A Quantitative Evaluation of Trust in the Quality of Cyber Threat Intelligence Sources

Thomas Schaberreiter
Veronika Kupfersberger
University of Vienna
Vienna, Austria
thomas.schaberreiter@univie.ac.at
veronika.kupfersberger@univie.ac.at

Konstantinos Rantos
International Hellenic University
Kavala, Greece
krantos@teiemt.gr

Arnolnt Spyros
Alexandros Papanikolaou
Innovative Secure Technologies
Thessaloniki, Greece
a.spyros@innosec.gr
a.papanikolaou@innosec.gr

Christos Ilioudis
International Hellenic University
Thessaloniki, Greece
iliou@it.teithe.gr

Gerald Quirchmayr
University of Vienna
Vienna, Austria
gerald.quirchmayr@univie.ac.at

ABSTRACT

Threat intelligence sharing has become a cornerstone of cooperative and collaborative cybersecurity. Sources providing such data have become more widespread in recent years, ranging from public entities (driven by legislatorial changes) to commercial companies and open communities that provide threat intelligence in order to help organisations and individuals to better understand and assess the cyber threat landscape putting their systems at risk. Tool support to automatically process this information is emerging concurrently. It has been observed that the quality of information received by the sources varies significantly and that in order to assess the quality of a threat intelligence source it is not sufficient to only consider qualitative indications of the source itself, but it is necessary to monitor the data provided by the source continuously to be able to draw conclusions about the quality of information provided by a source. In this paper, we propose a methodology for evaluating cyber threat information sources based on quantitative parameters. The methodology aims to facilitate trust establishment to threat intelligence sources, based on a weighted evaluation method that allows each entity to adapt it to its own needs and priorities. The approach facilitates automated tools utilising threat intelligence, since information to be considered can be prioritised based on which source is trusted the most at the time the intelligence arrives.

KEYWORDS

Cooperative and collaborative cybersecurity, cyber threat information sharing, cyber threat intelligence source evaluation, quality parameters, trust indicators

ARES '19, August 26–29, 2019, Canterbury, United Kingdom

ACM ISBN 978-1-4503-7164-3/19/08…$15.00
https://doi.org/10.1145/3339252.3342112

ACM Reference Format:
Thomas Schaberreiter, Veronika Kupfersberger, Konstantinos Rantos, Arnolnt Spyros, Alexandros Papanikolaou, Christos Ilioudis, and Gerald Quirchmayr. 2019. A Quantitative Evaluation of Trust in the Quality of Cyber Threat Intelligence Sources. In *Proceedings of the 14th International Conference on Availability, Reliability and Security (ARES 2019) (ARES '19), August 26–29, 2019, Canterbury, United Kingdom.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3339252.3342112

1 INTRODUCTION

The landscape of cyber threats is rapidly evolving, thus making it harder for security experts to deal with them. Cyber attacks rely on sophisticated methods in order to exploit the target system, while also being stealthy to avoid detection from defence mechanisms on the target system. Furthermore, attackers collaborate with each other by sharing tools and services, thus increasing the effectiveness of their attacks.

From the defenders' point of view, any information regarding the behaviour and the methods used by cyber attacks is very valuable for either preventing or confronting them. The cybersecurity community in the context of developing effective defence mechanisms, has started sharing Cyber Threat Information/Intelligence (CTI) to help organisations protect their assets against constantly evolving as well as emerging threats in the cyberspace. Cyber threat information is defined by NIST as any information that can help an organisation identify, assess, monitor, and respond to cyber threats (NIST 800-150) [11]. Moreover, cyber threat intelligence platforms are being built that are capable of consuming various cyber threat information feeds, analysing and enhancing the information (adding intelligence), and thus generating knowledge with regards to the status of the cyber threat information.

CTI is a way of expressing knowledge about cyber threats and vulnerabilities in a structured way, after relevant information has been collected, aggregated, evaluated, analysed, or enriched using appropriate analysis techniques [11]. A comprehensive definition of cyber threat intelligence is also provided by Gartner, according to which *"Threat intelligence is evidence-based knowledge, including context, mechanisms, indicators, implications and actionable advice, about an existing or emerging menace or hazard to assets that can*

be used to inform decisions regarding the subject's response to that menace or hazard" [14].

The type of information shared by these emerging CTI sources varies and might include suspicious domain names, hashes for malicious executables, IPs where malicious activity might originate, or even descriptions of techniques used by attackers, tools and exploits. In practice the level of analysis carried out by a source prior to sharing this information varies significantly. While some CTI sources might choose to share information without properly filtering or evaluating it to speed up sharing, others might adopt a different approach which requires a proper analysis prior to disseminating it. Moreover, CTI sources might provide similar information to other sources. Aggregators are types of CTI sources that typically combine information from various sources and of various types and share them as separate feed, while removing duplicates and potentially enriching the information with intelligence. Aggregators are therefore likely to provide more complete information, although not as fast as the original source.

When deciding to rely on a CTI source and integrate its feeds into security tools deployed in the organisation, one has to consider the above and other properties that can be used to evaluate CTI sources. In this work a methodology is presented to assess the trust in the quality of a CTI source by continuously re-evaluating the trust once new intelligence is shared by a source, and setting it in context with other relevant CTI sources. Trust in the context of this work stands for the computational trust evaluation as described by [2] conducted by the computation of a trust indicator as described in Section 4.2, while quality represents how good the information of a source is. This is achieved by following the closed world assumption, assuming that all the threat intelligence shared by the considered sources make up the entire world view of threat intelligence. A set of quantitative parameters is defined that is able to determine relevant quality aspects for threat intelligence, evaluating each new message against the information contained in the world view, and a trust indicator for each source is derived from those parameters at regular intervals. The advantage of this approach is that trust evaluation can be done in an online fashion, and the trust in the quality of a source considers each threat intelligence message shared by a source. No prior evaluation such as the establishment of a training set is required since the approach utilises the sources considered in the world view as baseline to validate against. It should be noted that while the goal of this research is to establish trust indicators, the core contribution of the research is to define the quantitative parameters at the core of the indicators. The actual calculation of trust is done by applying a previously defined trust model, and no contribution to research into computational trust is given in this work.

The methodology is built around the STIX (Structure Threat Information Expression) [20] standard, however without excluding sources that do not adopt the same standard as information provided in other formats can be converted to STIX. Other relevant standards like OpenIOC [19] provide a similar framework to structure threat intelligence, and the presented approach can be adapted to those standards easily.

The core of this paper is structured as follows: Existing approaches and relevant research is described in Section 2, followed by an overview of the current cyber threat intelligence landscape in Section 3. In Section 4 the methodology is introduced and described in further detail in Subsection 4.1 and 4.2. The benefits and challenges of this new approach to determining a trust value for CTI sources are outlined in Section 5, where an example calculation is also presented. Finally, a summary and outlook on future work is given in Section 6.

2 RELATED WORK

There are different possible approaches for evaluating cyber threat intelligence sources, depending on what the collected information is needed for and what type of source is considered. One such approach would be to analyse a source based on the quality of the data or information provided [3, 4, 22]. This becomes even more challenging in the era of big data and big data analysis. The authors in [4] introduce five dimensions (availability, usability, reliability, relevance and presentation quality), each divided into their individual elements, resulting in a big data quality assessment framework. In [22] the authors have demonstrated the necessity of information quality evaluation specific to big data in the intelligence community context. Their solution is to introduce a set of metrics to determine the quality of the source as well as an approach to validate said source. Another approach would be to identify metrics and indicators specific to the domain the information will be used for. An example for such a solution has been proposed by the authors in [3], who have structured their methodology into five assessment criteria: syntax accuracy, completeness, timeliness, situation certainty and consistency and relevance. This approach was designed specifically for the context of emergency situational awareness, aiming to improve critical information received by emergency response teams by assessing the sources based on above criteria. Evaluation methods in the area of cyber threat intelligence so far have been limited, those that have been devised are mostly theoretical and not yet publicly available [15] [18]. The criteria selected by the authors in [18] are useful for a general assessment of cyber threat intelligence sources, and must be evaluated manually by experts via a questionnaire, which determine weights for each criterion. These are then adjusted using a multi-objective algorithm, resulting in a final weight per criterion. This approach evaluates commercial cyber threat intelligence providers, stating that work on the micro level of threat intelligence is still to be undertaken. The authors in [13] have created an automated solution scanning thousands of blog entries and creating relevant cyber threat entities after matching information collected. This tool uses NLP (natural language processing) to identify threat information in unstructured text. While it focuses more on cyber threats on a micro level, it does not offer any analysis of the collected information. The approach presented in [15] uses the automated analysis of each single cyber threat message to derive an overall rating for the source it has come from. The authors based their ranking method on Google's PageRank [17], and provide an algorithm that *"ranks feeds according to the originality of their content and the reuse of entries by other feeds"*.

Another approach to evaluating information sources would be to focus on how much a source can be trusted. This research has been mostly applied to peer-to-peer information sharing platforms [16, 21] or data provided by commercial anti virus providers [1]. In the peer-to-peer community, where each peer can access and share

information, it is essential to trust said peers. Current peer-to-peer networks use personal, face-to-face validation of their peers and base their trust on personal experience. One of the attempts to automate trust distribution is based on four attributes, each weighted based on expert input: sharing activity, stakeholder rating, same source and same industry. The authors in [21] acknowledge remaining vulnerabilities of their approach in regards to collusion attacks or malicious stakeholders. Similar faults have been mentioned by the authors in [16], who have implemented a distributed data interaction model as well as a generic scoring model to evaluate the contributed attributes by each peer to the MISP (Malware Information Sharing Platform and Threat Sharing) community.

3 THE CYBER THREAT INTELLIGENCE LANDSCAPE

Cyber threat intelligence, or more generally cybersecurity-related information sharing is an essential part of the collaborative and cooperative cybersecurity efforts to effectively enhance security in cyber space, as laid out by the European cybersecurity strategy [7]. Two legislative actions that enforce this strategy are the Network and Information Security (NIS) Directive [8], as well as the General Data Protection Regulation (GDPR) [9] that require organisations to share cyber incident information as well as data breach information with the relevant authorities. Threat intelligence communities intend to utilise this information, gain a better understanding of the situation through analysis and share the added intelligence with the communities, organisations and the general public. In this collaborative and cooperative cybersecurity environment, both public as well as commercial and community-based actors have emerged that share CTI based on public mandate, for commercial profit or simply based on an interest to enhance security in cyberspace. Examples of such shareable information include:

- Indicators, i.e., system artefacts or observables that contain patterns which can help identify suspicious or malicious activity.
- Tactics, Techniques and Procedures (TTPs), i.e., (detailed) description of the behaviour of an actor.
- Security alerts, i.e., notification, usually human-readable regarding security issues, such as vulnerabilities.
- Threat intelligence reports, i.e., collections of threat intelligence for various topics, such as threat actors, malware, and attack techniques.
- Recommended security tool configurations, regarding automated collection and processing as well as healing of identified security issues.

Focusing on more automated solutions adopted for machine-to-machine CTI sharing, the exchanged CTI contain information in the form of Indicators of Compromise (IOC) which are forensic artefacts related to the security incident (e.g. malware file hashes, IP addresses, virus signatures). Given the complexity and the dynamic nature of the present cyber attacks, sharing and processing of IOCs has to be automated, so as to be completed within a reasonable time frame. Hence, a structured representation of CTI according to a common standard is essential, so as to maximise its usability potential. Two of the most popular and widely used standards for sharing CTI are the Structured Threat Information Expression (STIX) [20] and OpenIOC [19].

STIX is a language and serialisation format used to exchange cyber threat intelligence. It is expressive, as it includes IOCs and additional cyber threat information, such as techniques and procedures, indicators, cyber observables, campaigns and threat mitigations. Furthermore, it is flexible given that it can be extended with custom user-defined fields. The latest version of STIX (2.x) is defined using JSON schemas, thus rendering it easier to parse and expand than its XML-based predecessor (STIX 1.x).

OpenIOC, developed by Mandiant, is an extensible XML schema containing technical characteristics that identify a known threat, an attacker's methodology or other evidence of a compromise.

CTI can be obtained from either public sources, or from licensed ones by paying a fee. However, given the abundance of sources, suitable metrics are required for conducting an appropriate evaluation that will help towards minimising false alerts, as well as not missing any valuable threat information.

The literature review made it evident that limited metrics or scoring systems exist and that most of them were following a qualitative approach. Additionally, none of the qualitative approaches found in the literature fully captured the specific aspects the authors considered the most important when choosing a suitable threat intelligence source. As listed in Table 1, 6 categories of evaluation properties were defined that allow to assess and classify the types of sources available and how they share information. The *Quality* criterion indicates the type and complexity of information shared by a source, the *Provider Classification* assesses the distribution models, the *Licensing Options* identifies the potential restrictions of usage for provided data, the *Interoperability* criterion assesses different standards used to share CTI to see how well they support existing tools, *Advanced API support* assesses how data can be accessed by a consumer, and *Context Applicability* lists different classes of CTI that can be shared by sources to assess its applicability to specific contexts. The definition of the evaluation criteria presented in this work are based on the experiences of the authors in the H2020 project CS-AWARE[1], and a more extensive description of the methodology and expertise involved in defining those criteria is described in CS-AWARE project deliverable D2.2.

The properties examined at source level according to Table 1, which can be done in most cases by evaluating published properties of a source, may be sufficient for many use cases. However, considering that the quality of a source may vary over time, a quantitative approach that validates those aspects continuously will lead to more accurate results for cases where the quality of data to e.g. avoid misinformation or false positives is important. Taking the properties mentioned in Table 1 as a basis, a methodology based on purely quantitative parameters was introduced, resulting in the methodology presented in Section 4.

4 METHODOLOGY

While most of the current threat intelligence evaluation methods discussed in related work in Section 2 focus on a direct estimation of the quality of information provided by a source, we argue that such a quality assessment will always be incomplete and inaccurate

[1] https://cs-aware.eu/

Table 1: Properties for evaluating the quality of a CTI source.

1	Type of Information
1.1	Indicators
1.2	Sightings
1.3	Courses of Action
1.4	Vulnerabilities
2	Provider Classification
2.1	Data Feed Provider
2.1.1	Original Provider
2.1.2	Aggregator
2.2	Intelligence Platform
2.3	Report Provider
3	Licensing Options
3.1	Open
3.2	Restricted
3.3	Commercial
3.4	Information Reuse
3.4.1	Commercial: Allowed
3.4.2	Academic: Allowed
3.4.3	Personal: Allowed
4	Interoperability
4.1	*Supported formats*
4.1.1	STIX1
4.1.2	STIX2
4.1.3	PlainText
4.1.4	OpenIOC
4.1.5	RSS
4.1.6	JSON (non-STIX)
4.1.7	CSV
4.2	*Supported data exchange formats*
4.2.1	TAXII1
4.2.2	TAXII2
5	Advanced API Supported
5.1	Filtering based on dates
5.2	Filtering based on type of information
6	Context Applicability
6.1	Vulnerabilities
6.2	Threats
6.3	Campaigns
6.4	Hashes
6.5	Recommendations
6.6	Incidents (Sightings)

to a degree that can make it difficult to draw conclusions about the actual quality of information provided by a source. The main reason is that current methods are either qualitative indicators that require manual assessment and constant manual re-assessment in order to provide quality indications, or quantitative approaches that conduct one-time tests to assess the quality at specific points or frames in time. It is our assumption that for a comprehensive and meaningful evaluation of a threat intelligence source, constant monitoring and re-evaluation is required and that a single quality indicator is not the best way to represent this evaluation. The presented methodology is able to derive a trust value per source based on the quantitative evaluation of parameters for each message provided by a threat intelligence source.

Our approach is based on a closed world assumption, where it is assumed that the information provided by a fixed set of threat intelligence sources encompasses the complete set of information available in the world of threat intelligence. Thus, in order to validate information provided by a source, it only needs to be validated against the information provided by the other sources considered in this closed world. This assumption allows to draw concrete conclusions about specific parameters, such as how fast threat information is made available compared to other sources, or how original the provided information and added intelligence is compared to the other sources.

The methodology is composed of two main aspects: A definition of quantitative parameters and how they can be derived is described in Section 4.1, and a method of deriving a trust indicator from those parameters is described in Section 4.2.

Figure 1 shows an overview of the procedure of how the trust value is generated. Beginning with the collection of data from the pre-selected sources $s_1, \ldots, s_n$, of which each single message is analysed. Every message m_t by each source is directly evaluated in regards to the selected parameters as well as added to the database, which constitutes the overall world view of the evaluation. The parameter evaluation is applied to every new message provided by each of the sources as soon as the system receives it and results in $p_1 s_1, \ldots, p_m s_n$. The results of these parameters are then combined to determine the overall trust indicator, which is calculated in predefined time intervals and results in a final trust indicator for each source $TI_{s_1}, \ldots, TI_{s_n}$.

4.1 Quantitative evaluation parameters

In this Section the properties for evaluating source quality that have been defined in this work, as well as a tentative approach of deriving those parameters based on STIX 2.0 as a representative threat information exchange standard, are described. Table 2 summarises the aforementioned parameters. The output provided by them is restricted to the interval $[0, 1]$ (normalisation is applied, if necessary).

The **Extensiveness** parameter evaluates how much effort a threat intelligence source puts in describing specific aspects of an information that are not explicitly required. An indicator of this effort can be how many optional properties (in addition to the required properties) relating to a specific object defined by CTI sharing standards are filled by the source. In many cases a simple sum of filled-in optional properties will be a sufficient indicator, but depending on the specific object and/or protocol, a weighting of optional properties may be relevant in order to put more emphasis on those that are deemed more important than others. In STIX 2.x all objects have properties that are either required to be provided or optional. Moreover, these properties are grouped into common and object-specific ones. For example, when constructing an instance

Table 2: Quantitative parameters for cyber threat intelligence source evaluation.

	Parameter	Description
p_1	Extensiveness	Evaluates how many optional parameters are filled in.
p_2	Maintenance	Determines how often messages are updated.
p_3	False Positives	Determines how often messages of a source are invalidated.
p_4	Verifiability	Expresses how often a source verifies the information they provide by linking their source.
p_5	Intelligence	Indicates how much added value a source offers in their messages by linking it to other objects.
p_6	Interoperability	Based on which data format a source provides their data in.
p_7	Compliance	Determines how compliant a source is to the standard they use.
p_8	Similarity	Evaluates how similar specific entries of two sources are.
p_9	Timeliness	Analyses which source provides information the quickest.
p_{10}	Completeness	Indicates how much of the entire world a single source represents.

of the object *sighting*, it has the following object-specific properties that are required:

- `type`
- `sighting_of_ref`

whereas, the following are optional:

- `first_seen`
- `last_seen`
- `count`
- `observed_data_refs`
- `where_sighted_refs`.

The more of these a source has filled in, the better it is for the overall extensiveness value. Equation 1 details the evaluation of the extensiveness parameter p_1, where o_i is the sum of the filled-in optional properties in a specific message i, max y_i is the maximum number of optional properties defined for this specific type of message i by the CTI sharing standard like STIX, and z is the total number of messages shared by the source.

$$p_1 = \frac{1}{z} \sum_{i=1}^{z} \left(\frac{o_i}{\max y_i} \right) \tag{1}$$

The **Maintenance** parameter provides an indication of how information provided by a source evolves over time and can be evaluated by monitoring how often shared information is updated with new and refined information. When updating messages in STIX, the version number of the object is changed. Only the entity that created the object, can issue an update for the object. Equation 2 details the calculation of the maintenance parameter, where: u_i is the number of updated objects provided by source s, and $\text{avg}(p_2 s_1, \ldots, p_2 s_n)$ is the average maintenance parameter p_2 of the sources $1 \ldots n$ at the time. The parameter is normalised to result in a value between 0 and 1.

$$p_2 = \left\| \frac{\frac{1}{z} \sum_{i=1}^{z} u_i}{\text{avg}(p_2 s_1, \ldots, p_2 s_n)} \right\| \tag{2}$$

Analysing the average number of **False Positives** of a source allows an interpretation of how well a source deals with information that turns out to be wrong. STIX allows the inclusion of the optional property `revoke`, which if set to `true` marks the message as being invalid. The authors assume that the fewer false positives a source has, the better the information it provides is. However, in practice it may turn out that a source with a higher false positive rate indicates higher quality, since this source properly invalidates previous information if new evidence suggests a false positive. In that case this assumption may change in future. Equation 3 details the calculation of the false positive parameter, where F_{s_x} is the number of false positives provided by source s_x and F_{s_i} represents the number of false positives for source s_i in the world view.

$$p_3 = 1 - \left(\frac{F_{s_x}}{\sum_{i=1}^{n} F_{s_i}} \right) \tag{3}$$

The **Verifiability** parameter indicates whether threat information provided by a source is externally verifiable. Some threat intelligence standards allow to specify original sources utilised to derive the provided information. One way of evaluating the verifiability parameter is to check if a threat intelligence source tends to provide verifiable sources. In STIX this is done via `external_references`, which points to external sources that also provide information about the shared content. Equation 4 details the calculation of the verifiability parameter, where r_i represents the number of references provided by source s divided by the average of references per source in the world view. The parameter is normalised to result in a value between 0 and 1.

$$p_4 = \left\| \frac{\left\| \frac{1}{z} \sum_{i=1}^{z} r_i \right\|}{\text{avg}(p_4 s_1, \ldots, p_4 s_n)} \right\| \tag{4}$$

The **Intelligence** parameter indicates how much added value a threat intelligence source provides in addition to the basic threat information. A way of determining added intelligence is to identify how much additional information (like analysis, descriptions, mitigation actions) are added to the original information. A way of

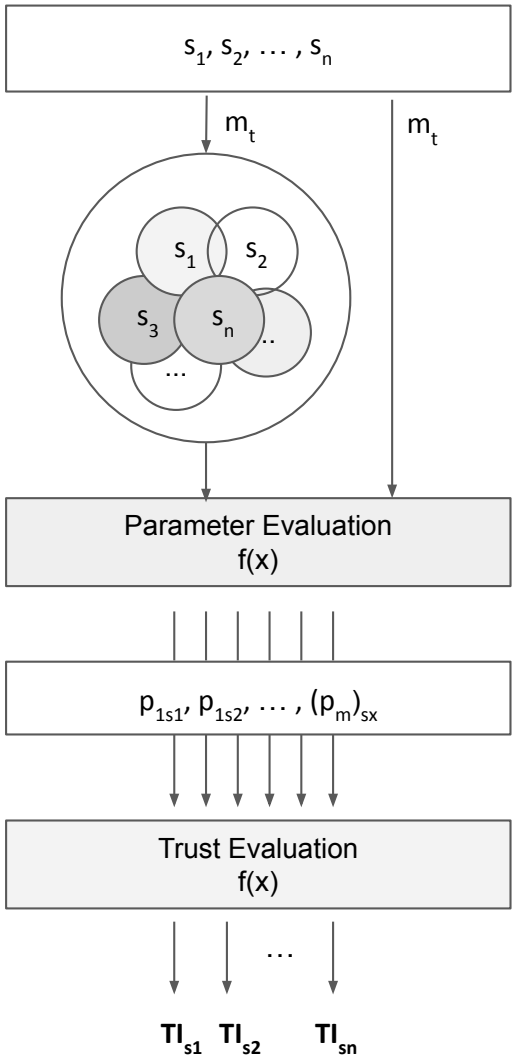

Figure 1: Methodology for CTI evaluation.

validating this parameter, according to state-of-the-art threat intelligence standards, is to evaluate how many additional objects are linked to a threat intelligence parameter. For STIX this can be done by evaluating the average number of relationships an object has of some type, such as `related-to` or `derived-from`. Equation 5 details the calculation of the intelligence parameter, where l_i is the number of links in objects from source s_i that are related to other objects divided by the average of linked objects per source in the world view. The computation of the average and the final parameter are normalised to be between 0 and 1.

$$p_5 = \left\| \frac{\left\| \frac{1}{z} \sum_{i=1}^{z} l_i \right\|}{\text{avg}(p_5 s_1, \ldots, p_5 s_n)} \right\| \tag{5}$$

The **Interoperability** parameter is used to evaluate how well CTI sources, as providers of information, work together with tools that act as information consumers. While this parameter may not be estimated as intuitively as the other parameters, one can derive an estimate of interoperability based on the format the data is exchanged, with commonly used and relevant standards like STIX having a bigger impact on interoperability than less well known or inadequate threat standards. This is done by firstly determining which format is being most commonly used in the world view, and ranking them accordingly. The sources are evaluated based on how high the format they use is ranked in the overall word-view rating. Equation 6 details the calculation of the interoperability parameters, where b_i is the number of sources in the world view providing information using the same standard divided by the total number of sources.

$$p_6 = \sum_{i=1}^{n} \left(\frac{b_i}{n} \right) \tag{6}$$

The **Compliance** parameter evaluates how well the messages provided by a threat intelligence source comply to the data standard used to model the threat information. Compliance can be checked by evaluating the messages using either the reference implementation validators (usually provided by standardisation organisations to validate compliance) or by implementing such a validator, should one not be available by default. The value of the parameter is calculated by determining how much percent of all messages over time have been 100% compliant to the corresponding standard, such as STIX. Equation 7 details the calculation of the compliance parameter, where c_i is the number of objects that originate from source s_i and are compliant to the CTI exchange standard used by the specific source. The parameter is normalised to result in a value between 0 and 1.

$$p_7 = \left\| \frac{\frac{1}{z} \sum_{i=1}^{z} c_i}{\text{avg}(p_7 s_1, \ldots, p_7 s_n)} \right\| \tag{7}$$

The **Similarity** parameter indicates how similar information provided by a source is to the same information provided by another source. The goal is to compare semantic text entries and see if a source would provide unique information relating to specific threat information. This can be evaluated using one of the existing text similarity algorithms such as the Jaccard or Cosine Similarity [10]. As with other comparison parameters, similarity also requires the evaluation of messages by comparing the type, followed by the similarity index computation. The similarity index will be a percentage on per message basis, which is then summed up for the overall similarity score of the source. Equation 8 details the calculation of the similarity parameter, where the similarity index y_i of all messages that are similar similar to others are added and an average is generated.

$$p_8 = \frac{1}{z} \sum_{i=1}^{z} (y_i) \tag{8}$$

The **Timeliness** parameter indicates which of the sources is usually the fastest at providing threat intelligence relating to the same event. This parameter can be evaluated by cross-referencing messages with messages from the other sources in the world view. Every incoming message ID is compared to all existing IDs in the database, to see if this exact information already exists or if the

source at hand is the first to publish. If there is no match of IDs, first the type of the object is compared to existing data, followed by the similarity evaluation. Depending on how similar two messages are, they are assumed to be equal and the message can be evaluated in regards to its timeliness. Equation 9 details the calculation of the timeliness parameter, where $\min t_i$ is the fastest source for object i, expressed in UTC timezone (using an integer representation, such as the UNIX Epoch time) and $(t_s)_i$ is the time that source s has shared the object with the same ID or a similar one.

$$p_9 = \frac{1}{z} \sum_{i=1}^{z} \left(\frac{\min t_i}{(t_s)_i} \right) \tag{9}$$

The **Completeness** parameter indicates how much of the information contained in the world view is covered by a single threat intelligence source. This parameter can be evaluated by cross-referencing the messages from a source with the information contained in the world view. The first step is to identify if a new message already exists in the world view and if so, how exact they match. This is determined by firstly comparing the IDs and if there is no match, comparing the type followed by the similarity evaluation. Based on how similar two messages are, equality can be assumed and will build a basis for the computation of completeness of a source. Equation 10 details the calculation of the completeness parameter, where $|A|$ is the number of objects found in the world view with different ids or not similar to those provided by a single source and $|B|$ is the total number of objects in the world view.

$$p_{10} = \frac{|B| - |A|}{|B|} \tag{10}$$

Some of the parameters can be combined in order to minimise computational effort, Figure 2 shows the combined procedural steps for determining *Timeliness*, *Completeness* and *Similarity*. The first step is to compare the ID of the incoming message to all existing message IDs in the world view. In the case of a match, the similarity of the two messages is evaluated immediately. Should there be no match, the next step is to select all prior messages of the same type and test for similarity.

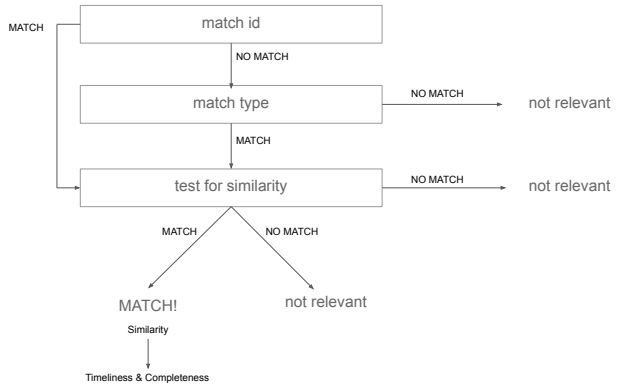

Figure 2: Calculating comparison parameters.

4.2 Trust-based quality indicator

Based on the parameters introduced in the previous section, the objective is to derive a trust indicator to assess the trust in the quality for each source contained in the world view. The key aspects of such a trust model need to be that it takes into account each parameter according to its importance for a specific use case, and that the trust can be updated if new evidence arrives, to account for the dynamically evolving cybersecurity environment. The goal of the trust model chosen for this purpose is to (a) provide a simple and effective way to combine the parameters according to their importance, and (b) has proven to be applicable in use cases similar to the one proposed in this research.

A model based on an averaged weighted sum was chosen, based on a trust model developed by Caldeira et al. in [5] and applied to the context of evaluating trust in risk estimates shared by potentially unknown or untrustworthy entities in [6]. The model is able to evaluate trust in a dynamic environment, and constantly update the trust level if new evidence arrives while at the same time discounting evidence further in the past to put more weight on recent events. It should be noted that there are many trust models available considering more complex trust concepts that may be relevant in this use case at some point, and it may be interesting to replace or adopt the trust model in a later phase of the research. At this point however, the chosen trust model is covering all the requirements of the use case.

The trust indicator derived in this work can be seen in Equation 11. For each threat intelligence source s_x in the world view, a trust indicator TI at time t can be evaluated by adding a weighted sum of the parameters at time t to the previous existing value of the trust indicator at time $t-1$. A recalculation of the trust indicator should only be triggered if the weighted sum of the parameters $\sum_{n=1}^{m} \omega_n * (p_n)_{s_x}(t)$ has changed during the time frame. The ageing factor D multiplied to the original trust indicator $TI_{s_x}(t-1)$ is used to put less emphasis on past events and highlight more recent events. This is to account for the fact that, while the past should not be dismissed completely, threat intelligence sources should be judged more on what quality they are able to deliver in the present.

In the trust parameter calculation each parameter p_n specific to source s_x is factored in at the relevant time t. The parameters are weighted with a factor ω according to their importance for a specific use case. In some use cases a parameter like *accuracy* may be more important, while in other cases *timeliness* may be more relevant. In Equation 11 those parameters with a higher ω value gain more relevance towards the overall trust in the quality of the threat intelligence source.

Finally, it is important to discuss the implications of choosing the time t, specifying at which intervals the trust indicator is recalculated. Since parameter calculation is performed independently from the time frame within which the trust indicator is recalculated, two major factors should be considered when choosing the time factor t:

- What is the expected rate in which the individual parameters are changing their values considerably. The rate of the most frequently changing parameter should be taken into consideration.

- What are the needs of the specific use case the trust indicators are evaluated for.

Based on those considerations, an applicable interval should be chosen to represent t (e.g. hourly, daily, weekly, monthly).

$$TI_{s_x}(t) = \frac{D * TI_{s_x}(t-1) + \sum_{n=1}^{m} \omega_n * (p_n)_{s_x}(t)}{D + \sum_{n=1}^{m} \omega_n} \tag{11}$$

5 APPLICATION EXAMPLE

Given the current status of cyber attacks and cyber threats worldwide, it is of high importance to have suitable defence mechanisms and mitigation procedures available, so as to protect effectively any valuable assets. As has already been presented in the previous sections, sharing CTI is vital for having effective defence against all modern kinds of cyber threats. One such defence mechanism is developed through the CS-AWARE project, the threat detection mechanism of which employs big data analytics and CTI [12]. This triggered a need for developing a suitable CTI source assessment, as getting data from reliable CTI sources would improve the system's threat detection and minimise the number of false positives. It was therefore necessary to devise suitable parameters for conducting a quantitative assessment, in order to be able to use the sources that would best meet our needs.

Within the scope of this research, several cyber threat intelligence sources were examined. Nevertheless, while there were many such sources available, it quickly became apparent that there were significant differences among them in terms of the provided information, ranging from volume size and refresh interval to CTI standard conformity. It would certainly be beneficial for such attributes to be considered in a potential assessment, since they may prove to be quite important factors affecting the final outcome. For instance, supposing that a given CTI source does not fully comply with the involved protocols and standards, any information it makes available requires additional effort during its parsing, before it becomes fully usable.

Perhaps the first challenge encountered was to formulate an assessment that would be able to work with a set of sources, semantically rich or not. For example, sources providing plain lists of IP addresses should somehow be compared to others that offered CTI in STIX format (to the degree that this is attainable).

Furthermore, the assessment parameters should have some degree of flexibility, that will be easy to configure. This approach will allow for dynamically adding or removing sources and will facilitate their use by entities with different requirements. For instance, while an entity may be more interested in the rate at which new CTI becomes available, another entity may regard information reliability to be of higher importance. An additional aim of the proposed assessment methodology was to keep it as simple as possible and avoid requiring expert intervention by data analysts, who would adjust weights or other parameters involved in the various calculations.

The major advantage of the approach presented in this paper is that an evaluation of sources can be carried out automatically without the need for interference by human experts. From an operational perspective this leads to several benefits. Firstly, the sources can be assigned priorities dynamically, according to their trust score over time. Secondly, each source will automatically be re-evaluated, as soon as it makes available new CTI data. Moreover, due to the absence of a human expert the introduction of any bias in the final results is avoided.

Another advantage of the proposed approach is that new sources can dynamically be added to the set that is subjected to evaluation. Whenever a new source is added, the methodology allows to evaluate the parameters as well as the trust in the quality of the source based on the information that is collected from the new source. For conducting the evaluation, historical messages can be taken into consideration, which would help in establishing a common reference point in the past for all sources and therefore conducting a 'more direct' comparison. Alternatively, the evaluation can be limited to future messages only, thus creating a reference 'point zero' at the time a source started being evaluated. The only restriction in order to ensure consistency in the knowledge base of the world view is to not add historical messages of the new source to the world view, because it would influence the parameter and trust validation of all sources if new past messages are added without triggering a re-evaluation of all sources.

From a run-time and scalability perspective the model primarily depends on the number of messages stored in the world view and on the parameters considered for the calculation of each quality parameter. We expect a linear run-time behaviour for each parameter and therefore a rough run-time estimate can be given as $O(m * n)$, where m is the number of messages contained in the world view and n is the maximum number of parameters that can be modelled according to threat intelligence standards for each message.

5.1 Computational example of the extensiveness parameter

In what follows we present an indicative example where the extensiveness quality indicator (p_1) is calculated according to Equation 1. The chosen data sample consists of 3 STIX messages obtained from MISP (event ID 24504, 21271 and 24362). The first message (event ID 24504) contains 7 SDO objects and for each of these objects, the number of optional parameters (filled-in and available in total) is presented in Table 3.

Therefore, for the first message:

$$\frac{o_1}{\max y_1} = \frac{10}{47}$$

Similarly, for the remaining two messages, the sums are calculated, respectively, as:

$$\frac{o_2}{\max y_2} = \frac{22}{193}$$

and

$$\frac{o_3}{\max y_3} = \frac{54}{172}$$

Hence,

$$p_1 = \frac{1}{3}\left(\frac{10}{47} + \frac{22}{193} + \frac{54}{172}\right) = 0.21$$

which means that this specific source, on average, has an extensiveness value of approximately 21% per message.

Table 3: Number of optional parameters for the SDO objects contained in MISP message with event ID 24504.

Object ID	Common		Object-specific	
	Filled-in	Max	Filled-in	Max
identity–5804fe16-fd94-4326-904f-07e6ac14012a	0	6	0	4
report–5ce65218-d350-423b-a303-339dac12042b	2	6	0	1
vulnerability–5ce652ae-ed7c-4d05-aee2-33a1ac12042b	1	6	1	2
observed-data–5ce65316-6b64-42b2-9549-0cd8ac12042b	2	6	0	0
observed-data–5ceaadd1-954c-464c-8a7f-142dac12042b	2	6	0	0
observed-data–5cee106b-2ea8-4846-8327-2f77ac12042b	2	6	0	0
marking-definition–dc2bb6b0-e841-46f0-bc97-77b025bce33b	0	0	0	4
Sum	9	36	1	11

6 CONCLUSION

In this paper we introduced a methodology for evaluating trust in the quality of threat intelligence sources based on a quantitative assessment of parameters defined to assess a wide range of aspects related to threat intelligence. The method is based on a closed world assumption, where each source (or the threat intelligence provided by each source) is evaluated against all other sources contained in the world view to assess how trusted the information provided by a source is expected to be in relation to the other sources in the world view. A set of parameters (Extensiveness, Maintenance, False Positives, Verifiability, Intelligence, Interoperability, Compliance, Timeliness, Completeness, Similarity) was introduced along with an approach to evaluate those parameters in a quantitative way for data compliant with state-of-the-art threat intelligence sharing standards like STIX 2.x. A method to derive a trust indicator for each threat intelligence source based on those parameters was also presented.

The main advantage of the proposed method as compared to other threat intelligence evaluation approaches is the possibility of online re-evaluation of the evaluation parameters every time new threat intelligence is shared by a source, allowing the dynamic adjustment of trust in the quality of sources. The utilisation of the closed world assumption allows interested parties to validate each source against other sources in the world view, representing an evaluation of how trusted a threat intelligence source is against the other relevant players in the field of threat intelligence. No expert intervention, such as the definition of a training set to establish a baseline, is required for the validation of sources.

Ongoing and future work is focused on the implementation of the methodology. An experimental setup for the collection of threat intelligence from sources supporting STIX 1.x, STIX 2.x and potentially OpenIOC standard, as well as the evaluation of parameters and trust indicators will be implemented to evaluate the validity of the methodology. The evaluation and validation results are expected to be published in a scientific publication.

ACKNOWLEDGMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 740723.

REFERENCES

[1] Omar Al-Ibrahim, Aziz Mohaisen, Charles Kamhoua, Kevin Kwiat, and Laurent Njilla. 2017. Beyond free riding: quality of indicators for assessing participation in information sharing for threat intelligence. arXiv preprint arXiv:1702.00552. (Feb. 2017).

[2] Donovan Artz and Yolanda Gil. 2007. A survey of trust in computer science and the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web* 5, 2 (2007), 58–71.

[3] Leonardo Castro Botega, Jéssica Oliveira de Souza, Fábio Rodrigues Jorge, Caio Saraiva Coneglian, Márcio Roberto de Campos, Vânia Paula de Almeida Neris, and Regina Borges de Araújo. 2017. Methodology for data and information quality assessment in the context of emergency situational awareness. *Universal Access in the Information Society* 16, 4 (2017), 889–902.

[4] Li Cai and Yangyong Zhu. 2015. The challenges of data quality and data quality assessment in the big data era. *Data Science Journal* 14 (2015), 2. DOI: http://dx.doi.org/10.5334/dsj-2015-002

[5] Filipe Caldeira, Edmundo Monteiro, and Paulo Simões. 2011. Trust and Reputation for Information Exchange in Critical Infrastructures. In *Critical Information Infrastructures Security*, Christos Xenakis and Stephen Wolthusen (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 140–152.

[6] Filipe Caldeira, Thomas Schaberreiter, Sébastien Varrette, Edmundo Monteiro, Paulo Simões, Pascal Bouvry, and Djamel Khadraoui. 2013. Trust based interdependency weighting for on-line risk monitoring in interdependent critical infrastructures. *International Journal of Secure Software Engineering (IJSSE)* 4, 4 (2013), 47–69.

[7] European Commission and High Representative of the European Union for Foreign Affairs and Security Policy. 2013. Cybersecurity Strategy of the European Union: An Open, Safe and Secure Cyberspace. JOIN(2013) 1 final. (2013).

[8] European Parliament and Council. 2016. Directive (EU) 2016/1148 concerning measures for a high common level of security of network and information systems across the Union. Official Journal of the European Union. (July 2016). https://eur-lex.europa.eu/eli/dir/2016/1148/oj.

[9] European Parliament and Council. 2016. Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union. (April 2016). http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679.

[10] Anna Huang. 2008. Similarity measures for text document clustering. In *Proceedings of the sixth New Zealand computer science research student conference (NZCSRSC 2008)*, Vol. 4. Christchurch, New Zealand, 49–56.

[11] Christopher S. Johnson, Mark Lee Badger, David A. Waltermire, Julie Snyder, and Clem Skorupka. 2016. *Guide to Cyber Threat Information Sharing.* Technical Report NIST SP 800-150. National Institute of Standards and Technology. DOI: http://dx.doi.org/10.6028/NIST.SP.800-150

[12] Veronika Kupfersberger, Thomas Schaberreiter, Chris Wills, Gerald Quirchmayr, and Juha Röning. 2018. Applying Soft Systems Methodology to Complex Problem

Situations in Critical Infrastructures: The CS-AWARE Case Study. *International Journal on Advances in Security* 11, 3 & 4 (2018), 191–200.

[13] Xiaojing Liao, Kan Yuan, XiaoFeng Wang, Zhou Li, Luyi Xing, and Raheem Beyah. 2016. Acing the IOC game: Toward automatic discovery and analysis of open-source cyber threat intelligence. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, Vienna, Austria, 755–766. DOI: http://dx.doi.org/10.1145/2976749.2978315

[14] Rob McMillan. 2013. *Definition: Threat Intelligence.* Technical Report G00249251. Gartner.

[15] R. Meier, C. Scherrer, D. Gugelmann, V. Lenders, and L. Vanbever. 2018. FeedRank: A tamper-resistant method for the ranking of cyber threat intelligence feeds. In *2018 10th International Conference on Cyber Conflict (CyCon)*. IEEE, Tallinn, Estonia, 321–344. DOI: http://dx.doi.org/10.23919/CYCON.2018.8405024

[16] Sami Mokaddem, Gerard Wagener, Alexandre Dulaunoy, and Andras Iklody. 2019. Taxonomy driven indicator scoring in MISP threat intelligence platforms. arXiv preprint arXiv:1902.03914. (Feb. 2019).

[17] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank citation ranking: Bringing order to the web.* Technical Report 422. Stanford InfoLab.

[18] Li Qiang, Jiang Zhengwei, Yang Zeming, Liu Baoxu, Wang Xin, and Zhang Yunan. 2018. A Quality Evaluation Method of Cyber Threat Intelligence in User Perspective. In *2018 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE)*. IEEE, New York, NY, USA, 269–276. DOI: http://dx.doi.org/10.1109/TrustCom/BigDataSE.2018.00049

[19] David Ross, Jason Shiffer, Tony Dell, William Gibb, and Doug Wilson. 2013. OpenIOC 1.1. Available online: https://github.com/mandiant/OpenIOC_1.1. (Sept. 2013).

[20] STIX 2017. Structured Threat Information Expression (STIX) version 2.0. OASIS standard https://www.oasis-open.org/standards#stix2.0. (July 2017).

[21] Thomas D Wagner, Esther Palomar, Khaled Mahbub, and Ali E Abdallah. 2018. A Novel Trust Taxonomy for Shared Cyber Threat Intelligence. *Security and Communication Networks* 2018, Article 9634507 (2018), 11 pages.

[22] Hongwei Zhu and Richard Y Wang. 2009. Information quality framework for verifiable intelligence products. In *Data Engineering*. Springer, Boston, MA, 315–333. DOI: http://dx.doi.org/10.1007/978-1-4419-0176-7_14