*applied sciences*

MDPI

*Article*

# Quality Dimensions for Automatic Assessment of Structured Cyber Threat Intelligence Data

**Algimantas Venčkauskas, Vacius Jusas * and Dominykas Barisas**

Department of Computer Science, Kaunas University of Technology, LT-51390 Kaunas, Lithuania; algimantas.venckauskas@ktu.lt (A.V.); dominykas.barisas@ktu.lt (D.B.)
* Correspondence: vacius.jusas@ktu.lt

**Abstract:** Cyber threat intelligence (CTI) has emerged as a promising approach to mitigating the effect of malicious activities. However, the potential usability of CTI data depends largely on their quality. The available CTI data quality assessment methods are either not fully automatic or deliver just a few dimensions. In this paper, we propose an automated CTI data quality assessment method that separately provides an assessment of CTI contents and confidence scores of CTI providers. Specifically, we introduce new dimensions to accommodate the requirements of the technical and tactical levels of CTI data. A comprehensive CTI quality assessment is proposed on CTI data provided in structured STIX 2.1 notation. Moreover, we present a visualization of the results to more easily interpret the obtained values of the quality dimensions. Extensive experiments on real datasets demonstrate that our proposed method can quantitatively and efficiently assess CTI data quality.

**Keywords:** cyber threat intelligence; CTI data quality; CTI quality dimensions; STIX

## 1. Introduction

Cyberattacks are part of our digital world. They disturb the normal functioning of the digital world. The effects of cyberattacks can cause great damage to participating organizations. To prevent cyberattacks, organizations share information available on the cyberattacks. A regular analysis of shared cyber threats must be performed since cyber threats are constantly evolving. The result of this analysis is called cyber threat intelligence (CTI). Sharing of CTI is very important because organizations cannot protect themselves in isolation from the changing landscape of cyber threats [1]. It also provides an opportunity for small organizations that do not have the resources to build an independent cyber threat intelligence program [2]. Analysis of shared CTI and proactive implementation of appropriate defensive measures can help to reduce the effects of cyberattacks.

A large number of sources providing information on emerging cyber threats are available to CTI consumers. These sources vary from open and commercial data feeds to open-source cyber threat intelligence platforms, e.g., the Malware Information Sharing Platform (MISP) [3–5]. The variety of cyber threat intelligence sources continues to grow, but the quality of the provided data remains in question since CTI can be posted online by anyone. Several studies [6–9] have highlighted that CTI data quality is one of the most common barriers to beneficial CTI exchange. The quality of the CTI data has a direct impact on the time required to respond to an incident [10]. Additionally, when the amount of available CTI increases, human analysts are not able to cope with it. There is a need to provide automatic assessments of CTI data quality. Rapid assessment of CTI data quality would help security experts to concentrate their attention on the CTI actually relevant to their interests.



Academic Editors: Juan-Carlos Cano and Jose María Alvarez Rodríguez

Received: 23 December 2024
Revised: 4 April 2025
Accepted: 12 April 2025
Published: 14 April 2025

**Citation:** Venčkauskas, A.; Jusas, V.; Barisas, D. Quality Dimensions for Automatic Assessment of Structured Cyber Threat Intelligence Data. *Appl. Sci.* **2025**, *15*, 4327. https://doi.org/10.3390/app15084327

Appropriately structured data and machine-readable formats are necessary for automatic assessment of data. Structured Threat Information Expression (STIX) is the most well-known language used to represent CTI in such a format [11]. The STIX format was proposed by S. Barnum in 2012 [12]. Two versions of STIX are currently available. STIX 1 is represented in Extensible Markup language (XML), while STIX 2 is represented in JavaScript Object Notation (JSON). STIX 2 is easier to use and it is preferred by the producers of CTI [13]. The latest available version of STIX is STIX 2.1. STIX 2.1 includes 18 different STIX Domain Objects (SDOs) and two different STIX Relationship Objects (SROs), which are used to express relationships between different SDOs [14]. In comparison to STIX 2, STIX 2.1 is complemented by numerous STIX Cyber Observable Objects (SCOs), which were previously represented in a different standard language called Cyber Observable Expression (CybOX). SDOs classify each piece of CTI with specific attributes to be populated. SCOs provide more technical details, such as IP or URL addresses.

The methodological approach employed in this article consists of a systematic literature review (SLR) followed by derivation of a set of literature-based quality dimensions.

This work makes a contribution to measuring structured CTI data quality automatically. We present an approach to assessing relevant CTI data quality dimensions in a standardized data format. For this purpose, we perform a thorough literature review and derive relevant literature-based CTI data quality dimensions. The dimensions are then adjusted to the STIX 2.1 format as their automatic calculation relies on its structure.

The distinguishable contributions of this paper are as follows:

- Providing a systematic literature review of CTI data quality dimensions;
- Defining separate calculations of CTI data quality for technical and tactical levels of using CTI. To the best of our knowledge, such an approach is novel for CTI data quality assessment;
- Defining a set of literature-based CTI data quality dimensions for structured CTI data provided in a machine-readable format;
- Creating measurement formulae for defined CTI data quality dimensions. Most of the formulae are original ones;
- Implementing defined formulae and providing experimental results of applications to real CTI data.

The remainder of this paper is organized in the following way: a review of the existing methods used to define CTI quality dimensions is provided in Section 2. Section 3 details a choice of quality dimensions for automatic assessment of CTI data and defines formulae for the chosen CTI quality dimensions. Section 4 considers implementation by presenting the results of the experiment and provides a discussion of the results obtained compared with related studies. Finally, Section 5 draws conclusions.

## 2. Review of Related Work

We performed a systematic literature review to define a set of literature-based CTI quality dimensions for CTI data. The search was limited to studies published between 2018 and 2024, with the exception of two studies. The exceptional studies are as follows: a seminal work on data quality by Wang and Strong (1996) [15], and ISO document ISO/IEC 25012 (2008) [16] on data quality models for data preserved in a structured format. The review is presented according to the chronological order of the publication of the studies.

Firstly, we consider the quality dimensions of general data. Then we present the evolution and application of these dimensions to CTI data.

Data quality dimensions, which are important to data consumers, were first presented in the seminal work of Wang and Strong [15]. Wang and Strong developed a hierarchical framework to capture aspects of data quality. Two surveys were used. The list of possible

data quality dimensions which came to mind when data consumers thought about data quality were produced during the first survey. The initial list included 118 data quality dimensions. The importance of these data quality dimensions to data consumers was assessed during the second survey. Finally, the data quality dimensions were divided into four categories and included 15 dimensions in total. However, the problem of calculating the values of the proposed quality dimensions was not considered.

Wang [17] introduced total data management methodology that was based on the data quality dimensions proposed by Wang and Strong [15]. Measurement of the data quality dimensions was not considered yet in this methodology either.

The next prominent data quality model according to the timeline is ISO/IEC 25012, presented in 2008 [16]. This model should not be considered as an alternative to the model proposed by Wang and Strong [15]. Actually, in conjunction with ISO 25024 [18], the model of ISO/IEC 25012 complements the model of Wang and Strong [15] by providing valuable insights into measurements methods for the data quality dimensions. ISO 25012 introduced fifteen data quality dimensions, which are called characteristics, that are divided into the following three groups: (1) inherent; (2) system dependent; and (3) inherent and system dependent.

Gao et al. [19] suggested a solution for the task of trust evaluation for threat intelligence sharing platforms (TISPs). A threat intelligence (TI) architecture was presented for trust evaluation of TISPs. The TI architecture model consists of three modules: TI collection and aggregation, TI graph construction, and TI trust evaluation. We are interested in the TI graph construction and TI trust evaluation modules. TI is presented in JSON format to construct graphs that relate to the objects in JSON format. Based on the constructed graphs, graph mining techniques to infer relationships among cyber threat infrastructures are applied. The following techniques are used: similarity-based inference (Simhash algorithm), correlation-based inference, and frequent pattern mining. The trust evaluation module extracts 24 features from the four dimensions: source/provenance, content, time, and feedback. We have to discard feedback, because it does not exist yet for the new TI. Calculation of the values of many features is not explained in detail. The model only needs feature definition. Consequently, many important details and results of the implementation of the proposed approach are left behind the scenes. Trust is very important for the evaluation of the data quality, since if there is no trust in the submitted TI, the data quality is not important. Therefore, we can borrow just ideas, concepts, and definitions, since some presented ideas and concepts seem to be appropriate and applicable for the data quality evaluation of cyber threat intelligence.

Li et al. [20] presented a quality evaluation method for cyber threat intelligence. An index system was designed consisting of three classes. The first class consists of five indexes, the second class consists of nineteen indexes, and the third class consists of more than fifty indexes. The typical properties of data quality, which include timeliness, relevance, accuracy, completeness, predictability, performability and customizability, can be observed among the indexes. However, the measurement of individual properties is not discussed. The evaluation is carried out manually by the experts via a questionnaire. The weights for each criterion are determined according to expert experience. Then the obtained values are adjusted using an algorithm of multi-objective optimization. Such an approach is more oriented to the evaluation of CTI data providers. The authors recognize that evaluation of CTI at the micro level is still an important research direction. The method presented is complex, based mainly on manual evaluation.

Meier et al. [21] introduced a CTI feed ranking approach. A prerequisite of the approach is the requirement of at least two snapshots of each considered feed. The key properties for evaluating feeds are as follows: completeness, accuracy, and speed. Com-

pleteness incorporates how many entries are covered. Accuracy includes how many entries are confirmed by other feeds. Speed includes the order in which the entries were covered. The naming of the first two properties is similar to the dimensions introduced by the other considered authors. However, the definitions of these properties are quite different. Therefore, the authors recognize that such criteria cannot be used to evaluate single entries from CTI feeds.

Li et al. [22] developed threat intelligence metrics to compare threat intelligence sources and described methodology for measuring them. The metrics included the following six indicators: volume, differential contribution, exclusive contribution, latency, coverage, and accuracy. The indicators of differential contribution, exclusive contribution, and latency are measured only by comparison of two feeds. The indicators of volume, coverage, and accuracy have meanings described in their names. The authors recognize that the accuracy or coverage of a feed cannot be measured precisely in a systematic way. In addition, the feeds to be compared were divided into six categories. The comparison was made according to these categories.

Schaberreiter et al. [23] proposed methodology for evaluating cyber threat information sources. The methodology is based on the standardized Structured Threat Information eXpression (STIX) format. Ten quantitative parameters were introduced. They are as follows: extensiveness, maintenance, false positives, verifiability, intelligence, interoperability, compliance, similarity, timeliness, and completeness. The presented parameters were used to assess the content and the providers of the content, with more attention paid to the providers. However, the parameters were not separated into different groups to differentiate between CTI data and CTI providers. The next limitation of the approach is that all sources were compared against other sources within the monitored community of CTI sources. Many formulae for the parameters include the definitions of values that are declared as "in the world view". Such an approach is not suitable to evaluate single new entry in CTI feeds.

Schlette et al. [24] proposed CTI quality metrics that rely on the standardized STIX format. The authors did not acknowledge the work by Schaberreiter et al. [23] and declared that they were the first ones to provide metrics on the standardized STIX format. The proposed CTI quality metrics follow an empirical approach [15]. However, the arguments for choosing specific dimensions from the approach by Wang and Strong [15] were not provided. In addition, input from a number of CTI practitioners and researchers was sought as well. The proposed metric was structured into three different levels: report, object, and attribute. The attribute level is based on evaluation of the available STIX objects. At the attribute level, six dimensions are considered. The object level dimensions (representational consistency and reputation) are not bound to the predefined attributes. Therefore, they are subjective. At the report level, a single indicator of the amount of data is considered. Consideration is presented for calculation of the amount of data. However, there is no final decision on how to compute this value. The obtained values at all levels are then aggregated into a combined quality indicator. The proposed quality dimensions were integrated into the STIX schema with the introduction of the custom object quality indicator. The following drawbacks of the approach can be observed. Firstly, the metrics proposed cannot be fully automatically calculated; the input of either data experts or data consumers is needed. Secondly, all the focus was concentrated on the quality of CTI products; the quality of the CTI providers was not considered separately. Thirdly, the metrics were tested in a small environment.

Mavzer et al. [25] defined an automatic method for the quality and trust metrics calculation. Quality included the following factors: completeness, freshness, timeliness, extensiveness, and relevance. However, the choice of factors was not discussed and was not

justified. For the final quality assessment, only three factors with weight coefficients were aggregated. They were as follows: completeness, freshness, and extensiveness. Moreover, Mavzer et al. [25] noticed that the calculation for completeness, freshness, and extensiveness required improvement. The trust metrics included unique factors in comparison with other studies considered. The factors are as follows: partner sharing activity, sector of activity, certified cybersecurity, previous ticket ratings, privacy, and partner sector. The choice of factors was not discussed and was not justified. The formula for the aggregation of the factors was not provided.

Zibak et al. [26] performed a systematic literature review to extract a set of quality dimensions for threat data, information, and intelligence. The search included studies published between 2014 and 2019. A total of 22 articles were selected in the field of cyber security. The search was limited, since a lot of new articles appeared in the field up to the publication of the paper, which are referenced in the paper but not included in the counting of dimensions. The most common dimensions were as follows (frequencies are shown in parenthesis): timeliness (14), completeness (13), accuracy (12), consistency (8), relevance (8), and reliability (7). Zibak et al. [26] decided to include only the dimensions that were mentioned in all three contexts: data, information, and intelligence. They are as follows: accuracy, completeness, interoperability (portability), relevance, reliability, timeliness, and uniqueness. The systematic literature review then was followed by a Delphi study that involved 30 threat intelligence experts around Europe. Two rounds of Delphi study were performed. During the first round, the experts reviewed the selected dimensions. The majority of experts agreed that the completeness and uniqueness dimensions should be dropped from the list. The reasons are as follows: pursuing intelligence completeness could undermine its timeliness and actionability; and uniqueness often calls for further investigation to be sure that it is not the product of faulty analysis. During the second round, the experts were asked to add dimensions that were overlooked in the literature. The dimensions of actionability and provenance were added to the list. The main drawback of the study is that a measurement of dimensions was not considered.

Yang et al. [27] presented an automatic CTI quality evaluation method that linked the assessment of the trustworthiness of CTI feeds to assessment of CTI content. The authors argue that CTI quality depends on the trustworthiness of CTI feeds. The original amount of information provided by the feed is considered to be a feature of CTI feed trustworthiness. The assessment of CTI content is composed of three categories of features: multi-source verification, completeness, and timeliness. The completeness of the content is defined by seven features. The features are outlined in the paper; however, the methods for the choice of the features are not discussed. Based on the features extracted from CTI content, a belief propagation neural network was designed to provide an assessment of the content quality. The assessment of the quality is divided into five levels: A, B, C, D, and E. Level A is the highest one. To join the different parts of the assessment, weights coefficients are used. The authors performed the experiment and established that the accuracy of quality classification reaches a peak when the weight coefficient for the trustworthiness is 0.3 and the weight coefficient for content is 0.7. The trustworthiness of CTI feeds is considered in correlation with other feeds. Such criteria cannot be used to evaluate single entries from CTI feeds. Finally, the authors recognize that the dimensions and features considered are not fully comprehensive yet.

Chen et al. [28] introduced a method for calculating threat scores of Indicators of Compromise (IoCs) only. The calculation of scores consists of two separate parts: severity and confidence scores. The severity scores are used to evaluate quality of content; meanwhile, confidence scores evaluate the credibility of the source. The severity score includes five indicators: timeliness, accuracy, completeness, relevance, and consistency. The choice

of the first four indicators is based on the outdated internet resources used. Therefore, such a justification is not reasonable. The fifth indicator was added on the initiative of the authors only. For the calculation of the values of the indicators, IoCs are presented in standardized STIX format to establish threat associations. Definitions of the indicators are provided at a high level of abstraction. The calculation of the relevance is not clear and it is not transparent. The calculation of the consistency is quite unusual, since it is measured through support by antivirus software. For confidence scores, source ranks and similarity scores are considered. The weight coefficient assigned to source rank is 70%. The weight coefficient assigned to similarity score is 30%. The data used for the experiment are confidential.

Sakellariou et al. [29] proposed a methodology to develop and assess CTI quality metrics. Pursuing their goal, the authors divided CTI quality factors into three groups: collected data, produced intelligence, and information sharing. Factors were selected for each group; however, the lists of factors are not complete and are misleading. The list of each group only includes a mixture of factors from different unrelated publications. The group of quality factors from the collected data is based on two publications [30,31] only. Moreover, Dalziel [31] wrote that the factors of relevance (actionability and value) characterize the produced intelligence. We presume that Sakellariou et al. [29] wrongly included these factors in the group of collected data. The methodology used to develop the CTI quality metrics is very generic, since the contents of steps 1–3 devoted to the development of CTI quality metrics are provided at a very high level of abstraction. The only process for the assessment of CTI quality metrics is detailed. To demonstrate the applicability of the proposed methodology, a weighted completeness metric was developed for structured CTI products (e.g., STIX v2.1). However, this metric is hypothetical and has no practical value, since many important intermediate values of the experiment are not provided. Moreover, a random function is used several times during the experiment. Therefore, there is no possibility of replicating the experiment and of comparing the results of an experiment applying different quality factors. Finally, the authors "recognize the importance of developing CTI quality metrics for CTI sources evaluation in the context of future research".

During the review of related works, we have noticed that different authors use different terms to denote CTI data quality. These terms are as follows: characteristic [16,32], property [20,21], feature [19], parameter [23], dimension [15,24,26–28,33,34], factor [28,29], indicator [28], metrics [22,25,35]. The most frequently used term is dimension. We will follow the choice of the majority of the authors.

In summary, we can conclude that CTI sources continue to grow, but the quality of the CTI data provided remains questionable. The quality of CTI data is a decisive factor in the usefulness of shared CTI data. Many authors have tried to suggest dimensions to measure the quality of CTI data. However, these dimensions vary quite substantially. Next, the methods of measurement of these dimensions also vary quite substantially. Moreover, not all the dimensions are automatically calculated [24]. The values of the dimensions sometimes must be decided by the user. In addition, many authors assess the quality of the submitted CTI data in comparison with available resources for CTI data. Such an approach is not possible for the assessment of new events. Only part of the proposed dimensions [23,24,28] is oriented to the presentation of CTI data in structured STIX language format, which is the best form to share CTI data. Only a few authors [21,27,28] separate the quality of shared CTI data and the trustworthiness of the CTI data provider. To overcome the shortcomings mentioned, we suggest fully automated measurement formulae for the dimensions of structured CTI data provided in STIX language.

## 3. Proposed Automated CTI Data Quality Assessment Method

### 3.1. Choice of the Dimensions

When developing a CTI quality assessment method, it is very important to follow design methodology. Ahlemann and Gastl [36] proposed a four-stage methodology: problem definition, construction of a frame of the model, core construction, and validation. The authors stated [36] that in this research area, where no widely accepted models exist, a model can be designed via analysis and abstraction of the elements and practices proposed in the literature. Based on the above, we follow a four-phase design methodology in this paper, as depicted in Figure 1. This is the foundation for developing our proposed CTI quality assessment method.

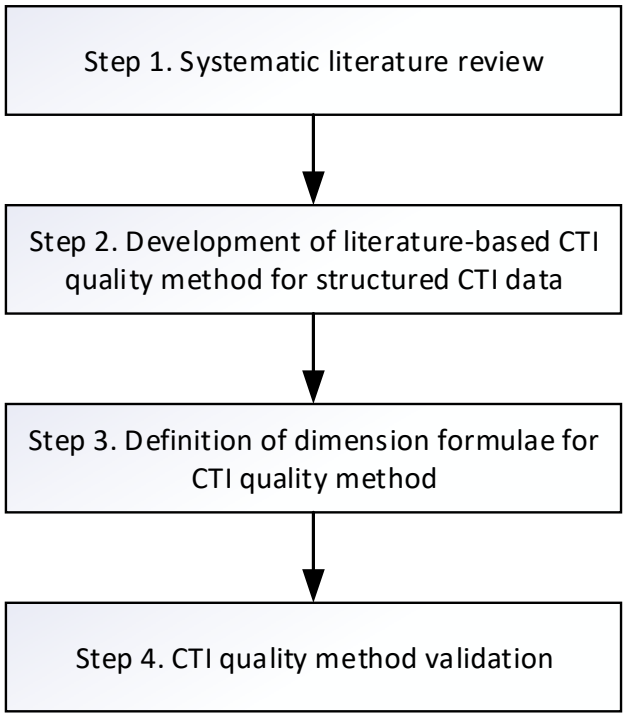


**Figure 1.** CTI data quality method design methodology.

The first step of the design methodology was completed in Section 2. Using the knowledge acquired from the systematic literature review, we will develop a literature-based automated CTI data quality assessment method and we will define the dimension formulae for this method in the current section. Sakellariou et al. [37] observed that validation of a developed method should be accomplished by solving real-world problems. We will perform an experiment and analyze results using real-world data in Section 4.

To make a choice of CTI data quality dimensions, they are extracted from the reviewed related studies and they are presented in Table 1. The naming of dimensions varies, as the meaning of dimensions also varies in the reviewed works. Therefore, to be as precise as possible, three different labels were used to show consideration of the dimensions in the particular related work. The meanings of the labels are explained in the bottom lines of the table. The last column of Table 1 shows the values of how frequently the dimensions were used in the reviewed works. The dimensions are arranged according to the values in the last column in decreasing order. These values show the popularity of the dimensions in the reviewed works. The obtained values of the popularity of the dimensions coincide with the results presented by Zibak et al. [26] for the years between 2014 and 2019; meanwhile, our review was performed for the years between 2018 and 2024. It is possible to observe that these sets of review years have only a short period of overlap; however, the popularity

of the dimensions shows very little change. This is especially true for the first two most frequent dimensions: timeliness and completeness. The third dimension (accuracy) had the same popularity; however, its popularity is lower in our review. Therefore, it is meaningful to choose the most popular dimensions for the automated CTI data quality assessment method.

**Table 1.** A summary of the CTI data quality dimensions.

| Dimension | Wang & Strong [15] | ISO/IEC 25012 [16] | Gao et al. [19] | Li et al. [20] | Meier et al. [21] | Grispos et al. [30] | Li et al. [22] | Schaberreiter et al. [23] | Griffioen et al. [35] | Tundis et al. [32] | Mavzer et al. [25] | Schlette et al. [24] | DeCastro-García & Pinto [33] | Wang et al. [34] | Zibak et al. [26] | Yang et al. [27] | Chen et al. [28] | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Timeliness | ** | + | ** | ** | ** | ** | + | ** | ** | ** | ** | ** | | ** | ** | ** | ** | 16 |
| Completeness | ** | ** | | ** | ** | ** | + | ** | | + | ** | ** | ** | + | | *+ | ** | 14 |
| Accuracy | ** | ** | | ** | ** | ** | ** | | | | | ** | | | ** | | ** | 9 |
| Reliability | + | + | + | | | | | | | + | | + | + | | ** | | | 7 |
| Relevance | ** | | | ** | | | | | | | ** | ** | ** | | ** | | ** | 7 |
| Reputation (Provenance) | ** | | + | | | | | + | | | | ** | | | ** | + | + | 7 |
| Similarity/Exclusivity | | | | | | | ** | ** | + | | | | ** | + | | + | ** | 7 |
| Consistency | ** | ** | | | | ** | | | | | | ** | ** | | | | *+ | 6 |
| Amount of data | ** | | + | | | | + | | | | | ** | + | | | | | 5 |
| Extensiveness | | + | | + | | | | ** | | | ** | | | | | | | 4 |
| Portability | | ** | | | | | | + | | | | | | | ** | | | 3 |
| Update frequency | | | ** | | | | | + | | | | | ** | | | | | 3 |
| Actionability | | | | | | | | | | ** | | | | | ** | | | 2 |
| Compliance | | ** | | | | | | ** | | | | | | | | | | 2 |
| Concise representation | ** | | | | | | | | | | | ** | | | | | | 2 |
| Differential contribution | | | | | | | ** | | | | | | + | | | | | 2 |
| Understandability | ** | ** | | | | | | | | | | | | | | | | 2 |
| Objectivity | ** | | | | | | | | | | | ** | | | | | | 2 |
| Security | ** | + | | | | | | | | | | | | | | | | 2 |
| Value-Added | ** | | | | | | | | + | | | | | | | | | 2 |
| Verifiability | | | | | | | | ** | | | | | | ** | | | | 2 |

** denotes use of the dimension in the paper. + denotes use of the dimension in the paper when the name of the dimension is different, but the meaning is the same. *+ denotes use of the dimension in the paper when the name of the dimension is the same, but the meaning is different.

To simplify the preliminary set of dimensions, we decided to include only dimensions that were mentioned more than once in the reviewed works. Only dimensions which are supported by at least by two researchers are provided in Table 1. After careful consideration of the dimensions provided in Table 1, and having in mind structured CTI data as were defined in the introduction section, eight dimensions were excluded from the list of dimensions to form a list for automatic assessment of CTI data quality. The thirteen remaining dimensions, which constitute the body of the automatic CTI data quality method, are as follows: timeliness, completeness, accuracy, reliability, relevance, reputation, similarity, consistency, amount of data, extensiveness, actionability, concise representation, and understandability. The eight dimensions to be excluded from further consideration are as follows: portability, update frequency, compliance, differential contribution, objectivity, security, value-added, and verifiability. Next, we will provide the reasons for the exclusion of the aforementioned dimensions.

The portability dimension, which is defined as an ability to move the provided CTI data from one system to another system [16], has no meaning because the documents are provided in structured STIX 2.1 language. The update frequency dimension is a characteristic of the CTI provider to define how frequently the submitted CTI data are updated [33]. For the new CTI data, such a dimension cannot be measured. The compliance dimension, which is defined as adherence to the standards [16], has no meaning since the documents are provided in structured STIX 2.1 language. The differential contribution dimension, which is defined as the provider indicators that are present in one provider but are not present in the other provider [22], is the opposite of the similarity dimension that is devoted to considering similarity between providers. We have included the similarity dimension since it is more popular than the dimension of differential contribution. Moreover, it is much easier to measure the value of the similarity dimension. Therefore, there is no need for the other related dimension. The objectivity dimension is defined as the extent to which data are unbiased and impartial [15], and to define its value requires detection methods that can follow a syntactical approach or center on semantics. Its implementation is too problematic. The security dimension, which is defined as the extent to which data access can be restricted and kept secure [15], is not needed since the implementation of the method is targeted to a private blockchain. The value-added dimension, which is defined as the extent to which data are beneficial and provide value for their use [15], cannot be implemented since it is not possible to measure an impact. The verifiability dimension, according to its definition [23,34], indicates whether threat information provided is supported by other sources. When new CTI data are provided, there is no expectation that this new information will be supported by other resources. Therefore, this dimension is not included in the automatic assessment of CTI data quality.

### 3.2. Definition of the Formulae for the Dimensions

We now start a presentation and consideration of measurement formulae for selected dimensions in their popularity order as they were introduced in Table 1. We have used the following common notations in the measurement formulae of the dimensions:

$v_t$ denotes a $t$-th CTI report from the set $V$;

$o_i$ denotes an $i$-th STIX object from the set $O$;

$p_i$ denotes an $i$-th STIX object property from the set $PV$.

*Timeliness* ($TM$). The importance of timeliness is recognized by the absolute majority of researchers; however, a measurement of timeliness differs in different research studies. A simple definition of timeliness is provided by Wang and Strong [15]. A definition of timeliness adapted to various situations is provided by Schlette et al. [24]. We will follow it and elaborate on it. Therefore, we will define several cases of the timeliness dimension adapted to specific situations.

(a) *Basic case*. This is the time difference between the current time and creation time of the objects in the CTI provided by the feed. The timeliness of CTI $v_t$ is calculated as follows (Equation (1)):

$$TM_{basic}(v_t) = \begin{cases} 1, \ if \ current\ time - max(modified(o_i),\ o_i \in O) \leq threshold \\ threshold \times \frac{1}{current\ time - \max(modified(o_i),\ o_i \in O)}, \ in\ opposite\ case \end{cases} \quad (1)$$

The value of threshold is applied to show how many days need to be considered for the CTI to be valuable. When the threshold is over, to slow down the decline of the dimension value, the obtained value is multiplied by the threshold value.

The CTI consists of several objects that are created at the same time. During the creation of the objects, the properties *created* and *modified* have the same values. When new

information concerning objects in CTI becomes available, the CTI can be updated. In such a case, a few objects are modified only. Consequently, the latest modification time of the object is chosen for the timeliness dimension.

(b) *Indicator object case*. The timeliness dimension can be based on specific properties of objects. The indicator object has unique properties *valid_from* and *valid_until* available only in this object. The value of property *valid_from* usually repeats the value of property *created*. However, the value of property *valid_until*, if it is available (since it is optional), defines the time till this indicator is valid. So, to know the time till the indicator is valid is very important. If this value is available, it must be used in the calculation of the timeliness dimension (Equation (2)). The *valid_until* property is usually defined for an indicator with the pattern "ipv4-addr:value". The threshold is applied for the indicator case, as well. The value of timeliness declines very quickly if the threshold is not applied.

$$TM_{indicator}(o_i) = \begin{cases} threshold \times \frac{1}{valid_{until(o_i)} - current\ time}, & if\ valid_{until(o_i)} > current\ time \\ 0, & in\ opposite\ case \end{cases} \quad (2)$$

If relationship objects are defined, the indicator object can be related to several types of other objects. These types are as follows: *attack-pattern*, *campaign*, *infrastructure*, *intrusion-set*, *malware*, *threat-actor*, *tool*, *course-of-action*, *grouping*. In this case, the timeliness of the objects must be corrected according to the related indicator object.

(c) *Statistical data case*. When statistical data about the decline of timeliness for specific CTI objects are available, the timeliness dimension must be adapted (Equation (3)). We know that the decline of certain CTI objects is higher than for others. File hashes used in the indicator object will likely have unchanged values, as malware binaries are often subject to slight modification, resulting in changed hash values. Meanwhile, information regarding object types such as malware analysis, threat actors, and tools might not change over time. For such object types, statistical decline must be taken into account.

$$TM_{statistical}(o_i) = \begin{cases} 1, & if\ modified(o_i) + decline(o_i) \geq current\ time \\ threshold \times \frac{1}{modified(o_i) + decline(o_i) - current\ time}, & in\ opposite\ case \end{cases} \quad (3)$$

*Completeness* (*CM*). Definitions of the completeness dimension also vary, as do definitions of the timeliness dimension. Several authors [19,24,28] have used graph theory to establish the connectedness of SDOs in CTI. Schlette et al. [24] applied the obtained values of the connectedness of SDOs to define the amount of data dimension. However, we suppose that such an approach is also directly suitable to define the completeness dimension (Equation (4)). Moreover, such an approach, which was expressed indirectly, was used by Chen et al. [28].

$$CMI(v_t) = \frac{|SRO \in O|}{\frac{|SDO \in O| \times (|SDO \in O| - 1)}{2}} \quad (4)$$

The completeness dimension sets the number of existing SROs in a given CTI in relation to the maximum possible number of SROs, as defined by the number of SDOs for this CTI. However, such a definition sets quite low values for the completeness dimension, since not all the SDOs can be grouped together. We have noticed that the extensiveness dimension used by several authors [23,25] and the completeness dimension used by Schlette et al. [24] are defined in the same way. The extensiveness dimension according to this

definition counts the number of filled in optional properties. The extensiveness dimension is defined at three levels: optional property of the object (Equation (5)); optional properties of the object (Equation (6)); and all the objects of report (Equation (7)).

$$EX(p_i) = \begin{cases} 1, \ if \ o(p_i) \ is \ present \\ 0, \ in \ the \ opposite \ case \end{cases} \tag{5}$$

$$EX(o_i) = \frac{\sum_{i=1,no} EX(p_i)}{no} \tag{6}$$

$$EX(v_t) = \frac{\sum_{o_i \in O} EX(o_i)}{|O|} \tag{7}$$

We decided to join the formulae for $CMI(v_t)$ and $EX(v_t)$ into a single formula of completeness (Equation (8)). The values of coefficients are defined in Equation (9). The CTI reports describing only indicators of compromise do not contain SROs. Such reports are usually submitted in a prompt way, and they are valuable for use at the technical level. However, the first term of Equation (8) for such reports is equal to 0. To assess the completeness of such reports as highly as possible, the value of coefficient $w_1$ is assigned as 30% and the value of coefficient $w_2$ is assigned as 70%.

$$CM(v_t) = w_1 \times CMI(v_t) + w_2 \times EX(v_t), \ w_1 \text{ and } w_2 \text{ are weight coefficients} \tag{8}$$

$$w_1 = 30\%, \ \ w_2 = 70\% \tag{9}$$

*Accuracy* (*syntactic*) (*AC*). The accuracy dimension according to our investigation was less popular than the dimensions of timeliness and completeness. Consequently, it was considered in fewer research studies and fewer definitions can be found. For accuracy, we chose the following definition: "The degree to which data has attributes that correctly represent the true value of the intended attribute of a concept or event" [16]. Having in mind this definition, we base the accuracy definition on the syntactic accuracy of the JSON schema. The JSON schemas for all object types with values for properties are defined by the OASIS consortium [14]. This enables objects to be automatically validated against those schemas to assess syntactic accuracy. The accuracy dimension is defined at three levels: for every property of the object (Equation (10)), for the properties of the object (Equation (11)), and for all objects in the report (Equation (12)). The extensiveness dimension was defined in a similar way. However, in this case, all the properties and their values are considered against the JSON schema standard of a specific object.

$$AC(p_i) = \begin{cases} 1, \ if \ value \ of \ o(p_i) \ is \ valid \\ 0, \ in \ the \ opposite \ case \end{cases} \tag{10}$$

$$AC(o_i) = \frac{\sum_{i=1,n} AC(p_i)}{n} \tag{11}$$

$$AC(v_t) = \frac{\sum_{o_i \in O} AC(o_i)}{|O|} \tag{12}$$

For example, if the *Description* property is either empty or absent, it is not filled in correctly. The Bundle object is not part of the SDOs in STIX. However, the Bundle object is recommended to be used when using CTI to join the provided objects into a single unit. Therefore, if the Bundle object is not present in the CTI, syntactic accuracy is lower by 10%.

*Reliability* (*RB*). Reliability means the reputation of data. The reliability dimension is less popular in research studies than the accuracy dimension. Moreover, the term "reliability" is used just in [26] from all the reviewed works. The terms "credibility" [16,32], "believability" [15,24], "source authority" [19], and "veracity" [33] were used to define the reliability of CTI in other works. The experts [38] were saying that the most important dimensions of CTI are timeliness and credibility. The importance of the timeliness dimension is recognized by the absolute majority of research studies considered in our review. However, the popularity of the reliability dimension is half as popular as the timeliness dimension. We presume that the main reason for the lesser popularity of the reliability dimension is that the reliability of CTI cannot be measured objectively [24]. Nevertheless, we and the other authors [27,28,39] recognize the importance of the reliability of CTI. The reliability of CTI can be expressed indirectly by the reputation of CTI providers [27,28,39]. The reputation of the provider is only a guarantee of the reliability of CTI. Consequently, our CTI system will accumulate the reputation of the provider and use it in the quality assessment of CTI.

*Relevance* (*RL*). Relevance must express the extent to which data are applicable and helpful for the specific user. This is an important aspect of the assessment, since this dimension shows the usefulness of CTI for the particular user. An incident which happens in a specific sector of an industry is likely to be less relevant to other industry sectors. Therefore, in contrast to all other dimensions, this dimension depends on the particular user. This dimension must be evaluated dynamically. The value of this dimension depends on the $PV_C$ properties possessed by the consumer, on the $PV_P$ properties possessed by the provider, and on the $PV_O$ properties of CTI. The properties of the consumer must coincide with the properties of the provider and the properties of CTI. Two cases are defined for the relevance dimension: basic case (Equation (13)) and augmentation of the value of the basic case if a sighting object is present (Equation (14)). The coefficients to join the two cases are presented in Equation (15). The presence of sighting objects is not a common case; therefore, 10% only are assigned the $w_2$ coefficient.

Basic case:

$$RL_{basic}(v_t) = \frac{\left|PV_c \cap \left(PV_p \cup PV_O\right)\right|}{\left|\left(PV_p \cup PV_O\right)\right|} \tag{13}$$

*Sighting object case.* Sighting objects are used to express a belief that something in CTI was seen. This fosters an assessment that frequently seen objects might indicate their versatility and applicability to different sectors of industry. Therefore, the value of relevance must be increased if a sighting object is present.

$$RL(v_t) = w_1 \times RL_{basic}(v_t) + w_2 \times \frac{|SO|}{|SDO|},\ w_1 \text{ and } w_2 \text{are weight coefficients} \tag{14}$$

$$w_1 = 90\%,\ \ w_2 = 10\% \tag{15}$$

*Reputation* (*provenance* (*provider*)) (*RP*). The reputation of a user is accumulated in the system. To make the value fractional the dimension sets the value of reputation of the current provider in relation to the maximal available reputation value of the user in the system (Equation (16)).

$$RP(p) = \frac{RP_u}{max(RP_u \in U)} \tag{16}$$

*Similarity/exclusivity* (*SM*). The reported event is more valuable if the event was not reported before by other producers, since the report provides new observed information. To compare the similarity of two CTIs $v_i$, $v_k$, we use a weighted similarity function, $SM(v_i, v_k)$, which can be expressed as follows (Equation (17)):

$$SM(v_i,\ v_k) = (w_1 \times s_{source} + \frac{\sum_{SDO,\ SRO,\ SCO\ \in v_i,\ v_k} cosine\ similarity,\ if\ object\ type \in v_i = object\ type \in v_k}{\sum_{SDO,\ SRO,\ SCO\ \in v_i,\ v_k} count,\ if\ object\ type \in v_i = object\ type \in v_k}) / 1.2 \quad (17)$$

$$s_{source} = \begin{cases} 1, \text{ if two CTIs are provided by the same provider} \\ 0,\ in\ the\ opposite\ case \end{cases} \quad (18)$$

The first term in the formula (Equation (17)) is devoted to consideration of whether two CTIs used for comparison are provided by the same provider. The calculation of the first term in Equation (17) is defined in Equation (18). The similarity value for the same provider is increased. The value of increase is regulated by the weight coefficient $w_1$. Its value was set to 0.2. The value of similarity is normalized since the maximal obtained value is 1.2. If the similarity is high, the inclusion of the new event must be considered. Calculation of the similarity related to the set of all provided CTIs, as it is in [28], is not the right way, since the increase in the set hides the effect of the particular CTI. We are interested in the maximal similarity with specific instances of CTI. The comparison of similarity is carried out with all available instances of CTI considering everyone separately and the maximal similarity then with the specific instance of CTI is chosen as the value of the dimension (Equation (19)).

$$SM(v_i) = \max\{SM(v_i,\ v_k),\ v_i,\ v_k \in V\},\ V\ is\ set\ of\ CTI \quad (19)$$

*Consistency* (*representational consistency*) (*CN*). The consistency dimension refers to the violation of semantic rules defined over a set of data items. With reference to the relational theory, integrity constraints are a type of semantic rules. Therefore, the consistency dimension enables us to address the issues that were not considered by the dimensions with higher popularity among the researchers. In the statistical field, data edits are typical semantic rules that allow for consistency checks. In relational theory, two fundamental categories of integrity constraints can be distinguished, namely intra-relation constraints and inter-relation constraints. Intra-relation constraints define the range of admissible values for an attribute's domain. Inter-relation constraints involve attributes from different relations [40]. So, these constraints are user defined. We define three constraints for the inter-relation of STIX objects. The first constraint is as follows: referenced objects of embedded relationships must exist. Special attention must be given to SROs, since they connect two SDOs. The second constraint is as follows: the time of object creation must follow the chronological order of objects' creation. Hence, SROs can connect two SDOs only after their creation. The creation time of SROs must be later or equal to the creation time of corresponding SDOs. The third constraint is as follows: all SCOs must be connected to some existing SDOs. We define a single constraint for the intra-relation category. The constraint is as follows: the modification time of any object is later or equal to the time of creation.

We define a set of constraints C. Equation (20) is defined for the particular object and particular constraints. Equation (21) defines the formula for all the objects and all the constraints.

$$CN(o_i) = \begin{cases} 1,\ if\ o_i\ satisfies\ constraint \\ 0,\ in\ the\ opposite\ case \end{cases} \quad (20)$$

$$CN(v_t) = \sum_{i=1}^{|C|} \frac{\sum_{j=1}^{|O|} CN_i(o_j)}{|O|} \quad (21)$$

*Amount of data* (*AD*). Appropriate amount of data: the extent to which the quantity or volume of available data is appropriate [15]. Domain experts recommend including

the dimension of the appropriate amount of data [24]. The definition of the dimension of the amount data provided by Schlette et al. [24] is more suitable for the definition of completeness. The report, which contains homogeneous SDO types and very few relationships, does not seem to be very helpful. To distinguish between a report with homogeneous object types and one with more different object types we apply the dimension of amount of data. This is achieved by counting the instances of the different SDO, SCO, and SRO types within a report. The dimension sets the number of existing different SDO, SCO, and SRO types in a given STIX report in relation to the total number of existing objects for this report (Equation (22)).

$$AD(v_t) = \frac{|\{Types \ \in O\}|}{|O|} \tag{22}$$

*Actionability* (*AT*). According to the European Union Agency for Cybersecurity, CTI's actionability is related to timeliness, accuracy, completeness, ingestibility, and relevance to the recipient [41]. We suppose that the ingestibility dimension, which was not found in the research works, can be replaced by the understandability dimension. Experts [26] were suggesting inclusion of the actionability dimension in CTI quality assessment. However, the actionability becomes less critical on a tactical or strategic level [42]. Therefore, to accommodate different requirements of CTI categories, we suggest calculating two values of actionability, for a technical level and for a tactical level. The formula of actionability for both levels is the same (Equation (23)). The difference is in the values of the weight coefficients. A trade-off is needed between timeliness and both completeness and accuracy [38,41,43]. For the technical level, priority is afforded to timeliness versus completeness and accuracy (Equation (24)). In order for CTI data to be useful, they must be relevant. So, the highest value is assigned to the relevance of the information. Next, the CTI data must be delivered in a timely manner. Information about security incidents older than several hours can be considered irrelevant due to rapid changes in threat characteristics. As a consequence, 28% are assigned to timeliness and 28% are divided in two for completeness and accuracy. For the tactical level, priority is assigned to completeness and accuracy versus timeliness (Equation (25)).

$$AT = w_{TM} \times TM + w_{cm} \times CM + w_{AC} \times AC + w_{RL} \times RL + w_{UN} \times UN \tag{23}$$

$$w_{TM} = 28\%, \ \ w_{CM} = 14\%, \ \ w_{AC} = 14\%, \ \ w_{RL} = 38\%, \qquad w_{UN} = 6\% \tag{24}$$

$$w_{TM} = 12\%, \ \ w_{CM} = 22\%, \ \ w_{AC} = 22\%, \ \ w_{RL} = 38\%, \qquad w_{UN} = 6\% \tag{25}$$

*Concise representation* (*CR*). This dimension is devoted to the measurement of similarity within the report (Equation (26)). Despite the fact that this dimension was not popular among researchers, the dimension of concise representation is important, since it considers the inner consistency of the report. The application of methods for semantic similarity is needed. The *Simhash* algorithm is one way to approach this problem. An object $o_1$ is considered unique in a set of objects $O$ if its *similarity* to any other object $o_2 \in O$ is below a threshold $t_{CR}$ (Equation (27)). In the course of the experimentation, the value of $t_{CR}$ (Equation (27)) was chosen to be equal to 0.3. The value of concise representation for the report is calculated according to Equation (28). The lower the value of concise representation, the better the report is presented. Therefore, we apply negation (subtraction) in Equation (28).

$$CR(o_1, \ o_2) = similarity \ (o_1, o_2), \ o_1, o_2 \in O \tag{26}$$

$$CR(o_i) = \begin{cases} 0, \ \{\max(o_i, o_j), \ j = 1, \ |O|\} \leq t_{CR} \\ \{\max(o_i, o_j), \ j = 1, \ |O|\} > t_{CR} \end{cases} \quad (27)$$

$$CR(v_t) = 1 - \frac{\sum_{i=1}^{|O|} o_i}{|O|} \quad (28)$$

*Understandability* (*UN*). The large amount of flexibility of the STIX language is by itself a weakness. Three different threat intelligence analysts came up with three different representations of the same information [44]. The STIX language allows extensions that can introduce new object types and new properties to existing object types. The type "extension_definition" is used to include custom objects and custom properties. If either custom objects or custom properties are used, the understandability of the presented STIX description is decreased. Therefore, understandability must be measured. Next, the presence of not empty fields of description in the objects adds a lot to the understandability. Therefore, the formula for calculation of understandability (Equation (29)) consists of two terms joined by weight coefficients $w_1$ and $w_2$. The calculation of the understandability value for a single object is presented in Equation (30). The values for coefficients $w_1$ and $w_2$ are presented in Equation (31). The values for coefficients $w_1$ and $w_2$ are chosen to be equal.

$$UN(v_t) = w_1 \times \left(1 - \frac{|type \text{ extension\_definition}|}{|different \ object \ types|}\right) + w_2 \times \frac{\sum_{i=1}^{|O|} UN(o_i)}{|O|} \quad (29)$$

$$UN(o_i) = \begin{cases} 1, \text{ if description property contains an information} \\ 0, \ in \ the \ opposite \ case \end{cases} \quad (30)$$

$$w_1 = 50\%, \ \ w_2 = 50\% \quad (31)$$

*Total score of the CTI data quality* (*TQ*). All the dimensions, except the actionability, reputation, and similarity dimensions, are aggregated using weight coefficients into a generic quality score of CTI data dimensions (Equation (32)). As in the case of the actionability dimension, we define two values of the total score of CTI data quality: for technical level and for tactical level. The technical level prioritizes timeliness versus completeness and accuracy (Equation (33)). The tactical level prioritizes completeness and accuracy versus timeliness (Equation (34)). All the remaining dimensions are considered of equal importance.

$$QS = w_{TM} \times TM + w_{CM} \times CM + w_{AC} \times AC + w_{RL} \times RL + w_{CN} \times CN + w_{AD} \times AD + w_{CR} \times CR + w_{UN} \times UN \quad (32)$$

$$w_{TM} = 22\%, \ w_{CM} = 11\%, \ w_{AC} = 11, \ w_{RL} = 32\%, \ w_{CN} = 6\%, \ w_{AD} = 6, w_{CR} = 6\%, \ w_{UN} = 6\% \quad (33)$$

$$w_{TM} = 8\%, \ w_{CM} = 18\%, \ w_{AC} = 18, \ w_{RL} = 32\%, \ w_{CN} = 6\%, \ w_{AD} = 6, w_{CR} = 6\%, \ w_{UN} = 6\% \quad (34)$$

*Confidence score of the provider* (*CS*). Interviews with experts revealed that the trustworthiness of CTI providers plays an essential role in assessing the quality of CTI for nine experts from the group of twenty-five experts [42]. Moreover, the reputation of the CTI provider is critical at the tactical level of CTI [42]. Therefore, we provide an assessment of the confidence score as a separate quality dimension. The same view was taken in [28] as well. The dimensions of reputation and similarity are aggregated into a dimension of confidence score for the provider [21,27,28,35] (Equation (35)). We invert the similarity value SM in the confidence score of CTI provider calculation. The reason is that less similarity with available CTI in the dataset means more confidence for the CTI provider, since the

CTI provider delivers CTI reports for a new event. Such an approach is presented in [27] as well. More originality in provided CTI data means more trustworthiness of the feed. The values of weight coefficients are chosen using an experience shared in [28] (Equation (36)).

$$CS = w_{RP} \times RP + w_{SM} \times (1 - SM) \tag{35}$$

$$w_1 = 70\%, \;\; w_2 = 30\% \tag{36}$$

## 4. Results Analysis

### 4.1. Experiment

To summarize the consideration of the dimensions provided in the previous section, the architecture of the proposed CTI quality assessment solution is presented in Figure 2. The architecture of the proposed solution includes four major components: parser of structured CTI data, content assessment, provider assessment, and visualization. The parser of structured CTI data parses the file, extracts the STIX objects and calculates the values of indicators required for the measurement of the used dimensions. Initially, thirteen dimensions were selected. After design, the assessment was separated into two distinct parts: content and provider. The content assessment is based on eight dimensions that are enumerated in the second component. The initial dimensions of completeness and extensiveness were joined into a single dimension of completeness. The results of content assessment are presented in four aggregating dimensions that are as follows: actionability technical, actionability tactical, quality score technical, and quality score tactical. The initial dimension of actionability is presented as the finalizing indicator for CTI data quality, since it is an aggregator of several initial dimensions and it is defined as one of the fundamental building blocks of successful incident responses [41]. Actionability is defined at two levels to solve an everlasting dispute between timeliness and completeness, whichever dimension is more important. The quality score of CTI data is defined in a similar way to the actionability dimension. A confidence score for the provider is based on two dimensions: reputation and similarity. The initial dimension of reliability was truncated to the reputation of the provider. It means that the reliability of the provided CTI data is directly associated with the value of the reputation of the CTI provider. Finally, the calculated values are visualized. Visualization eases the understanding and interpretation of the obtained results.

Our research indicates a lack of benchmark datasets related to structured CTI products. This problem was noticed by several authors [27–29] as well. However, only Sakellariou et al. [29] constructed a dataset of structured CTI products and placed it on GitHub [45]. We have used the dataset geosakel77 [45] (42.34 MB), constructed by Sakellariou et al. The purpose of using the public dataset is to provide the results of our method to enable the possibility of comparison for the future related studies.

To assess the latest CTI products, we have constructed a dataset called bari24 (980 KB) that includes the newest pulses provided by AlienVault Open Threat Exchange [46] in the last quarter of 2024.

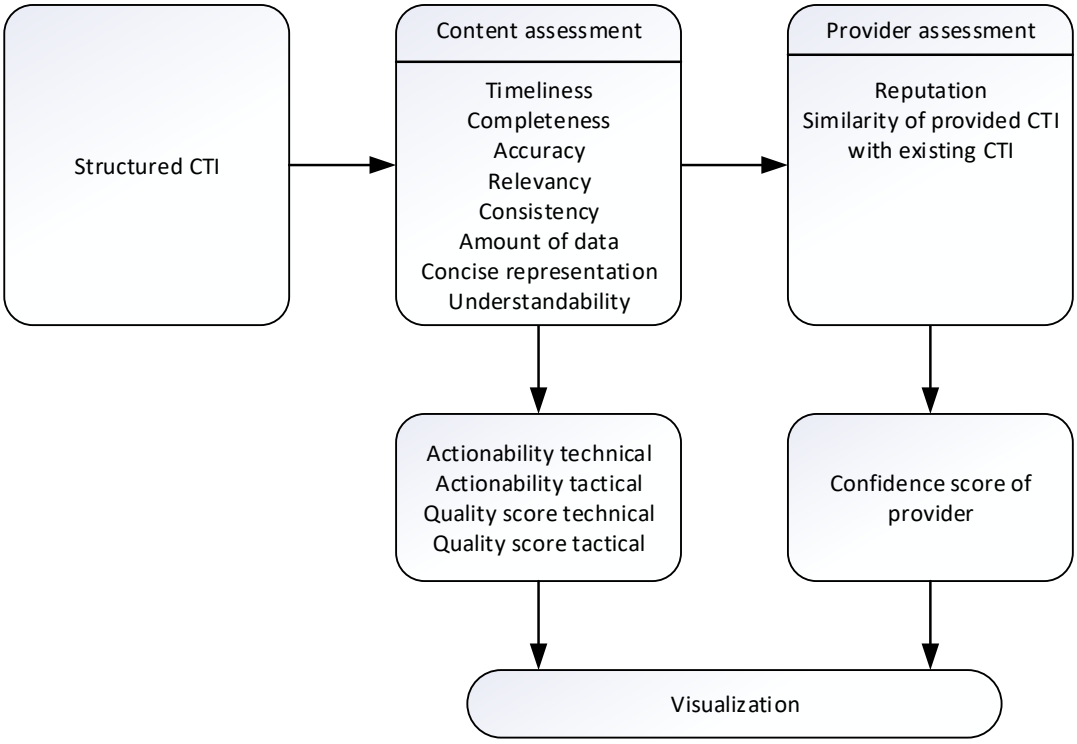


**Figure 2.** System architecture of the proposed CTI quality assessment solution.

To produce defined adversary tactics and techniques, some CTI data from MITRE ATT&CK knowledge base [47] are used to construct a dataset called bariMit (10.1 MB). The description of these CTI data differs greatly from the descriptions used in two previous datasets, since these descriptions include a lot of custom developed objects. Moreover, these descriptions are very extensive; one description on average occupies one 1 MB. So, using the dataset bariMit allows us to test the scalability of our method. In advance, we could say that our method passed the test of scalability, since the processing of 1 MB of CTI data description required 54 s of time (Intel Core i5-8500t processor (ANAFRA, Prague, Czech Republic), 16 GB RAM, SSD NVMe hard drive and Windows 11 operating system).

All CTI data are presented in STIX v2.1 schema. Table 2 presents the statistics of used and constructed datasets. We can observe that the datasets are quite diverse if we look at the statistics of the datasets.

The Java application was implemented to evaluate the cyber threat intelligence data quality by calculating key quality dimensions on STIX data files. While using the app, the user can upload STIX files in JSON format. The uploaded STIX file is parsed and objects are extracted. Then, dimensions are calculated for each file and detailed results are provided. The app includes a graph visualization feature, where each STIX object is represented as a node, while relationships and sighting objects are represented as edges between nodes. When clicking on a single node, a detailed view of individual attributes and dimensions is provided. Exports of dimensions to Excel are also provided and can be used for further analysis of data. This tool is devoted to measuring the quality of CTI data from different sources and it is used for experimental evaluation.

**Table 2.** Dataset statistics.

| Dataset | Objects | Number |
|---|---|---|
| geosakel77 | identity | 50 |
| | indicator | 67,307 |
| | report | 50 |
| | threat-actor | 4 |
| | vulnerability | 4 |
| | bundles | 50 |
| bari24 | identity | 30 |
| | indicator | 1445 |
| | report | 30 |
| | threat-actor | 15 |
| | vulnerability | 14 |
| | bundles | 30 |
| bariMit | attack-pattern | 853 |
| | course-of-action | 140 |
| | identity | 10 |
| | intrusion-set | 17 |
| | malware | 465 |
| | marking-definition | 10 |
| | relationship | 4125 |
| | tool | 13 |
| | x-mitre-collection | 10 |
| | x-mitre-matrix | 16 |
| | x-mitre-tactic | 105 |
| | bundles | 10 |

Implementation details include the following:

- Gson Java library [48]—used to convert a JSON string to an equivalent Java object representing a STIX file.
- Apache POI library [49]—to export data in Excel format.
- Java version—JDK 23 (build 23.0.1+11-39)

The application was developed and run on a computer with an Intel Core i5-8500t processor (ANAFRA, Prague, Czech Republic), 16 GB RAM, SSD NVMe hard drive and Windows 11 operating system. The graphical user interface (GUI) of the app is presented in Figure 3. All the discussed aggregated dimensions are presented in this view. In addition, the dataset files, the calculated values of the initial dimensions for selected files, and the content of selected STIX files are presented in this view. The GUI was developed using standard Java Swing graphical components and several custom designed components drawn on JPanel components. The custom developed components include circular bars for percentage visualization and progress bar-like components to show dimension values for each individual object.

Additional functions for visualization were created. To visualize a STIX graph, a custom component `StixGraphVisualizer` was developed (Figure 4).

In Figure 4, it is possible to observe all objects from a single STIX file and their relationships. Drag-and-drop functionalities are enabled. When clicking on a single STIX object in the graph, all individual dimensions for that object (on the right) and available attributes (on the left) are displayed.

An example of a STIX file is provided in Listing 1. This example of STIX file can be found in [50]. The dimension values (Table 3) are calculated for the provided example using the implemented application. The purpose of providing the STIX file and the calcu-

lated values is to demonstrate the obtained specific values for the particular file. Such a demonstration would enable comparison of obtained values for future research studies.

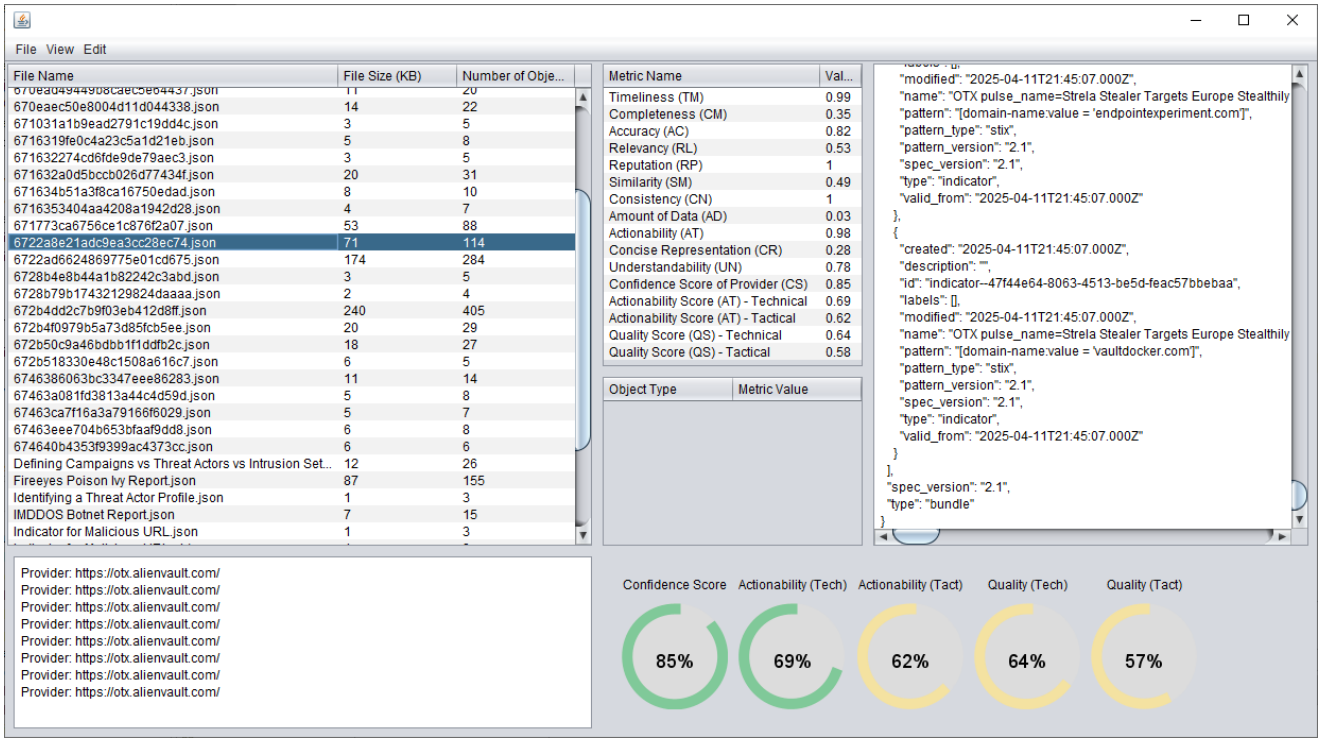


**Figure 3.** Graphical user interface of dimension calculation app.

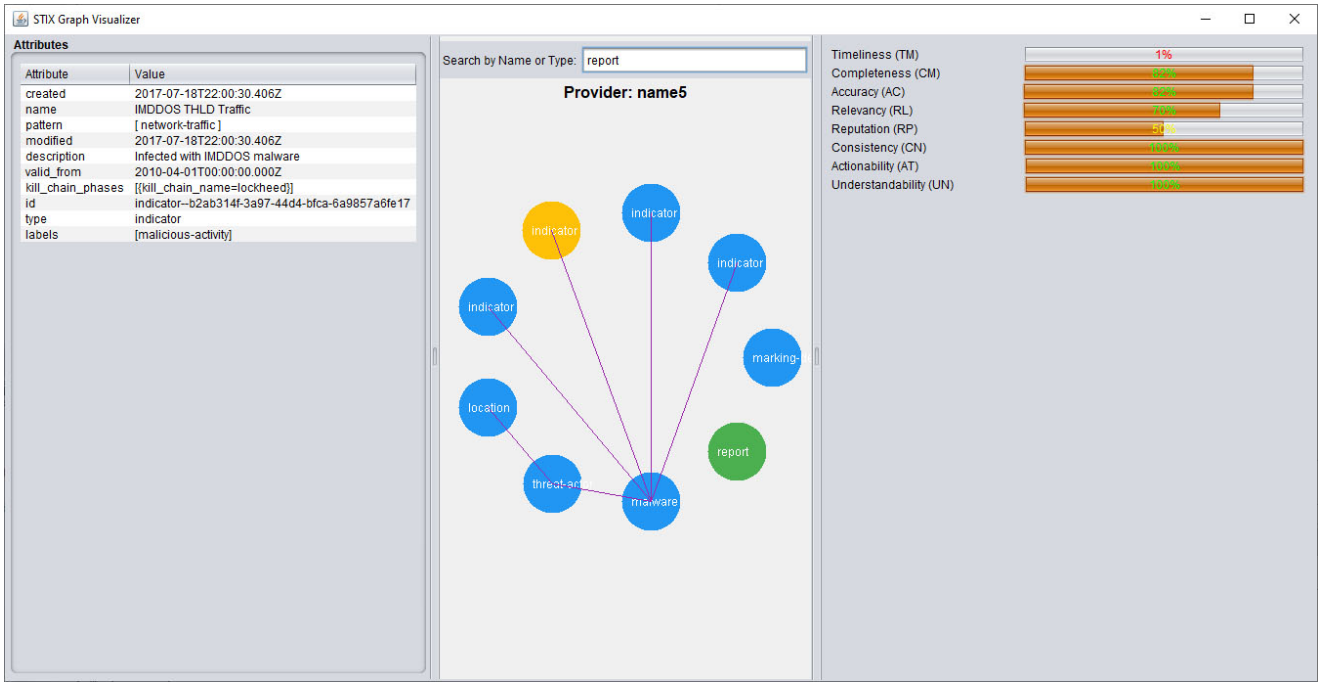


**Figure 4.** Visual graph of selected STIX file.

**Table 3.** Dimension values of the provided example.

| Dimension | Indicator | STIX File |
|---|---|---|
| Timeliness (TM) | 0.93 | 0.93 |
| Completeness (CM) | 0.65 | 0.49 |
| Accuracy (AC) | 0.82 | 0.69 |
| Relevance (RL) | 0.55 | 0.49 |
| Reputation (RP) | - | 0.99 |
| Similarity (SM) | - | 0.24 |
| Consistency (CN) | 1.00 | 1.00 |
| Amount of Data (AD) | - | 1.00 |
| Concise Representation (CR) | - | 0.47 |
| Understandability (UN) | 1.00 | 0.92 |
| Confidence Score of Provider (CS) | - | 0.92 |
| Actionability Score (AT) Technical | - | 0.67 |
| Actionability Score (AT) Tactical | - | 0.61 |
| Quality Score (QS) Technical | - | 0.70 |
| Quality Score (QS) Tactical | - | 0.65 |

**Listing 1.** Example of STIX file.

```
  {
      ‘‘type’’: ‘‘bundle’’,
      ‘‘id’’: ‘‘bundle--56be2a3b-1534-4bef-8fe9-602926274089’’,
           ‘‘objects’’: [
               {
                     ‘‘type’’: ‘‘indicator’’,
                     ‘‘spec_version’’: ‘‘2.1’’,
                     ‘‘id’’:
‘‘indicator--d81f86b9-975b-4c0b-875e-810c5ad45a4f’’,
                     ‘‘created’’: ‘‘2024-10-22T13:49:37.079Z’’,
                     ‘‘modified’’: ‘‘2024-10-22T13:49:37.079Z’’,
                     ‘‘name’’: ‘‘Malicious site hosting downloader’’,
                     ‘‘description’’: ‘‘This organized threat actor group
operates to create profit from all types of crime.’’,
                     ‘‘indicator_types’’: [
                         ‘‘malicious-activity’’
                     ],
                     ‘‘pattern’’: ‘‘[url:value =
‘http://x4z9arb.cn/4712/’’’,
                     ‘‘pattern_type’’: ‘‘stix’’,
                     ‘‘valid_from’’: ‘‘2024-10-22T13:49:37.079Z’’
                 },
                 {
                     ‘‘type’’: ‘‘malware’’,
                     ‘‘spec_version’’: ‘‘2.1’’,
                     ‘‘id’’:
‘‘malware--162d917e-766f-4611-b5d6-652791454fca’’,
                     ‘‘created’’: ‘‘2024-10-22T09:15:17.182Z’’,
                     ‘‘modified’’: ‘‘2024-10-22T09:15:17.182Z’’,
                     ‘‘name’’: ‘‘x4z9arb backdoor’’,
```

**Listing 1.** Cont.

```
                    ‘‘description’’: ‘‘This malware attempts to download
remote files after establishing a foothold as a backdoor.’’,
                    ‘‘malware_types’’: [
                        ‘‘backdoor’’,
                        ‘‘remote-access-trojan’’
                    ],
                    ‘‘is_family’’: false,
                    ‘‘kill_chain_phases’’: [
                        {
                          ‘‘kill_chain_name’’:
‘‘mandiant-attack-lifecycle-model’’,
                          ‘‘phase_name’’: ‘‘establish-foothold’’
                    }
                    ]
                },
                {
                    ‘‘type’’: ‘‘relationship’’,
                    ‘‘spec_version’’: ‘‘2.1’’,
                    ‘‘id’’:
‘‘relationship--864af2ea-46f9-4d23-b3a2-1c2adf81c265’’,
                    ‘‘created’’: ‘‘2024-10-22T18:03:58.029Z’’,
                    ‘‘modified’’: ‘‘2024-10-22T18:03:58.029Z’’,
                    ‘‘relationship_type’’: ‘‘indicates’’,
                    ‘‘source_ref’’:
‘‘indicator--d81f86b9-975b-4c0b-875e-810c5ad45a4f’’,
                    ‘‘target_ref’’:
‘‘malware--162d917e-766f-4611-b5d6-652791454fca’’
                }
            ]
}
```

The provided STIX file contains three objects: indicator, malware, and relationship. The dimensions are calculated for a single STIX object, indicator, and for the whole STIX file, which includes three objects.

The dimensions such as timeliness, completeness, and others, which depend on the attributes of individual objects, can be calculated for each STIX object independently. The dimensions of similarity, amount of data, and others require aggregation or referencing of other objects within the STIX file. Therefore, such dimensions are possible for the whole STIX file only.

The experiment was carried out on three datasets introduced in Table 2. A comparison of obtained median values for every singular dimension on every dataset is provided in Figure 5.

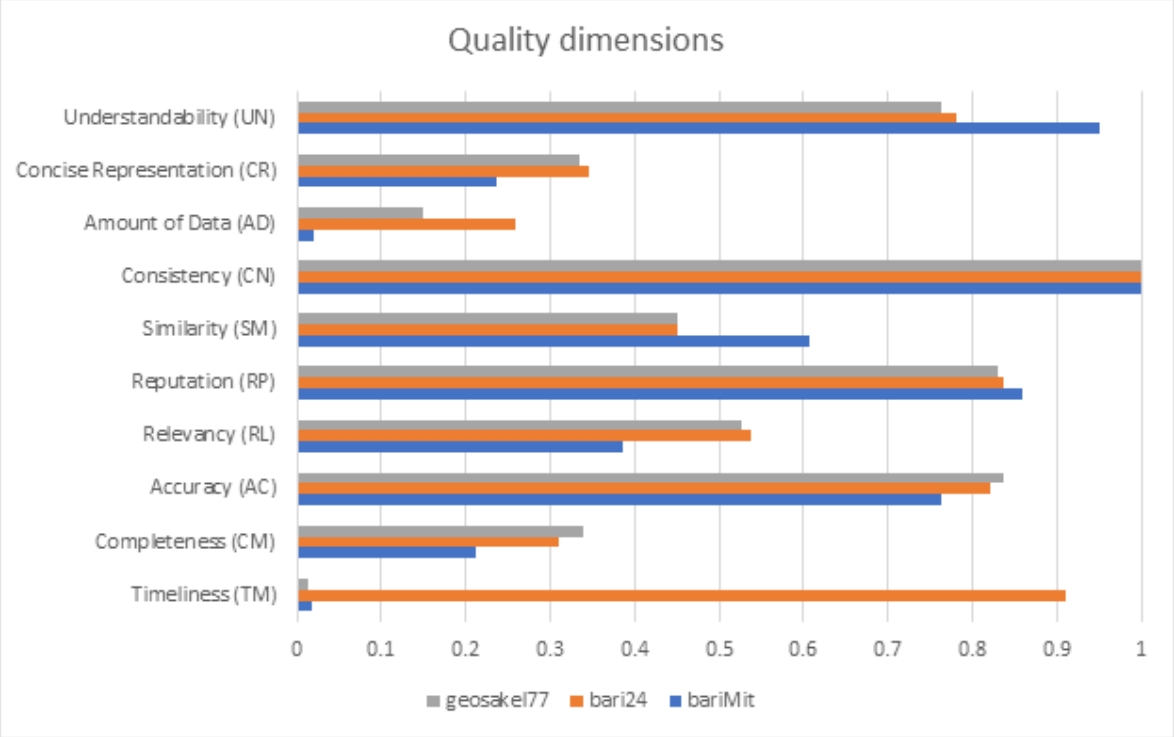


**Figure 5.** Median values of quality dimensions.

The following observations regarding Figure 5 can be made. STIX files from dataset bari24 stood out at the dimension amount of data, which means that it has more unique object types without repeating too many of the same type objects. The dataset bari24 was created using the latest STIX files obtained on the AlienVault site, meanwhile, bariMit and geosakel77 datasets were created or modified a long time ago. Therefore, their timeliness value is low compared to the dataset bari24. Furthermore, this low timeliness value significantly influences the final quality dimensions of these two sources, as shown in Figure 6. The similarity dimension shows that the dataset bariMit provides data with better differentiation and it has less overlapping or redundant entries. The bariMit dataset also has higher understandability which makes data more intuitive, having more descriptions and explanations compared to other datasets.

The following observations regarding Figure 6 can be made. The aggregated values of dataset bari24 were consistently slightly higher overall in most cases compared to the other two sources. As indicated earlier, the dataset bari24 excels in most aggregating dimensions, mostly due to its timely data. The dataset bariMit is more suitable for tactical level insights and shows weaker technical quality and technical actionability. The results of dataset geosakel77 are similar to the results of dataset bari24, since dataset geosakel77 was built from AlienVault files as well.

To conclude the comparison of the datasets, we can make the following observations. The datasets geosakel77 and bari24 are more suitable for use at the technical level. The dataset bariMit is more suitable for use at the tactical level. We expected similar results. Therefore, it is possible to state that the dimensions and the values of the weight coefficients in aggregating the dimensions are chosen correctly.

The aim of another experiment is to analyze the impact of the number of objects on the values of the dimensions. The dimensions, which are related to the number of STIX objects, are shown in Table 4. This table provides an informal definition of the dimensions closely related to the number of objects.

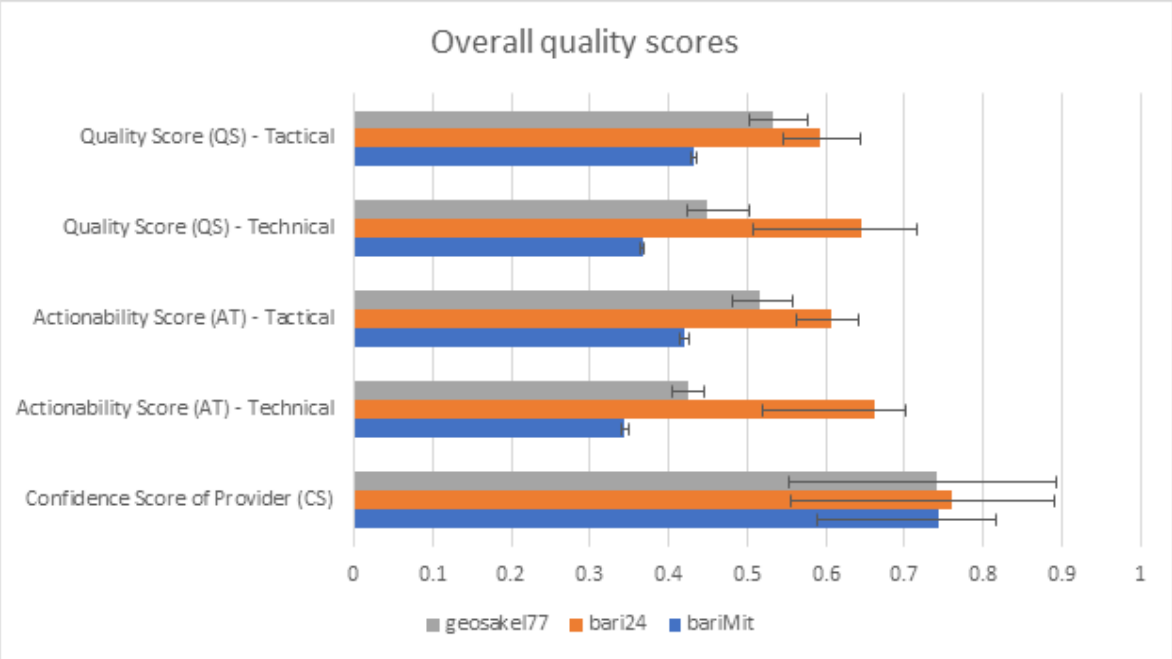


**Figure 6.** Final quality dimensions.

**Table 4.** List of dimensions that are dependent on the number of objects.

| Dimension | How It Relates to the Number of Objects |
|---|---|
| Completeness (CM) | STIX Domain Objects (SDOs) and STIX Relationship Objects (SROs) are counted. |
| Consistency (CN) | Constraints for all objects are checked for higher numbers of objects, higher probabilities of inconsistencies, incorrect relationships, or missing attributes. |
| Amount of Data (AD) | The number of object types is divided by the total number of objects. |
| Concise Representation (CR) | Related to similarity between objects. A larger number of objects introduces more opportunities for overlapping attributes and affects the dimension value. |
| Understandability (UN) | Related to the number of unique object types. |
| Similarity (SM) | Object pairs are compared looking for the maximum similarity, therefore files with more objects will have an impact on final value. |
| Relevance (RL) | The intersection of consumer required attributes with provider available attributes is compared across all objects. |

The relationships between quality dimensions and number of objects are shown in Figures 7 and 8. The STIX files from dataset bari24 were used to define the relationships shown in Figure 7. The STIX files from dataset geosakel77 were used to define the relationships shown in Figure 8. It is not meaningful to explore these relationships for dataset bariMit, since the number objects is very large; the smallest number of objects in the STIX file is close to 400.

The following observations regarding Figure 7 can be made. The completeness dimension appears to have lower values while the number of objects is small, and the dimension value stabilizes when the number of objects increases. A stable high value of the consistency dimension shows that data are valid and meet the schema and standards. The value of the actionability dimension increases with an increase in the number of objects, which means that larger files are more likely to have actionable attributes. The value of the amount of data dimension decreases with an increase in the number of objects, which means that the percentage of unique object types decreases.

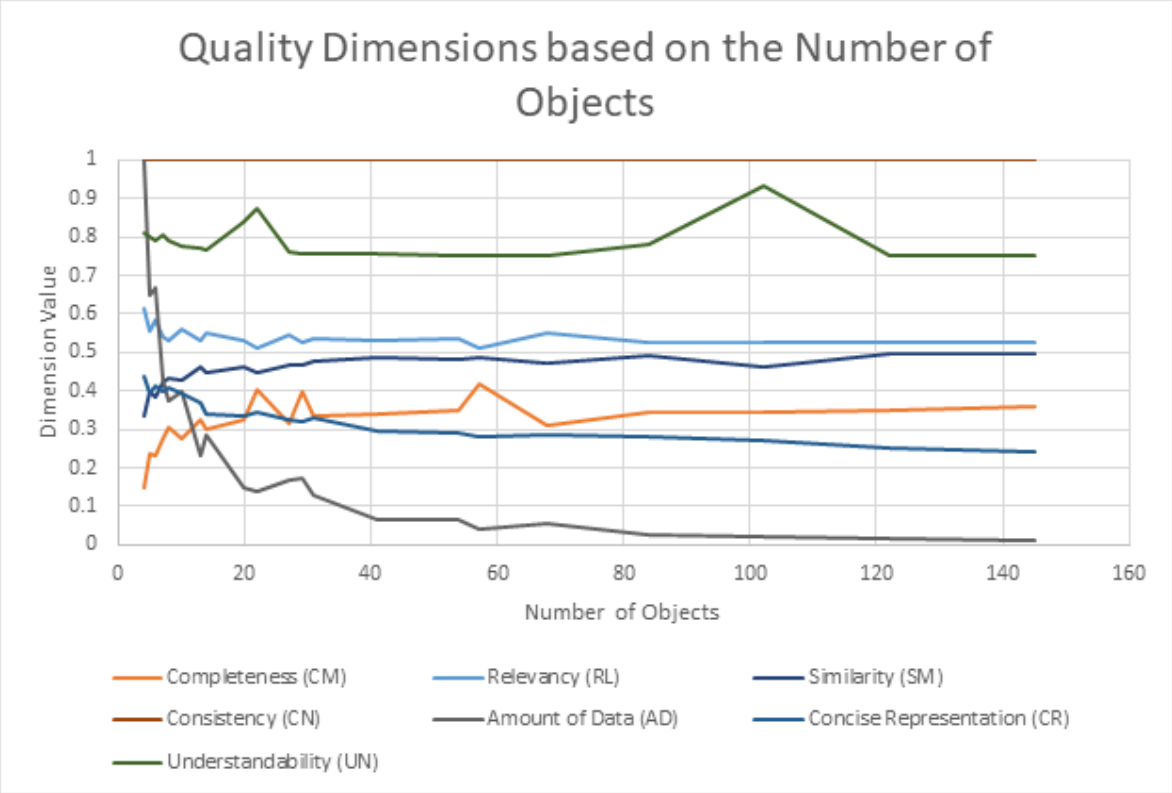


**Figure 7.** Relationship between quality dimensions and number of objects (dataset bari24).

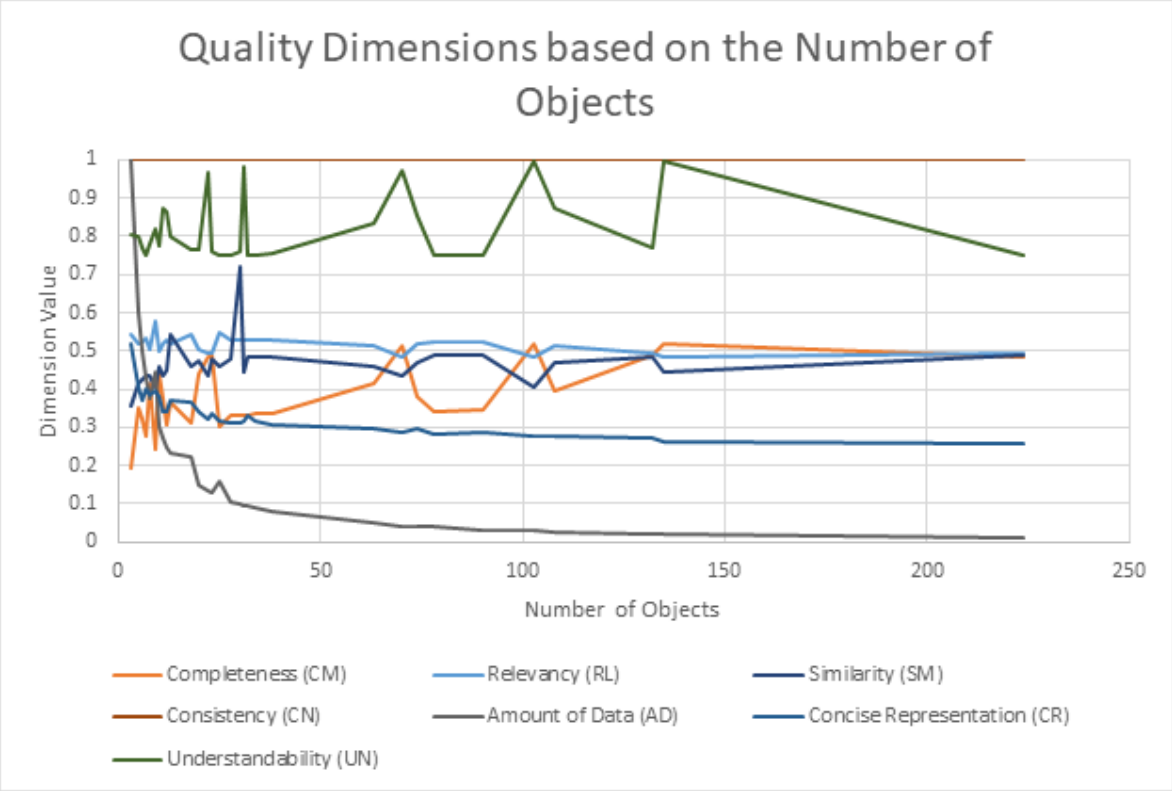


**Figure 8.** Relationship between quality dimensions and number of objects (dataset geosakel77).

If we compare the tendencies of the relationships between the dimension values and the number of objects in Figures 7 and 8, we can observe that these tendencies are almost the same. The curve of the consistency dimension is the same. The next highest values

are obtained for the understandability dimension in both figures. The lowest values are obtained for the amount of data dimension in both figures. Small differences are observed in the relevance, similarity, and completeness dimensions. Their curves overlap in Figure 7, and their curves do not overlap in Figure 8.

### 4.2. Discussion

Our approach possesses two features distinguishable to all the approaches presented so far. Firstly, the actionability dimension is presented as a summarizing feature of CTI data quality. Secondly, the assessment of the actionability dimension and of overall CTI data quality is separated into two values. The naming of separated values corresponds to the different levels of threat intelligence in use. By introducing two values for the assessment, we have solved an everlasting dispute between timeliness and completeness [38,41,43] regarding which dimension is more important, since these two dimensions have an inversely proportional relationship. It is obvious that the high value of timeliness is important for technical actionability, when CTI data are provided in the form of indicators of compromise, indicators of attack, forensic evidence, and technical description. The high value of completeness is important for tactical actionability, when tools, tactics, techniques, and procedures are described to offer context to attack analysis and information regarding the actors. Based on these observations, we split calculation of the value of actionability into two dimensions, technical actionability and tactical actionability. In the same way, we provide two scores for the final assessment of CTI data quality, technical quality and tactical quality. For technical quality, we assign the highest value of the weight coefficient to the timeliness dimension. For tactical quality, we assign the highest value of the weight coefficients to the completeness and accuracy dimensions.

We have provided a separate calculation for actionability as a finalizing indicator of CTI data quality, because actionability is identified as one of the fundamental building blocks of successful incident response [41]. Actionability shows the extent to which provided CTI data can be used directly to support an organization's security objectives [38].

A confidence score for the provider is presented as a separate finalizing dimension of quality assessment. A similar approach has been applied only in the study by Chen et al. [28]. The literature review revealed that the reliability of CTI data is the fourth most important indicator of quality of CTI data. However, the value of the reliability dimension is not assessed in our presented quality assessment. We have decided that the reliability of CTI data can be assessed according to the confidence score of the provider. This decision has to be made by the consumer of CTI data. Therefore, the confidence score of the provider is a separate finalizing dimension of the quality assessment.

To provide a many-sided comparison of the proposed method with methods used in related studies, we separate a comparison into two parts. These parts are as follows: a comparison of weights in calculating CTI quality scores (Table 5) and a comparison of the distinguishable features possessed by the methods (Table 6). Tables enable a rigorous quantitative evaluation of the methods.

The comparison of weights is possible only if we desire to compare the numeric values of obtained results with similar results declared in related studies, since obtained specific values of dimensions, which could be used for comparison, are not provided in the related studies.

We propose that relevance is the most important dimension in all the cases since there is no meaning to considering the CTI data if they are not relevant to the particular organization. Similar positions were demonstrated by Chen et al. [28] and Zibak et al. [26]. Relevance was assigned second place in the study by Chen et al. [28]. Zibak et al. [26] assigned only ranks to dimensions. They assigned the highest rank to relevance in the CTI

assessment. Zibak et al. [26] differentiated between two cases of CTI data; however, their terminology is confusing (CT data and CTI data). We propose the more understandable terminology of quality score technical and quality score tactical.

**Table 5.** Comparison of weights of dimensions.

| | CTI Quality Score | | | | | | | | Confidence Score of the Provider | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Method** | **Timeliness** | **Completeness** | **Accuracy** | **Relevance** | **Consistency** | **Amount of Data** | **Concise Represent.** | **Understandability** | **Reputation** | **Similarity** |
| Chen et al. [28] | 9% | 12% | 3% | 36% | 40% | - | - | - | 70% | 30% |
| Zibak et al. [26]. Ranks for CT data | First | - | Fourth | Second | - | - | - | - | Sixth | - |
| Zibak et al. [26]. Ranks for CTI | Third | - | Fourth | First | - | - | - | - | Fifth | - |
| Schlette et al. [24] | The weights are defined by the user. Initially, they are equal. | | | | | | | | | |
| DeCastro-García & Pinto [33] | The weights are defined by the user. The default values are not disclosed. | | | | | | | | | |
| Our CTI Quality Score Technical | 22% | 11% | 11% | 32% | 6% | 6% | 6% | 6% | 70% | 30% |
| Our CTI Quality Score Tactical | 8% | 18% | 18% | 32% | 6% | 6% | 6% | 6% | 70% | 30% |

Chen et al. [28] assigned first place to consistency; however, they used a few dimensions only in comparison to our approach. We additionally presented the dimensions amount of data, concise representation, and understandability, which are related to consistency. If we were to sum the coefficients of all these four dimensions, this accumulated dimension would take second place in our assessment.

Schlette et al. [24] and DeCastro-García & Pinto [33] did not consider the importance of the dimensions. They left this decision to users. Such an approach requires a profound knowledge of the dimensions and deep understanding of the domains by the user. The user can always make mistakes. We propose that our approach, which presents a finalized assessment, is preferable.

Next, we compare the distinguishing features possessed by our method and by the methods of related works (Table 6).

**Table 6.** Comparison of features possessed by the methods.

| **Features** | **Wang & Strong [15]** | **Li et al. [20]** | **Schaberreiter et al. [23]** | **Tundis et al. [32]** | **Mavzer et al. [25]** | **Schlette et al. [24]** | **DeCastro-García & Pinto [33]** | **Wang et al. [34]** | **Zibak et al. [26]** | **Yang et al. [27]** | **Chen et al. [28]** | **Our Method** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Comprehensive literature analysis | | | | | | | | | + | | | + |
| Fully automatic assessment of the dimensions | | | + | + | + | | + | + | | + | + | + |
| Ability to assess new CTI data | + | | | | + | + | + | | | | | + |
| Way to calculate dimensions | | | + | + | + | + | | + | | + | + | + |
| Number of summarizing dimensions | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 2 | 5 |
| Structured CTI oriented | | | + | + | + | + | + | + | + | | + | + |
| Visualization | | | | | | + | | | | | + | + |
| Delphi study | + | | | + | | | | | + | | | |
| Total number of possessed features | 3 | 1 | 4 | 5 | 5 | 5 | 4 | 4 | 4 | 3 | 5 | 7 |

We can observe (Table 6) that our method significantly surpasses all the methods considered by the number of possessed distinguishable features. Next, we consider explicitly the advantageous features provided by our method. Our proposed method possesses all the features that were enumerated in Table 6, except for a Delphi study. To cover this shortcoming, we have used insights provided by experts in the other studies (Wang and Strong [15], Tundis et al. [32], Zibak et al. [26]). The next feature, which distinguishes our method from related methods, is the number of summarizing dimensions. The importance of this peculiarity of our method has already been introduced at the beginning of this section. Wang and Strong [15] only suggested four categories for data quality assessments in their seminal work in 1996. However, this view was not supported in later research studies. The next distinguishing features are as follows: comprehensive literature review, visualization, and ability to assess new CTI data. Many research studies apply multi-source verification during assessment of CTI data quality. Therefore, such methods are not suitable for assessing new CTI data that are not supported by other sources.

The unifying features of the considered methods are orientation to structured data, provision of ways to calculate the dimension values, and automatic assessment of the dimensions. The features of orientation to structured data and automatic assessment are closely related, since automatic assessment of structured data is possible only. The structured data also provide more benefits. The benefits are as follows: a template to fill in the data, machine readable format for further interpretation and analysis, improved quality of provided information, easy comprehension of provided data, and others.

## 5. Conclusions

A review of the related literature revealed that many authors have tried to suggest dimensions to measure the quality of CTI data. However, these dimensions vary quite substantially. Furthermore, the measurement methods used for these dimensions also vary quite substantially. Not all the dimensions are automatically calculated. Many research studies apply multi-source verification. Such an approach is not suitable for assessing CTI data that are presented for the first time. To overcome the shortcomings of the available methods of CTI data quality assessment, we selected from the literature the quality dimensions recognized by a majority of related studies and suggested the fully automated measurement formulae of the dimensions for structured CTI data provided in the STIX language. We have created formulae to calculate values for the chosen dimensions using the experiences of related studies and augmenting our own views. Therefore, many of our formulae are original ones.

To accommodate the requirements of the technical and tactical levels of using CTI data, we have introduced new dimensions as follows: technical actionability, tactical actionability, technical CTI data quality, and tactical CTI data quality. In such a way, we have solved the eternal dispute regarding the contradicting dimensions of timeliness and completeness to show which dimension is more important in assessing CTI data quality. The next distinguishing feature is the separation of the assessment of CTI data quality and the confidence scores of CTI providers. At the tactical level, confidence scores for CTI providers are very important. Therefore, the user himself, seeing a value of confidence score for a CTI provider, can decide whether to trust or not to trust the provided CTI data. Next, we have visualized the calculated values of CTI data quality dimensions to make the values easier to interpret for the user.

For this experiment, we have used three datasets: a publicly available dataset geodakel77, to enable the possibility of comparison for future related studies; our constructed dataset bari24 to test the CTI descriptions of the latest available threats; and our constructed

dataset bariMit to test large CTI descriptions with numerous custom-developed user objects. The results of the experiment showed the viability and applicability of our method.

Finally, it is possible to observe that we have kept our promise announced in [51], which regarded automatic CTI quality assessment methods. Our next research direction involves the development of a method for zero-day ransomware detection using static features of portable executable files.

**Author Contributions:** Conceptualization, V.J. and A.V.; methodology, V.J.; software, D.B.; validation, V.J., D.B. and A.V.; formal znalysis, V.J.; investigation, V.J.; resources, V.J. and D.B.; data curation, V.J. and D.B.; writing—original draft preparation, V.J.; writing—review and editing, V.J.; visualization, D.B.; supervision, A.V.; project administration, A.V.; funding acquisition, A.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Economic Revitalization and Resilience Enhancement Plan "New Generation Lithuania" as part of the execution of Project "Mission-driven Implementation of Science and Innovation Programmes" (No. 02-002-P-0001).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** One part of the dataset is available publicly [44]. The other part of the dataset will be made available by the authors on request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Chatziamanetoglou, D.; Rantos, K. Cyber Threat Intelligence on Blockchain: A Systematic Literature Review. *Computers* **2024**, *13*, 60. [CrossRef]
2. Alaeifar, P.; Pal, S.; Jadidi, Z.; Hussain, M.; Foo, E. Current approaches and future directions for Cyber Threat Intelligence sharing: A survey. *J. Inf. Secur. Appl.* **2024**, *83*, 103786, ISSN 2214-2126. [CrossRef]
3. Preuveneers, D.; Joosen, W. Sharing Machine Learning Models as Indicators of Compromise for Cyber Threat Intelligence. *J. Cybersecur. Priv.* **2021**, *1*, 140–163. [CrossRef]
4. Bandara, E.; Shetty, S.; Mukkamala, R.; Rahaman, A.; Liang, X. LUUNU—Blockchain, MISP, Model Cards and Federated Learning Enabled Cyber Threat Intelligence Sharing Platform. In Proceedings of the 2022 Annual Modeling and Simulation Conference (ANNSIM), San Diego, CA, USA, 18–20 July 2022; pp. 235–245. [CrossRef]
5. Gillard, S.; David, D.P.; Mermoud, A.; Maillart, T. Efficient collective action for tackling time-critical cybersecurity threats. *J. Cybersecur.* **2003**, *9*, tyad021. [CrossRef]
6. Menges, F.; Putz, B.; Pernul, G. DEALER: Decentralized incentives for threat intelligence reporting and exchange. *Int. J. Inf. Secur.* **2021**, *20*, 741–761. [CrossRef]
7. Ali, H.; Ahmad, J.; Jaroucheh, Z.; Papadopoulos, P.; Pitropakis, N.; Lo, O.; Abramson, W.; Buchanan, W.J. Trusted Threat Intelligence Sharing in Practice and Performance Benchmarking through the Hyperledger Fabric Platform. *Entropy* **2022**, *24*, 1379. [CrossRef]
8. Sun, N.; Ding, M.; Jiang, J.; Xu, W.; Mo, X.; Tai, Y.; Zhang, J. Cyber Threat Intelligence Mining for Proactive Cybersecurity Defense: A Survey and New Perspectives. *IEEE Commun. Surv. Tutor.* **2023**, *25*, 1748–1774. [CrossRef]
9. Chatziamanetoglou, D.; Rantos, K. Blockchain-Based Cyber Threat Intelligence Sharing Using Proof-of-Quality Consensus. *Secur. Commun. Netw.* **2023**, *2023*, 3303122. [CrossRef]
10. Purohit, S.; Neupane, R.; Bhamidipati, N.R.; Vakkavanthula, V.; Wang, S.; Rockey, M.; Calyam, P. Cyber Threat Intelligence Sharing for Co-Operative Defense in Multi-Domain Entities. *IEEE Trans. Dependable Secur. Comput.* **2023**, *20*, 4273–4290. [CrossRef]
11. Dimitriadis, A.; Lontzetidis, E.; Mavridis, I. Evaluation and Enhancement of the Actionability of Publicly Available Cyber Threat Information in Digital Forensics. In Proceedings of the 2021 IEEE International Conference on Cyber Security and Resilience (CSR), Rhodes, Greece, 26–28 July 2021; pp. 318–323. [CrossRef]
12. Barnum, S. Standardizing Cyber Threat Intelligence Information with the Structured Threat Information Expression (STIX™). 2014. Available online: https://stixproject.github.io/about/STIX_Whitepaper_v1.1.pdf (accessed on 18 October 2024).
13. Ainslie, S.; Thompson, D.; Maynard, S.; Ahmad, A. Cyber-threat intelligence for security decision-making: A review and research agenda for practice. *Comput. Secur.* **2023**, *132*, 103352, ISSN 0167-4048. [CrossRef]
14. Introduction to STIX. Available online: https://oasis-open.github.io/cti-documentation/stix/intro.html (accessed on 18 October 2024).

15. Wang, R.Y.; Strong, D.M. Beyond accuracy: What data quality means to data consumers. *J. Manag. Inf. Syst.* **1996**, *12*, 5–33. [CrossRef]
16. *ISO/IEC 25012*; Software Engineering-Software Product Quality Requirements and Evaluation (Square)-Data Quality Model. The International Organization for Standardization: Geneva, Switzerland, 2008. Available online: https://www.iso.org/standard/35736.html (accessed on 7 October 2024).
17. Wang, R.Y. A product perspective on total data quality management. *Commun. ACM* **1998**, *41*, 58–65. [CrossRef]
18. *ISO 25024*; Systems and Software Engineering—Systems and Software Quality Requirements and Evaluation (SQuaRE)—Measurement of Data Quality. The International Organization for Standardization: Geneva, Switzerland, 2015. Available online: https://www.iso.org/standard/35749.html (accessed on 11 April 2025).
19. Gao, Y.; Li, X.; Li, J.; Gao, Y.; Guo, N. Graph Mining-based Trust Evaluation Mechanism with Multidimensional Features for Large-scale Heterogeneous Threat Intelligence. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 1272–1277. [CrossRef]
20. Li, Q.; Jiang, Z.; Yang, Z.; Liu, B.; Wang, X.; Zhang, Y. A Quality Evaluation Method of Cyber Threat Intelligence in User Perspective. In Proceedings of the 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), New York, NY, USA, 1–3 August 2018; pp. 269–276. [CrossRef]
21. Meier, R.; Scherrer, C.; Gugelmann, D.; Lenders, V.; Vanbever, L. FeedRank: A tamper-resistant method for the ranking of cyber threat intelligence feeds. In Proceedings of the 2018 10th International Conference on Cyber Conflict (CyCon), Tallinn, Estonia, 29 May–1 June 2018; pp. 321–344. [CrossRef]
22. Li, V.G.; Dunn, M.; Pearce, P.; McCoy, D.; Voelker, G.M.; Savage, S. Reading the tea leaves: A comparative analysis of threat intelligence. In Proceedings of the 28th USENIX Security Symposium (USENIX Security 19), Santa Clara, CA, USA, 14–16 August 2019; pp. 851–867.
23. Schaberreiter, T.; Kupfersberger, V.; Rantos, K.; Spyros, A.; Papanikolaou, A.; Ilioudis, C.; Quirchmayr, G. 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, 26–29 August 2019; pp. 1–10. [CrossRef]
24. Schlette, D.; Böhm, F.; Caselli, M.; Pernul, G. Measuring and visualizing cyber threat intelligence quality. *Int. J. Inf. Secur.* **2021**, *20*, 21–38. [CrossRef]
25. Mavzer, K.B.; Konieczna, E.; Alves, H.; Yucel, C.; Chalkias, I.; Mallis, D.; Cetinkaya, D.; Sanchez LA, G. Trust and QualityComputation for Cyber Threat Intelligence Sharing Platforms. In Proceedings of the 2021 IEEE International Conference on Cyber Security and Resilience (CSR), Rhodes, Greece, 26–28 July 2021; pp. 360–365. [CrossRef]
26. Zibak, A.; Sauerwein, C.; Simpson, A.C. Threat intelligence quality dimensions for research and practice. Digit. *Threat. Res. Pract.* **2022**, *3*, 44. [CrossRef]
27. Yang, L.; Wang, M.; Lou, W. An automated dynamic quality assessment method for cyber threat intelligence. *Comput. Secur.* **2024**, *148*, 104079. [CrossRef]
28. Chen, S.-S.; Hwang, R.-H.; Ali, A.; Lin, Y.-D.; Wei, Y.-C.; Pai, T.-W. Improving quality of indicators of compromise using STIX graphs. *Comput. Secur.* **2024**, *144*, 103972. [CrossRef]
29. Sakellariou, G.; Fouliras, P.; Mavridis, I. A Methodology for Developing & Assessing CTI Quality Metrics. *IEEE Access* **2024**, *12*, 6225–6238. [CrossRef]
30. Grispos, G.; Glisson, W.B.; Storer, T. How good is your data? Investigating the quality of data generated during security incident response investigations. In Proceedings of the 52nd Hawaii International Conference on System Sciences Scholar Space Hawaii International, Maui, HI, USA, 8–11 April 2019; pp. 7156–7165. Available online: https://hdl.handle.net/10125/60152 (accessed on 24 October 2024).
31. Dalziel, H. A Problem Well-Defined is Half-Solved. In *How to Define and Build an Effective Cyber Threat Intelligence Capability*; Elsevier: London, UK, 2015; pp. 3–6.
32. Tundis, A.; Ruppert, S.; Mühlhäuser, M. On the Automated Assessment of Open-Source Cyber Threat Intelligence Sources. In *Lecture Notes in Computer Science, Proceedings of the International Conference on Computational Science, Amsterdam, The Netherlands, 3–5 June 2020*; Krzhizhanovskaya, V.V., Závodszky, G., Lees, M.H., Dongarra, J.J., Sloot, P.M.A., Brissos, S., Teixeira, J., Eds.; Springer: Cham, Switzerland, 2020; Volume 12138. [CrossRef]
33. DeCastro-García, N.; Pinto, E. Measuring the Quality Information of Sources of Cybersecurity by Multi-Criteria Decision Making Techniques. In *Lecture Notes in Computer Science, Proceedings of the Hybrid Artificial Intelligent Systems, Salamancaa, Spain, 5–7 September 2022*; Bringas, P.G., García, H.P., de Pisón, F.J.M., Flecha, J.R.V., Lora, A.T., de la Cal, E.A., Herrero, Á., Álvarez, F.M., Psaila, G., et al., Eds.; Springer: Cham, Switzerland, 2023; Volume 13469. [CrossRef]
34. Wang, M.; Yang, L.; Lou, W. A Comprehensive Dynamic Quality Assessment Method for Cyber Threat Intelligence. In Proceedings of the 2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W), Baltimore, MD, USA, 27–30 June 2022; pp. 178–181. [CrossRef]

35. Griffioen, H.; Booij, T.; Doerr, C. Quality evaluation of cyber threat intelligence feeds. In Proceedings of the International Conference on Applied Cryptography and Network Security: 18th International Conference, ACNS 2020, Rome, Italy, 19–22 October 2020; Proceedings, Part II. Springer: Berlin/Heidelberg, Germany, 2020; pp. 277–296. [CrossRef]
36. Ahlemann, F.; Gastl, H. Process model for an empiracally grounded reference model construction. In *Reference Modeling for Business Systems Analysis*; IGI Global Scientific Publishing: Hershey, PA, USA, 2007; pp. 77–97. [CrossRef]
37. Sakellariou, G.; Fouliras, P.; Mavridis, I.; Sarigiannidis, P. A Reference Model for Cyber Threat Intelligence (CTI) Systems. *Electronics* **2022**, *11*, 1401. [CrossRef]
38. Tundis, A.; Ruppert, S.; Mühlhäuser, M. A Feature-driven Method for Automating the Assessment of OSINT Cyber Threat Sources. *Comput. Secur.* **2022**, *113*, 102576. [CrossRef]
39. Gong, S.; Cho, H.; Lee, C. A Reliability Comparison Method for OSINT Validity Analysis. *IEEE Trans. Ind. Inform.* **2018**, *14*, 5428–5435. [CrossRef]
40. Batini, C.; Cappiello, C.; Francalanci, C.; Maurino, A. Methodologies for data quality assessment and improvement. *ACM Comput. Surv.* **2009**, *41*, 16. [CrossRef]
41. European Union Agency for Cybersecurity, Actionable Information for Security Incident Response. 2015. Available online: https://www.enisa.europa.eu/publications/actionable-information-for-security (accessed on 28 October 2024).
42. Geras, T.; Schreck, T. The "Big Beast to Tackle": Practices in Quality Assurance for Cyber Threat Intelligence. In Proceedings of the RAID '24: 27th International Symposium on Research in Attacks, Intrusions and Defenses, Padua, Italy, 30 September–2 October 2024; pp. 337–352. [CrossRef]
43. Yucel, C.; Chalkias, I.; Mallis, D.; Karagiannis, E.; Cetinkaya, D.; Katos, V. On the Assessment of Completeness and Timeliness of Actionable Cyber Threat Intelligence Artefacts. In *Communications in Computer and Information Science, Proceedings of the Communications in Computer and Information Science, Kraków, Poland, 8–9 October 2020*; Dziech, A., Mees, W., Czyżewski, A., Eds.; Springer: Cham, Switzerland, 2020; Volume 1284. [CrossRef]
44. Bromander, S.; Swimmer, M.; Muller, L./P.; Jøsang, A.; Eian, M.; Skjøtskift, G.; Borg, F. Investigating Sharing of Cyber Threat Intelligence and Proposing a New Data Model for Enabling Automation in Knowledge Representation and Exchange. *Digit. Threat. Res. Pract.* **2021**, *3*, 6. [CrossRef]
45. Cyber Threat Intelligence (CTI) Quality Metrics. Available online: https://github.com/geosakel77/s2 (accessed on 28 October 2024).
46. AlienVault Open Threat Exchange (OTX). Available online: https://otx.alienvault.com/dashboard/new (accessed on 28 October 2024).
47. Mitre-Attack/Attack-Stix-Data. Available online: https://github.com/mitre-attack/attack-stix-data/ (accessed on 28 October 2024).
48. Google/Gson. Available online: https://github.com/google/gson (accessed on 27 September 2024).
49. Apache POI. Available online: https://poi.apache.org/download.html (accessed on 7 October 2024).
50. Indicator for Malicious URL. Available online: https://oasis-open.github.io/cti-documentation/examples/indicator-for-malicious-url.html (accessed on 28 October 2024).
51. Venčkauskas, A.; Jusas, V.; Barisas, D.; Misnevs, B. Blockchain-Based Model for Incentivized Cyber Threat Intelligence Sharing. *Appl. Sci.* **2024**, *14*, 6872. [CrossRef]