Trust and Quality Computation for Cyber Threat Intelligence Sharing Platforms

Kadir Burak Mavzer*, Ewa Konieczna*, Henrique Alves*,
Cagatay Yucel†, Ioannis Chalkias†, Dimitrios Mallis†, Deniz Cetinkaya †
Luis Angel Galindo Sanchez ‡
*VisionSpace Technologies GmbH
Email: ewa.konieczna@visionspace.com
†Bournemouth University
Email: cyucel@bournemouth.ac.uk
‡Telefonica Global Services GmbH
Email: luisangel.galindosanchez@telefonica.com

***Abstract*—Information sharing has been considered a critical solution against the ever-increasing complexity of cyber-attacks. In this effort Cyber Threat Intelligence is undergoing a process of increasing its maturity levels. The quantification of the quality of shared information and the assessment of trust amongst information sharing entities is an important part of the process. The Trust and Quality Tool has been designed as a tool with the aim of improving the trust in the relevancy of shared information by enabling an option to assess its trustworthiness and defining a set of metrics for trust and quality.**

Keywords: Trust in Cyber Threat Intelligence, Cyber Threat Intelligence Quality, Information Sharing.

I. Introduction

Sharing of the Cyber Threat Intelligence (CTI) has been presented and utilised as a rewarding solution to the ever-increasing complexity of cyberattacks [1]–[3]. The maturity level of CTI communities and applications acting on this highly valued - actionable threat information increases along with the complexity of cyberattacks. It has also been of paramount importance to quantify the quality of the shared information and assess trust amongst information sharing entities [4]–[6]. Several issues and challenges have been identified so far by the CTI community and data quality measurement has been reported as one of them [7].

The idea behind defining trust and quality metrics for the threat intelligence data that is shared amongst CTI communities is to decrease the level of information overload as well as reduce false positives, which are common in most CTI sharing platforms. A metric is defined as a consistent standard of measurement. It allows us to measure attributes and behaviours of interest. In order to improve the trust between partners and the quality of the threat intelligence they share, we first need to be able to measure these two parameters. These are also interdependent, adding a layer of complexity to the task. Establishing trust between partners heavily depends on improving the quality of information shared among them. However, measuring the quality of threat intelligence data is not an easy task as it involves many variables and requires thorough research beforehand. Also, it is difficult to determine how organisations define high-quality threat intelligence.

The methodology described in this paper aims to propose another approach for the definition of these metrics and to demonstrate how they are to be used to rate the quality of threat intelligence data shared between partners. By doing so, this methodology will aid improving the trust in the relevancy of information shared among them. This will also enable an option to assess the trustworthiness of received information and its source, based on metrics directly attributed to the trust level. This research presents an implementation of the calculation of these metrics by a software plugin named Trust and Quality Tool (TQM) leveraging a cyber threat intelligence sharing software named ECHO - Early Warning System (E-EWS) developed within the efforts of the acknowledged ECHO project. Within the activities of the project, four cybersecurity tabletop exercises have been organised amongst the partners of the project and TQM has been applied to the CTI that has been created during these exercises. The collected CTI has also been evaluated by security experts, cybersecurity software developers and cybersecurity academics within the ECHO network and the metrics have been improved and calibrated accordingly.

The contributions of this research are as follows:

- To define trust and quality metrics for CTI
- To present TQM tool for calculation and evaluation of these metrics in the E-EWS environment
- To present the quality and trust calculations of the produced CTI of the tabletop exercises
- To present the calibration of the tool using expert views within the ECHO network

The rest of the paper is organised as follows: Section 2 presents the literature on information quality on cyber threat intelligence and conducts a brief survey towards them, Section 3 defines the methodology, introduces the E-EWS, tabletop exercises and the metrics. Section 4 explains the requirements, structure and development of the TQM. Moreover, Section 5 presents the evaluation and calibration of TQM guided by the CTI collected with tabletop exercises while Section 6 concludes the paper.

II. Background

A. CTI Quality and Trust

According to the ENISA report "Actionable Information for Security Incident Response" [8], the five criteria that an information should meet to be actionable and support decision makers are accuracy, relevance, timeliness, completeness, and ingestibility. A recent survey shows that contemporary CTI feeds do not make it up to the requirements and expectations of IT security practitioners [8], especially regarding the aforementioned criteria. Threat Intelligence Sharing Platforms (TISPs) are trying to achieve the goal of meeting these requirements by creating policies and software that produces and shares actionable information.

The quality of data of the shared feeds can be seen as a vector compiled by the criteria of timeliness, accuracy, scope, relevance and completeness are used in references [9], [10] to measure the quality of the data and further evaluate the available threat intelligent feeds. In [11], a threat score function is introduced to evaluate Indicators of Compromise (IoC) collected from various sources in order to support Security Operations Centre (SOC) analysts prioritise the incidents' analysis. Another study [10] investigates the data quality dimensions of IoCs which are collected by several open sources in order to assess their effectiveness.

The challenges of data quality in TISPs were also researched and presented in the works of Sillaber et al. [4], aiming to address the factors affecting data quality of CTI at each of the levels of gathering, storing, processing, and sharing data. Their analysis was based on the data quality dimensions mentioned above with consistency added as another factor.

The quality of the data generated by incident response teams during investigations is discussed in [5]. According to this analysis, there is still a lot of future work to be done towards enhancing the quality of data generated by incident response teams in order to facilitate and support CTI. The same metrics were suggested by S. Sadiq in his handbook [9] under the category titled 'data values' which is one of the three main categories defining the dimensions of data quality in his work.

The list of challenges extends over to the information security features of privacy and trust. Applying privacy preserving mechanisms introduces a trade-off between the effectiveness of the process of information sharing and compliance with privacy preserving guidelines and policies [10], [11]. Moreover, trustworthiness of a CTI source is evaluated in the studies of [12], [13] in the aim of providing a more rigorous trust-model.

Diverse data models, tools and standards might affect the actionability, timeliness and consistency in regards of interoperability between TISPs [14], [15]. For that reason, the standardisation of CTI is crucial for the improvement of analytical and management capabilities in order to further increase the quality of the shared CTI. An implementation of this ideas has been done recently in MISP [16], an overall score is given based on tags present in the CTI and the source reliability, also an analysis is done per type of attribute of the CTI using a decay function to quantify its quality. Although the use of a trust in a source in the scoring is used, the article mentions the historical data sharing should be part of future research. Another TISP project developed in the scope of H2020 PROTECTIVE [17] assess the quality of data as well as the entity producing the CTI using a computational trust method which aggregates multiple properties to generate an overall reputation score of Threat intelligence feeds. In the end the CTI quality is calculated using a default configuration that uses schema completeness, freshness, relevant associated tags and source reputation.

There are several approaches to determine Trust, further described in [12]. To the best of our knowledge, the TISPs that are mentioned in this study do not establish trust using automated mechanisms.

There are also different approaches to what should define the quality of information shared. Based on the analysis described above, many CTI sharing platforms do not provide any extensive description related to the relevance of the threat intelligence other than the decay function related to timeliness.

III. Methodology

A. Enter E-EWS: ECHO's Early Warning System and Tabletop Exercises

The answer to the ever-increasing complexity and speed of cyber attack campaigns from cybersecurity communities is the fast and accurate sharing of actionable cyber threat intelligence. E-EWS is designed as a security operations support tool for increasing the incident handling capacities and enabling members of a community to share and coordinate with actionable CTI in near real-time level.The tabletop exercises (TTX) are part of the evaluation of the E-EWS. The purpose of the TTXs is to evaluate the development of EWS' platform and its added features, identify possible bugs, while at the same time monitoring the information-sharing policies during its operation.

During the exercises, the participating teams are exposed to a series of injects that require an incident response and allow them to engage by sharing information, according to their decision-making policies, the situational awareness of the incident and its nature.

The main concept behind the creation of the scenarios is the development of a main theme that is addressed by all the participants and some secondary/complementary incidents that will be delivered to the participants at any given moment during the exercise. The main theme of the exercise is not a collection of regular incidents, the other incidents are common incidents that incident response teams face at an almost daily rate. In the 2nd and 3rd TTX, each participant received two secondary incidents which revealed the main topic and the way they were convoluted in order to affect the organisation.

The scenarios involve complex and interrelated situations that provide the teams with motives to share tickets with the rest of the EWS. Since the purpose of a TTX is the generation and circulation of information, the users can share information internally (either to the whole team or specific users) or externally (information can be forwarded to specific entities

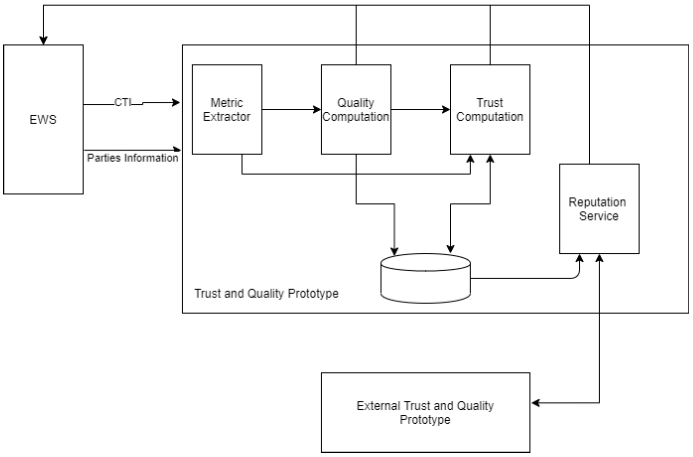

Fig. 1. Component View of the TQM Tool

or whole constituencies). The injections for the exercises are prepared from events in incident response, contemporary cybersecurity news, newly released vulnerabilities, and the characteristics of the organisations and of ECHO as a consortium. On occasion, an incident can be designed to specifically test the use of a new feature or tool or to fulfil a request of a team for specific testing.

Another useful addition to the scenarios is the "hourlies", tickets created by the organisers and shared in the constituencies to increase cyber situational awareness. The tickets contain information that can be relevant to the incidents and are useful to maintaining a flow during the exercise.

B. TQM Requirements and Structure

The TQM tool runs as a standalone application, an outside entity which is responsible for providing the input data following the API provided. The Prototype uses the fields of the CTI and information provided of both sender and receiver to determine the value of metrics. The calculation of the Quality and Trust is based on available metrics, which are weighted and combined. The computed scores are outputted in a report so it can be easily integrated with a Threat Intelligence Sharing Platform (TISP). The Trust score can also be shared with other entities using the prototype to later compute overall Trust score called Reputation. The overall components and flow is depicted in Figure 1. As can be seen from this figure, the metrics are extracted and computed from the E-EWS, stored in a MongoDb database. The stored values are then being utilised in the Reputation Service to compute the reputation scores of each organisation in the community of E-EWS.

The Figure 2 depicts a course the ticket follows, high-level view of both quality and trust computation and the postulated metrics that could be extracted. A ticket from the E-EWS is received with the metadata of the ticket, the contents and additionally, the metadata of the sharing organisation. These information is then fed to the separate metric extractors as described in detail in Table I.

C. Trust And Quality Computation

TQM Prototype aims to use several metrics to compute the Quality score of the information shared based on several aspects. Derived Quality score should enable organisations to prioritise the information shared with them, that is relevant, complete, and actionable. Since past experience is imperative to building and maintaining Trust, TQM Prototype aims to use both, computed Trust score together with collective experience of all organisations reputation, to determine the trustworthiness of the organisation sharing the information with it.

IV. Evaluation & Calibration

For the evaluation of the tool and the computation of the metrics, CTI produced during the aforementioned tabletop exercises is leveraged.

Tabletop exercise data and the current computation of the metrics have some limitations which are thoroughly discussed in the next section. Therefore, The computation of four metrics Completeness, Extensiveness, Freshness and Quality scores is executed for over 66 intelligence items from the data taken from four tabletop exercises.

In the Table I, the computation of these metrics can be found. As it can be seen, quality metric is defined as a function of other metrics with the given weights.

In addition to the evaluation of these tickets with the TQM tool, the produced CTI is evaluated by seven cybersecurity experts from the participants of the tabletop exercises. For these 4 metrics, the average value of the metric is calculated and the computation is adjusted by taking these considerations into account. Adjusted computation can also be found in the adjusted computation column of Table II.

The results of the adjusted computation, along with the expert's views and the initial computation are presented in the figures from 3 to 6. As can be seen from these figures these adjustments calibrated the tool to produce values that represent and reflect on the expert's views compared to the initial computation.

V. Discussion & Limitations

We are aware that the complete development of a tool requires substantial work; therefore the tool has limitations, especially related to the different metrics considered, which have been evolving in an agile manner as soon as a limitation is detected.

In the table of parameters that TQM measures, it can be seen that extensiveness is one of the significant limitations. Extensiveness is evaluated based on the number of optional parameters which have been filled in; however a modification in order to weigh not only the number of optional parameters filled in by the user but also the quality of the information included by the user can be implemented as future work. To undertake this new definition of this metric, it will be necessary to resort to another series of techniques that understand and assess the quality of the information included by analysing, among other techniques, the language used by the author of the CTI.

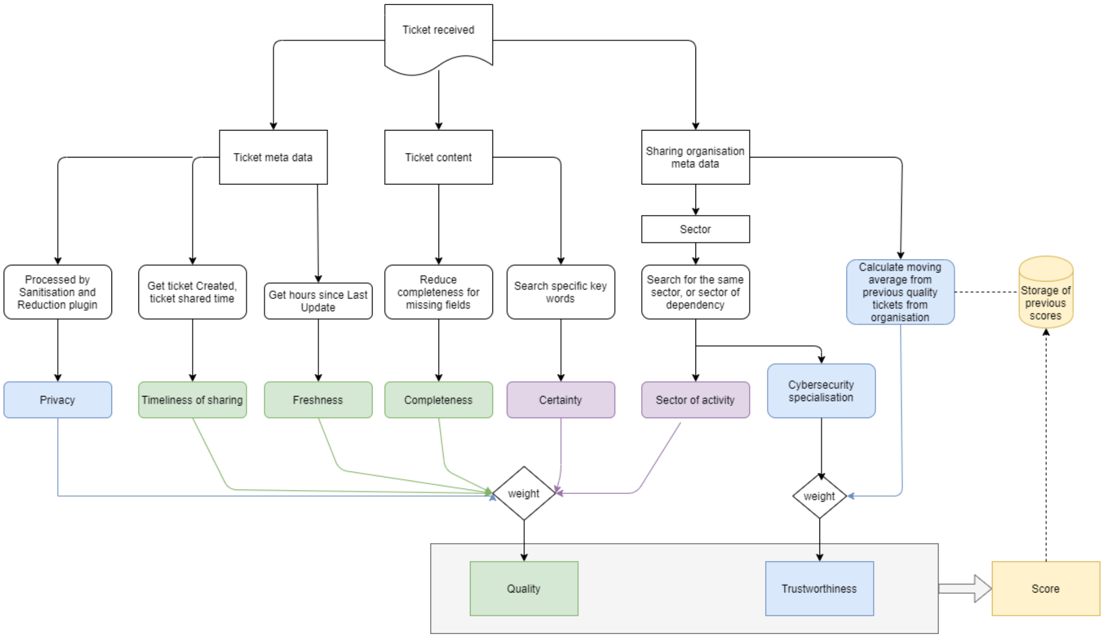
Ticket received
Ticket meta data
Ticket content
Sharing organisation meta data
Sector
Processed by Sanitisation and Reduction plugin
Get ticket Created, ticket shared time
Get hours since Last Update
Reduce completeness for missing fields
Search specific key words
Search for the same sector, or sector of dependency
Calculate moving average from previous quality tickets from organisation
Storage of previous scores
Privacy
Timeliness of sharing
Freshness
Completeness
Certainty
Sector of activity
Cybersecurity specialisation
weight
weight
Quality
Trustworthiness
Score

Fig. 2. Computational flow of a ticket

TABLE I
CTI TRUST AND QUALITY METRICS AND THEIR POSTULATED COMPUTATIONS

Acronym	**Metric**	**Description**	**How is it determined**
CTI.R.	CTI rating (=T&Q rating)	Based on reputation of the source (T) and the attributes of specific ticket/ warning (Q)	T + Q
T	Trustworthiness	Calculated as weighted sum of p.s.a., s.c., c.c., p.t.r., p., p.s.	
p.s.a.	Partner Sharing Activity	The number of CTI contributions may not be a direct indicator for trust. Nevertheless, the activity may signal the stakeholder whether someone is a free-rider or actively interested in a collaboration.	From partner's activity logs on the E-EWS.
s.c.	Sector of Activity	This is defined by the sector that the partner operates; e.g. competence centre, company specialising in the cybersecurity area	From partner's metadata when registering for the E-EWS system.
c.c.	Certified Cybersecurity	This parameter is defined whether the partner has a certification for cyber security; e.g. ISO 27001	From partner's metadata when registering for the E-EWS system.
p.t.r	Previous Ticket Ratings	This metric shows the way that past tickets of this source were rated. Accumulation of quality ratings of tickets is given by all organisations.	From aggregated feedback for all tickets shared by this organisation.
p.	Privacy	The metric is defined by an external tool to check if the content includes personally identifiable information	From ticket data.
p.s.	Partner Sector	The trustworthiness increases when two information-sharing entities belong to the same sector. The user can preselect the sectors he/she considers relevant/ is dependent on.	from ticket metadata.
Q	Quality	Calculated as weighted sum of t.c., t.f., t.t., t.e., t.r.	
t.c.	Completeness	Evaluation The metric shows how many mandatory parameters are filled in when sharing a ticket.	From ticket data.
t.f.	Freshness	The time passed since the information is shared; it decays as time passes.	From ticket data. Time between the day of calculation and the last activity date.
t.t.	Timeliness	Time between ticket created and ticket shared	Time between ticket created and ticket shared.
t.e.	Extensiveness	Evaluates how many optional parameters are filled in.	From ticket data.
t.r.	Relevance	For each type of alert source, a list of specific keywords is predefined along an associated relevance value. The relevance score is based on the number of keyword occurrences in the alert's field.	search keywords in the ticket using predefined tags.
R	Reputation	Reputation of the source can be considered as a collective measure of trustworthiness (in the sense of reliability) based on the referrals or ratings from members in a community.	Sum of Trust scores of a community.

In relation to the organisation opinion, this is an issue that has a certain sensitivity since rating a ticket by users is fine as long as the users are trustworthy when it comes to making the rating. An added improvement could identify users and rate them using an artificial intelligence engine that is capable of detecting bad ratings that are not true but have been made

TABLE II
COMPUTATIONS OF THE METRICS BEFORE AND AFTER ADJUSTMENTS

	Initial Computation	Adjusted computation
Completeness	Base Score (+50), At least one attachment (+25), At least one reference (+25)	Base Score (+40), At least one attachment (+20), At least one reference (+15), At least one facet(+25)
Extensiveness	Base Score(0), One Attachment (+25), Two Attachments (+50), Three Attachments (+75), Four or more (+100))	Attachments (Three or more attachments(+35), Two attachments (+25), One attachment (+10) References (Five or more references (+30), Three or four references (+20), Two references (+15), One reference (+10)) Facets (Three or more facets (+35), Two facets (+25), One facet (+15))
Freshness	less than 7 days: 100 between 7 days and 30 days: 80 between 30 days and 90 days: 60 between 90 days and 180 days: 40 between 180 days and 365 days: 20 more than 365 days: 0	the last 3 days: 100 between 3 and 7 days: 90 between 7 and 15 days: 80 between 15 and 30 days: 70 between 30 and 90 days: 60 between 90 and 120 days: 50 between 120 and 150 days: 40 between 150 and 180 days: 30 between 180 and 270 days: 20 between 270 and 365 days: 10 more than 365 days: 0
Quality	(0.2)*Completeness + (0.4)*Freshness + (0.4) Extensiveness	(0.2)*Completeness + (0.4)*Freshness + (0.4) Extensiveness

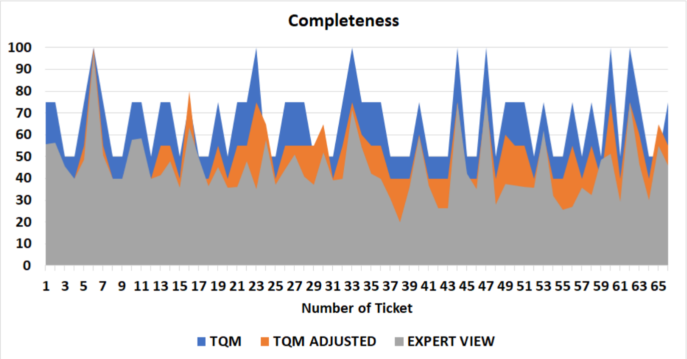

Fig. 3. Original and Calibrated Completeness Values with Expert's View on Individual Tickets

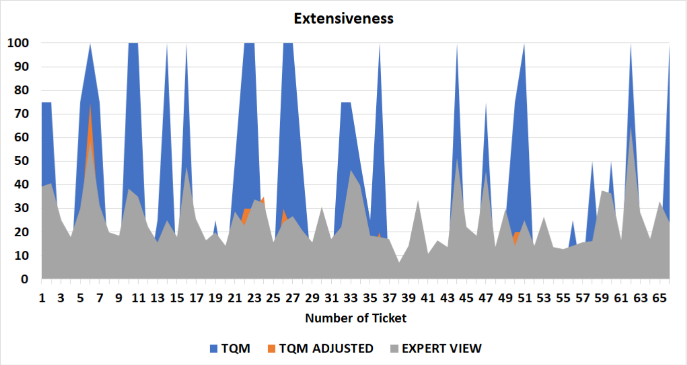

Fig. 4. Original and Calibrated Extensiveness Values with Expert's View on Individual Tickets

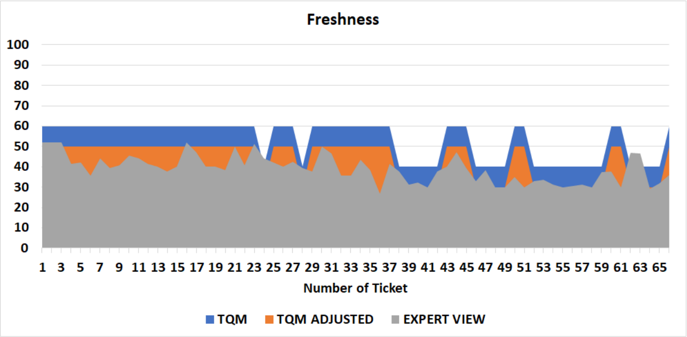

Fig. 5. Original and Calibrated Freshness Values with Expert's View on Individual Tickets

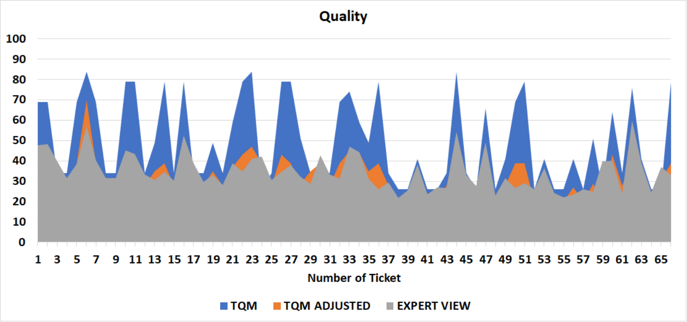

Fig. 6. Original and Calibrated Quality Values with Expert's View on Individual Tickets

with some kind of intention other than the quality of the CTI.

The calculation of freshness is based on the measurement between the current time and the last activity date. Using this metric on the Tabletop data has lead to the conclusion that since all the tickets had been generated several months ago, the metric indicated a low freshness. However, we believe that a possible improvement of this metric could be made by combining the current measurement with the subsequent relevance of the ticket. If a ticket is subsequently viewed by a certain number of users or used to consider the aspects which are reflected in the ticket; it could inspire possible new tickets. Based on that information, it would also indicate part of that freshness or timeliness of the information that the ticket contains.

In relation to the completeness metric, we realised that not only we had to measure whether the information or certain fields had been filled in by the user but also that this completeness should measure that the information is reflected in the appropriate section following the ticket structure defined. For example, the references should be added in the references field and not in the body of the ticket. We also had to modify the metric so that in the cases that the user had put the references in the body of the ticket, it was taken into computation by the tool, even if the user had not followed the defined structure.

The limitations of the prototype are rooted from the preferred metrics devised to measure quality and based on the effort to capture and formulate the different aspects of quality. Therefore, the next step for the development of this tool is to create complex and combined metrics so that TQM could automatically contrast the quality of the information entered by a user and that the quality score is more accurate and instrumental when displaying or sorting tickets within the E-EWS. Provision of this will be giving much more useful context information for users who view and analyse the tickets shared by different organisations in the E-EWS.

VI. Conclusion

In this paper, an evaluation tool for the quality and the trustworthiness merits of CTI is presented. The metrics and dimensions of the CTI are postulated and presented in a structured configuration. By leveraging sixty six intelligence items produced from four tabletop exercises to improve the capturing of the characteristics of these items, the results of the evaluation show that with the use of computed metrics, the overall trustworthiness and quality of CTI content can eventually be improved and contribute to the increase of the maturity of information-sharing within the E-EWS and(or) other information-sharing platforms.

Acknowledgement

This work has received funding from the European Union's Horizon 2020 research and innovation program under the grant agreement no 830943 (ECHO).

References

[1] R. Bourgue, J. Budd, J. Homola, M. Wlasenko, and D. Kulawik, "Detect , SHARE , Protect Solutions for Improving Threat Data Exchange among CERTs," *European Network and Information Security Agency (ENISA)*, no. October, p. 51, 2013. [Online]. Available: https://www.enisa.europa.eu/publications/detect-share-protect-solutions-for-improving-threat-data-exchange-among-certs

[2] ENISA, "Exploring the opportunities and limitations o current Threat Intelligence Platforms," no. December, p. 42, 2017.

[3] ——, *ENISA Threat Landscape Report 2018 15 Top Cyberthreats and Trends About ENISA Contributors Editors*, 2018, no. January. [Online]. Available: www.enisa.europa.eu

[4] C. Sillaber, C. Sauerwein, A. Mussmann, and R. Breu, "Data quality challenges and future research directions in threat intelligence sharing practice," in *WISCS 2016 - Proceedings of the 2016 ACM Workshop on Information Sharing and Collaborative Security, co-located with CCS 2016*, 2016, pp. 65–70.

[5] G. Grispos, W. Glisson, and T. Storer, "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*, 2019.

[6] T. Schaberreiter, V. Kupfersberger, K. Rantos, A. Spyros, A. Papanikolaou, C. Ilioudis, and G. Quirchmayr, "A quantitative evaluation of trust in the quality of cyber threat intelligence sources," *ACM International Conference Proceeding Series*, 2019.

[7] "Cyber threat intelligence – Issue and challenges," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 10, no. 1, pp. 371–379, 2018.

[8] ENISA, *Actionable Information for Security Incident Response*, 2014, no. November.

[9] S. Sadiq, *Handbook of Data Quality*, S. B. Heidelberg, Ed., 2013.

[10] F. Giubilo, A. Sajjad, M. Shackleton, D. W. Chadwick, W. Fan, and R. De Lemos, "An architecture for privacy-preserving sharing of CTI with 3rd party analysis services," *2017 12th International Conference for Internet Technology and Secured Transactions, ICITST 2017*, pp. 293–297, 2018.

[11] J. M. Fuentes, L. Gonzalez-Manzano, J. Tapiador, and P. Peris-lopez, "PRACIS: Privacy-preserving and Aggregatable Cybersecurity Information Sharing," vol. 69, pp. 127–141, 2017.

[12] T. D. Wagner, E. Palomar, K. Mahbub, and A. E. Abdallah, "A Novel Trust Taxonomy for Shared Cyber Threat Intelligence," *Security and Communication Networks*, vol. 2018, 2018.

[13] G. Fisk, C. Ardi, N. Pickett, J. Heidemann, M. Fisk, and C. Papadopoulos, "Privacy principles for sharing cyber security data," *Proceedings - 2015 IEEE Security and Privacy Workshops, SPW 2015*, pp. 193–197, 2015.

[14] K. Rantos, A. Spyros, A. Papanikolaou, A. Kritsas, C. Ilioudis, and V. Katos, "Interoperability Challenges in the Cybersecurity Information Sharing Ecosystem," *Computers*, vol. 9, no. 1, p. 18, 2020.

[15] "Tactics, Techniques and Procedures (TTPs) to Augment Cyber Threat Intelligence (CTI): A Comprehensive Study," Master's thesis, 2018.

[16] S. Mokaddem, G. Wagener, A. Dulaunoy, and A. Iklody, "Taxonomy driven indicator scoring in MISP threat intelligence platforms," 2019. [Online]. Available: http://arxiv.org/abs/1902.03914

[17] "Horizon 2020 PROTECTIVE project homepage." [Online]. Available: https://protective-h2020.eu/