A workflow and toolchain proposal for analyzing users’ perceptions in cyber threat intelligence sharing platforms

Borce Stojkovski  
SnT, University of Luxembourg  
borce.stojkovski@uni.lu

Gabriele Lenzini  
SnT, University of Luxembourg  
gabriele.lenzini@uni.lu

Abstract—Cyber Threat Intelligence (CTI) sharing platforms are valuable tools in cybersecurity. However, despite the fact that effective CTI exchange highly depends on human aspects, cyber behavior in CTI sharing platforms has been notably less investigated by the security research community.

Motivated by this research gap, we ground our work in the concrete challenge of understanding users’ perceptions of information sharing in CTI platforms. To this end, we propose a conceptual workflow and toolchain that would seek to verify whether users have an accurate comprehension of how far information travels when shared in a CTI sharing platform.

We contextualize our concept within MISP as a use case, and discuss the benefits of our socio-technical approach as a potential tool for security analysis, simulation, or education/training support. We conclude with a brief outline of future work that would seek to evaluate and validate the proposed model.

I. INTRODUCTION

There exists a broad consensus on the benefits of cyber threat intelligence (CTI) sharing among a diverse set of stakeholders and communities [1]. Furthermore, CTI exchange is believed to be an effective countermeasure to the growing number and sophistication of attacks in a number of different cyber security scenarios [2]. However, researchers and practitioners have highlighted that effective CTI exchange is complicated by a significant number of barriers and multidisciplinary challenges that entail a range of technical, organizational, legal, economical, and social aspects [3], [4], [5], [6].

While significant efforts from the research community have typically been directed towards technical aspects and challenges, such as defining CTI exchange formats, standards or automating CTI sharing, in recent years there has also been an increased interest around the human role in CTI sharing [7], [8]. For instance, researchers investigated the impact of extrinsic and intrinsic motivation on employees’ attitudes toward information security knowledge sharing intention and found that earning a reputation, gaining promotion as well as satisfying curiosity all had positive effects on employees’ attitudes, which in turn affected CTI sharing behavior [9].

User experience (UX) aspects have also been highlighted as very important in the context of threat intelligence sharing platforms, where getting UX design and human motivation right have been considered as pivotal success factors [10].

Furthermore, usability has been included as key evaluation criteria in recent frameworks for comparing the state-of-the-art in CTI sharing platforms [11]. Nevertheless, despite this acknowledgment on the importance of UX in the context of CTI sharing platforms, empirical evidence on their usability, or perceived UX is scarce to non-existent. Consequently, we lack UX insights into the enabling and constraining factors of security information sharing as well as how much effective CTI sharing is impacted by usability problems or UX challenges encountered by different information security workers.

Especially, we see a knowledge gap regarding users’ perceptions of key tasks pertaining to the consumption and use of indicators, their organization and storing as well as their production and publishing. While we find a number of questions around indicator prioritization such as the relative importance, perceived value, and actionability, worth exploring in this direction, in this paper we focus on the related problem of understanding users’ perceptions of the extent of information sharing. In other words, their understanding of how far does information that is shared in a CTI platform travel and who does it reach.

Having an accurate understanding of how far does shared information travel in a CTI platform can help towards ensuring agreements and rules of the sharing community are not violated, facilitate the prioritization of threat intel, or support the establishment of trust among the sharing community members. In particular, this is important to:

- avoid accidental leakage of sensitive information to entities beyond the intended recipients within a sharing community;
- avoid under-sharing i.e. to maximize the reach of shareable information with desired entities so that the members of the community can build a better situation awareness picture of the possible threats;

Consequently, our paper is motivated by the following research question:

RQ1 Do users of a CTI sharing platform have an accurate understanding of the extent of information sharing i.e. how far does information travel when it is shared in a CTI sharing platform?
While we do not perform and report a user study, in this paper we present a conceptual model of a workflow and toolchain, consisting of several technical and social components, that could be deployed to answer this question. We intend to conduct a subsequent evaluation and validation of our blueprint within a full-fledged user study, nevertheless, we believe that the potential of this socio-technical approach can already be recognized as a useful complement to analysis, simulation, or education/training efforts in CTI sharing.

II. CONTEXT AND RELATED WORK

Our model is inspired by similar work we did in the domain of secure email, where we sought to detect misalignments between system security and user perceptions i.e. identify situations where users might have a false sense of security or insecurity, potentially impacting the secure use and UX of the E2E email encryption system [12].

In both the secure email and in the current CTI sharing context, the fundamental building blocks are the same i.e. we need insights about both the technical aspects of a system and the users perceptions, opinions or behavior, that we take in for a joint analysis with respect to a specific question.

Thus, in order to answer the above-stated research question, we need to obtain and compare (i) users’ perceptions of how far information travels when shared in a CTI sharing platform, with (ii) the ground truth i.e. how far information travels in a CTI sharing platform in reality. This necessitates breaking down our motivating research question into the following:

RQ2 How can we obtain users’ perception (i.e. understanding) about how far information travels when it is shared in a CTI sharing platform?

RQ3 How can we obtain the ground truth i.e. how far does information that is shared in a CTI sharing platform travel in reality?

A. MISP as a CTI sharing use case

In order to ground our theoretical concept within a practical CTI sharing setting, we demonstrate how the workflow and toolchain could look like within the context of one leading CTI sharing platform i.e. MISP [13].

Since its inception within military circles, MISP has grown into one of the leading open-source sharing platforms used by over 6000 organizations worldwide1. MISP is also one of the most studied platforms [14], [15], [16], characterized as holistic and applicable in diverse scenarios as well as flexible considering the compatibility with different formats [11].

III. MISP SHARING AND EVENT REPRESENTATION

The core functionality of MISP is to enable consumption and/or contribution of information within a specific community of users. Thus, a MISP instance can be considered as an independent server (administered by a host organization) that facilitates this process among a defined set of participating organizations.

1https://www.misp-project.org/, accessed on April 19, 2021

As depicted in Figure 1, MISP instances can be standalone or interconnected between each other using different synchronization mechanisms, allowing for shared information to flow between instances in one or both directions.

New data entries in MISP are called event objects, which can be described with different levels of granularity of information as per the user’s wish [13]. Furthermore, the sharing model in MISP, which relies on voluntary action of its community to share information and indicators, allows those events to be shared under various scenarios [13].

Figure 2 provides an abstract representation of an event and its optional sub-components, whereas Figure 3 shows how a hypothetical event might look like in MISP.
Some of the information in the event core, attachments, attributes, and/or objects may be sensitive, thus ensuring that it is not shared with specific entities, groups of entities or sharing communities is of paramount importance. To this end, we propose the conceptual framework depicted in Figure 4.

IV. PROPOSED WORKFLOW AND TOOLCHAIN

A. Definitions

The first step involves the definition or formalization of a data model for a simplified CTI event representation. Thus, events and their components, properties, and sharing in MISP, could be more formally expressed as follows.

Distributions

\[ D = \{d_1, \ldots, d_n\} \]  \hspace{1cm} (1)

is a set of distribution options. For instance:

\[ D = \{\text{All communities, This Community, Connected Communities, Sharing Group, Your organization, Inherit Event}\} \]

Attributes

\[ a = (\text{data, metadata, distribution}) \]  \hspace{1cm} (2)

is an attribute that contains the data and related metadata that can be text, binaries, images, etc., and where \( \text{distribution} \in D \).

\[ A = \{a_1, \ldots, a_n\} \]  \hspace{1cm} (3)

is a set of attributes, which can also be an empty set.

Objects

\[ o = (\text{data, metadata, distribution}) \]  \hspace{1cm} (4)

is an object that contains the data and related metadata that can be text, binaries, images, etc., and where \( \text{distribution} \in D \).

\[ O = \{o_1, \ldots, o_n\} \]  \hspace{1cm} (5)

is a set of objects, which can also be an empty set.

Attachments

\[ t = (\text{data, metadata, distribution}) \]  \hspace{1cm} (6)

is an attachment that contains the data and related metadata that can be text, binaries, images, etc., and where \( \text{distribution} \in D \).

\[ T = \{t_1, \ldots, t_n\} \]  \hspace{1cm} (7)

is a set of attachments, which can also be an empty set.

Published status

\[ S = \{s_1, \ldots, s_n\} \]  \hspace{1cm} (8)

is a set of statuses applicable for an event denoting whether an event has been published or not, and how. For example:

\[ S = \{\text{Not published, Published, Published (no email)}\} \]

Events

\[ e = (\text{data, metadata, distribution, status, } A, O, T) \]  \hspace{1cm} (9)

is an event that consists of core information (i.e. the event data and metadata that can be text, binaries, images, etc.) with a distribution \( \in D \), and a published status \( \in S \). The event can contain zero or more attributes (each with individual distribution options), zero or more objects (each with individual distribution options), and zero or more attachments (each with individual distribution options).

\[ E = \{e_1, \ldots, e_n\} \]  \hspace{1cm} (10)

is a set of events, which can also be an empty set.

Fig. 4. Workflow for analyzing users’ perceptions in cyber threat intelligence sharing platforms
Entities

\[ N = \{n_1, \ldots, n_n\} \]  \hspace{1cm} (11)

is a set of entities i.e. organizations that are part of a sharing community.

MISP instances

\[ m = \{\text{admin}, N, G, E, \text{syncs}\} \]  \hspace{1cm} (12)

is an instance that consists of an admin organization that hosts and manages the instance, a set of entities N, a set of sharing groups G, a set of CTI events E, and a designated set of synchronization users that act as links with other instances.

\[ M = \{m_1, \ldots, m_n\} \]  \hspace{1cm} (13)

is a set of MISP instances.

B. System Master Model

While we could specify additional components depending on our research goals, the defined constructs above provide the underlying structure around which we can now build the rules, interaction and interdependency between the components. In other words, we can build the business logic of the CTI platform. Some examples of the business logic are:

• Every entity \( n \) belongs to one MISP instance \( m \in M \).
• There is a link between two entities \( n_1 \) and \( n_2 \) if and only if they belong to the same MISP instance \( m \), or the MISP instance of \( n_1 \) e.g. \( m_1 \), is directly or indirectly connected to the MISP instance of \( n_2 \) e.g. \( m_2 \) via sync users.
• Data from entity \( n_1 \) can travel to entity \( n_2 \) if they are linked.
• Data from \( n_1 \) cannot travel to another entity if the event distribution setting is This organization only.

The above statements are only for illustration purposes as we do not provide an extensive specification nor a more formal representation due to space constraints. Depending on the desired level of model replication of the CTI sharing platform, constructing the system master model can be a task of varying complexity and investment of resources. However, building the system master model is fundamentally related to RQ3 i.e. once we have the necessary ingredients to construct the master model, we should know what happens to CTI information that is shared based on the business logic of the platform.

C. Event Generation and Configuration

Once the theoretical master model that defines how data is shared within and between MISP instances is complete, we can generate CTI events with different configurations in order to see the effects of the different distribution possibilities within the platform. We propose three modes of event generation i.e. configuration, as displayed in Figures 5a, 5b and 5c:

1) Random Event Generator: “With a click of a button” the simulator generates a random event with a different number of attributes, objects, attachments, as well as a published status and different distribution settings. These are hypothetical, but valid combinations, representative of real MISP events.

2) Event Block Builder: Following the concept of drag-and-drop block programming we can construct a hypothetical event by combining the desired blocks, or simulating what-if scenarios by substituting targeted blocks which represent the different distribution and publishing settings. The detached example pieces in Figure 5b can be seen as interface components that allow for more controlled modifications to the event’s configuration in comparison to the Random Event Generator.

3) Replicator of existing MISP events: Existing events from MISP instances could be exported and automatically imported into the simulator in order to generate the tables of truth (described below) for real MISP events. Such events, could also be replicated manually using the Block Builder.

D. Entities of Interest

The next step involves the specification of entities relevant for our investigation. These can refer to the intended recipients of shared CTI information, or contrary, parties with which CTI should not be shared, thus the entities that might unintentionally have access to the shared data. For simplicity, we consider all data and metadata here to be sensitive or equally important and the overall purpose is to see who in the sharing community is able to see it i.e. who could it reach.

While one could aim to specify individual or very specific entities, the general application would be to analyze events against a bulk set of entities that share common characteristics e.g. organizations that are hosted on the same instance, organizations that are hosted on a connected instance, organizations in a specific sharing group, etc. These options should correspond to all valid distribution settings generated for the events and their subcomponents.

An investigation of specific entities, such as a particular organization, would require additional modeling or specification of the available entities and associations between them so that the system master model knows e.g. which organizations are hosted on the current MISP instance, which instances are connected to the current instance, etc.

E. Event Simulation

The next step is also linked to RQ3 and involves the automatic generation of the tables of truth which represent the actual i.e. factual distribution of information from the technical perspective as defined by the business logic or system master model. Figure 6 displays an example table of truth generated for the hypothetical MISP event from Figure 3. It can be regarded as the output of a query sent to the system master model with the generated event and entities of interest as parameters to that query. The simulator output, thus, tells us how data is shared in MISP based on the distribution and publication settings defined for that event i.e. which entities can see what information?

The table of truth can be constructed either (i) selectively, focusing only on the set of entities \( N \) that we are interested in, or alternatively, (ii) exhaustively, meaning all possible entities are considered, beyond the ones that we are interested in or that we had explicitly defined.
If we focus only on $N$, for each entity in the set, we can iteratively ask the simulator which data and metadata of $e$ and its components is it able to see. E.g.

- Can $n_k$ see $e.data$?
- Can $n_k$ see $e.metadata$?
- Can $n_k$ see $e.attribute1.data$?
- Can $n_k$ see $e.attribute1.metadata$?

We can similarly run queries for a bulk set of entities, e.g.:

- Can other organizations on current instance see $e.data$?

To illustrate, we can refer back to our hypothetical event from Figure 3. We can see that the distribution setting for the event is set to “This community only”, while the distribution setting for its only attribute is set to “All communities”. According to the MISP business logic, the most restrictive setting wins, thus the distribution of the event and its components is limited to “This community only”. Furthermore, the event is not published which further limits propagation to directly connected instances via pushing. Even without an overview of specific entities and associations between them, the system master model is able to respond to the queries exemplified earlier, and produce the appropriate table of truth in Figure 6.

| EVENT | Your organization on current MISP instance | Other organizations on current MISP instance | Other organizations on directly connected MISP instances (via push sync) | Other organizations on indirectly connected MISP instances |
|-------|----------------------------------------|-------------------------------------------|------------------------------------------------|--------------------------------------------------|
| Event Core Data | “Very important information” | ✓ | ✓ | x | x |
| Event Core Metadata | ID, UUID, Creator org, Creator user, Tags, Date, Analysis, Threat Level | ✓ | ✓ | x | x |
| Attribute Data | file.txt | ✓ | ✓ | x | x |
| Attribute Metadata | Attribute data, Category, Type | ✓ | ✓ | x | x |

Fig. 6. A table of truth indicates which entities can see what information. It is a summary of Y/N answers to queries about the CTI distribution specific to that event and entities of interest.
Having obtained the ground truth i.e. the technical aspects necessary for our comparison, we can now focus on the second crucial component.

F. Obtaining the user’s perceptions

Numerous user research methods could be deployed to investigate RQ2. Interviewing users, doing user observations or surveys could help us as a first step to understand how users learn about the sharing process in MISP. Is it via trial-and-error as they click through MISP or the numerous virtual machines made available for testing and training purposes? Do they read the documentation? Do they use the visual aids and widget in the User Interface that is supposed to facilitate the understanding of sharing? Do they follow a training session?

Any insights gathered could be helpful in coming up with initial assumptions about users perceptions that we would seek to validate in dedicated user studies. For instance, we could conduct a number of experiments wherein we could reuse the components from the previous steps. For a defined set of arbitrary events that have their unique components and distribution settings as well as a list of entities of interest relevant for the investigation, instead of running the simulator and queries, we could ask participants to tell us the extent of the data reach based on their understanding. Example prompt:

“Please have a look at the following event, its components and distribution settings, and indicate which of the components can be seen by the entities listed in this table?”

We can, thus, complete the table with user generated Y/N values similar to Figure 6 against which we are going to compare the values.

Depending on the study objective and format, such inquiries could take place in-situ while users are working on an actual MISP instance, or in an out of context investigation e.g. using a questionnaire administered online or paper-based.

G. Comparison

Having obtained both the table of user perceptions as well as the table of truth, we can perform an automatic check whether there are alignments or misalignments. This could be realized via direct comparisons of the outputs akin to methods comparing expected vs observed values.

V. DISCUSSION

A. Purpose

We see three main applications of such a socio-technical approach to comparing users’ perceptions (i.e. understanding) of how far information travels when shared in MISP (the user generated table) against what happens in reality (i.e. the table of truth).

1) Security Analysis and Audit: Our approach could assist organizations in the identification of specific misperceptions or misunderstandings among their staff members with respect to CTI sharing. For instance, studies could show that users from Organization A predominantly share events with less entities than supposed to, whereas users from Organization B fail to realize that they are sharing beyond their community only. Once such misalignments are identified, a subsequent automated investigation could be performed to get an estimate of the exposure or extent of such already exchanged CTI. Under the assumption that we can easily feed into the Simulator specific MISP events of interest, an automated audit could quickly highlight all the transactions (i.e. exchanged events) where a misalignment regarding specific entities exists between how data was shared in reality and how we model the users’ perception (even though we did not ask the users’ input for that transaction specifically).

2) Simulation: As indicated earlier, such an approach could be a useful simulation tool for projecting and experimenting how CTI sharing could be impacted by tweaks to the numerous distribution options and settings on the event-level, object-level as well as attribute-level. While the security analysis/audit aspect of the tool is geared towards events that were already shared, the simulation aspect is geared towards minimizing the negative impact of sharing future CTI events with wrong or suboptimal distribution settings.

3) Training: In addition to simulating and generating the table of truth for a specific event & entities combination, a number of other tasks could be performed. For instance, the inverse. Participants could see a filled-out table of truth, and would be asked to (re)construct an event with a possible sharing configuration that will correspond/satisfy the table values. Furthermore, we could ask participants to construct an event and choose a sharing configuration where the objective would be, for instance, to allow the maximum reach of the data, while making sure that certain “sensitive data” is not shared beyond the instructions.

B. Other applications

We believe that in addition to the purposes of comparing users’ perceptions to what happens in reality, as described above, the simulator could also be useful in the following scenarios:

- Verifying the correctness of the implementation in MISP i.e. checking whether information shared in MISP instances really matches the sharing specification of the theoretical model.
- Establishing the accuracy of the visual infographic / widget available in MISP that is supposed to facilitate users’ understanding of the different distribution options.

Verifying the correctness would require additional experiments in a test-bed with connected MISP instances. Similarly, investigating the usability of the widget would be a separate user study.

VI. FUTURE WORK

Despite being deeply rooted in a specific CTI information sharing platform, the applicability of the presented workflow and toolchain is limited by the theoretical nature of our work. Therefore, our intention is to instantiate the proposed model by obtaining the necessary social and technical components (RQ2 and RQ3) by means of:
• Looking at the available user documentation, conducting experiments in the MISP VM, and talking to lead MISP developers from the Computer Incident Response Center Luxembourg, in order to construct and validate a minimalistic theoretical master model that defines how data is shared within and between MISP instances.
• Conducting experimental user studies with existing MISP users, participants of MISP trainings, or prospective MISP users.

As our focus is on investigating the distribution, less attention is paid here on the actual content of these CTI events. One can consider all data and metadata to be dummy values and the purpose of our investigation would be to see whether a certain entity could see the dummy values. Nevertheless, for the purpose of identifying inadvertent disclosure or under-sharing, it is good to designate which dummy values should be considered as sensitive or important for the investigation. At the moment, this is left to the investigator to designate mentally or outside of the system which dummy values should be considered as such. The current approach is, thus, geared towards capturing the worst-case scenarios when a misalignment happens i.e. as if all data and metadata was sensitive, urgent, actionable, etc. In reality, not all data and metadata may be relevant and not all situations where a misalignments happen may be problematic. An extension of this work could be to designate inside the model which values are specifically important for the investigation at hand.

We believe that the master model of the system, along with additional modeling of the instances, entities, user perceptions, and other variables of interest, could be extended to a formal model. The application of model checking and formal methods could yield additional insights as well as helpful learning inputs to the participants as to where their understanding is wrong. What to consider and how to create such a formal model is left as an open question at this point.

VII. CONCLUSION

Effective CTI information sharing is complicated by a number of multi-disciplinary challenges, yet human aspects and behavior in CTI sharing platforms is largely unexplored. In this paper we propose a theoretical concept of a workflow and toolchain that seeks to verify whether users have an accurate comprehension of how far information travels when shared in a CTI sharing platform. Our socio-technical approach, presented in the context of MISP, argues to be helpful in the analysis, simulation and training efforts in CTI sharing. Validation of the proposed model would require performing additional technical and user research that we leave as future work.

ACKNOWLEDGMENT

Authors are supported by the Luxembourg National Research Fund through grant PRIDE15/10621687/SPsquared. We would like to thank the anonymous reviewers for their useful comments and suggestions on how to improve the paper.