A different cup of TI? The added value of commercial threat intelligence

Xander Bouwman, *Delft University of Technology, the Netherlands;* Harm Griffioen, *Hasso Plattner Institute, University of Potsdam, Germany;* Jelle Egbers, *Delft University of Technology, the Netherlands;* Christian Doerr, *Hasso Plattner Institute, University of Potsdam, Germany;* Bram Klievink, *Leiden University, the Netherlands;* Michel van Eeten, *Delft University of Technology, the Netherlands*

https://www.usenix.org/conference/usenixsecurity20/presentation/bouwman

This paper is included in the Proceedings of the 29th USENIX Security Symposium.

August 12–14, 2020

978-1-939133-17-5

Open access to the Proceedings of the 29th USENIX Security Symposium is sponsored by USENIX.

A different cup of TI? The added value of commercial threat intelligence

Xander Bouwman[1], *Harm Griffioen*[2], *Jelle Egbers*[1],
Christian Doerr[2], *Bram Klievink*[3], *and Michel van Eeten*[1]

[1]*Delft University of Technology, the Netherlands*
[2]*Hasso Plattner Institute, University of Potsdam, Germany*
[3]*Leiden University, the Netherlands*

Abstract

Commercial threat intelligence is thought to provide unmatched coverage on attacker behavior, but it is out of reach for many organizations due to its hefty price tag. This paper presents the first empirical assessment of the services of commercial threat intelligence providers. For two leading vendors, we describe what these services consist of and compare their indicators with each other. There is almost no overlap between them, nor with four large open threat intelligence feeds. Even for 22 specific threat actors – which both vendors claim to track – we find an average overlap of only 2.5% to 4.0% between the indicator feeds. The small number of overlapping indicators show up in the feed of the other vendor with a delay of, on average, a month. These findings raise questions on the coverage and timeliness of paid threat intelligence.

We also conducted 14 interviews with security professionals that use paid threat intelligence. We find that value in this market is understood differently than prior work on quality metrics has assumed. Poor coverage and small volume appear less of a problem to customers. They seem to be optimizing for the workflow of their scarce resource – analyst time – rather than for the detection of threats. Respondents evaluate TI mostly through informal processes and heuristics, rather than the quantitative metrics that research has proposed.

1 INTRODUCTION

Cyber threat intelligence (TI) has acquired a strong presence in the market for security services. TI is, simply put, information on attacker behavior that can be used to adapt one's defenses to the threat landscape. A well-known form of TI are indicators of compromise (IOCs): machine-readable data feeds with resources – typically IP addresses, domains, or file hashes – that have been observed in malicious behavior. Commercial vendors also regularly release analyst-focused reports which go beyond indicators and paint a picture of the tactics, techniques and procedures (TTPs) of specific actors.

While organizations can generate limited forms of TI from their internal systems, they increasingly turn to procuring TI from external sources. Roughly speaking, there are three sources of TI: open, shared and paid. Open TI (OTI) typically consists of public lists of indicators, such as Abuse.ch [1], AlienVault [3], and Malwaredomains.com [12]. Here, threat intelligence is often another name for abuse feeds and blacklists. Shared TI (STI) is sourced via trusted communities where members exchange their own threat information, and where there is no payment associated with this exchange. The community can be formalized, such as the membership in an Information Sharing and Analysis Center (ISAC), or it can be informal, where membership is based on personal trust relationships. The third source is paid TI (PTI). A 2019 survey amongst 1,908 IT and security professionals in North America and the U.K. found that 44% of respondents say that the primary source of threat intelligence in their organization is purchased [31]. The commercial market for TI products and services is valued at over USD 5 billion globally and predicted to triple in the next five years [22].

Despite its importance, the commercial market for TI is largely uncharted territory in academic research. In light of the high fees and license restrictions associated with PTI, this is understandable. The nearest work, conducted by Li et al. [21], was not focused on PTI specifically, but did include two edge cases of paid services. These services did not provide original TI sources, but helped curate and aggregate otherwise free or low-end indicator feeds. They were priced in the range of USD 1-10k per year. In this paper, we focus on high-end, original TI sources from market leaders, which charge around USD 100-650k per year (Section 5.2). Except for the edge cases, prior research has analyzed only open sources in the form of abuse feeds and blacklists. Recent work [21, 15] confirmed the results of earlier studies [39, 25, 19]: the feeds are highly heterogeneous, overlap among any two feeds is often very small and never more than 10%, and the feeds are dominated by 'singletons'; entries that do not appear in any other feeds. This also explains why, in order to achieve better coverage of their threats, network defenders combine on average 7.7 TI sources [31].

No prior work has analyzed what TI services the high-

end vendors in the commercial market offer, how their data compares to OTI and how customers evaluate PTI. This lack of insight also affects firms facing security investment make decisions. The subscription fees of the market leaders usually range in the hundreds of thousands of dollars per year [20]. Buyers have to decide if these substantial costs are worth it compared to open or shared TI. According to research firms Gartner and Forrester, purchasers are struggling to compare services [20, 14]. Prior work [21, 33, 29, 24, 27, 15] assumed that buyers are in need of quantitative metrics on criteria like coverage, volume, accuracy, and timeliness. This assumption, however, has not been empirically validated via a user study.

This paper reports on a mixed-method study that sheds light on the market for paid TI and reduces the information asymmetry currently confronting buyers. We present the first qualitative and quantitative analysis of commercial TI services. Using grounded theory, we analyze 14 interviews with PTI customers to understand what sources they are buying, how they use them and how they evaluate added value in the absence of independent analyses. We complement this user study with a high-level quantitative analysis of services of two market leaders, comparing these to several open TI feeds.

With this approach, we aim to answer the following questions: *(i)* What do paid TI services consist of? *(ii)* How is paid TI different from open TI? *(iii)* How do customers use TI and perceive value? We make the following contributions:

- We present the first empirical analysis of paid TI from market leaders, comparing the data of two leading vendors to each other and to OTI.
- We demonstrate that there is almost no overlap between paid and open TI sources, signaling that they capture a different part of the threat landscape. Surprisingly, there is also little overlap among the two paid feeds, although they focus on the same topic areas. Even when tracking the same 22 threat actors, only 2.5 % to 4.0% of indicators are found by both vendors, depending on the type of indicator. Timeliness, as measured on the small overlap among sources, shows delays of more than a month between sources. These findings suggest serious issues with coverage and timeliness of commercial TI.
- We find that customers use TI less exclusively for network detection than is often assumed. Other use cases include understanding of the threat landscape, informing business decisions, awareness programs, and threat hunting.
- Where prior work has assumed that customers work with metrics like coverage in order to evaluate TI, we find a different logic in practice. Customers value the better curated and more selective paid TI sources over other sources – especially the larger and potentially more noisy open sources – because they consume less analyst time. Surprisingly, the fact that smaller sets may also imply more false negatives, i.e., limited coverage, is much less a concern. Costs hardly play a role at all.

2 BACKGROUND

A major challenge in an organization's risk management process is to identify and understand all relevant threats. For many cybersecurity threats, their existence, likelihood and impact are not known to the organization. Threat intelligence services claim to address this by providing the necessary information to identify risks, aid in their quantification, guide the selection of controls, provide indicators to detect adversaries, and show possible courses of action.

To a limited extent, TI can be extracted from an organization's own security controls. Think of firewall logs that observe external IP addresses involved in brute-forcing SSH passwords or of spam filters that contain emails with phishing URLs. Such IP addresses and URLs are typically referred to as indicators of compromise (IOCs). The downside of producing in-house TI is that any single organization will only observe a small fraction of the threat landscape. Another challenge is that extracting the most relevant signals, rather than the most obvious ones, requires resources and expertise. For this reason, many organizations acquire external sources of threat intelligence, like open sources or sharing communities.

Compared to OTI and STI, paid threat intelligence makes a different value proposition. It does not only contain indicators and information observed from an ongoing threat somewhere else (e.g., an IP address that has brute-forced or phished a different organization), but insights based on active research, proprietary vantage points, and potentially insider information by a specialized provider. PTI is often perceived as being of higher quality and providing better and earlier warning. To protect value and exclusivity, vendors typically vet their customers, so that adversaries cannot readily see that their activities were detected. Vendors also typically provide integration into products, like a malware detection middlebox. This provides them with unique visibility across the networks of their clients, where they have probes for monitoring and middleboxes for protection. The vendors in this market typically claim that with these vantage points and cloud-based aggregation and analysis of data, they can track advanced attacks and threat actors. Vincenzo Iozzo, Senior Director at CrowdStrike, a key player in this market, articulated this advantage as follows: "If you [the attacker] get detected on one machine, all of your offensive infrastructure has to be scrapped" [17]. Later in this paper, we explore the extent to which this advantage allows PTI vendors to uncover offensive infrastructure.

In general, assessing quality of TI data through metrics is very hard, as there is no ground truth on global maliciousness [25]. Below we briefly describe some quality characteristics and metrics that were developed in earlier work, as these concepts return at various points in the paper.

Coverage pertains to the proportion in which the TI actually observes the attacks it promises to observe – i.e., the proportion to which the intended indicators are actually con-

tained in the feed [21, 29]. The opposite is how much relevant information it fails to provide, which is the rate of false negatives [24, 27, 25].

Accuracy is the proportion of indicators in a feed that actually belong in the feed. This pertains to the degree of true positives. Its opposite is the degree of false positives [21, 24, 27] and this a major factor in the value of a data-source. Depending on the organization, even a relatively low number of false positives may lead to notification fatigue and thereby reduce the feed value.

Timeliness of information in a feed pertains to the time gap between an attack vector occurring and its associated indicators being included in the feed [28]. Some authors refer to this as latency [21], or speed [24]. Timeliness may be essential for the value of a feed used in active defense mechanisms, e.g., intrusion detection middle boxes, but for forensics purposes this is less critical.

Ingestibility relates to the structure and consistency of structure in a feed [28], i.e., how well it can be automatically processed.

Relevance of a TI feed [28], also referred to as fitness [29], describes how well the indicators and contents of a feed fit an organization's use case. A TI feed rarely attempts to cover *all* malicious activity but often focuses on a certain type of threat. A feed may hence be of low value if it is not relevant to an organization.

In summary, assessing the quality of TI feeds is a difficult subject for OTI, and–due to limited availability–even more so for PTI. In the remainder of the paper, we will produce some of these metrics for PTI, replicating the approach of [21] (Sections 5 & 6), while also using interviews with users to see how organizations assess the value of PTI with or without such metrics (Section 6).

3 ETHICS

Research on PTI data is hindered by high fees and license restrictions. Firms who buy a subscription are not allowed to share the data with third parties, including researchers. We were able to overcome this barrier when one of the authors was set up as an intern in an organization that has a subscription to the TI of two market leaders. These two vendors are included in Gartner's market overview [20] and they are positioned among the most expensive suppliers. The analysis of the data was conducted on the organization's premises and within the conditions of their vendor license agreements. Only the aggregate results of the analysis were shared with the rest of the author team. The organization was willing to collaborate with our study on condition that we would not name them nor the vendors included in the study, and that we would not include characteristics of the feeds that would make the vendors easily identifiable. Hence, some numbers are reported as ranges rather than exact counts.

Our second data source consists of interviews with 14 security professionals who work with paid threat intelligence. We received approval from our Institutional Review Board for this human-subjects research. All respondents explicitly gave their consent to have their interview transcribed and used in this study. To enable our respondents to talk about their use and evaluation of PTI without risking reputational repercussions for themselves or their organizations, we have anonymized their identities. To provide context for specific quotes, we describe a respondent's role and sector. Respondents were provided with information on the research objectives and the interview protocol before the interview. Afterwards they could check and correct quotes attributed to them.

4 METHODOLOGY

We use a mixed-methods approach, combining a qualitative user study with a quantitative analysis of the TI data. To answer the first question – what does paid TI consist of? – we report on the answers from our respondents, rather than impose our own definition of TI. We complement these answers with a high-level description of the feeds and reports that were provided by two market leaders from 2013-2018. For the second question – how PTI compares to OTI – we analyzed indicator feeds of the two market leaders and four open feeds. The third question – how do customers use TI and perceive its value? – we answered based on our interviews. We answer the questions consecutively in Sections 5, 6, and 7. Here, we describe the data collection and analysis. A high-level overview is presented in Table 1.

4.1 Threat intelligence data

As described in Section 3, an internship of one of the authors allowed for access to the TI services of two market leaders, both included in Gartner's market overview [20]. The organization which provided us with the access chose these vendors because they are among the market leaders and were deemed to have the most relevant TI. It did not conduct any analysis of the overlap in indicators among the two vendors before we started our research.

The offerings of the vendors consisted of 5-10 subsets around specific topic areas, e.g., 'financial industry', 'cyberespionage' or 'cybercrime'. Customers typically subscribe to the subsets most relevant for them, rather than to all. As explained in Section 3, we cannot identify the vendors, nor can we list the exact topic areas the customer organization subscribed to. We can only say that we had access to 3-5 subsets for each of the vendors and that these subsets focus on the same topic areas. The selection of these topic areas likely influences how the indicators are distributed over target industries, as visualized in Figure 4. The degree of indicator overlap between vendors might also vary across focus areas.

Type	Data	Source	Contents	Period
PTI	Paid TI services	Two leading providers	7,308 reports; and 420,173 indicators (IPs, domains, MD5)	2013/01/01–2018/12/31
OTI	Alienvault OTX	Community-aggregator	59,290 IPs	2018/10/01–2018/10/31
OTI	Blocklist.de	Independent	121,540 IPs	2018/10/01–2018/10/31
OTI	CINSscore	Security firm	55,906 IPs	2018/10/01–2018/10/31
OTI	Emergingthreats	Security firm	876 IPs	2018/10/01–2018/10/31
TI	14 interviews	Professionals using PTI	Qualitative findings	2019/08/27–2019/12/23

Table 1: Data sources for this mixed-methods study.

We assessed only the indicators that were packaged with the TI reports that the vendors release. These reports analyze the developments in the threat landscape and actor groups. One vendor also had a bulk feed of indicators that were not associated with reports. We did not include this in our comparison, because the other vendor provided no such feed.

We assessed the overlap between the paid TI sources as follows. Vendors label their intelligence products with metadata, referring to a specific threat actor for 35% of their reports and 60% of all indicators. Vendors use their own naming schemes for threat actors. The same actor may thus appear as Deep Panda, APT19, or KungFu Kittens across different vendor reports. We mapped the names used by the vendors to a common set of threat actors using an overview maintained by well-known security researcher Florian Roth [32]. About 30% of all indicators could be mapped to a common threat actor listed in the overview. These indicators form the basis for the analysis of the overlap visualized in Figure 3. In sum, we measure overlap specifically where the vendors claim that they are tracking the same actor groups. The results are reported in Section 5.1. In Section 6.2, we also report on the overall overlap across the feeds.

To map the distribution of indicators across targeted industries, we mapped 179 labels from both vendors to a common set of 16 categories. These are listed in Figure 4.

To represent OTI, we collected four freely available sources. Three indicator feeds, for the high degree to which they were reused by other open sources [15] – which we expected might lead to reuse in paid TI– and one community-based aggregator that enables its users to extract indicators from blogs and reports, which vice versa we expected might lead to reuse of PTI indicators in OTI. This way, we selected for an upper bound of overlap between the two types. We compare one month of data from these collected OTI feeds (October 2018) against five months months of indicators from the two PTI vendors (July to December 2018), in order to compensate for the higher churn in the OTI feeds and, again, to find an upper bound in overlap. The preparation of the OTI data consisted of deleting all duplicate IPs except for the first occurrences, then removing all IPs present on the first measurement day of each set, as it was unsure when those were first received. To remove inconsistencies from the threat intelligence, we normalized them as follows. For URLs and domains, we removed `http(s)://`, but kept prefixes that are part of the domain – including the `www` domain which may point to another location. For file hashes in indicators, we compared only MD5 hashes for each file to prevent duplicate results. For IPs, no normalization was necessary, as the format was consistent between sources. For the reports, we removed all punctuation and casing from labels.

When matching indicators, we had to assume a time-to-live period during which indicators could be reasonably assumed to remain valid. Prior work chose periods of 30 days [21] or does not make such assumptions explicit. Since overlap among sources is an important indicator of coverage, we proceeded conservatively and chose a time-to-live period of 360 days. This again to provide us with an upper bound, i.e. over- rather than under-estimating the amount of overlap.

4.2 Interviews

As the field of PTI is marked by ambiguous terms and complex practices, we opted for a data collection method that can help us unpack the ambiguity, namely semi-structured interviews in a grounded-theory approach – similar to [41]. Grounded theory means that the researchers draw conclusions through a reflexive process of inductive reasoning [10, 9]. This approach means that our findings can help form an understanding about how the market for PTI functions, but not about how the reported views are distributed across the global population of security professionals who use PTI.

Between August and December of 2019, we conducted 14 interviews with professionals who work with PTI. Participants were selected from different sectors in business and government (Table 2). A requirement was that their organization purchased commercial threat intelligence. We contacted participants via their personal networks, as well as via LinkedIn. Geographically, the participants were located in The Netherlands (11) and Japan (3). They fulfilled positions from analyst to management in both security operations and threat intelligence teams. Two of the respondents worked at Managed Security Service Providers (MSSPs) which use TI to protect the networks of their clients. In the sample, the

INDUSTRY	Respondents *n=14*
Finance	4
R&D	3
Government	2
Managed Security Service Provider	2
Infrastructures	2
Oil and gas	1

Table 2: Interviewees by industry, all experienced with PTI.

financial industry supplies the most interviewees. While this might simply be an artefact of our recruitment effort, it is consistent with the fact that financial sector firms have the highest investment levels in cybersecurity [40] and are thus more likely to acquire expensive PTI subscriptions. One of our respondents, a teamlead TI at an MSSP, remarked: "I would say that the financial industry is one of the most mature sectors and can do more with threat intel."

The interviews we conducted with security professionals were semi-structured, meaning that respondents' views were central and conversations were open-ended [16]. The researchers did not ask loaded questions to avoid steering the conversation [5, 9]. A simple interview protocol was used in which participants were asked about: *(i)* Their definition of threat intelligence; *(ii)* What commercial sources their organization pays for and, for each of the sources, discuss costs, source properties, use-cases, and valuable organizational outcomes; *(iii)* Their experiences with use of non-commercial sources; *(iv)* Their reasons for having discontinued a source, if ever. The full interview protocol is included in Appendix A.

This approach was chosen deliberately to minimize the influence of pre-conceived ideas about what TI is and how it should be evaluated. Due to the nature of open questions it also means that participants' answers cannot be seen as an exhaustive description of their opinion. Rather, outcomes of the interviews are the perceptions that participants have prioritized, that were 'top of mind' for them at the time.

We transcribed the interviews and coded them using the ATLAS.ti software. Analytic codes were drawn from the interviews and used as labels to identify recurring answers. The codebook was iteratively saturated over the course of 9 interviews [9]; it reached the point where no new codes could be observed from the interviews (see Figure 1). One researcher carried out initial coding. We developed the codebook through meetings with co-authors and two other researchers, each time independently coding and then refining codes as needed. According to [23], this is a suitable way to ensure reliability of findings for our purposes. This eventually led to analytic codes for TI service types, sources, use cases, and value perceptions. The codebook is reflected in Tables 3 and 4, and is included in Appendix B.

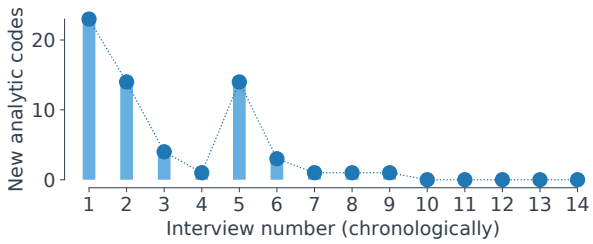


Figure 1: We reached analytical saturation after 9 interviews, as no more new codes were uncovered.

TI SOURCE TYPES	Respondents *n=14*
Paid threat intel providers (PTI)	100%
Open sources (OTI)	79%
Shared sources (STI)	64%
Government	50%
With product or service	50%
Collective procurement	36%
Own research	36%

Table 3: Prominent TI source types amongst our respondents are paid TI, open TI, and shared TI. We selected for the first.

5 DESCRIPTION OF PAID TI

We now address the first sub-question: What do paid TI services consist of? We coded the different answers to the interview question of what TI meant to the respondents. While prior work has focused on the indicator feeds, respondents also mentioned other PTI services, such as reports, requests for information, portals and custom alerts. For some, these services were more important than the indicators. We describe these different forms of PTI. For the two main services, indicators and reports, we also take a look at the offerings of the two PTI vendors in our study. We end with a brief exploration of price levels in the commercial market for threat intelligence.

5.1 TI services

Customers might subscribe to multiple TI services [31]. The variety of services reflects different needs in the market. As stated by a Team Lead TI at a bank: "Intelligence requirements differ per department. The SOC [Security Operations Center] would like to see indicators of compromise and to know TTPs [tactics, techniques and procedures], in order to understand what criminals are targeting, while the Risk department wants to know if these criminals have the capability and intent to disrupt our business, and if we are in control of those risks."

INTELLIGENCE PRODUCTS	Respondents *n=14*
Indicators	71%
Reports	71%
Requests for information	57%
Portal	50%
Data mining and aggregation	29%
Custom alerts	14%

Table 4: Most respondents name indicators and reports as intelligence products that they receive.

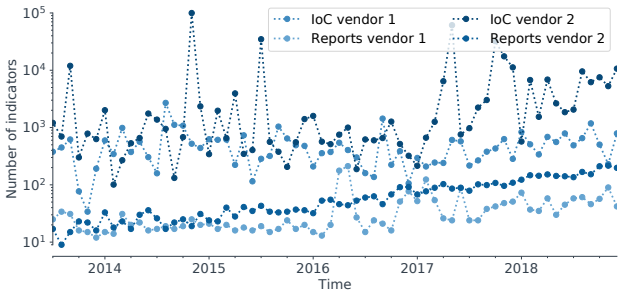
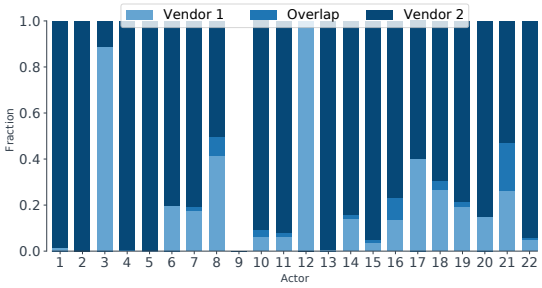


Figure 2: An upward trend is visible in the indicators and reports published by two leading paid TI providers between 2013-2019. Note that the Y-axis uses a log scale. Over 2018, a customer of one of the vendors might receive some 100 reports and 2500 indicators per month. The distribution of indicators over reports (and thus over months) is highly irregular.

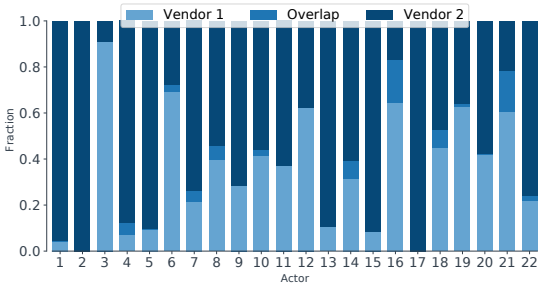
Vendors of paid TI attach metadata to their intelligence products. It describes what industry a report relates to, as well as what threat actors the provider believes are involved. If the provider attaches indicators to the report, they will place in the metadata the degree of confidence they have that the indicator is malicious. This information is used to interpret TI and determine how it can be used.

■ **Indicators** are signals of attacker presence on a network. They are also referred to indicators of compromise (IOCs). Examples are an IP address of known attacker infrastructure, the hash of a piece of malware or a domain associated with a phishing campaign. Indicators are provided in proprietary formats via an API, making them ingestable by detection systems such as a SIEM or IDS. Indicators might be used for network-based or host-based detection, but also in different business processes, from security engineering to various business decisions.

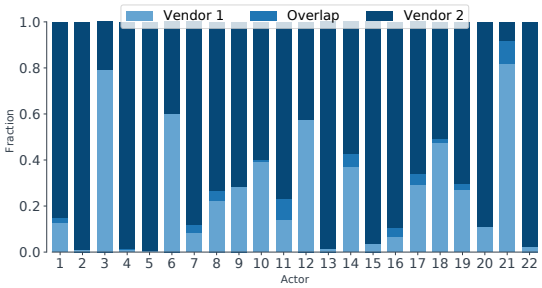
The two PTI vendors in our study attach indicators to their reports. We find that the volumes of new indicators and reports of both vendors have steadily increased over the course of five years (Figure 2). The publication of indicators is



(a) IP indicators, avg. overlap 2,8%



(b) Domain indicators, avg. overlap 4,0%



(c) MD5 indicators, avg. overlap 2,5%

Figure 3: Indicators for threat actors tracked by both vendors from 2013-2018. Overlap is tiny and concentrated on a handful of actors.

unevenly distributed over time. For example, two hikes of the indicator volume can be explained by a vendor's introduction of new report types, leading them to to publish many of the corresponding indicators at once. Indicators are also very unevenly distributed over the reports they are published with.

Coverage is the extent to which a TI source actually includes indicators for all the threats that the source intends to capture. In the absence of ground truth on all ongoing threats, we look at the overlap among the two vendors – a similar approach to Li et al. [21]. Less overlap means that each vendor is observing unique indicators that are missed by the other providers, suggesting limitations in the coverage of threats.

We analysed the overlap for the same 22 actors that both

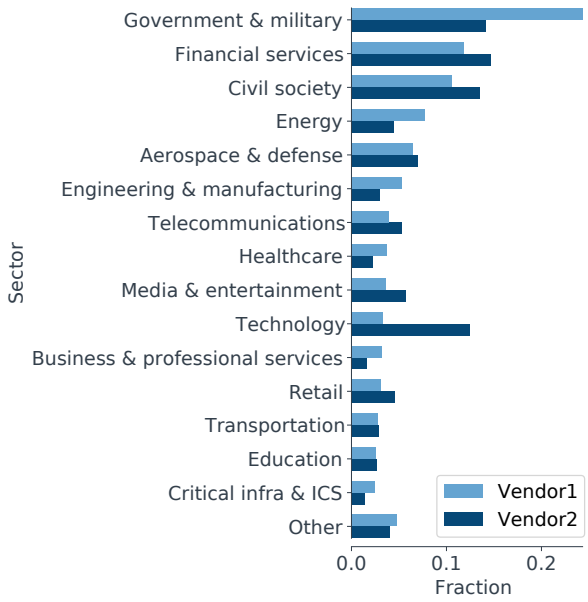


Figure 4: Government and financials represent the main industries as reported on by two leading paid TI providers from 2013-2018. Civil society is the third most targeted sector observed, which is surprising if we assume that these organization are not really in a position to pay for and use high-end TI. We mapped vendor labels to a common structure.

vendors claim to track. The results are summarized in Figure 3. We find very low overlap – on average between 2.5 to 4.0%, depending on the indicator type. The overlap is unevenly distributed and mostly concentrated on a handful of actors. The highest overlap is 21.0% for the IP addresses of threat actor 21. This overlap is low, considering that these indicators are supposed to provide coverage of the activity of the same threat actors. These results beg the question what coverage these indicators actually provide of all malicious activity, by these actors and otherwise. We extend the analysis of overlap in Section 6 with OTI.

■ **Reports** come in various flavors. Malware reports describe the technical results of reverse-engineering captured binaries. Threat reports describe the goals and modus operandi of threat actors. Advisories and alerts describe current events, such as a software vulnerability being targeted or a threat group expanding their focus to another geographic region.

As can be seen in Figure 4, paid TI reports primarily consider threats to government, military, and financial services. These sectors likely form an important part of the customer base of PTI providers. A surprising third in subject matter was civil society, including NGOs and international organizations. Possibly this prominence could be explained by the political significance of the targeting of civil society, or by the relatively large number of such campaigns being observed.

In our interviews, respondents describe that reports of PTI providers are helpful on "all decks of the organisation", from SOC analyst to CISO. A manager at an MSP described reports as the most important form of PTI, as they allowed his analysts to provide context for the clients of why certain alerts happened and what they mean: "We need to inform the customer if we get a critical alert. It may be 2 AM and we call the customer's mobile phone, only to find out that it's a false positive – that happens very often. Then we need to explain why this [security appliance] product says this, so we need some kind of reasoning. Then the customer asks us to filter out the event, or keep monitoring it."

■ **Requests for information (RFIs)** are inquiries from customers to the vendor's analysts. These were described by eight of our respondents as an important form of PTI. A request might work as follows. From reading a threat report, an analyst could be wondering about the relevance of the described threat for their organization. They might inquire with the report's authors if campaigns in some specific sector or geographic area had been observed. The vendor would then search their own data and report back with information.

Our respondents described that requests for information are budgeted as part of the contract at around 10 inquiries per year. One respondent explained that, in practice, the PTI providers were willing to share information if they could, even without a formal RFI. The quota for inquiries were mostly a formality. Another respondent described that as part of their contract, they had been assigned an analyst at the vendor for 0.5 FTE. They would be always in touch with the same person, who over time came to understand their information needs.

■ **Portals** provide access to information that a PTI vendor has delivered over time. They consist of websites with historic data on threat actors and their campaigns, overviews of reports by target sector, as well as other data that the vendor may provide, such as indicators or malware samples. Our respondents describe that portals contain most of the information a vendor has, sometimes more than what can be requested from the API.

■ **Data mining platforms and aggregators** are effectively OTI as-a-service. These are subscription-based platforms on which customers can run queries, sometimes coming pre-loaded with open source security data. Some TI aggregators focus on the analysis tool (TI platform) as a product, and have a curated OTI feed as an additional service. Customers can plug in their own PTI data sources in these platforms. The categorization of such trade tools as 'intelligence products' is up for debate, but our respondents did indeed name them as part of their paid threat intelligence.

■ **Custom alerts** notify customers of specific risks to their organizations. An example is when a domain is registered that is similar to the customer domain, which could be used for typosquatting. Another example is when compromised credentials of the customer organization occur in credential

dumps. We note that the two leading providers that we have analyzed do not offer this service. Providers that offer it may be targeting a different audience with their services, possibly customers that are in need of a managed TI capability, rather than external data sources as input to their existing TI team.

5.2 Pricing of PTI

Public information on pricing of paid threat intelligence information is sparse. We did not identify a single instance of a PTI provider transparently providing pricing information on its website. In general, a recent Gartner report lists the services of the market leaders as upwards of USD 100,000 [20].

We collected 38 price points for 6 popular PTI providers [14] as well as 2 smaller providers. These are displayed in Figure 5. The data points were derived from publicly available schedule price lists [11, 8], as well as by requesting quotes from these vendors. Note that services offered by vendors are not directly comparable and that therefore this figure gives only a rough indication of pricing in the market as a whole.

On the right side ($100,000-$650,000 per year), we find high-end vendors which sell their own TI, while on the left ($30,000-$100,000 per year) we find paid aggregators, whose services primarily consist of providing a platform to integrate TI from other sources and to support analysts with analytics.

The wide bandwidth may be explained by pricing models and negotiations. Our respondents describe that pricing models are sometimes based on per-user licensing, where costs increase as the number of analysts that have access to the provider's portal grows. Furthermore, pricing can be negotiated. A TI analyst at a major bank, said: "Vendor pricing is arbitrary. It's based on the size of the customer organization in most cases." Another respondent described a negotiation with a TI provider in which the asking price was lowered by a factor ten.

Although there is leeway and room for negotiation, PTI currently seems restricted to enterprise organizations with large budgets for information security. A Team Lead TI at a bank shared: "The global costs of a CTI team is around €1-1.5 million per year, including 6-7 staff members and tooling costs. [...] Purchasing two or three feeds at €115,000 to €135,000 per year each may sound like a lot, but is actually not so bad for a bank."

6 COMPARISON WITH OPEN TI

We now address the second question: *How is paid TI different from open TI?* We define sources of OTI by the simple fact that they are freely available, but they are very diverse in nature. A news article may be seen as open source intelligence, as may a thread on a message board on which criminal activity is discussed. According to one analyst we spoke to, open TI included the cybersecurity podcast he listened to on his commute to work.

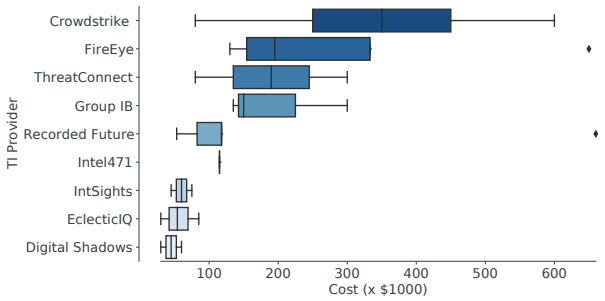


Figure 5: Subscription costs of various PTI providers. Pricing is often based on the size of the customer organization, with major vendors demanding upwards of $100,000 per year.

We then compare indicators from two paid providers and four prominent open source feeds, and establish that there is almost no overlap between the two types of sources.

6.1 Reports

Security researchers, also those at PTI vendors, post blogs, write-ups, and tweets that are accessible to everyone. These are basically the OTI analogue to the reports from PTI vendors. In fact, most vendors make a small fraction of their reports publicly available for marketing, moving them into the OTI domain.[1] These are the same reports as their customers receive, but without the associated indicator sets.

Open and paid sources are by no means decoupled in terms of the attackers they describe. Vendor reports often refer to work posted by OTI researchers, in some cases copying in screenshots of tweets that first observed certain attacker behaviour. We find that paid providers often draw on such open sources, but their reports are much more complete in their description of the context, implications, and possible mitigation options. A manager at an R&D institute in Japan had the view that this property makes paid TI useful especially for larger organizations: "[Paid] threat intelligence is useful for organizations wanting to know much more details about an attack, but using OSINT should be enough for the purposes of a SOC in a medium-sized company."

6.2 Indicators

Respondents collect indicators from open, shared, and paid sources, often loading these in their TI platform or SIEM system. In these systems, indicators are labeled with their origin, allowing analysts to interpret them, e.g. based on their source. Three respondents stated that they did not distinguish between paid and open sources in their detection processes – indicators are indicators.

[1] Freely available vendor reports are indexed by the APTNotes project and can be referenced and searched on https://threatminer.org.

Indicators from open source feeds are more commonly referred to as blacklists or abuse feeds. Respondents discussed the confidence they place in TI sources. In this context, PTI was thought to be more 'accurate' than those from OTI. When pressed, it seems that respondents actually meant: more curated and smaller feeds, rather than more accurate. Smaller sets produce, by definition, fewer false positives. Respondents were emphasizing the element of accuracy that impacts the analyst's workflow, namely minimizing the number of false positives. The other side of accuracy, the rate of false negatives, was not mentioned, even though this rate might actually be higher for smaller and more curated feeds. Because of this perceived accuracy, PTI indicators are used with more confidence in detection and other use cases.

■ **Overlap** among paid and open sources is negligible, even though the OTI lists are vastly more voluminous than the lists of PTI indicators, as can be seen in Figure 6. This is relevant, because overlap helps us understand the coverage of a source – the level to which it captures the intended threats – as described in Section 5. The overlap between the individual sources is shown in Figure 7 as a fraction of the total volume of that source for that period. Vendor 1 and vendor 2 (PTI) share some indicators amongst each other – 1,3% and 13,0% respectively. This seems low, considering that the feeds are focused on the same topic areas in the overall landscape. It appears that PTI has the same pattern as OTI: a lack of overlap among feeds and the dominance of 'singletons'. Less than 1% of the PTI indicators overlap with any of the OTI sources. Vice versa, the OTI sources share indicators with each other, quite substantially in some cases, but there is basically no overlap with PTI: 0,0% of all OTI indicators are also observed in the PTI sources. As we describe in the Methodology, the uncommonly large amount of overlap between the OTI sources [21] is explained by the fact that we selected specifically for OTI that is often re-used [15] in hopes that this would increase overlap with PTI sources – to no such effect.

■ **Timeliness** of indicators means that the information is available to the customer early enough to actually detect and stop an attempt at compromise. While there is no ground truth as to when a particular resource (domain, IP, binary) was first used by a threat actor, we can assess timeliness via pairwise comparisons of the different feeds and measuring the delays in which indicators are made available to customers. For this analysis, we work with the small number of indicators that occur in more than one set.

We first compare the two PTI vendors (n=16 and n=28, respectively). In Figure 8, we can see drastic delays in when PTI vendors observe indicators. On average, it take more than a month before an indicator observed by one vendor is also observed by the other vendor. In terms of defending against sophisticated threat actors, this is a very long delay. We find almost no instances where threat intelligence is distributed by other PTI within the same week. It seems

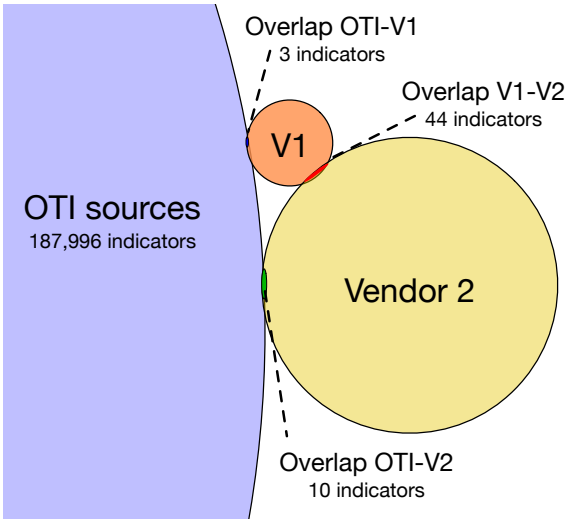


Figure 6: Overlap between OTI indicators published in one month (October 2018) and PTI indicator sets published in the enclosing five months (Aug-Dec). The latter time frame allows for variations in timing when the indicators are first reported and thus provides us with an upper-bound estimate of overlap. The overlap turns out to be negligible, suggesting that these types of intelligence are different in kind, as well as in volume. Areas in the diagram are proportional to the number of indicators.

that vendors do not use the TI from their competitors, or at least not successfully, to find the same indicators in their own telemetry.

We conducted the same analysis for the indicators of the two vendors versus the four OTI sources. The latter comparison is only based on the tiny overlap of 2 to 7 cases, which means that we cannot draw any strong conclusions. One might expect that paid providers would be faster in all cases, but PTI sources were faster in only half of the cases we have analysed. So it seems PTI is not faster in finding indicators than OTI. The same appears to hold vis-à-vis STI, in the words of a team lead TI at an oil and gas company: "We also have our own networks with companies in [our sector] and I must honestly say, we quite often get the information earlier there than from our [paid] intelligence providers."

While we have no ground truth as to when a particular resource (domain, IP, binary) was first used by an actor, the fact that we see major delays in all pairwise comparisons of our PTI and OTI sources suggests timeliness is a major problem.

7 USES AND VALUE OF TI

This section addresses our final question: *How do customers use TI and how do they perceive its value?* Given that customer organizations of PTI are paying a substantial amount

Figure 7: Indicator overlap as a percentage of the row's total volume. From Vendor 1's indicators, 13% is also listed by Vendor 2, and 0.3% is listed by by AlienVault. Overall, PTI and OTI sources hardly share any indicators, at most 0.9% relative to the PTI set, and at most 0.02% relative to the OTI set. The same subsets are used as for Figure 6.

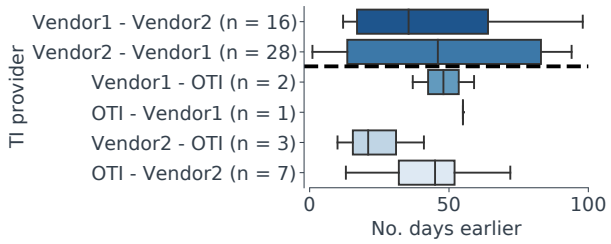


Figure 8: Timeliness comparison between sources. There is often a delay of over a month before an indicator is listed by a second source. We show three pairwise comparisons. To illustrate: we found 16 instances where Vendor 1 was earlier than Vendor 2, to an average of 45 days. Note that the number of indicators (n) is low, because it is based on the small overlap between sources. We aggregate the 4 OTI sources into a single set. For the comparison between the two vendors, we use the period of 2013-2018. For the comparison between vendors and OTI, we used the same subset as for Figure 6.

of money compared to OTI and STI, they apparently value PTI to be worth the asking price. We tried to understand their perceptions of this value in two ways: by asking them how they used PTI (use cases) and by asking about what they see as strengths and weaknesses of their sources. These two ways are aligned with the economic distinction between 'stated preferences' versus 'revealed preferences'. The former infers preferences from what people explicitly state as preference, the latter from their actual choices and behaviors.

7.1 Use cases of TI

Based on the analysis of the interviews, we found 9 use cases for TI (Table 5). The percentages refer to what percentage of the respondents mentioned this use case.

The top three use cases are central to SOC operations. **Network detection** (93% of respondents mentioned this) is still the main use of threat intelligence. This includes all instances in which TI is used to reduce attacker dwell time in an automated fashion, including correlating TI to logs, ingesting it in a SIEM or IDS, or using it in host-based detection controls.

Situational awareness was mentioned in two out of three interviews (64%) as a use case. This is the ability of TI or SOC analysts to have a general understanding of their organization's threat environment and risk profile. Situational awareness is broader than detection, It is relevant in the planning and direction phases of the intelligence cycle [18]. . A Team Lead TI at a bank said it is not just about cybersecurity: "I believe that is too limited. It's about understanding who is a threat to my organization, not just the technical channel used for the attack."

SOC prioritization (50%) is a more practical use of TI, e.g. to assess how critical alerts are or to direct threat hunting efforts. This way, resources – especially attention of analysts – can be allocated toward most relevant threats.

Informing business decisions (36%) concerns uses of TI to improve organizational decision-making. For example, a CISO used TI to evaluate the return on various options to invest in security controls. But there are also organizations that use paid TI to assess the risks associated with a potential acquisition of international competitors, to gain a 'business decision advantage'. An analyst at a bank said: "Threat intelligence means engaging with various business units in order to understand their information needs, and then developing a way to answer those in a timely way. As a consultative practice within [our organization] we could provide information on geopolitical affairs, and intelligence on physical risks in a country that [we are] operating in. It doesn't necessarily have to relate to cyber." This use case, like situational awareness, also underlines the value of reports compared to raw indicators.

Enrichment of own threat intelligence (36%) where PTI is processed with the aim to improve the organization's own services – e.g., by managed SOC providers and government CERTs, as well as by TI teams to internal stakeholders. As an MSSP manager stated: "We can give our customer the reason why [their security appliance] generates this alert. We

can get background information to understand why [the PTI vendor] detects this as suspicious."

Improving end user awareness (29%) is about using TI to educate the wider employee population of the organization, e.g., security-awareness based on reports about recent phishing campaigns.

Threat hunting (29%) is active investigation using TI. This is the type of research which requires human creativity and is currently hard to automate. Combining TI and other data can generate insights for an analyst on where and how to search for attacker activity in systems and networks.

Informing security engineering (21%) includes using TI to organize vulnerability management as part of maintaining the organizations own systems. It also includes the prioritization of developer tasks, e.g., on a customer-facing app, based on observed attacker tactics.

Reducing financial fraud (14%) is a specific use case for banks. Their PTI vendors are supplying them with lists of compromised credit cards. Based on this data, they can decide to block cards or investigate accounts for money laundering. A Team Lead TI at a bank shared: "We are not in a position to buy something from dark web criminals, but [the PTI vendors] are."

7.2 Value perception of TI

Prior work has assumed that users would like to evaluate TI based on quality criteria such as volume, overlap, timeliness, accuracy, and coverage [21, 33], as discussed in Section 2. The assumption is that those criteria capture what users value. Rather than work from this assumption, we have followed an inductive approach. When our respondents made evaluating comments about TI, we labelled them. This resulted in wider set of 16 separate codes of the properties that made TI to be perceived as more valuable. We describe these labels, printed here in italics, and list them in Table 6 in Appendix B.

First of all, only three respondents (21%) mentioned anything about price or affordability. If price is not a key factor, this begs the question: what is? We distinguished three clusters of values: confidence, relevance and actionability.

Confidence relates to how much the user trusts the TI to provide useful results. This was primarily interpreted by respondents as not wasting the time of analysts. A head of SOC at an ISP described: "One of our commercial sources actually even has a negative value for us, because it costs us time to look into alarms that it generates, which turn out to be mostly false positives." Most of our respondents desire low 'noise' in the TI they receive, which we labeled as to *automatability* (79% of respondents named this property). Loaded into systems such as firewalls or IDSes, low-quality information will immediately have drastic operational impacts. Closely related to automating is the use of TI is trust in the vendor. A head of SOC at an ISP described wanting to be able to verify the origin of intelligence they receive, or what one might call the vendor's transparency: "For us, a provider's ability to answer questions about their intel is an indication of the confidence we can place in them. Of course I understand if they sometimes cannot name their sources. But we need some understanding of the process that led to an indicator being placed on a list in order to use it." Respondents perceive a source as more valuable if it providers an *original contribution* (50%), as demonstrated by this quote by a head of SOC: "We notice a lot of re-use between providers and people in the community. As a rule, we prefer original intelligence over curated or aggregated intelligence because you can cut out the middleman and directly ask questions about the assessment or provide feedback on the intelligence." Confidence is evidently related to *accuracy* (43%) of TI, but respondents also mentioned *selectiveness* (29%). As at analyst at a bank told us: "From [this trusted community] we get emails so often that we filter and tend to ignore them. Whereas [another source] only emails us twice per year. In that case we are likely to look into it." Selectiveness indicates a preference for a low volume, which is seen as an indication of accuracy.

These perception of value reflect an intriguing implicit trade-off that users are making; smaller, more curated TI sources are valued higher, as these require fewer organizational resources and may prevent information overload for analysts [4]. But they also imply more false negatives – something our analysis of the overlap of sources (Section 5) has confirmed. The risk of having high false negative rates was, remarkably, much less of a worry.

The second cluster of properties perceived as valuable is on **relevance**. Here, we find properties that value the degree in which the TI is tuned to the specific situation of the organization. Two respondents mentioned ending a contract with a PTI vendor because their intelligence mostly covered a *sector* (64%) not relevant to them. *Geographic focus* (50%) is a valued property because our respondents seem to understand attacker groups to choose targets based on earlier successful campaigns in a given country. Furthermore, a certain bias may be related to the geographic focus, as a respondent in Japan described the market for PTI as too US-centric, with certain information not being usable for their organization. A Team Lead TI at an oil and gas company warned: "Be aware of biases of your intelligence providers. For example, a US-based provider will never report on US spying activities." Hence, coverage of one source may have the *ability to correct bias* (14%) in another. *Coverage of relevant threats* (50%) was mentioned by respondents, but not in terms of if the indicator feeds exhaustively contained all the relevant infrastructure of the relevant threat groups. Rather, respondents seemed to interpret this in terms of coverage of their threat landscape, i.e. providing information about threats relevant for their organization. A Team Lead TI at a bank explained how at one point they were confronted with an advanced threat actor. They then separately asked four PTI vendors to tell them what they knew about this threat actor and validated this with their own

observations. They noted: "One vendor had nothing, the other three came with a theory. Based on the data that we could observe ourselves, we saw that [vendor X] was totally wrong. The other two vendors were right. It might have been a coincidence, but we did this a few more times and then decided to work with those two vendors." Again, these value perceptions reflect a way of thinking that aims to reduce impact of TI on analysts, in this case by valuing TI that reduces the inputs into the workflow to what is considered the most relevant.

Finally, there is a cluster of values around **actionability**. This was defined by a Team Lead TI at a bank as: "intelligence which you can use to influence your business." The *capability to provide context* (100%) means that TI helps the user to understand and explain events and alerts. Paid sources are seen by respondents as better at providing context than open sources. One analyst from the Netherlands said: "Intelligence is about context, about putting threats into perspective for your organization." *Timeliness* of TI (50%) is a valued property because indicators lose their relevance rapidly. Once TI is outdated, it is no longer actionable. In Section 6, we compared the timeliness of OTI and PTI and did not find a significant difference. This was consistent with the remarks of our respondents. Some said that they do see a difference with STI: they receive certain TI earlier from trusted communities. Yet, even timely TI is only actionable if it is *comprehensive* (50%) enough to be able to base decisions on it, such as suggesting possible mitigation strategies. Some respondents spoke about 'rich' information and about the difference between 'raw' vs. 'polished' intelligence, where PTI is deemed of higher quality because it is more polished towards use. An analyst as a bank stated: "A [colleague] at another bank is just going to post some IOCs to you, or it will be a small write-up, because their time is limited. It will not be of the same quality [as that of PTI providers]. That's the key difference: you're paying for polished intel rather than what we would call raw intel." *Interpretability* (50%) refers to the property that the analyst can make sense of the information, e.g., it has good meta-data. *Data visualisation* (14%) is related property that aids in putting the TI to use by making it accessible.

In sum, we find that TI is evaluated on a much broader set of criteria than prior work assumed. Furthermore, an underlying logic in the properties that respondents value is that they are optimizing the workflow of their organization – most notably their analysts – rather than the detection of threats. This is one of the key reasons why they value the smaller and more curated PTI sources. The fact that these smaller sources might have limited coverage and uncertain timeliness (Section 5 & 6) is not described as a major problem.

7.3 Evaluating TI

Customers of TI found it difficult to compare sources, which corroborates findings in market research [20, 14]. Evaluation happens mostly in informal processes and based on tacit criteria and heuristics. One research manager described: "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 intel daily, and they can feel if they are comfortable with it." Six out of fourteen respondents did define some criteria or intelligence requirements in order to evaluate TI sources, often in the form of information gaps in the organization – i.e., what questions the TI team needed to answer.

One analyst described using metrics within the network detection use case: "You can [demonstrate the effectiveness] by generating a metric on IOC feeds. For example, how many times does this commercial IOC feed purchased from [vendor X] create security events within our organization? And then, what is the outcome of those security events? Is it a false positive? In which case, that means that IOCs sent by that vendor are inaccurate. We can feed that back to the vendor when it comes that negotiation about the contract." On calculating metrics for the use case of informing business decisions, he added: "That's slightly more difficult to develop metrics around and quantify. But really, what we're looking for from stakeholders [in our organization] is very simple feedback: Was this useful? Did this aid your decision? [...] That is good enough to say if [my team's] reporting is having an impact."

8 DISCUSSION

In this paper, we have attempted to lift the veil of paid TI services. We confirm that, indeed, paid TI seems to be a different cup of tea, with distinct intelligence products and low overlap with open TI sources. The interviews we conducted display an apparent contradiction in the practice of TI use: professionals discuss at length the properties that they believe make TI valuable to them, yet hardly attempt to measure or validate these beliefs.

This contradiction questions if threat intelligence metrics, as proposed by [21] and others, can actually capture the right value properties. Research has focused on developing metrics that could be used to understand the coverage, accuracy and timeliness that PTI providers can provide. In our interviews we found, however, that customers are much more pragmatic in how they evaluate the added value of TI, namely through the impact it has on their analysts and security operations. To optimize the analysts' workflow, poor coverage is not necessarily a big problem, while the number of alerts is. This drives customers to smaller, curated sets – the opposite direction of where a coverage metric would point them. In detection parlance: one might expect customers of PTI to select sources for low false negatives, while actually they seem to be selecting for low false positives. These are two distinct goals that are both part of the concept of accuracy. Another limitation of these metrics is that although quantification make sense for network detection – events can be measured – it make

less sense for the other uses of PTI services that customers described. Thoughtfully composed threat actor reports do not lend themselves to quantitative analysis. Further, just counting network events does not tell much about organizational outcomes: an event may occur without it having much relevance or impact. Analyst skill and experience therefore remain essential for triage in the SOC [4]. Carefully prepared analysis reports could contribute to answering strategic questions in organizations [35], yet TI is currently used mostly in operational processes. That being said, metrics could help to optimize the selection of TI sources for event detection and to understand the potential for false negatives by looking at coverage and overlap. Metrics for TI are useful in this, more narrow, context.

Currently we lack a good understanding of the coverage of PTI vendors due to secrecy around their methods. While that is understandable in order to maintain operational security in the face of advanced attackers, it does make it harder for customers to evaluate what they are actually buying. This paper seeks to address this by describing overlap and timeliness through the comparison of indicators. We find that even when looking at the same actor groups, two of the leading PTI providers have diverging information with very small overlap. The secrecy around their methods to observe threat actors also benefits vendors economically: as long as their methods remain opaque, myths will live on about how TI providers may offer some special degree of TI coverage, possibly through an exclusive skillset, 'hacking back', or by means of access to restricted information. As described by Shires [36], vendors use "cyber noir" symbols that portray their work as deploying unconventional tactics in mythical battles between good and evil, often aligned with national security. Such stories and symbols give rise to an understanding of detection and attribution capabilities of PTI vendors that currently cannot be substantiated nor vetted.

As a consequence of the low transparency, the market for paid TI shows signs of asymmetric information, in which the vendors know what they are selling, but customers don't know what they are buying. Consumers in the market for TI therefore find it hard to compare services [20, 14]. As Metcalf concluded already in 2015 for blacklists: "secrecy does not benefit the operational analyst who must must decide which lists to apply" [25]. And indeed, five years later, our respondents say it is still "mostly guesswork" to understand the visibility and methods of paid TI providers, and with that the value of the services they offer.

Under conditions of information asymmetry, buyers rely on signals. One such signal is whether the firm is seen as a market leader, which is partially signalled via a high price for its services. In this sense, the phrase 'nobody ever got fired for buying IBM' also rings true for threat intelligence. Customers are incentivized to purchase from leading providers – the safe choice under uncertainty. This way, economic value is linked to vendor reputation. In the longer run though, structural information asymmetry holds the risk for vendors that customers may lose trust in the value of PTI services, which would decrease the willingness to pay. This effect is known as a 'market for lemons' [2]. Grigg [38] went one step further and argued that even vendors might lack reliable information on the quality of their products. Providers of PTI might know what data they collect and how, but they do not know – and can't know, Grigg would argue – how effective their product is in improving the security of their clients. Our analysis suggests that, in light of lacking ground truth and low overlap in indicators, vendors themselves may not even know how well they are able to track specific threat actors. When both seller and the buyer lack reliable information on the quality of a product, this creates – in Grigg's analysis – a market for 'silver bullets', where herding behavior and arbitrary best practices triumph over rational purchasing decisions.

Finally, we note that through their forensic work, TI vendors have profound influence on how the general public and the political leadership understands security incidents. Reporting on such incidents is not just neutral technical analysis but also requires interpretation and 'sense-making', as Stevens (2019) showed for the analysis of Stuxnet by Symantec [37]. Indeed, public understanding of such incidents is shaped by the political and economic prisms of the experts who carry out the analysis [13, 42]. Information asymmetry in the market for paid threat intelligence is therefore not only of economic, but also of political significance.

9 RELATED WORK

There is a rich line of research that has studied the properties of open threat intelligence, also known as abuse feeds and blocklists – e.g., [39, 25, 19]. Problems in coverage, timeliness and accuracy have consistently been observed in these studies.

In recent years, proposals have been put forward to formalize and measure the quality of TI [21, 33, 15, 29, 30, 27]. This includes metrics on features such as coverage, accuracy, timeliness, relevance, overlap, latency, and volume. [34] has investigated how to present TI quality to analysts. Applications of these approaches to measure quality of TI have been limited to OTI, also in the recent studies by Li et al. [21] and Griffioen et al. [15].

Aside from the availability of high quality information, it is essential how this information is used. [31] identifies that organizations have issues interpreting threat intelligence, triaging large volumes of threat information or dealing with large numbers of false positives. In this sense, TI has similar operational issues as blacklists of IP addresses and domain names, which have an established history in computer security. While TI as contextualized, high-level information has the potential to remediate these issues [6], a 2019 SANS survey nonetheless found (low-level) indicators of compromise to be valued higher by respondents than information about (high-

level) tactics, techniques and procedures (TTPs) [7]. The authors attribute this to the fact that most of their respondents were security operations analysts, who might view the value proposition of TI primarily as enriching alerts with technical details. Our study provides a detailed analysis of how threat intelligence is actually being used within organizations, and how the value is perceived by those directly affected by it.

We go beyond the related work in two key ways. First, we present the first empirical study of the PTI of market leaders. The nearest study is [21], which was not focused explicitly on PTI, but did include two edge cases of paid services. These services were not providing original high-end TI sources, but helped curate and aggregate otherwise free or low-end indicator feeds, and were priced in the range of USD 1-10k per year, as kindly confirmed to us by one of the paper's co-authors. We followed the measurement approach developed by [21] but provide the first application to 'real' PTI: services of commercial threat intelligence providers which operate their own detection network and perform forensic analysis to generate original data about threats. With this value proposition, vendors justify pricing between USD 100-600k per year. A common sentiment in the TI industry is that 'real' high-quality threat intelligence may only be obtained from these exclusive closed-source commercial providers, and [31] finds PTI sources are used twice as often as OTI in industry.

Second, we contextualize these quantitative approaches to measuring quality by conducting a user study of PTI customers and identify their perceptions of value. This has enabled us to find that users use and evaluate TI differently than the measurement approaches developed by researchers assume. In reality, users hardly calculate the proposed metrics. Their perception of value is determined by various use cases in which this quantification is not only missing, but sometimes points in conflicting directions – as around the issues of accuracy and coverage.

10 LIMITATIONS

Our mixed-methods approach introduces several limitations. First, we only analyzed the services of two PTI vendors. As they are among the market leaders at the high end of the market, we assume that our findings are representative for that market, but future work is needed to corroborate this.

Second, our analysis was based on data of a single customer of these two vendors. This customer acquired 3-5 subsets of indicators of each vendor in the same topic areas, of a total offering of 5-10 subsets that each vendor offers. Other subsets might show somewhat different results for the target industries (Figure 4) or the overlap between vendors. Given that the available selection of subsets form a significant portion of all subsets, we expect that they provide a valid basis for comparison. The exact numbers, however, are likely to vary across other subsets.

Third, our analysis of PTI has to contend with a lack of ground truth. We followed the approach from prior work on OTI [21] and conducted a comparative analysis among different feeds. For the analysis of indicators on different threat actors, we relied on the well-known mapping developed by Florian Roth across the different threat actor naming schemes used by PTI vendors. We cannot ascertain how reliable this mapping is, other than the fact that Roth is an expert in the field, his mapping is well known, and he is collaborating on it with other industry insiders – so mistakes would presumably be corrected.

Fourth, the comparison with OTI was limited to a single month of four feeds. While these feeds were chosen because they are actually re-used by many other feeds in the OTI landscape [15], a broader set of feeds will provide a more reliable result. That said, the lack of overlap with PTI was quite stark and unlikely to change when analyzing other feeds. In OTI research, the low overlap among any two feeds has been a consistent finding for years.

Fifth, regarding our user study, our main limitations stem from a small sample size (n=14). Our sample contains a variety of organizations, but it may contain selection bias as respondents are geographically located in the Netherlands and Japan only, were working with TI (rather than choosing not to), and willing to talk about this in an interview. All of this makes that we do clearly do not claim that our findings are generalizable for all organizations using TI. Given that no prior work existed, neither on the PTI feeds nor on users of PTI, we chose to do an in-depth exploration of the views of such users using the grounded theory method. For more generalizable results, a survey could be designed based on our findings.

11 CONCLUSIONS

This study explored services in the market of commercial threat intelligence. We analyzed the indicators of two paid TI vendors and found 13.0% of vendor 1's indicators appear in vendor 2's set and – vice versa – a mere 1.3% of vendor 2's indicators in vendor 1's set. If we drill down to the 22 threat actors for which both vendors have indicators, we find an average overlap of these indicators of no more than 2.5 to 4.0% per group, depending on the type of indicator. Further, this overlap occurs primarily with a handful number of actors. The fact that the indicators of two vendors are largely separate sets, even when assessed for specific threat actors that they both track, raises questions on the coverage that services of these vendors actually provide.

Reports produced by paid TI providers describe the tactics of threat actors, the results of malware reverse-engineering, or give advisories for current events, amongst other things. The reports concern primarily government, military, and financial institutions – important customers of TI vendors – but surprisingly also pay a lot of attention to campaigns targeting

civil society, possibly due to their political significance.

Besides indicators and reports, paid TI services also consist of requests for information from analysts, portals with historic information, data mining platforms, and custom alerts. These services are expensive, with subscription costs of major vendors often upwards of $100,000 per year.

Whereas paid sources offer 'polished' TI, open sources contain 'raw TI' as one respondent described it. We find that this statement holds for the two types of paid intelligence products that we have compared with open TI, namely indicators and reports. In terms of substance, paid reports are for example similar to open source blog posts and tweets of security researchers that are freely available online, but the paid reports are more comprehensive in their descriptions of context and recommendations. Further, paid reports are packaged with machine-readable indicators. We compare these to open TI indicators feeds (which are much larger in volume), and find less than 1% overlap between them, suggesting that PTI providers successfully differentiate themselves and are capturing a different part of the threat landscape. In terms of timeliness, we find no evidence that PTI is faster than OTI, surprisingly enough, although this is based on the small sample of overlapping indicators. There is a delay of around one month before indicators from one set are found in another.

The main use case for TI is network detection, followed by situational awareness – which we understand to mean: informing your threat profile – and prioritization of resources in the SOC. We find that one-third of respondents use threat intelligence to improve organizational decision-making: to inform security engineering, to reduce financial fraud, but in one instance also for risk management around international mergers and acquisitions. Asked what makes TI valuable, respondents name properties related to actionability, relevance, and confidence. All respondents describe valuing the ability for TI to provide context, which suggests that they view TI as a reference. Further, the ability to automate using TI is important for respondents: almost all name valuing a low false-positive ratio and interoperability with their detection systems. Importantly, only half of our respondents discuss coverage as something they value in threat intelligence – reducing false-negatives or misses seems to be much less of a concern. We conclude that TI consumers evaluate threat intelligence mostly through the impact on their organization's detection processes.

Evaluation of TI sources is done mostly through informal processes by our respondents. When a subscription renewal comes up, TI professionals decide if to continue largely based on implicit criteria and tacit understanding of the value of that source. This is surprising, given that research has focused on developing metrics and heuristics that could enable a quantitative understanding of value of TI. In practice, metrics or intelligence requirements are used by less than half of the professionals we interviewed.

There is this promise that leading paid TI vendors would be able to overcome the persistent problem of sharing threat information among defenders. They can aggregate and analyze threat data at scale from vantage points across different clients and networks – as was also argued by Crowdstrike Senior Director Vicenzo Iozzo (see Section 2). We do not dispute that important advances are being made. That being said, our study raises doubts as to the extent in which this promise has been fulfilled today. Even when the vendors claim to track the same threat actor, they each see only a tiny fraction of the associated indicators. The fact that almost all PTI indicators are unique to one vendor, is a pattern we know all too well from OTI sources. So the pay-off of aggregating data across clients and networks, as claimed by Iozzo, is not very clear in terms of detection capability, to say the least. Even when a client would be willing to pay the steep price of simultaneously acquiring the feeds of all market leaders – a proposition that would cost them millions each year – it is likely that this strategy would reproduce the pattern that we know from OTI: feeds contain mostly singletons and more feeds still get nowhere close to comprehensive coverage.

The sharing of indicators across vendors would still be a first step to improve coverage and the detection of attackers. The current state of affairs in paid TI resembles the market for anti-phishing services about a decade ago. The lack of data sharing meant that each anti-phishing company thought it had strong coverage and could protect its client brands well, while the truth was that they missed most of the attacks they were hired to detect [26]. In that market, like in the market of malware detection, sharing across vendors was eventually recognized as a superior security strategy. Until that happens for commercial threat intelligence, the problem of information sharing persists.

We are grateful to the anonymous reviewers and the shepherd for their comments and advice. The work presented in this paper was supported by funding from the Ministry of the Interior and Kingdom Relations of the Netherlands, Delft University of Technology, and Leiden University, under ref. number M75B07.

REFERENCES

[1] *Abuse.ch*. [Online; accessed 3. Feb. 2020]. Feb. 2020. URL: https://abuse.ch.

[2] George A. Akerlof. "The Market for "Lemons": Quality Uncertainty and the Market Mechanism". In: *Quarterly Journal of Economics* 84.3 (Aug. 1970).

[3] *AlienVault Open Threat Exchange (OTX)*. [Online; accessed 3. Feb. 2020]. Feb. 2020. URL: https://cybersecurity.att.com/open-threat-exchange.

[4] Tiffany Bao and Gail-joon Ahn. "Matched and Mismatched SOCs: A Qualitative Study on Security Operations Center Issues". In: *ACM CCS*. 2019.

[5] Beth L Beech. "Asking questions: Techniques for semi-structured interviews". In: *Political Science* 35.4 (2009), pp. 665–668.

[6] David Bianco. *The Pyramid of Pain*. [Online; accessed 3. Feb. 2020]. 2014. URL: http://detect-respond.blogspot.nl/2013/03/the-pyramid-of-pain.html.

[7] Rebekah Brown and Robert M. Lee. *The Evolution of Cyber Threat Intelligence (CTI): 2019 SANS CTI Survey*. Tech. rep. SANS Institute, 2019.

[8] *Caharasoft GSA Pricelists*. [Online; accessed 13. Feb. 2020] https://web.archive.org/web/20200213230809/https://static.carahsoft.com/concrete/files/1414/3223/9520/CMASPricelist.pdf, https://web.archive.org/web/20200213231518/https://static.carahsoft.com/concrete/files/4114/5029/6739/Compiled_new_DIR_Pricelist_-_Revised_12-10-2015.pdf, . Feb. 2020.

[9] Kathy Charmaz. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. London: SagePublication Ltd, 2006.

[10] John W. Cresswell. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. Second edi. 2007.

[11] *DLT NCPA Pricelists*. [Online; accessed 13. Feb. 2020] https://www.dlt.com/sites/default/files/contract-attachments/NCPA%20Pricelist%20-%20PDF_0.pdf. Feb. 2020.

[12] *DNS-BH Malware Domain Blocklist*. [Online; accessed 3. Feb. 2020]. Feb. 2020. URL: https://www.malwaredomains.com.

[13] Florian J Egloff. "Contested public attributions of cyber incidents and the role of academia". In: *Contemporary Security Policy* 0 (Oct. 2019), pp. 1–27.

[14] Forrester Research. *External Threat Intelligence Services, Q3 2018*. Tech. rep. 2018. URL: https://www.fireeye.com/content/dam/fireeye-www/products/pdfs/pf/intel/rpt-forrester-threat-intel-services.pdf.

[15] Harm Griffioen, Tim M. Booij, and Christian Doerr. "Quality Evaluation of Cyber Threat Intelligence Feeds". In: *ACNS*. 2020.

[16] Dean Hammer and Aaron Wildavsky. "The Open-Ended, Semistructured Interview: An (Almost) Operational Guide". In: *Craftways*. Routledge, Feb. 2018, pp. 57–101.

[17] Vincenzo Iozzo. *The Case for Scale in Cyber Security*. [Conference talk; accessed 4. Feb. 2020]. Dec. 2019. URL: https://media.ccc.de/v/36c3-11220-the_case_for_scale_in_cyber_security.

[18] Joint Chiefs of Staff of the United States. *JP 2-01 Joint and National Intelligence Support to Military Operations*. Tech. rep. Washington, D.C.: Department of Defense, 2017. URL: http://www.jcs.mil/Doctrine/Joint-Doctrine-Pubs/2-0-Intelligence-Series/.

[19] Marc Kührer, Christian Rossow, and Thorsten Holz. "Paint It Black: Evaluating the Effectiveness of Malware Blacklists". In: *RAID* (2014).

[20] Craig Lawson, Ruggero Contu, and Ryan Benson. *Market Guide for Security Threat Intelligence Products and Services*. Tech. rep. Febuary. Gartner Research, 2019. URL: https://www.gartner.com/en/documents/3902168/market-guide-for-security-threat-intelligence-products-a.

[21] Vector Guo Li et al. "Reading the Tea Leaves : A Comparative Analysis of Threat Intelligence". In: *Proceedings of the USENIX Security Symposium*. 2019.

[22] *MarketWatch Forecast to 2025*. [Online; accessed 3. Feb. 2020]. Feb. 2020. URL: https://www.marketwatch.com/press-release/threat-intelligence-market-size-share-application-analysis-regional-outlook-growth-trends-key-players-and-competitive-strategies---forecast-to-2025-cagr-of-189-2019-08-13.

[23] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. "Reliability and Inter-rater Reliability in Qualitative Research". In: *ACM on Human-Computer Interaction* 3 (2019).

[24] Roland Meier et al. "FeedRank: A Tamper- resistant Method for the Ranking of Cyber Threat Intelligence Feeds". In: *2018 10th International Conference on Cyber Conflict*. Tallinn: NATO CCD COE Publications, 2018, pp. 321–344.

[25] Leigh Metcalf and Jonathan M. Spring. "Blacklist Ecosystem Analysis". In: *ACM Workshop on Information Sharing and Collaborative Security*. 2015.

[26] Tyler Moore and Richard Clayton. "The Consequence of Non-Cooperation in the Fight Against Phishing". In: *APWG eCrime Researchers Summit*. 2008.

[27] Paweł Pawliński and Andrew Kompanek. "Evaluating Threat Intelligence Feeds FIRST Technical Colloquium for Threat Intelligence". In: *FIRST Technical Colloquium for Threat Intelligence*. Munich, 2016. URL: https://www.first.org/resources/papers/2016#munich2016.

[28] Paweł Pawliński et al. *Actionable Information for Security Incident Response*. Tech. rep. January. ENISA, 2015, pp. 1–79. URL: https://www.enisa.europa.eu/publications/actionable-information-for-security.

[29] Alex Pinto. "Determining the Fit and Impact of CTI Indicators on Your Monitoring Pipeline (#tiqtest2)". In: *FIRST Conference*. Kuala Lumpur, 2018. URL: https://www.first.org/conference/2018/program#pdetermining-the-fit-and-impact-of-cti-indicators-on-your-monitoring-pipeline-tiq-test-2-0.

[30] Alex Pinto and Kyle Maxwell. "Measuring the IQ of your Threat Intelligence Feeds". In: *Defcon 22*. 2015.

[31] Ponemon Institute. *The Value of Threat Intelligence: The Second Annual Study of North American & United Kingdom Companies*. Tech. rep. February. Ponemon Institute, 2019.

[32] Florian Roth. *The Newcomer's Guide to Cyber Threat Actor Naming*. May 2018. URL: https://medium.com/@cyb3rops/the-newcomers-guide-to-cyber-threat-actor-naming-7428e18ee263.

[33] Thomas Schaberreiter et al. "A Quantitative Evaluation of Trust in the Quality of Cyber Threat Intelligence Sources". In: *ARES*. 2019.

[34] Daniel Schlette et al. "Measuring and visualizing cyber threat intelligence quality". In: *International Journal of Information Security* (2020).

[35] Andreas Sfakianakis. *Let's make CTI great (again): a 5-year lookback in CTI*. [Online; accessed 14. Feb. 2020]. Nov. 2018. URL: https://www.enisa.europa.eu/events/2018-cti-eu-event/cti-eu-2018-presentations.

[36] James Shires. "Cyber-noir: Cybersecurity and popular culture". In: *Contemporary Security Policy* 0.0 (Sept. 2019), pp. 1–26.

[37] Clare Stevens. "Assembling cybersecurity: The politics and materiality of technical malware reports and the case of Stuxnet". In: *Contemporary Security Policy* 0.0 (Oct. 2019), pp. 1–24.

[38] *The Market for Silver Bullets*. [Online; accessed 7. Jun. 2020]. Mar. 2008. URL: https://iang.org/papers/market_for_silver_bullets.html.

[39] Kurt Thomas et al. "The Abuse Sharing Economy: Understanding the Limits of Threat Exchanges". In: vol. 7462. 2016, pp. 143–164.

[40] *UK businesses: average investment in cyber security 2019*. [Online; accessed 1. Jun. 2020]. July 2019. URL: https://www.statista.com/statistics/586587/iinvestment-in-cyber-security-by-businesses-in-the-uk-by-sector.

[41] Daniel Votipka et al. "Hackers vs. Testers: A Comparison of Software Vulnerability Discovery Processes". In: *IEEE Symposium on Security and Privacy*. 2018.

[42] JD Work. "Evaluating Commercial Cyber Intelligence Activity". In: *International Journal of Intelligence and CounterIntelligence* (Jan. 2020), pp. 1–31.

A INTERVIEW PROTOCOL

We conducted our interviews along the following questions:

- What does 'cybersecurity threat intelligence' mean to you?
- Which threat intelligence sources does your organization pay for?
- Which non-commercial sources are important for your organization? Are these as important as the paid sources?
- Did you ever discontinue a source?
- For each paid TI source, discuss:
  - What are the costs of this source to your organization?
  - What does this source consist of? How do you receive it?
  - How is this source used in your organization? Which systems or processes rely on it?
  - Which results of the use of this source are most valuable? How often does this occur? What are strengths and weaknesses of this source?

B CODEBOOK

The codebook is composed of four tables. Tables 3 & 4 in Section 5, and the two tables below that reflect Section 7.

USE CASES OF TI	Respondents *n=14*
Network detection	93%
Situational awareness	64%
SOC prioritization	50%
Informing business decisions	36%
Enrichment of more intel	36%
Improving end user awareness	29%
Threat hunting	29%
Informing security engineering	21%
Reducing financial fraud	14%

Table 5: Network detection is the most prominent use case of TI, though more than half of respondents also describe situational awareness and SOC prioritization.

TI VALUE PERCEPTIONS	Respondents *n=14*
Actionability	
Providing context	100%
Timeliness	50%
Comprehensiveness	50%
Suitable abstraction level	36%
Interpretability	21%
Visualized well	14%
Relevance	
Sectoral focus	64%
Geographic focus	50%
Coverage of relevant threats	50%
Ability to correct bias	14%
Confidence	
Automatability	79%
Confidence in vendor	71%
Original contribution	50%
Accuracy	43%
Selectiveness	29%
Affordability	21%

Table 6: We identify actionability, relevance, and confidence as three aspects of value of a TI source. All respondents share the view that TI should provide context.