22]	Semantic extraction, Threat ontology
		[23]	PoS tagging
	Logs	[24]	Open source tool for IoC parsing

TABLE 2. CTI extraction related work-II.

Level	Extract What	Study	Attribution	Performance	Technique
High Level	Threat Actions	TTPDrill [22]	✓	Accuracy: 82 %	NLP & IR
		ActionMiner [23]	X	Precision: 93%; Recall: 92%	NLP
	TTPs	ATHRNN [27]	X	Micro F1 score	Attention based Transformer
		DT and RL [26]	X	0.92 F1	Modified BERT
	Others	HinCTI [28]	X	Micro F1 score	HIN
		Vulcan [29]	X	0.972 F1	NER & RE
Low Level	IoCs	Acing the IOC Game [17]	X	Precision: 95%; Coverage: 90%	Graph mining, NLP

TABLE 3. CTI evaluation related work.

Subject	Study	Accuracy	Relevance	Completeness	Timeliness	Dynamic Weights	Multiple Platforms
Platforms and Feeds	Platform Evaluation [30]	X	X	X	X	X	✓
	Feed Evaluation [31]	X	X	X	✓	X	✓
	User Perspective [32]	✓	X	✓	X	X	X
Threat Reports	Weighted Evaluation [33]	X	X	✓	✓	X	X
	Visualization [34]	✓	✓	✓	✓	X	X
Platforms and IoCs	Ours	✓	✓	✓	✓	✓	✓

response teams, to validate CTI data and ensure that it is actionable and relevant. Organizations typically evaluate CTI quality using a set of defined metrics, such as the completeness of the data, the relevance of the data to the organization's specific security needs, and the reliability of the data source. By evaluating the quality of CTI data, organizations can improve their overall cybersecurity posture, enhance their incident response capabilities, and proactively mitigate potential cyber threats before they can cause harm.

III. RELATED WORK

Table 1 documents existing studies focused on CTI extraction. These studies are grouped into two categories: those targeting IoC extraction and those focused on TTP extraction. Although not exhaustive, the studies represent the current state of the field. The studies are further classified based on their mining sources: threat reports, forums, blogs, and logs. The table also describes the approaches the authors have used for IoC and TTP extraction. These studies aim to automate the extraction of relevant threat intelligence from diverse sources to help detect and respond to threats. Cited approaches include CTI Miner for IoC Parsing [10], named entity recognition (NER) [11], sequence labelling [12], collection/aggregation/clustering [13], topic classification [16], semantic extraction [18], threat ontology [22], and POS tagging [23]. The approaches used in these studies aim to automate the extraction of actionable threat intelligence by identifying relevant terms, extracting semantic relationships, and labeling data for analysis. Collectively, these approaches contribute to improved organizational cybersecurity through timely and relevant CTI. It is evident that most works utilize threat reports as their primary data source, due to their structured content. Therefore, we also select threat reports as our mining source.

We also evaluated these studies in terms of attribution, accuracy, and techniques, as documented in Table 2. Attribution refers to the identification of attackers responsible for the attack [25]. The table summarizes studies and techniques used to extract threat intelligence at both high and low levels. High-level extraction focuses on identifying threat actions and TTPs associated with cyber threats. The studies cited include *TTPDrill* [22], *ActionMiner* [23], *ATHRNN* [27], *DT and RL* [26], *HinCTI* [28], and *Vulcan* [29]. The techniques used in these studies include NLP (Natural

Language Processing), IR (Information Retrieval), attention-based transformers, modified BERT, and HIN (Heterogeneous Information Network). The reported accuracy of these studies varies, with some achieving higher precision and recall rates as compared to others. Low-level extraction focuses on IoCs (Indicators of Compromise). The primary study cited at this level is *Acing the IOC Game* [17], which uses graph mining and NLP techniques to achieve a precision rate of 95% and a coverage rate of 90%. As shown in Table 2, although these studies focus on high-level TTPs extraction, most do not provide attribution to threat actors, which is an important part of CTI. In addition, most documented studies utilize Natural Language Processing and BERT for CTI extraction. In our work, we also employ an open-source CTI extractor based on NLP.

Table 3 documents different studies that provide an evaluation of IoCs, platforms, intelligence feeds, and threat reports. The table divides these studies into three categories: those that evaluate platforms and feeds, those that evaluate threat reports, and those that evaluate platforms and IoCs extracted from threat reports. The table indicates whether these studies evaluate CTI in terms of accuracy, relevance, completeness or coverage, and timeliness. It also notes whether these studies employ dynamic weighting and assess multiple platforms. It can be seen from Table 3 that most documented studies provide accuracy, completeness, and timeliness evaluation. Many of these studies do not include relevance evaluation. It can also be seen that none of these studies focus on the dynamic allocation of weights to calculate accuracy, and most do not use multiple platforms for the evaluation of CTI. References [30] and [31] provide an evaluation of various platforms, but use different parameters for the evaluation process. Reference [32] provides accuracy and completeness evaluation but lacks relevance and timeliness. Reference [33] provides completeness and timeliness evaluation but lacks accuracy and relevance evaluation. Reference [34] is closest to our work as it provides accuracy, relevance, completeness, and timeliness evaluation, but lacks dynamic weights and multiple platforms. In addition, [34] only evaluates individual threat reports, whereas our work evaluates both the extracted IoCs and the CTI platforms themselves.

In the literature, various studies have proposed extraction and evaluation using various parameters and sources. A gap remains for a method that provides a comprehensive CTI perspective while also evaluating CTI platforms using accuracy, timeliness, completeness, and relevance. We propose a method that not only extracts TTPs and IoCs from the unstructured CTI reports but also assesses the quality of these IoCs across multiple CTI platforms, provides an accurate verdict using various platforms, and offers a comparative evaluation of these platforms. Our proposed method also provides dynamic weighting to the parameters so that organizations have the flexibility to choose which parameters are more important to them. To our knowledge, no other study provides CTI extraction, evaluation, and enhancement in one place.

IV. PROBLEM FORMULATION

This work focuses on the extraction, quality evaluation, and enhancement of CTI. Each stage of CTI extraction, evaluation, and enhancement presents distinct challenges. Accurate extraction of TTPs and IoCs from threat reports is essential. After extraction, selecting appropriate evaluation parameters for IoCs and CTI platforms becomes critical.

The core problem is to accurately extract TTPs and IoCs from CTI reports and design an evaluation system with dynamic weighting to evaluate the accuracy of extracted IoCs across multiple CTI platforms. The core problem is that the existing approaches for CTI analysis and evaluation often lack accuracy in identifying and verifying IoCs from publicly available CTI platforms and reports. This issue arises because different platforms and CTI reports often provide conflicting or inconsistent verdicts for the same IoC. Furthermore, the evaluation of CTI quality requires a method that incorporates dynamic weighting. The problem can be stated as follows:

Given: Publicly available CTI platforms and CTI reports.

Objectives:

- To accurately extract IoCs and TTPs.
- To evaluate CTI accuracy across multiple platforms.
- To enhance the overall quality of CTI.

Output:

- TTPs and IoCs
- CTI evaluation results
- CTI platform evaluation results

We divided our problem into two sub-problems as follows:

A. SUB-PROBLEM 1 - EXTRACTION

Sub-problem 1 addresses the accurate extraction of CTI from unstructured CTI reports, and it can be stated as:

Given: Publicly available CTI reports CR = CR_1, CR_2, CR_3,.....CR_n.

Objectives: The objective is to accurately extract TTPs and IoCs.

Output:

- TTPs $T_1, T_2, T_3, \ldots T_n$ and IoCs $IoC_1, IoC_2, IoC_3, \ldots IoC_n$ extracted from CTI reports.

B. SUB-PROBLEM 2 - EVALUATION

Sub-problem 2 involves the evaluation of the extracted CTI and associated CTI platforms, and it is stated as follows:

Given: IoCs IoC_1, IoC_2, $IoC_3, \ldots IoC_n$ from CTI reports and three CTI platforms.

Objectives:

The objective is to enhance the quality of CTI by evaluating it against various platforms and also to evaluate multiple CTI platforms in terms of accuracy, timeliness, completeness, and relevance.

Output: The output consists of the following:

- A (IoC_1, IoC_2, $IoC_3, \ldots IoC_n$) where A = Accuracy Verdict
- T (IoC_1, IoC_2, $IoC_3, \ldots IoC_n$) where T = Timeliness (Freshness in terms of days)

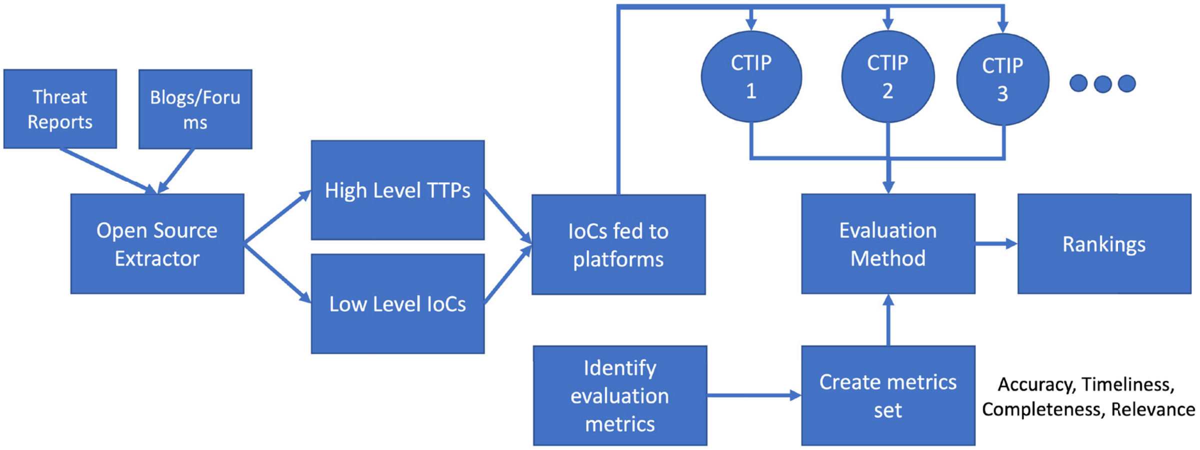


FIGURE 2. Proposed solution.

- CS (IoC_1, IoC_2, IoC_3,...IoC_n) where CS = Completeness Score Percentage
- RS (IoC_1, IoC_2, IoC_3,...IoC_n) where RS = Relevance Score Percentage

V. PROPOSED SOLUTION DESIGN

accurately extracting IoCs and TTPs from CTI reports, as well as their subsequent evaluation and enhancement, we propose a solution that aggregates CTI from multiple sources, including forums, blogs, and vendor-published CTI reports. The proposed solution consists of an extraction mechanism and an evaluation mechanism, as shown in Figure 2. The solution comprises several stages, beginning with the data collection from relevant sources. We collect data from threat reports published by CTI vendors, and we also collect data available on online blogs and forums. Then, we use an open-source extractor to extract various IoCs during the extraction stage. We categorize these IoCs into two types: high-level TTPs and low-level IoCs. In the enrichment stage, we submit the extracted IoCs to CTI platforms to retrieve comprehensive information about them. Then, we pass the extracted information to the evaluation stage. The evaluation stage applies metrics to assess the accuracy, completeness, relevance, and timeliness of the information provided by each CTI platform. In this way, we rank the information obtained from various CTI platforms.

The logic flow of the proposed solution is as follows:

- ***Extraction:*** The open-source extractor ingests CTI reports and online sources such as blogs and forums to extract low-level IoCs and high-level TTPs. We not only take into account CTI obtained from the Cyber Threat Intelligence Platforms (CTIP), but also extract IoCs and TTPs from the CTI documents published by vendors.
- ***Enrichment:*** The enrichment process takes the extracted IoCs and finds relevant information about these IoCs from publicly available CTI platforms. The retrieved data is then aggregated to enrich the IoCs.
- ***Evaluation:*** The evaluation metrics consist of the *Accuracy*, *Timeliness*, *Completeness*, and *Relevance*. The evaluation method takes the enriched IoCs, evaluates them according to the defined metrics, and ranks CTI platforms according to these metrics.

A. ACCURACY

To assess the accuracy of determining whether an IoC is malicious, we employ three publicly available CTI platforms: VirusTotal (VT), AlienVault (AV), and MetaDefender (MD). We chose these platforms because of their free Application Programming Interface (API), widespread use, comprehensive sources, and distinct approaches to evaluating IoCs. This diversity allows for a more comprehensive assessment of IoCs. We extract an IoC from the CTI report and use VirusTotal, AlienVault, and MetaDefender APIs to obtain information about the extracted IoC. Each platform provides unique information that can be used to evaluate the reputation of an IoC.

1) VirusTotal: We analyze the "last_analysis_stats" from the platform's response to count the number of sources that categorized the IoC as "harmless," "malicious," "suspicious," or "undetected." An IoC is considered malicious if two or more sources flag it as such.
2) AlienVault: We extract the number of "pulses" (indicators of activity or interest) from the response of the platform. An IoC is marked as malicious if it has two or more pulses.
3) MetaDefender: Here, the evaluation is stricter; if even one source flags the IoC as malicious, it is considered malicious. MetaDefender typically uses fewer sources for evaluation (approximately 8), compared to VirusTotal (approximately 88).

Finally, we combine the verdicts from all three platforms, applying different weights to each based on the number of sources they use. The rationale for this weighting method is that platforms aggregating verdicts from a larger number of sources are considered more reliable, and therefore, their verdicts are weighted more heavily. This is different from a simple majority vote, as we assign weights to the verdicts based on the number of sources a platform uses. This weighted approach takes into account the level of reliability of the platform based on the diversity of its data sources. For each IoC, the final verdict (malicious or not) is derived by calculating a weighted average of the individual verdicts as follows:

$$\text{Final Verdict} = W_1 \cdot V_1 + W_2 \cdot V_2 + W_3 \cdot V_3$$

where $V_1, V_2, V_3 \in \{0, 1\}$ are the verdicts (1 for malicious, 0 for non-malicious) provided by three platforms, and W_1, W_2, W_3 are the corresponding weights based on the number of sources used by each platform. The weights are calculated as follows:

$$W_i = \frac{N_i}{N_1 + N_2 + N_3} \quad \text{for} \quad i = 1, 2, 3$$

Here N_1, N_2, N_3 represent the number of sources used by each platform. An IoC is classified as malicious if the final verdict score is greater than or equal to 0.5; otherwise, it is deemed non-malicious.

B. FRESHNESS

The freshness of an IoC is crucial as it reflects how recently the information was updated. CTI platforms typically provide the "Last Seen" timestamp, used to calculate the freshness of the IoC. We calculate freshness by subtracting the Report Time (T_R) from the Current Time (T_C), as shown below:

- T_R = Report Time: $day/month/year(hour : minute : second)$,
- T_C = Current Time: $day/month/year(hour : minute : second)$,
- Freshness = $T_C - T_R$.

Each CTI platform provides the IoC information differently, and therefore, the freshness calculation must be adapted for each platform.

1) VirusTotal: We use the "last_analysis_date" from the response data of an IoC to calculate the freshness.
2) AlienVault: We get multiple pulses for each IoC, and each of these pulses provides us with two values that are relevant to the Report Time (T_R): "modified_timestamp" and "created_timestamp". We calculate the freshness using the "modified_timestamp" and average the results across all pulses.
3) MetaDefender: We calculate the freshness score from the "update_time" of each source. The freshness scores are then averaged across sources to determine the overall freshness.

C. COMPLETENESS

Completeness assesses the extent to which IoC information is comprehensively provided by the CTI platforms. We calculate the completeness of the information provided by each platform based on its fields in the response data. These fields are documented in Table 4 for IP addresses where VirusTotal provides various fields in the response of an IoC, including but not limited to `network`, `tags`, `whois`, etc. AlienVault has different fields in the response, such as `id`, `name`, `description`, etc., while MetaDefender provides fields such as `provider`, `assessment`, `location`, etc. We calculate the completeness score as follows:

- Completeness Score of an IoC I_n is: $(Objects) : CS_{I_n} = (Fields_{non-empty}/Fields_{total}) * 100$

It is important to note that $Fields_{total}$ varies from platform to platform. Each CTI platform prefers its own set of fields that it provides information about. For example, the fields used to calculate the completeness of the information provided by AlienVault are different from the fields used in VirusTotal. Therefore, the completeness score is inherently subjective; hence, a relevance score is also necessary.

D. RELEVANCE

Relevance measures the extent to which the provided information aligns with client needs. Since each CTI platform offers different fields, we map them to the essential fields defined in the STIX 2.1 format. Then, we select key fields from STIX as "S_O" (STIX Objects) and categorize the important fields of interest to clients as "V_C" (Important values for clients). The relevance score is derived from how well the information aligns with the client's needs, using the following formula:

- V_C: Important values for clients
- S_O: STIX Objects
- Relevance Score: $(V_C \cap S_O)/S_O$

VI. RESULTS AND EVALUATION

We extracted multiple IoCs from various text reports, selected 600 of them, and used a Python script to retrieve information about these IoCs from multiple CTI platforms such as AlienVault, VirusTotal, and MetaDefender in February 2025. Then, we used our proposed evaluation method to calculate the accuracy, freshness, completeness, and relevance results. We provide Python scripts for IoC extraction, weighted verdict calculation, and IoC evaluation on GitHub[1]. We also include a README.md file in the repository, which contains instructions for executing the proposed code and replicating the results.

A. ACCURACY RESULTS

We extracted multiple IoCs from CTI reports and submitted 600 of them to VirusTotal, AlienVault, and MetaDefender to check if they are malicious or not. We provide a comparison of the verdicts provided by three different CTI platforms for 10 IoCs in Table 5, along with our proposed weighted verdict. The table illustrates a representative snapshot of the verdicts and demonstrates that the weighted verdict approach

[1] https://github.com/asadalibajwaa/IoCEvaluator

TABLE 4. Fields in platforms' response.

Number	VirusTotal Fields	AlienVault Fields	MetaDefender Fields
1	network	id	provider
2	tags	name	assessment
3	whois	description	detect_time
4	last_analysis_date	modified	update_time
5	as_owner	created	status
6	last_analysis_stats	tags	country
7	asn	references	city
8	whois_date	adversary	location
9	reputation	targeted_countries	subdivisions
10	last_analysis_results	malware_families	-
11	country	attack_ids	-
12	last_modification_date	industries	-
13	regional_internet_registry	cloned_from	-
14	continent	groups	-
15	total_votes	-	-

TABLE 5. Verdict comparison.

IoC	VirusTotal	AlienVault	MetaDefender	Ours
1	Malicious	Malicious	Not Malicious	Malicious
2	Malicious	Malicious	Not Malicious	Malicious
3	Malicious	Not Malicious	Not Malicious	Not Malicious
4	Malicious	Malicious	Not Malicious	Malicious
5	Malicious	Malicious	Not Malicious	Malicious
6	Malicious	Malicious	Malicious	Malicious
7	Malicious	Malicious	Malicious	Malicious
8	Malicious	Malicious	Not Malicious	Malicious
9	Malicious	Malicious	Malicious	Malicious
10	Not Malicious	Not Malicious	Not Malicious	Not Malicious

takes verdicts provided by various platforms and outputs a more reliable final verdict. We only provide results for 10 IoCs in Table 5 for the depiction of the results, but we tested 600 IoCs in our experiment. The complete set of IoCs, along with their verdicts, is available via the provided GitHub link.

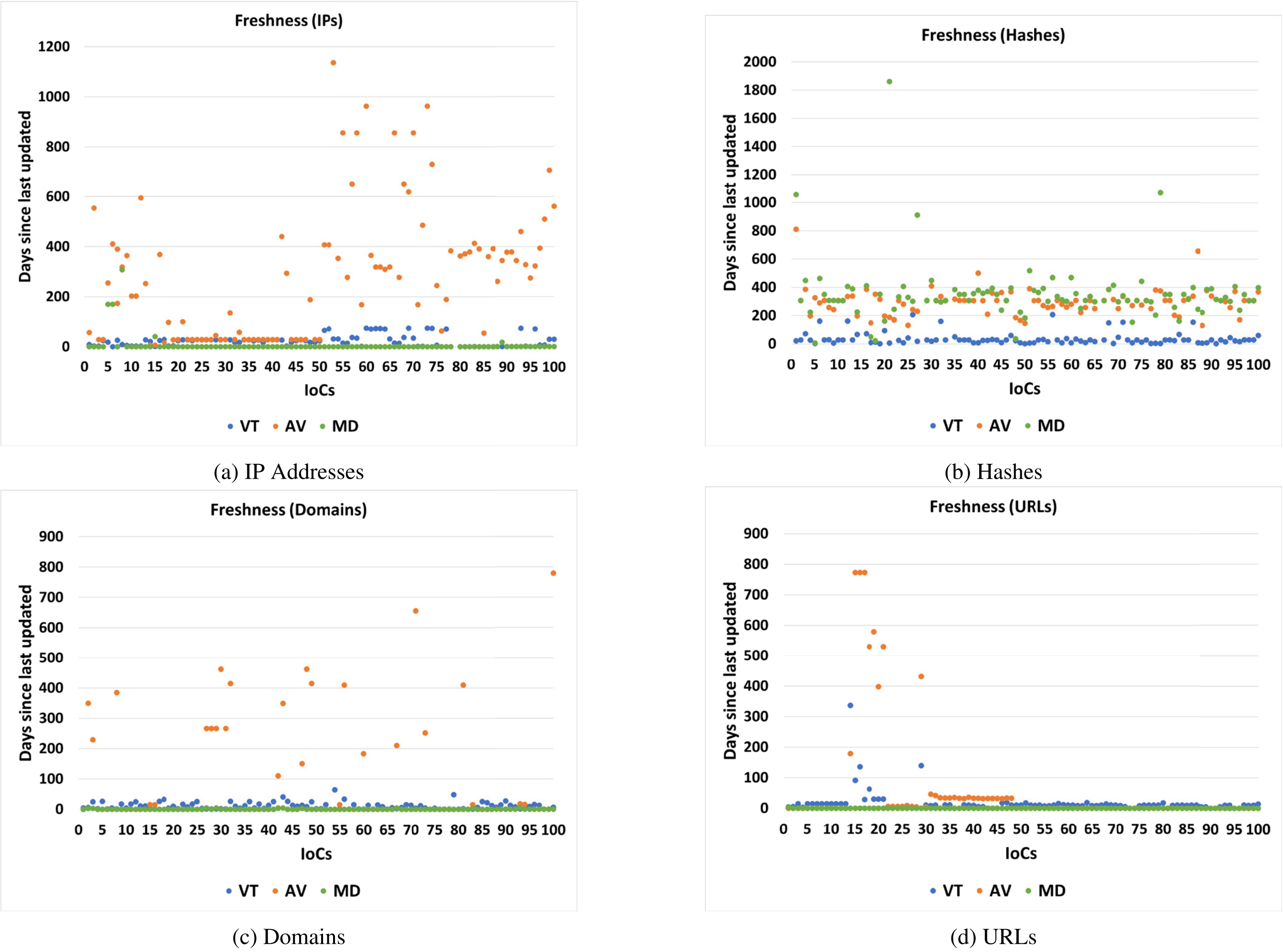
Table 5 shows that VirusTotal mostly provides the malicious verdict, as opposed to AlienVault and MetaDefender. Also, it can be seen that the verdicts of VirusTotal and MetaDefender are different from our weighted verdict in some cases, while the verdicts of AlienVault are the same as our weighted verdict in Table 5. We tested the results for multiple IoCs and found that VirusTotal marks 83.7%, AlienVault marks 74.1%, and MetaDefender marks 0.02% of IoCs as malicious. The verdict of VirusTotal matches our weighted verdict for 79.4% of the IoCs, AlienVault for 87.39% of the IoCs, and MetaDefender for 39.6% of the IoCs. This indicates that AlienVault is more reliable than the other two platforms in terms of accuracy.

We also compare our work with related works, such as [30] and [31], and find that these works focus primarily on the evaluation of platforms and feeds using different parameters. Our study not only examines the accuracy of information provided by platforms, but also provides a weighted verdict by integrating platform verdicts using a dynamically weighted evaluation mechanism. Our work also provides more complete information on IoCs in terms of accuracy, via multiple platforms, compared to [34], which evaluates the accuracy of individual threat reports.

B. FRESHNESS RESULTS

We calculated the freshness of 600 IoCs provided by each CTI platform by analyzing the time elapsed since the last updates for specific attributes mentioned in Section V-B. Although our results are based on a representative sample, the purpose is to provide a comparison for the readers to see which of the tested platforms updates their provided information more frequently, which platform provides consistent freshness results for different IoCs, and which ones provide better maintenance of IoCs. We provide freshness results for 600 IoCs that belong to 4 different IoC types to see how different platforms respond to different input types. The mean freshness for different types of IoCs is 32 days (standard deviation: 46 days), 248 days (standard deviation: 197 days), and 173 days (standard deviation: 230 days) for VirusTotal, AlienVault, and MetaDefender, respectively.

This shows that VirusTotal provides the freshest information compared to other platforms, as it has a low mean and standard deviation. The high standard deviation for other platforms calls for detailed analysis according to different types of IoC. Therefore, we compare the freshness of different IoCs provided by each platform. We tested 600 IoCs, but the number of IoCs belonging to each IoC type is different. For example, we have more IP addresses (200) than hashes (184),



(a) IP Addresses

(b) Hashes

(c) Domains

(d) URLs

FIGURE 3. Freshness comparison.

domains (104), and URLs (112). Therefore, we chose 100 IoCs of each type to provide a fair comparison and symmetric presentation of the freshness of each IoC type. It can be seen in Figure 3 that each type of IoC has different freshness. We analyze them one by one as follows.

1) IP ADDRESSES

It can be seen in Figure 3a that VirusTotal provides similar freshness for most IP addresses. The mean freshness provided by VirusTotal for IP addresses is 23.19 days, and the standard deviation is 19.38 days. AlienVault, on the other hand, is not consistent and has a mean of 218.89 days with a standard deviation of 249.58 days, which is the highest among all platforms, and shows inconsistency. The figure also shows that MetaDefender provides similar freshness results for most IP addresses, with a few exceptions. The mean freshness of MetaDefender is 5.22 days, and the standard deviation is 32.97 days, which is high compared to VirusTotal, which shows that MetaDefender is consistent, but not as consistent as VirusTotal.

2) HASHES

Figure 3b shows that VirusTotal provides similar freshness for most hashes, with a few exceptions. The mean freshness provided by VirusTotal for hashes is 43.52 days, and the standard deviation is 59.4 days. AlienVault shows a bit more consistency for hashes as compared to IP addresses, with a mean of 297.11 days and a standard deviation of 86.06 days. The figure also shows that MetaDefender provides the least fresh information among all 3 platforms for hashes. The mean freshness for MetaDefender is 362.65 days, and the standard deviation is 218.96 days, which is high compared to the other platforms. This also shows that the high standard deviation for AlienVault and MetaDefender stems from hashes.

3) DOMAINS

Figure 3c shows that VirusTotal is also consistent in providing similar freshness for domains like IP addresses and hashes. The mean freshness provided by VirusTotal for domains is 10.2 days, and the standard deviation is 11.38 days.

AlienVault shows inconsistency for domains just like it did for IP addresses, with a mean of 276.46 days and a standard deviation of 195.97 days. Another point worth mentioning here is that the missing values for AlienVault in the figure show that it does not provide any information about many domains for which other platforms provide information. The figure also shows that MetaDefender provides the freshest information among all 3 platforms for domains. The mean freshness for MetaDefender is 1.77 days, and the standard deviation is 1.78 days.

4) URLs

Figure 3d shows that VirusTotal provides consistent freshness for URLs with few exceptions. The mean freshness provided by VirusTotal for URLs is 18.35 days, and the standard deviation is 38.15 days, which shows exceptions. AlienVault shows inconsistency for URLs just like it did for IP addresses and domains, with a mean of 54 days and a standard deviation of 163.49 days. The missing values for AlienVault show that it does not have information about some URLs that are recorded by other platforms. The figure also shows that MetaDefender provides the freshest information among all 3 platforms for URLs. The mean freshness for MetaDefender is 0.03 days, and the standard deviation is 0.31 days.

It can be concluded from the analysis that VirusTotal is the most consistent platform with a low standard deviation for all IoCs, while AlienVault is the most inconsistent platform for all IoCs. MetaDefender is consistent for most IP addresses, domains, and URLs, but it does not provide consistency for hashes. Overall, VirusTotal can be considered the platform with the freshest information, and if we exclude hashes, then MetaDefender becomes the platform with the freshest information. AlienVault provides the least fresh information among all 3 platforms. This can cause collateral damage if outdated information about an IoC is used. Figures indicate that VirusTotal's freshness scores fall between those of AlienVault and MetaDefender for IP addresses, domains, and URLs. It should be noted that although MetaDefender provides more fresh information, the number of sources used by VirusTotal (around 88) is much higher than that of MetaDefender (around 7-8). Therefore, it can be concluded that VirusTotal provides information that is most recently updated and is more consistent compared to the other 2 platforms.

We also provide a comparison of our work with related work in terms of freshness. Our evaluation of freshness stands out from previous studies, such as [30] and [32], which lack freshness evaluation. Additionally, studies such as [33] and [34] do not consider the calculation of freshness from multiple platforms. Our work offers a detailed comparison that showcases the inconsistency in freshness between platforms and the potential impact on the reliability of CTI.

C. COMPLETENESS RESULTS

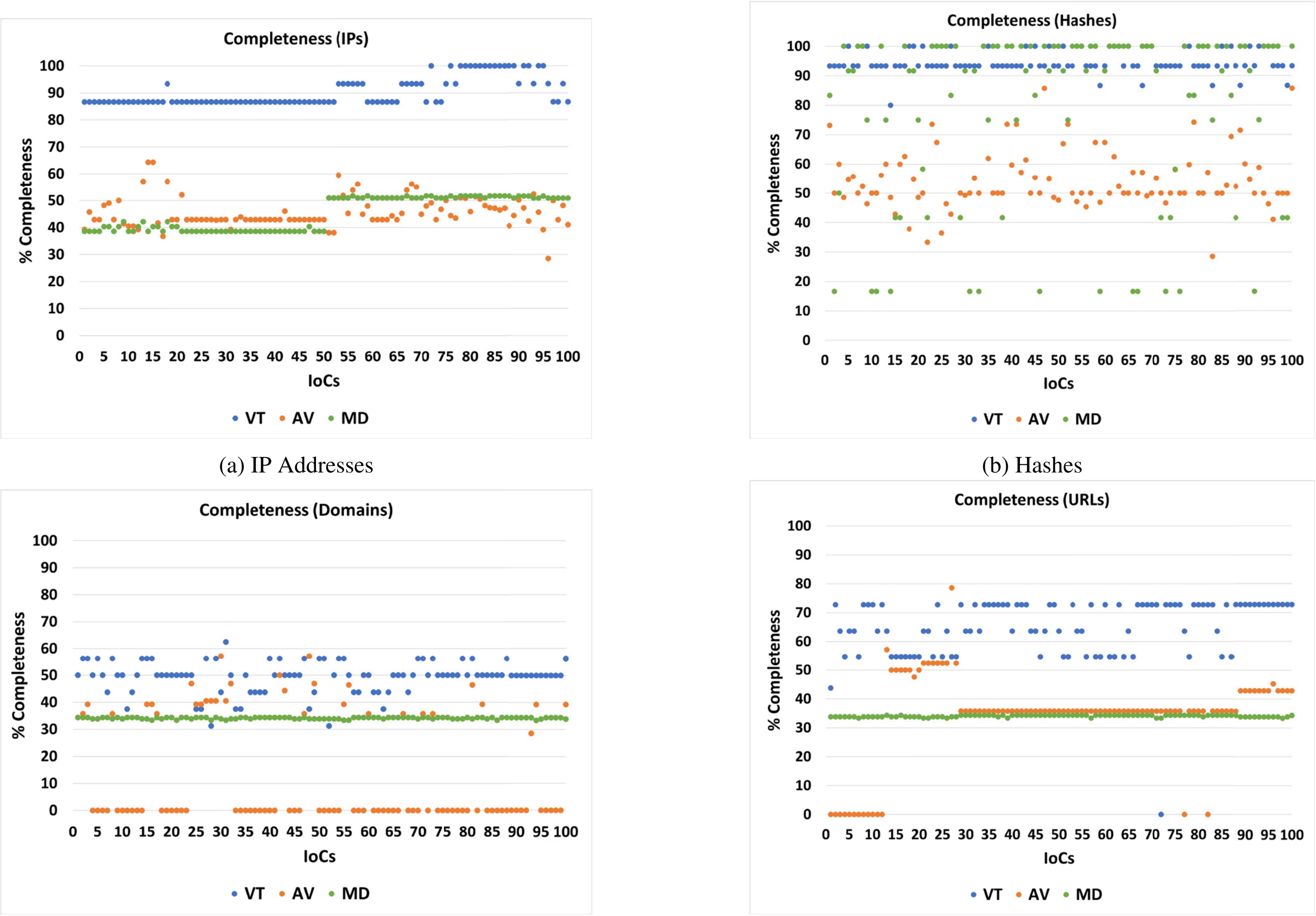
We also calculated the completeness of the information provided by each CTI platform for 600 IoCs that belong to 4 different IoC types. VirusTotal demonstrates consistent completeness, with a few exceptions. The mean completeness percentage is 84%, and the standard deviation is 21.52% for VirusTotal. AlienVault shows inconsistency for different IoCs where the completeness percentage varies drastically, and we also see that some IoCs have 0% completeness, showing that AlienVault does not have any information on some IoCs. The mean completeness percentage for AlienVault is 48%, and the standard deviation is 22.35%. MetaDefender shows inconsistent results for various IoCs, where the mean is 55% and the standard deviation is 24.7%. We perform further analysis to find the means and standard deviations for different types of IoCs. As stated in section VI-B, the number of IoCs that belong to each IoC type is different. Therefore, we chose 100 IoCs of each type to provide a fair comparison and symmetric presentation of the completeness of each IoC type, as shown in Figure 4.

1) IP ADDRESSES

It can be seen in Figure 4a that all 3 platforms provide consistent completeness for most IP addresses. VirusTotal outperforms other platforms again as it provides the most consistent results, with a standard deviation of 5.21%. AlienVault and MetaDefender are not as consistent and have standard deviations of 5.55% and 6.14%, respectively. Although the difference between standard deviations is not great, VirusTotal outperforms others in terms of mean completeness, which is 90.23% compared to 45.6% and 45.2% for AlienVault and MetaDefender. VirusTotal usually provides information on all fields that are used to calculate completeness, with a few exceptions such as `tags` and `reputation`, which are absent from the responses of many IP addresses. AlienVault does not provide information about some fields, such as `id`, `name`, `description`, `modified`, `created`, and `tags`. MetaDefender usually lacks information about `assessment` and `detect_time`.

2) HASHES

Figure 4b shows that VirusTotal provides consistent freshness for hashes, while the other two platforms show inconsistency. The mean completeness provided by VirusTotal for hashes is 94.5% (highest among 3 platforms), and the standard deviation is 3.73%. VirusTotal only lacks information about `reputation` and `tags` for a few hashes. AlienVault shows a bit more completeness for hashes as compared to IP addresses, with a mean of 54.55% and a standard deviation of 9.91%. AlienVault does not provide information about some fields, such as `id`, `name`, `modified`, `created`, and `tags`. The figure also shows that MetaDefender is the least consistent among the 3 platforms for hashes. The mean completeness of MetaDefender is 76% and the standard deviation is 31%, which is high compared to the other platforms. MetaDefender lacks information about `process_info`, `malware_family`, `malware_type`, `threat_name`, `last_start_time`,



(a) IP Addresses

(b) Hashes

(c) Domains

(d) URLs

FIGURE 4. Completeness comparison.

sanitized, votes, scan_result_history_length, data_id, and file_id for some hashes.

3) DOMAINS

Figure 4c shows that VirusTotal is not as consistent in providing completeness for domains as it was for IP addresses and hashes. The mean completeness provided by VirusTotal for domains is 49.7%, and the standard deviation is 6.07%. VirusTotal usually lacks information about network, tags, whois_date, as_owner, country, reputation, expiration_date, and regional_internet_registry. AlienVault also shows inconsistency for domains with a mean of 40.86% and a standard deviation of 6.24% (we excluded zero values to calculate the standard deviation) and provides no information for some fields such as id, name, modified, created, tags, and references. The figure also shows that AlienVault does not provide any information on many domains for which other platforms provide information. The figure highlights the consistency of MetaDefender, which provides the most consistent information among all 3 platforms for domains. The mean completeness of MetaDefender is 34.1% and the standard deviation is 0.32%. MetaDefender does not provide information about detect_time, start_time, detected_by, sources, assessment, and category for domains.

4) URLS

Figure 4d shows that VirusTotal provides the least consistent completeness for URLs (standard deviation: 8.06%) compared to IP addresses, hashes, and domains. Although the standard deviation is high, VirusTotal still outperforms other platforms, with a mean completeness of 65.68%. VirusTotal usually lacks information about tags, reputation, title, category, and referrer_samples. AlienVault has a mean completeness of 40.13% and a standard deviation of 7.25% (after excluding zeros). AlienVault fails to provide information about id, name, modified, created, tags, and references. The figure also shows that AlienVault does not have any information on some URLs that are recorded by other platforms. The figure also shows that MetaDefender is the most consistent among the 3 platforms for URLs. The mean completeness of MetaDefender is 34.05% and the standard deviation is 0.35%. MetaDefender does not provide information about `detect_time`, `start_time`, `detected_by`, `sources`, `assessment`, and `category` for URLs.

The analysis indicates that VirusTotal offers complete information about all types of IoCs compared to other platforms. Although the average percentage completeness for AlienVault and MetaDefender is quite similar, it should be noted that AlienVault considers 15 fields as opposed to 9 used by MetaDefender, although MetaDefender demonstrates better consistency for domains and URLs than the other 2 platforms. In general, it can be concluded that VirusTotal is a better choice when the purpose is to obtain more complete and consistent information for all types of IoCs.

Documented related studies either overlook the completeness of the IoC information or assess it using a limited number of fields. Our work stands out by comprehensively evaluating the completeness of the provided IoC across all fields provided by individual platforms and highlighting how each platform performs. Unlike [34], which evaluates completeness at a report level, our study dives deeper into the completeness of individual IoCs across multiple platforms, revealing critical gaps in the coverage offered by various platforms.

D. RELEVANCE RESULTS

We also calculated the relevance of the information provided by each CTI platform for 600 IoCs. Figure 5 shows that MetaDefender does not provide relevant information for any IoC, since none of the fields in the obtained response match with STIX 2.1 objects. VirusTotal provides a consistent relevance score of 4% for IP addresses only, but does not provide any relevance for other types of IoCs. AlienVault shows inconsistency for different types of IoCs, where the relevance percentage varies drastically. The mean relevance percentage for AlienVault is 30.08% and the standard deviation is 6.34%. Here, the results show an interesting finding that both VirusTotal and MetaDefender perform poorly when it comes to providing information about the objects defined in STIX 2.1. Figure 5 shows that AlienVault outperforms the other two because it provides the most relevant information.

It is also observed that the gap between the percentage completeness and relevance of the IoC information provided by AlienVault is smaller than that of VirusTotal. The gap between the percentage completeness and relevance of the IoC information exists because, although VirusTotal provides more information than other platforms, much of it is not relevant to the SDOs. This gap between the completeness and relevance of the information provided by VirusTotal needs to be reduced by including more SDOs. We can conclude that when it comes to the provision of information as per SDOs, AlienVault performs better than other platforms. Therefore, it is up to the users to choose platforms depending on which parameter is more important to them.

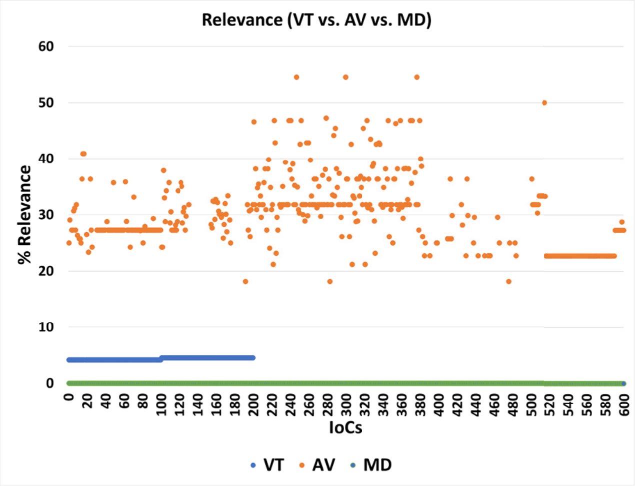


FIGURE 5. **Relevance comparison.**

Relevance evaluation is a metric often neglected in documented related work, as seen in [32] and [33], which do not address how well the provided information aligns with standardized threat intelligence frameworks such as STIX 2.1. Our research addresses this gap by evaluating the relevance of IoC data in the context of STIX 2.1, as opposed to [34], which provides relevance in terms of the user's perspective, offering a more nuanced understanding of how well these platforms contribute to actionable threat intelligence.

VII. SYSTEM IMPLEMENTATION

This research can be implemented as a system that accepts user prompts about an IoC or CTI report, retrieves information from all platforms, and generates a comprehensive threat report. This section presents the implementation of the proposed solution as a system. The system preprocesses input CTI reports available in various formats (HTML, DOCX, PDF) and converts them into text. The system does this by using the script pre_process.py, which checks the file extension and extracts the text based on the file extension. The system then extracts the IoCs from the text (Extract_IoCs.py) and feeds the extracted IoCs into multiple platforms, using the provided Python script in the referenced repository. The system then checks the number of sources in each platform, assigns weights to each platform based on the number of sources, and provides a combined verdict on whether an IoC is malicious or not using the Weighted_Verdict_General.py script. This enhances user confidence in the IoC verdict by incorporating judgments from multiple sources. The system also gathers information about the completeness, relevance, and freshness of each IoC from different platforms using AV_Eval_General.py, VT_Eval_General.py, and MD_Eval_General.py. Finally, the system provides this collected information about the IoC to ChatGPT with the prompt of writing a CTI report and returns a report containing the IoC's weighted and individual verdicts, along with evaluation results from all platforms, indicating which platform provides more complete, fresh, and relevant information.

VIII. CONCLUSION AND FUTURE WORK

This study presents an automated method for CTI extraction and evaluation, including the assessment of multiple CTI platforms. We tested 600 IoCs, and the results show that the evaluation of VirusTotal matches our weighted evaluation for 79.4%, AlienVault for 87.39%, and MetaDefender for 39.6% of the IoCs. Overall, VirusTotal classified 83.7%, AlienVault 74.1%, and MetaDefender 0.02% IoCs as malicious. VirusTotal provides a consistent freshness evaluation of all types of IoCs, whereas AlienVault shows inconsistency. MetaDefender provides consistent freshness for IP addresses, domains, and URLs, and performs poorly for hashes. The results also showed that the average freshness of the information provided by VirusTotal (32 days) is better than AlienVault (248 days) and MetaDefender (173 days) for all IoCs. If we exclude hashes, MetaDefender provides better freshness compared to VirusTotal, but the number of sources used by VirusTotal (around 88) is much higher than that of MetaDefender (around 7-8), positioning VirusTotal as the most reliable platform for freshness evaluation. VirusTotal also shows consistency in the completeness evaluation of various types of IoCs and provides more complete information about IoCs (84%) compared to AlienVault (48%) and MetaDefender (55%). AlienVault outperforms the other two in terms of relevance, as it provides the most relevant information in terms of STIX 2.1 objects. A noticeable gap exists between the completeness and relevance of information across platforms. Future work will focus on reducing this gap by enhancing the relevance of the provided information. We also intend to increase the number of platforms to increase confidence in the provided evaluation. While this study focused solely on the evaluation of IoCs, future work will include the evaluation of TTPs.

REFERENCES

[1] N. Sun, M. Ding, J. Jiang, W. Xu, X. Mo, Y. Tai, and J. Zhang, "Cyber threat intelligence mining for proactive cybersecurity defense: A survey and new perspectives," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 3, pp. 1748–1774, 3rd Quart., 2023, doi: 10.1109/COMST.2023.3273282.

[2] C. Shin, I. Lee, and C. Choi, "Exploiting TTP co-occurrence via glove-based embedding with MITRE ATT&CK framework," *IEEE Access*, vol. 11, pp. 100823–100831, 2023, doi: 10.1109/ACCESS.2023.3315121.

[3] MITRE. *MITRE ATT&CK Groups*. Accessed: Nov. 20, 2024. [Online]. Available: https://attack.mitre.org/groups/

[4] MITRE. *MITRE ATT&CK Techniques*. Accessed: Nov. 20, 2024. [Online]. Available: https://attack.mitre.org/techniques/enterprise/

[5] M. G. Gaber, M. Ahmed, and H. Janicke, "Malware detection with artificial intelligence: A systematic literature review," *ACM Comput. Surv.*, vol. 56, no. 6, pp. 1–33, Jun. 2024, doi: 10.1145/3638552.

[6] AT&T Cybersecurity. *How Does AlienVault Respond to Zero-Day Threats?*. Accessed: Jul. 27, 2024. [Online]. Available: https://success.alienvault.com/s/article/How-does-AlienVault-respond-to-Zero-Day-threats

[7] S. Ainslie, D. Thompson, S. Maynard, and A. Ahmad, "Cyber-threat intelligence for security decision-making: A review and research agenda for practice," *Comput. Secur.*, vol. 132, Sep. 2023, Art. no. 103352, doi: 10.1016/j.cose.2023.103352.

[8] H. Zhang, G. Shen, C. Guo, Y. Cui, and C. Jiang, "EX-action: Automatically extracting threat actions from cyber threat intelligence report based on multimodal learning," *Secur. Commun. Netw.*, vol. 2021, pp. 1–12, May 2021, doi: 10.1155/2021/5586335.

[9] M. R. Rahman, R. M. Hezaveh, and L. Williams, "What are the attackers doing now? Automating cyberthreat intelligence extraction from text on pace with the changing threat landscape: A survey," *ACM Comput. Surveys*, vol. 55, no. 12, pp. 1–36, 12th Quart., Dec. 2023, doi: 10.1145/3571726.

[10] D. Kim and H. K. Kim, "Automated dataset generation system for collaborative research of cyber threat analysis," *Secur. Commun. Netw.*, vol. 2019, pp. 1–10, Sep. 2019, doi: 10.1155/2019/6268476.

[11] Z. Zhu and T. Dumitras, "ChainSmith: Automatically learning the semantics of malicious campaigns by mining threat intelligence reports," in *Proc. IEEE Eur. Symp. Secur. Privacy (EuroS P)*, Mali, Apr. 2018, pp. 458–472, doi: 10.1109/EUROSP.2018.00039.

[12] Z. Long, L. Tan, S. Zhou, C. He, and X. Liu, "Collecting indicators of compromise from unstructured text of cybersecurity articles using neural-based sequence labelling," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8, doi: 10.1109/IJCNN.2019.8852142.

[13] R. Azevedo, I. Medeiros, and A. Bessani, "PURE: Generating quality threat intelligence by clustering and correlating OSINT," in *Proc. 18th IEEE Int. Conf. Trust, Secur. Privacy Comput. Communications/13th IEEE Int. Conf. Big Data Sci. Eng. (TrustCom/BigDataSE)*, Aug. 2019, pp. 483–490, doi: 10.1109/TRUSTCOM/BIGDATASE.2019.00071.

[14] N. Kim, M. Kim, S. Lee, H. Cho, B.-I. Kim, J. Park, and M.-S. Jun, "Study of natural language processing for collecting cyber threat intelligence using SyntaxNet," in *Proc. 3rd Int. Symp. Inf. Internet Technol. (SYMINTECH)*. Cham, Switzerland: Springer, 2019, pp. 10–18, doi: 10.1007/978-3-030-20717-5_2.

[15] M. Macdonald, R. Frank, J. Mei, and B. Monk, "Identifying digital threats in a hacker web forum," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2015, pp. 926–933, doi: 10.1145/2808797.2808878.

[16] R. Williams, S. Samtani, M. Patton, and H. Chen, "Incremental hacker forum exploit collection and classification for proactive cyber threat intelligence: An exploratory study," in *Proc. IEEE Int. Conf. Intell. Secur. Informat. (ISI)*, Nov. 2018, pp. 94–99, doi: 10.1109/ISI.2018.8587336.

[17] X. Liao, K. Yuan, X. Wang, Z. Li, L. Xing, and R. Beyah, "Acing the IOC game: Toward automatic discovery and analysis of open-source cyber threat intelligence," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2016, pp. 755–766, doi: 10.1145/2976749.2978315.

[18] A. Niakanlahiji, J. Wei, and B.-T. Chu, "A natural language processing based trend analysis of advanced persistent threat techniques," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 2995–3000, doi: 10.1109/BIGDATA.2018.8622255.

[19] A. Bose, V. Behzadan, C. Aguirre, and W. H. Hsu, "A novel approach for detection and ranking of trendy and emerging cyber threat events in Twitter streams," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2019, pp. 871–878, doi: 10.1145/3341161.3344379.

[20] G. Ayoade, S. Chandra, L. Khan, K. Hamlen, and B. Thuraisingham, "Automated threat report classification over multi-source data," in *Proc. IEEE 4th Int. Conf. Collaboration Internet Comput. (CIC)*, Oct. 2018, pp. 236–245, doi: 10.1109/CIC.2018.00040.

[21] R. R. Ramnani, K. Shivaram, and S. Sengupta, "Semi-automated information extraction from unstructured threat advisories," in *Proc. 10th Innov. Softw. Eng. Conf.*, Feb. 2017, pp. 181–187, doi: 10.1145/3021460.3021482.

[22] G. Husari, E. Al-Shaer, M. Ahmed, B. Chu, and X. Niu, "TTPDrill: Automatic and accurate extraction of threat actions from unstructured text of CTI sources," in *Proc. 33rd Annu. Comput. Secur. Appl. Conf.*, Dec. 2017, pp. 103–115, doi: 10.1145/3134600.3134646.

[23] G. Husari, X. Niu, B. Chu, and E. Al-Shaer, "Using entropy and mutual information to extract threat actions from cyber threat intelligence," in *Proc. IEEE Int. Conf. Intell. Secur. Informat. (ISI)*, Nov. 2018, pp. 1–6, doi: 10.1109/ISI.2018.8587343.

[24] F. Sadique, S. Cheung, I. Vakilinia, S. Badsha, and S. Sengupta, "Automated structured threat information expression (STIX) document generation with privacy preservation," in *Proc. 9th IEEE Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, Nov. 2018, pp. 847–853, doi: 10.1109/UEMCON.2018.8796822.

[25] E. Irshad and A. Basit Siddiqui, "Cyber threat attribution using unstructured reports in cyber threat intelligence," *Egyptian Informat. J.*, vol. 24, no. 1, pp. 43–59, Mar. 2023, doi: 10.1016/j.eij.2022.11.001.

[26] X. Wang, R. Chen, B. Song, J. Yang, Z. Jiang, X. Zhang, X. Li, and S. Ao, ''A method for extracting unstructured threat intelligence based on dictionary template and reinforcement learning,'' in *Proc. IEEE 24th Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, May 2021, pp. 262–267, doi: 10.1109/CSCWD49262.2021.9437858.

[27] C. Liu, J. Wang, and X. Chen, ''Threat intelligence ATT&CK extraction based on the attention transformer hierarchical recurrent neural network,'' *Appl. Soft Comput.*, vol. 122, Jun. 2022, Art. no. 108826, doi: 10.1016/j.asoc.2022.108826.

[28] Y. Gao, X. Li, H. Peng, B. Fang, and P. S. Yu, ''HinCTI: A cyber threat intelligence modeling and identification system based on heterogeneous information network,'' *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 2, pp. 708–722, Feb. 2022, doi: 10.1109/TKDE.2020.2987019.

[29] H. Jo, Y. Lee, and S. Shin, ''Vulcan: Automatic extraction and analysis of cyber threat intelligence from unstructured text,'' *Comput. Secur.*, vol. 120, Sep. 2022, Art. no. 102763, doi: 10.1016/j.cose.2022.102763.

[30] S. Bauer, D. Fischer, C. Sauerwein, S. Latzel, D. Stelzer, and R. Breu, ''Towards an evaluation framework for threat intelligence sharing platforms,'' in *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, Apr. 2020, pp. 1–10.

[31] H. Griffioen, T. Booij, and C. Doerr, ''Quality evaluation of cyber threat intelligence feeds,'' in *Proc. 18th Int. Conf. Appl. Cryptography Netw. Secur. (ACNS)*, Jun. 2020, pp. 277–296, doi: 10.1007/978-3-030-57878-7_14.

[32] L. Qiang, J. Zhengwei, Y. Zeming, L. Baoxu, W. Xin, and Z. Yunan, ''A quality evaluation method of cyber threat intelligence in user perspective,'' in *Proc. 17th IEEE Int. Conf. Trust, Secur. Privacy Comput. Commun./12th IEEE Int. Conf. Big Data Sci. Eng. (TrustCom/BigDataSE)*, Aug. 2018, pp. 269–276, doi: 10.1109/TRUSTCOM/BIGDATASE.2018.00049.

[33] T. Schaberreiter, V. Kupfersberger, K. Rantos, A. Spyros, A. Papanikolaou, C. Ilioudis, and G. Quirchmayr, ''A quantitative evaluation of trust in the quality of cyber threat intelligence sources,'' in *Proc. 14th Int. Conf. Availability, Rel. Secur.*, Aug. 2019, pp. 1–10, doi: 10.1145/3339252.3342112.

[34] D. Schlette, F. Böhm, M. Caselli, and G. Pernul, ''Measuring and visualizing cyber threat intelligence quality,'' *Int. J. Inf. Secur.*, vol. 20, no. 1, pp. 21–38, Feb. 2021, doi: 10.1007/s10207-020-00490-y.

ASAD ALI received the B.S. degree in electrical engineering from the University of Engineering and Technology, Taxila, in 2012, the master's degree in electrical engineering from the National University of Science and Technology (NUST), Pakistan, in 2015, and the Ph.D. degree in electrical engineering and computer sciences from National Yang Ming Chiao Tung University (NYCU), Taiwan, in 2022. He is currently a Researcher with the National Institute of Cyber Security (NICS), Taiwan. His research interests include cyber threat intelligence, network security, network protocols, wireless communications, cellular networks, 4G/5G communications, artificial intelligence wireless, network design, and optimization.

REN-HUNG HWANG (Senior Member, IEEE) received the Ph.D. degree in computer science from the University of Massachusetts, Amherst, MA, USA, in 1993. He is currently 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 received the Outstanding Technical Achievement Award of the IEEE Tainan Section in 2022 and the Outstanding Research Award of TACC, Taiwan, in 2023. He is also on the editorial boards of IEEE Communications Surveys and Tutorials, IEEE Transactions on Vehicular Technology, and *IEICE Transactions on Communications*. His research interests include deep learning, network security, wireless communications, the Internet of Things, and cloud and edge computing.

YING-DAR LIN (Fellow, IEEE) received the Ph.D. degree in computer science from the University of California at Los Angeles (UCLA), in 1993. He is currently a Chair Professor of computer science with National Yang Ming Chiao Tung University (NYCU), Taiwan. His research interests include network softwarization, cybersecurity, and wireless communications. His work on multi-hop cellular was the first along this line and has been cited over 1000 times. He is an IEEE Fellow and an IEEE Distinguished Lecturer. He served or is serving on the editorial boards for several IEEE journals and magazines and was the Editor-in-Chief of IEEE Communications Surveys and Tutorials, from 2016 to 2020.