2023 IEEE 14th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) | 979-8-3503-0413-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/UEMCON59035.2023.10315725

Defining Metrics for Comparing Threat Intelligence Solutions Through the Lens of the Analyst

Alexander Plaza
dept. of Computer and Information Sciences
Fordham University Graduate School of Arts and Sciences
New York, USA
aplaza4@fordham.edu

Thaier Hayajneh
Fordham Center for Cybersecurity
Fordham University
New York, USA
thayajneh@fordham.edu

***Abstract*—Threat intelligence can be a valuable tool for any cybersecurity team. However, paid solutions are often costly. Furthermore, it is difficult to determine what solution is suitable for the organization, especially when the perspective of the analyst implementing the solution is often not consulted. This paper derives metrics from comparing three threat intelligence solutions with free options through the analyst's lens to rank them. The metrics enumerate the challenges an analyst may face when adopting the solutions. Three main criteria are identified when evaluating a solution: the quality of the user interface, the quality of the programmatic interface, and the quality of the data for threat intelligence operations. Drawing from existing research on these topics, further criteria and questions are developed for each metric to "score" the solution. The scores are combined to rank the solutions. These scores and the criteria that inform them lead to a baseline that an employer or vendor can utilize to evaluate the successful implementation of a threat intelligence solution.**

***Index Terms*—Cybersecurity, Cyber Threat Intelligence, Inteligence Feeds**

I. Introduction

Cyber threat intelligence can be an essential part of a mature organization. Cyber Threat Intelligence is proactive and can come from many sources and provide insight and context into an organization's daily activities and events. A good threat intelligence practice within an organization allows an organization to learn from others or the past to prevent future attacks. Using cyber threat intelligence can be very valuable because, without adequate threat intelligence or learning from past exploitation, a threat actor could reuse the same methods and resources on various targets [1]. There is a wide range of sources now that provide various levels of information. These can range from simple indicators of compromise (IOCs) like IPs or URLs the attacker uses to perform an attack to higher-level tactics, techniques, and procedures (TTPs) that provide insight into how a threat actor goes about their attacks.

Several challenges arise when starting to utilize cyber threat intelligence sources. Organizations need to parse through the data and sources to derive their usefulness. However, this is difficult given the wide range of threat intelligence services provided. A methodology needs to be defined for investigating different sources and providing a way for organizations to navigate various solutions. The research in this paper devises metrics, and thus a methodology, for evaluating different threat intelligence solutions from the perspective of the analyst or equivalent end user who will be using and implementing the solution. This work differs from others in that it considers the user's perspective that will be working with the data. Other work tends to focus either on the intelligence source and the data quality or on the organization's higher-level considerations when choosing a threat intelligence solution.

This paper is structured in the following way. Section II describes related work in the field. Section III describes the background and challenges with cyber threat intelligence that provide context for how the metrics in the paper are derived. Section IV describes the methodology used to develop the metrics for evaluating threat intelligence solutions. Section V provides the metrics' results on three threat intelligence solutions. Section VI contains a discussion of the results that are provided in Section V. Finally, Section VII provides the conclusion based on the metrics and results indicated in the paper.

II. Related Work

Several different papers have been researched to examine and compare various threat intelligence feeds. These papers have used a variety of methodologies to compare threat intelligence solutions and quantify their results. These papers can be divided into two main groups, papers that focus on the quality of data or some aspect of the vendor to compare threat intelligence sources and papers that focus on the consumer perspective to compare sources. The groupings of the papers can be seen in Table 1.

Many papers have compared threat intelligence sources based on the intelligence provided or the service provider. FeedRank was introduced as a metric for ranking cyber threat intelligence feeds, as presented in Meier, R., et al. [2]. FeedRank used the originality of content and the reuse of entries to rank threat intelligence feeds. Sauerwein, C., et al. [3] conducted a study that compared 22 threat intelligence platforms and described several key findings about the landscape of vendors. Wagner, T. D., et al. [4] noted the importance of trust in threat intelligence and proposed a trust taxonomy to create a trusted threat-sharing environment. Their research compared 30 threat intelligence platforms on their trust functionality. Schlette, D., et al. [5] proposed many relevant "quality dimensions" to create

979-8-3503-0413-8/23/$31.00 ©2023 IEEE

TABLE I
RELATED WORKS PERSPECTIVE TABLE

Focus on the Vendor or Data	Focus on the Customer Perspective
[2] Meier, R., et al.	[8] Berndt, A., & Ophoff, J.
[3] Sauerwein, C., et al.	[9] Bouwman, X., et al.
[4] Wagner, T. D., et al.	[10] Noor, U., et al.
[5] Schlette, D., et al.	
[6] Li, V. G., et al.	
[7] Griffioen, H., et al.	

metrics to examine the dimensions in the context of cyber threat intelligence. Li, V. G., et al. [6] defined a set of metrics for characterizing data feeds and used those metrics to characterize several open-source and commercial sources. Griffioen, H., et al. [7] looked at the quality of several open-source cyber threat intelligence feeds based on an existing taxonomy derived from a presentation by Pawliński and Kompanek [11].

Some published papers shifted their approach to comparison and analysis towards the implementation and the customer of the threat intelligence source. Berndt, A., & Ophoff, J. [8] used thematic analysis to examine how each organization used cyber threat intelligence and the challenges they faced, and the benefits they saw. Bouwman, X., et al. [9] provided an empirical assessment of many threat intelligence providers, raising questions of intelligence timeliness and coverage. The researchers also interviewed professionals using the services, showing that respondents evaluated threat intelligence not necessarily through quantitative metrics but through "informal processes and heuristics" [9]. The interviews showed that many professionals valued sources that consumed less analyst time. Noor, U., et al. [10] proposed a framework that assigned weights to key performance indicators based on the customer's requirements and then ranked the threat intelligence providers.

These papers are informative and valuable, but to our knowledge, our research will differ in a few specific ways. Our research differs specifically because of our perspective on the investigation, which informs how we compare the threat intelligence solutions. We will be comparing sources from the perspective of the SOC analyst, or equivalent end user, that will be using and implementing the threat intelligence solutions. Since paid intelligence sources may be omitted in a tight budget, our investigation will focus on three free services or services with free components. Considering this perspective, we can compare the sources to examine how beneficial the information is in our day-to-day operations. Furthermore, comparing intelligence solutions can provide several challenges that employees face when implementing these threat intelligence solutions and, thus, create a general checklist of criteria for adopting a threat intelligence solution.

To our best knowledge, the current literature on comparing threat intelligence lacks the perspective of the end user that will implement the solution and lacks the analyst's consideration for what challenges must be addressed to choose and adopt a solution. At the moment, some papers may examine a variety of threat intelligence sources, but many focus mostly on paid sources instead of freely accessible ones [9], [10], or they focus on different free solutions than we do [6], [7]. When comparing threat intelligence feeds, the researchers often consider the quality of the data itself [2], which of course is essential, but it is not the only factor that comes into play when choosing a solution. Furthermore, while some papers do consider the customer perspective [9], [10], they cover mostly high-level considerations, such as relevance to the business or compatibility with existing tools. While these works contribute to our research, our analysis is grounded in the analyst's perspective, which we believe to be unique to this paper.

III. Challenges With Cyber Threat Intelligence

Many challenges need to be addressed when implementing threat intelligence, and those challenges, combined with the existing literature on the subject, drive our research toward the analyst's perspective. There are also a number of different choices we make to guide our research. We assume that we are acting from a similar position to a smaller organization, so cyber threat intelligence is an extra and not being paid for. Thus, while most papers focus on the data itself, the main factor that will dictate the success of implementing a threat intelligence solution, in the long run, is its usefulness to the analyst. Furthermore, there are a number of baseline challenges that need to be addressed to incorporate threat intelligence into an organization. An organization needs to deal with the sheer number of sources, the need to combine feeds or solutions, determine the actual relevance, and deliver the intelligence to the appropriate stakeholder [1]. Given challenges such as training, funding, and time constraints, we reason that an analyst will ultimately choose the tool they are most comfortable with [8].

IV. Methodology

We reason that for the low-level analyst, there are three categories of quality that determine the likelihood of use and the long-term success of implementing a solution. These categories are the constituent parts that define the metrics for comparing threat intelligence solutions, and they ultimately define what a threat intelligence solution should strive to achieve. The main categories include the quality of the user interface (using the graphical interface in scenarios such as incident investigations), the quality of programmatic implementation (using the API in automation projects), and the quality of data itself (the data is useful for business processes).

We focus on these aspects because our main consideration is the analyst. Past research has generally assumed that when evaluating a threat intelligence solution, the data quality is most pertinent [9]. However, Bouwman, X., et al. [9] conducted interviews with security professionals who used paid threat intelligence to assess how they found value, and the researchers found that threat intelligence is evaluated on a broad scale while optimizing for their analyst and not necessarily for the best detection of the threat [9]. One respondent was quoted saying, “So far, we don’t have any kind of scientific evaluation process or method. Just a feeling of the analysts. They are using the threat intelligence daily, and they can feel if they are comfortable with it.” [9]. Respondents evaluated threat intelligence sources mostly in a less structured and more qualitative manner compared to quantitative metrics that most research proposes [9]. For this reason, our metrics are centered around the analyst, and they bridge the gap for businesses by structuring an assessment for a threat intelligence solution. The metrics are generally simple to understand so that any user could use the metrics to assess a threat intelligence solution.

Through the investigation of the three categories, we can provide a number of metrics for evaluating a threat intelligence solution. These metrics are challenges that employees face when implementing these threat intelligence solutions, and furthermore, they are a checklist of criteria from the analyst’s perspective for a threat intelligence solution to be adopted.

Within each core metric, there are 8 subcategories or metrics. Each of these sub-metrics will be assessed for each threat intelligence solution on a scale of 0 to 2, where 0 indicates that the solution does not satisfy the metric, 1 indicates that the solution satisfies the metric somewhat, and 2 indicates that the vendor satisfies the metric exceptionally well. These totals will be tallied to provide insight for ranking the metrics based on each individual metric and their scores overall.

A. *Quality of the User Interface*

The user interface is the primary way an analyst interacts with a threat intelligence solution. Using the graphical interface of a threat intelligence solution is a quick way to gain information on an indicator of compromise (IOC), likely about an investigation to quickly decide whether the activity is suspicious or not. When it works, an exemplary user interface is hardly noticeable, but when it doesn’t work, it impedes the success of the overall threat intelligence function. For example, Sauerwein, C., et al. [12] noted that threat intelligence platforms require many manual tasks. Users need to share intelligence to contribute to the greater success of the threat intelligence platform. However, researchers note that ”most threat intelligence sharing platforms lack convenient user interfaces for quickly adding new data records and require many user interactions to achieve the desired goal [9].” In this case, bad user interface design can exacerbate existing problems in a threat intelligence solution, leading to a worse experience and product overall. For this reason, we need to consider various metrics in order to assess the quality of a threat intelligence solution from the perspective of UI.

Guntupalli, R. C. C. [13] provided a number of metrics for evaluating the quality of UI. We will be deriving some of these metrics as applicable to the threat intelligence solutions in order to compare the threat intelligence solutions based on their user interfaces. The criteria of interest include system consistency, feedback systems, promptness of action, efficiency, documentation, focus, sufficient information presented, and relevance. These metrics are defined below and provided in Table 2, which illustrates a UI metrics table with scores evaluated for three solutions.

- System consistency - The design choices and language should be consistent throughout the platform. Deviations from a standard can confuse the purpose and use of the software.
- Feedback system - Users should be involved in the process so that the product can consistently improve and evolve with customer needs.
- Promptness of action towards a request - Information that is requested or actions that are taken should be presented back to the user quickly.
- Efficiency - The amount of information presented at a given time is adequate. This concerns how information is presented and user expectations for where that information is. The information is sorted and presented in a way that the user expects in the place they expect.
- Documentation - Help content should be available to the user to aid the user with any possible problem that can arise while using the solution.
- Focus - The user interface is designed with the purpose of the solution in mind. It is readily apparent how the product functions and what each feature on the user interface does.
- Sufficient information presented - The information is adequate and matches expectations. This has to do with the amount of information so the user has everything they need, without being overwhelmed. They also do not lack valuable information and spend time trying to find it in a confusing interface.
- Relevance - The content displayed is relevant to what the user is looking for and desires. Irrelevant data is not displayed for the sake of presenting it.

B. *Quality of the Programming Interface*

The programmatic interface is how an analyst will interface with the threat intelligence solution to automate tasks. Automation is vital in cybersecurity and threat intelligence to keep up with the constantly evolving landscape and the sheer amount of data being captured and processed at any given time. For that reason, the API needs to be helpful for the analyst to implement and use a threat intelligence solution effectively. Piccioni, M., et al. [14] provided a study of API usability and addressed a number of key aspects underlying usability. The researchers highlighted four fundamental aspects that underlie API usability. These include understandability, abstraction level, reusability, and learnability [14]. Understandability involves how much effort is required to understand the semantics and nature of the API features [14]. The abstraction level refers to how well the abstraction of API function caters to usability [14]. Reusability means the API is designed to allow easy reuse and adaptation of code [14]. Finally, learnability means the API can be learned quickly and incrementally [14]. Moreover, the researchers created several questions addressing these core usability aspects. We will be using these aspects as we deem relevant to the threat intelligence solutions to compare the solutions based on their APIs. These guiding questions will inform our scores for the subcategories of the API. These metrics are defined below and provided in Table 4, which illustrates an API metrics table with scores evaluated for three solutions.

- Mapping - Refers to how well the API maps to the concepts. Do they map in a way that is expected?
- Additional Code Needs - Is additional work required to keep track of information not represented by the API in order to solve a problem or task?
- Adequate Abstraction - Is the level of abstraction adequate for the program task?
- Blackbox - Is there a need to know more about the implementation in order to use the API?
- Code Effort - Is the amount of code required adequate for the program task?
- Discretization - Is there a lot of code to change in order to write another query, or is code easily repackaged?
- Learning Curve - Is it easier to use the API once one query is used?
- Prerequisite - Is there a need to learn other underlying dependencies in order to use the API?

C. *Quality of the Data Itself*

The quality of the intelligence returned from a threat intelligence solution is essential to evaluate how helpful a solution is. Brown, S., et al. [1] examines the world of cyber threat intelligence as organizations need to gather data, process it, and share that information with others. Information can come from a variety of sources but a key part of intelligence is that it is actionable [1]. The paper presents the challenges of building out the threat intelligence management platform, and it highlights four main challenges that need to be addressed. The challenge of interest to us is the need to determine relevance. Brown, S., et al. [1] note that there are a few different indicators of data quality, including timeliness, accuracy, precision or consistency, and data relevance.

When testing the threat intelligence solutions for accuracy and consistency, we will be programmatically feeding the three services with 100 known malicious and known benign sample IPs, 200 total. We can evaluate them using accuracy, precision, and recall, to assess how well the solution can positively identify malicious and benign sample. These can be evaluated using a confusion matrix as presented in Table 6. The benign samples are a number of IPs selected from a dataset provided by Singh, A. K. [15]. The malicious samples are a number of IPs selected from a dataset downloaded from Feodo Tracker [16]. Each of the threat intelligence solutions works in slightly different ways and some calculate risk or reputation scores but all of them return how many scan sources, vendors, etc. returned a "suspicious" classification, or in the case of AlienVault, how many "pulses" the IOC shows up in. We can then classify them as malicious or benign using a threshold for how many sources correctly evaluated the IOC. For the purposes of our comparison, we will use a threshold of 4, so if 4 or more sources classify an IOC as malicious, we then consider the IOC malicious. In a real business setting these thresholds might be tuned based on false positives, alerting noise, and business needs.

Within the realm of data relevancy, there are also a number of criteria that are used to evaluate how relevant the data content is to the consumer. We have compiled the relevant indicators of data quality from the suggestions put forth by Brown, S., et al. [1]. Notably, we combined similar indicators and removed indicators that would require knowledge of the specific industry the consumer is in. This is a large part of assessing data relevance for a consumer but in order to keep the metrics focused on the solution and in keeping the metrics general, we choose to remove those indicators. These metrics are defined below and provided in Table 7, which illustrates a data quality metrics table with scores evaluated for three solutions.

- Timeliness - Up to date Intelligence must be broadcast quickly.
- Accuracy - Intelligence provides an accurate picture about a threat. Evaluated with the Accuracy metric.
- Consistency - The intelligence should provide results that are expected and are inline with expectations for noise from false positives and missed coverage from false negatives. Evaluated using precision and recall.
- Uniqueness of Collection - The source of the intelligence is unique and unavailable from other sources.
- Credibility/Reliability - The sources can be trusted.
- Depth of Collection - Multiple sources are used to compile the intelligence.
- Analyst Time - The intelligence is clear and analyzed. It does not require much manual work or extra analysis from the analyst.
- Usability - The likelihood of successful utilization of the source. The intelligence will provide the expected results if used and will be helpful to the analyst and business.

V. Results

We used our metrics to evaluate three different threat intelligence solutions with accessible free components in order to demonstrate the implementation of the metrics and the ranking of the threat intelligence solutions that a smaller organization may choose to pursue. The services were evaluated in April 2023 and the versions tested reflect what the current version was at that time.

TABLE II
UI Metrics Table

Vendor	Consistency	Feedback	Promptness	Efficiency	Documentation	Focus	Information presented	Relevance
AV	2	1	2	1	1	1	0	1
VT	2	2	2	2	2	2	2	2
IPV	2	1	2	1	1	1	2	2

TABLE III
UI Metrics Scores

Vendor	AlienVault	VirusTotal	IPVoid
Score	**9/16**	**16/16**	**12/16**

A. Explanation of results for AlienVault

AlienVault Open Threat Exchange is a platform that allows sharing from researchers and other threat data sources to help investigate threats [17]. The service provides “pulses” which are collections of IOCs that are reported by the community. Examining AlienVault with the metrics we developed above provided a number of results.

AlienVault scored 9 out of 16 for UI, the worst of the three threat intelligence solutions compared. The scores can be viewed in Table 2 and Table 3. The design choices of AlienVault are consistent, and there are no sudden changes, so it received a 2. It also received a 2 for the prompt response when querying for information on an IOC. However, it received a 1 for the feedback system because while there is a support forum, it is not easily accessible or upfront and does not appear to be fully integrated with the user experience. It received a 1 for efficiency and a 0 for sufficient information presented because the information is sorted and presented in a logical way but requires time to fully interpret because the amount of information presented is overwhelming, and it is not readily apparent how useful it is. Documentation is accessible but hidden behind some clicks so it receives a 1. For focus, the solution receives a one because the interface is sufficient for compiling information on an IOC but requires more to be easily interpretable. Finally, it receives a 1 for relevance because the content displayed is relevant, but it is not readily apparent what is useful at first glance.

AlienVault received a score of 13 out of 16, for the API metric, again the lowest, though within one point of the top two solutions. The scores can be viewed in Table 4 and Table 5. The API received a 2 for mapping because it maps intuitively to the concepts. API endpoints fall under main categories such as indicators or pulses and branch off from there so getting information about an IP would look like /api/v1/indicator/IPv4/ip/. Most work is handled by the API but some things are omitted, for example, it does not seem as though a list of IPs can be input in bulk to provide a list of output, so it receives a 1 for additional code needed. Alien Vault’s level of abstraction is adequate and provides an easy way to programmatically access the data, so it receives a 2 for adequate abstraction. There is no need to know the underlying implementation to use the API, so it receives a 2 for discretization. There is not much code needed to make a query, and the API is mapped intuitively, so little needs to change to make more API calls, thus it receives a 2 for code effort and discretization. Finally, the API is able to be learned gradually, but there is no interactive documentation, so it receives a 1 for the learning curve. There is a base familiarity needed with coding and some Python libraries to use the API so prerequisite knowledge is marked as a 1.

AlienVault received a score of 9 out of 16 for data quality. The scores can be viewed in Table 7 and Table 8. Since the data is heavily dependent on the quality of contribution from the community, the solution receives a 1 for the timeliness, uniqueness of collection, and credibility. The vast community means that there is a good depth of collection, and it receives a 2. The results of evaluating the confusion matrix results in Table 9 show that it did slightly worse than VirusTotal at evaluation IOCs, with 5 false negatives, so it receives a 1 for accuracy and consistency. For analyst time, it receives a 1 because unless there is a good “pulse” found, ad hoc queries for intelligence do not provide clear results for answering whether an IOC is malicious or not. Intelligence requires extra work to adequately use. The likelihood of successful utilization is low. Usage is dependent on the quality of pulses, and ad hoc queries do not provide readily accessible and usable reputation information so usability receives a 1.

B. Explanation of results for VirusTotal

VirusTotal is a platform that compiles information about IOCs from vendors and the community to evaluate the disposition of an IOC [18]. The service provides information about how many vendors determined an IOC to be malicious and provides community input about the IOC. Examining VirusTotal with the metrics we developed above provided a number of results.

VirusTotal scored 16 out of 16 for UI, the best of the three threat intelligence solutions compared. The scores can be viewed in Table 2 and Table 3. The service received a 2 for all categories because the design is clean and consistent. The function is readily apparent from the design. There is a “contact us” link at the bottom, and it is easy to provide feedback and user input for the function of the solution. Information is returned to the user quickly and presented in an easy-to-understand manner. Support is provided through links and an interactive bot. The content is focused on easily determining whether an IOC is malicious or not, and relevant information is displayed upfront.

TABLE IV
API METRICS TABLE

Vendor	Mapping	Additional Code	Adequate Abstraction	Blackbox	Code Effort	Discretization	Learning Curve	Prerequisite
AV	2	1	2	2	2	2	1	1
VT	2	1	2	2	2	2	2	1
IPV	2	1	2	2	2	2	2	1

TABLE V
API METRICS SCORES

Vendor	AlienVault	VirusTotal	IPVoid
Score	**13/16**	**14/16**	**14/16**

TABLE VI
CONFUSION MATRIX

CONFUSION MATRIX		Predicted	
		Malicious	Benign
Actual	Malicious	TP	FN
	Benign	FP	TN

Reading the Confusion Matrix:
TP - True Positive - Positively Identified Malicious IOC
FP - False Positive - Falsely Identified Malicious IOC
TN - True Positive - Positively Identified Benign IOC
FN - True Positive - Falsely Identified Benign IOC

VirusTotal scored 14 out of 16 for API, tied for the best of the three threat intelligence solutions compared. The scores can be viewed in Table 4 and Table 5. It received a 1 for additional code because, while the API handles most work, some functionality may require outside work. For example, a list of IPs cannot be input in bulk to produce a list of responses. Also, some things need to be calculated on the programmers' side, such as calculating the total number of vendor sources compiled to get detections on an IOC. Also, similar to AlienVault's API there is a base familiarity needed with coding and some Python libraries to use the API so prerequisite knowledge is marked as a 1. Otherwise, VirusTotal receives a 2 for all categories under abstraction and reusability. Learnability is marked as a 2 as well because learning is easier as time goes on and there is interactive documentation where the user can interact with the API in the browser to practice and test.

VirusTotal received a score of 13 out of 16 for data quality, tied for the best of the three threat intelligence solutions compared. The scores can be viewed in Table 7 and Table 8. Since data depends on vendors and the community, the timeliness, uniqueness of the collection, and credibility all receive a 1. Accuracy and consistency receive a 2 because the solution perfectly evaluated malicious and benign IOCs, as shown in Table 9. The depth of the collection is strong due to the number of vendors and community sourcing, so it receives a 2. It is very easy to use and determine, based on vendors and community, whether an IOC is malicious or not, so analyst time and usability are marked as 2.

C. Explanation of results for IPVOID

IPVoid is a platform that compiles information about IOCs from several sources [19]. The service has many IP tools that can be used to evaluate an IP. The primary tool used is the IP Blacklist check which allows checking an IP against several block lists to generate a "reputation" for the IP. Examining IPVoid with the metrics we developed above provided a number of results.

IPVoid scored 12 out of 16 for UI, in between VirusTotal and AlienVault. The scores can be viewed in Table 2 and Table 3. The design is consistent throughout the application so it receives a 2 for system consistency. Providing feedback is difficult, it requires looking for a "contact us" link and requires emailing the developers so the feedback system is marked as 1. Information is quickly returned on a given IP so promptness is a 2. The information presented is adequate, especially when selecting the tool to use, but before then the user needs to figure out what tool to use. Other solutions work simply by taking the IOC and doing the rest of the work for the user, so efficiency and focus are marked as 1. There is help available but it is mostly in the form of linking to the source of the threat intelligence so documentation is marked as 1. The amount of information returned is marked as 2 because it is useful and not overwhelming. Relevance is marked as 2 because the content displayed is catered toward detecting the reputation of an IP and it works well for that purpose.

IPVoid scored 14 out of 16 for API, tied with VirusTotal for the highest score. The scores can be viewed in Table 4 and Table 5. The same problems that hurt VirusTotal's API hurt IPVoid. Much of the work is handled by the API. Fields like risk score, number of sources used or queried, number of detections, and detection rate are returned in an object directly to the user. Nonetheless, passing in multiple IPs at once is not supported. The learning curve is gradual and once one query is used it is very easy to move on to another one based on similar structures and formatting. Similar to previous APIs there is a base familiarity needed with coding and some Python libraries to use the API so prerequisite knowledge is marked as a 1.

IPVoid scored 7 out of 16 for data quality, the worst solution. The scores can be viewed in Table 7 and Table 8. Data again depends on the community and the blocklists sourced so the timeliness, uniqueness of the collection, and credibility all receive a 1. Accuracy and Consistency receive a 0 because the solution did poorly in evaluating the IOCs, as shown in Table 9. The collection depth is vital due to the number of

TABLE VII
DATA QUALITY METRICS TABLE

Vendor	Timeliness	Accuracy	Consistency	Uniqueness of Collection	Credibility/Reliability	Depth of Collection	Analyst Time	Usability
AV	1	1	1	1	1	2	1	1
VT	1	2	2	1	1	2	2	2
IPV	1	0	0	1	1	2	1	1

TABLE VIII
DATA QUALITY METRICS SCORES

Vendor	AlienVault	VirusTotal	IPVoid
Score	**9/16**	**13/16**	**7/16**

TABLE IX
CONFUSION MATRIX FOR EACH SERVICE

Actual		Predicted	
VT		Malicious	Benign
	Malicious	100	0
	Benign	0	100
AV		Malicious	Benign
	Malicious	95	5
	Benign	0	100
IPV		Malicious	Benign
	Malicious	16	84
	Benign	0	100

VT - Accuracy = 100% Precision = 100% Recall = 100%
AV - Accuracy = 97.5% Precision = 100% Recall = 95%
IPV - Accuracy = 58% Precision = 100% Recall = 16%

blocklists sourced, so it receives a 2. The inconsistency with evaluating IOCs means that extra analyst time would be needed to filter the most valuable and up-to-date sources. Analyst time and usability are marked as 1.

VI. DISCUSSION

The results show that VirusTotal had the best UI, VirusTotal and IPVoid had the best API, and VirusTotal had the most quality data. Overall, VirusTotal was the best solution. This is somewhat expected as VirusTotal is generally the most well-known of the solutions. Given the results, VirusTotal excelled in its UI and quality of data. AlienVault dropped off with a cluttered and generally confusing UI that made analyzing an IOC tedious. AlienVault held the analyst back with its UI compared to the other two. Furthermore, data quality hurts IPVoid. This could be due to the timeliness of the blacklists that source IPVoid's evaluation. Finally, the API component is comparable among the solutions. This is also somewhat expected since mapping the concept of getting information about an IOC translates well to an API. VirusTotal was able to rank the best because it consistently performed well on each metric. AlienVault and IPVoid could each be further improved with these metrics in mind to create a threat intelligence ensemble that could be useful for analysts.

VII. CONCLUSION

Cyber threat intelligence can be an incredibly useful tool, but the number of solutions in the field and the differences in quality require a defined methodology for evaluating it. This paper devises the metrics needed to compare and rank threat intelligence solutions. The paper compares three threat intelligence solutions, AlienVault OTX, VirusTotal, and IPVoid, to illustrate the usefulness of the metrics and demonstrate how the threat intelligence solutions compare to each other. The three main criteria that are identified when evaluating a solution are the quality of the user interface, the quality of the programmatic interface, and the quality of the data. By utilizing these scores, an organization can easily identify the best threat intelligence solution to pursue.

REFERENCES

[1] Brown, S., Gommers, J., & Serrano, O. (2015, October). From cyber security information sharing to threat management. In Proceedings of the 2nd ACM workshop on information sharing and collaborative security (pp. 43-49).

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

[3] Sauerwein, C., Fischer, D., Rubsamen, M., Rosenberger, G., Stelzer, D., & Breu, R. (2021, August). From threat data to actionable intelligence: an exploratory analysis of the intelligence cycle implementation in cyber threat intelligence sharing platforms. In Proceedings of the 16th International Conference on Availability, Reliability and Security (pp. 1-9).

[4] Wagner, T. D., Palomar, E., Mahbub, K., & Abdallah, A. E. (2018). A novel trust taxonomy for shared cyber threat intelligence. Security and Communication Networks, 2018.

[5] Schlette, D., Böhm, F., Caselli, M., & Pernul, G. (2021). Measuring and visualizing cyber threat intelligence quality. International Journal of Information Security, 20, 21-38.

[6] Li, V. G., Dunn, M., Pearce, P., McCoy, D., Voelker, G. M., & Savage, S. (2019). Reading the tea leaves: A comparative analysis of threat intelligence. In 28th USENIX security symposium (USENIX Security 19) (pp. 851-867).

[7] Griffioen, H., Booij, T., & Doerr, C. (2020). Quality evaluation of cyber threat intelligence feeds. In Applied Cryptography and Network Security: 18th International Conference, ACNS 2020, Rome, Italy, October 19–22, 2020, Proceedings, Part II 18 (pp. 277-296). Springer International Publishing.

[8] Berndt, A., & Ophoff, J. (2020). Exploring the value of a cyber threat intelligence function in an organization. In Information Security Education. Information Security in Action: 13th IFIP WG 11.8 World Conference, WISE 13, Maribor, Slovenia, September 21–23, 2020, Proceedings 13 (pp. 96-109). Springer International Publishing.

[9] Bouwman, X., Griffioen, H., Egbers, J., Doerr, C., Klievink, B., & Van Eeten, M. (2020). A different cup of TI? the added value of commercial threat intelligence. In 29th USENIX security symposium (USENIX security 20) (pp. 433-450).

[10] Noor, U., Anwar, Z., Altmann, J., & Rashid, Z. (2020). Customer-oriented ranking of cyber threat intelligence service providers. Electronic Commerce Research and Applications, 41, 100976.

[11] P. Pawliński and A. Kompanek, "Evaluating Threat Intelligence Feeds ," in FIRST Technical Colloquium for Threat Intelligence, Accessed: Apr. 11, 2023. [Online]. Available: https://www.first.org/resources/papers/munich2016/kompanek-pawlinski-evaluating-threat-ntelligence-feeds.pdf

[12] Sauerwein, C., Sillaber, C., Mussmann, A., & Breu, R. (2017). Threat intelligence sharing platforms: An exploratory study of software vendors and research perspectives.

[13] Guntupalli, R. C. C. (2008). User interface design: methods and qualities of a good user interface design.

[14] Piccioni, M., Furia, C. A., & Meyer, B. (2013, October). An empirical study of API usability. In 2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (pp. 5-14). IEEE.

[15] Singh, A. K. (2020). Malicious and benign webpages dataset. Data in brief, 32, 106304.

[16] “Botnet C2 IP Blocklist ,” Feodo Tracker. [Online]. Available: https://feodotracker.abuse.ch/blocklist/. [Accessed: 13-Mar-2023].

[17] “AlienVault - Open Threat Exchange,” AlienVault Open Threat Exchange. https://otx.alienvault.com/

[18] “Virustotal.” VirusTotal, https://www.virustotal.com/.

[19] “IP Address Tools, Network Tools, DNS Tools.” IPVoid, https://www.ipvoid.com/.