Modeling Trust in Cyber Threat Intelligence within a Zero Trust Architecture

Mariam Wehbe
LIFO
INSA CVL
Bourges, France
mariam.wehbe@insa.cvl.fr

Laurent Bobelin
LIFO
INSA CVL
Bourges, France
laurent.bobelin@insa-cvl.fr

Sabine Frittella
LIFO
INSA CVL
Bourges, France
sabine.frittella@insa-cvl.fr

***Abstract*—Traditionally, network security has been based on the principle of segmentation, dividing the network into different segments or perimeters to secure resources. The security of computer systems based on network segmentation has a drawback as attacks can propagate within sub-networks. To mitigate this risk, Zero Trust (ZT) concept, giving no trust for any entity by default, has been adopted to evaluate the trust level of each communication flow based on the knowledge about various information such as user information, network security policies and Threat Intelligence sources. Cyber Threat Intelligence (CTI) is essential to detect cyberattacks. It identifies potential threats to a computer system. This research aims to formalize the trust in each piece of CTI information gathered to make decisions in cybersecurity.**

***Index Terms*—Trust, Multivalued Logic, Cybersecurity, Zero Trust, Threat Intelligence.**

I. Introduction

Distributed systems security, such as an enterprise or institutional network, is usually based on network segmentation. In such systems, the whole network is divided into different network segments where devices and users accessing it are supposed to have the same roles in the organization. Indeed, within a segment, anyone is implicitly granted the same trust and may have the implicit right to access the same resources. It means that any compromised resource within a segment is granted full access to this segment, so it can use any possible means to compromise other resources within the same segment.

Organizations are increasingly turning to Zero Trust Architecture (ZTA) to bolster their cybersecurity defenses. ZTA relies on the motto that no trust is implicitly given to any entity (either devices, users, or services) within the network. Zero Trust (ZT) adopts a proactive risk-based approach. It necessitates continuous authentication, authorization, and validation of users and systems, irrespective of their location, and it has a holistic identity-centric approach to safeguarding digital assets. Communication flows are allowed only if an entity named trust engine (TE) agrees to let this flow occur. This entity evaluates the trust it can have in the different entities involved in a flow based on its knowledge about the user, device, and involved service.

European Union, Marie Skłodowska-Curie grant No. 101007627.

An organization needs to identify real threats to defend against cyberattacks. Hence, the need to harness Cyber Threat Intelligence (CTI). It identifies potential threats to a computer system. Any organization with an IT system must be able to defend it against threats, detect and anticipate possible attacks, decide on countermeasures if there is an attack, apply them and evaluate their effectiveness.

Decision-making is modeled by the OODA loop (Observe, Orient, Decide, Act) as in [1]. During the Observe phase, CTI information can be gathered by agents or retrieved from threat intelligence sources like CTI platforms. This information is structured as flows coming from sources and is usually formatted in the form of a STIX ontology. We chose the platform OpenCTI [2], which is recommended by the French National Agency for the Security of Information Systems (ANSSI) and used to aggregate different flows into a single ontology. To help cybersecurity decisions, it is then necessary to estimate the trust in each information taking into account the various dimensions that can make up trust: reliability of the source, its competence, plausibility of the information, credibility of the information, for example. In this paper, we address how to model trust in CTI environment and aggregate information dimensions to compute a trust score that will help make cybersecurity decisions in ZTA.

The rest of this paper is organized as follows: section II-A gives an overview of the ZT architecture and then section II-B defines the CTI. Section III presents the problem statement, our contribution and the framework we have implemented. Finally, we conclude in section IV.

II. Context

A. Zero Trust

Zero Trust is a security model relying on the idea that perimeter-based security is inefficient when the so-called perimeter is breached. A user will likely compromise at least one of the resources enclosed within this perimeter, and by doing so, will compromise the whole system. It is then mandatory to never grant trust to other resources by default, as in perimeter-based defense. Therefore, Zero Trust model relies on this simple motto “Never Trust, Always Verify”.

ZT guidelines are derived from the motto. There is a kind of consensus on the set of guidelines to follow [3]:

- All network flows MUST be authenticated before being processed.
- All network flows SHOULD be encrypted before being transmitted.
- Authentication and encryption MUST be performed by the endpoints in the network.
- All network flows MUST be enumerated so that access can be enforced by the system.
- The strongest authentication and encryption suites SHOULD be used within the network.
- Authentication SHOULD NOT rely on public PKI providers. Private PKI systems should be used instead.
- Devices SHOULD be regularly scanned, patched, and rotated.

To comply with these guidelines, the National Institute of Standards and Technology (NIST) [4] recommended the logical architecture depicted in figure 1. It presents the Zero Trust Network in three separate planes: the data plane represented by the PEP, the control plane represented by the PDP, and the Policy Information Points (PIPs). In the figure, a subject identifies a user or an application that requests access to a protected resource (e.g., data or service) using a device that can host an agent (i.e., part of the ZT architecture) that will secure the asset and information provided by this device. This communication takes place via the data plane. Policy Enforcement Point (PEP) is often implemented as a gateway: it enforces decisions about whether or not to grant trust to a flow by the Policy Decision Point (PDP). The separation between PEP and PDP relies on the separation between the control plane (that makes decisions on how to handle the traffic) and the data plane (widely used in other domains such as Software Defined Networking (SDN)). The PDP retrieves data from the PIPs, such as information about the system state, users, deployed policies, threat intelligence, and other factors to verify subjects and their endpoints and decide whether to authorize access to a resource. In the NIST standard, PDP includes the Policy Engine (PE) component, which is the decision-making component. Some authors and companies add a TE component. TE is responsible for running a Trust Algorithm (TA), interacting with the different data sources to evaluate risk. In this case, PE makes its decision based on the risk evaluation returned by TE and the policies applied to the system. It is then not responsible for evaluating the risk per se. Google ZT solution BeyondCorp has pioneered the use of TE: it helps to maintain a lower complexity of the system policy by discarding edge cases and other unknown/unaddressed cases. The PIPs encompass Identity, Credential and Access Management (ICAM), Endpoint Detection and Response (EDR), Security Analytics, and Data Security systems. It retrieves information from PDPs and continuously assesses access to resources. PIPS also processes access requests sent by PDPs issuing approvals or denials.

B. Cyber Threat Intelligence (CTI)

ZT works with cyberthreat databases which are often based on ontologies such as openCTI. OpenCTI is a threat intelligence platform designed to help organizations collect, manage, analyze and share threat information. CTI relies on information gathering by an agent in charge of intelligence. This information is gathered from various sources and is produced by the CTI process [5]. This process is a cycle comprising five phases: planning and orientation, collection, processing, analysis and dissemination. One of the most classic types of information produced in the analysis phase is the indicators of compromise (IoC). A compromise occurs when an attacker takes control of a machine by introducing malicious software (i.e., malware). We are assessing the information gathered from the dissemination phase in particular.

III. Evaluating Trust

A. Problem Statement

An in-depth understanding of the specific risks faced by an organization is required before security measures can be taken. This will improve the security of information systems. Perceiving, detecting, and mitigating risks and taking appropriate security measures to improve organizations' security posture are the day-to-day tasks of the cybersecurity teams at a Security Operation Center (SOC). In cybersecurity, false positives security alerts (i.e., misidentifying legitimate activity as malicious) can lead to unnecessary alerts, while false negatives (i.e., failing to identify a genuine threat) can lead to missed attacks. Inaccuracies can lower trust not only in security tools but also in the analysts who rely on these tools to make critical decisions. The goal is to aggregate information automatically from various sources to help with decision-making. Such an automation is mandatory in some cybersecurity paradigms such as Zero Trust. It is a great help to leverage SOC analysts' alert fatigue caused by the flow of information they face. This helps by reducing the amount of information an analyst has to handle. To defend information systems against attacks, an organization relies on CTI information. This information will be used to defend against threats through either intrusion detection or the deployment of countermeasures (stopping/interdiction of data flows, address range bans, for example). Then, it is mandatory to estimate the trust one can have in each piece of information (IoC).

B. Contribution Overview

The trust estimation in a piece of information is usually done by considering various dimensions related to trust. The dimensions considered may be related to the information source (for instance its reliability and competency), plausibility with regard to the knowledge of the agent, or credibility of the new information when compared to other information already gathered. Then, the freshly gathered piece of information has to be aggregated with the previously available information to consolidate the knowledge of an agent.

Recent progress has been made in the theory behind trust modeling and decision-making. Many works focus on these approaches based on trust scales such as the one defined by NATO forces. Among the various models, the multivalued logic framework provides interesting features. In [6] [7] [8],

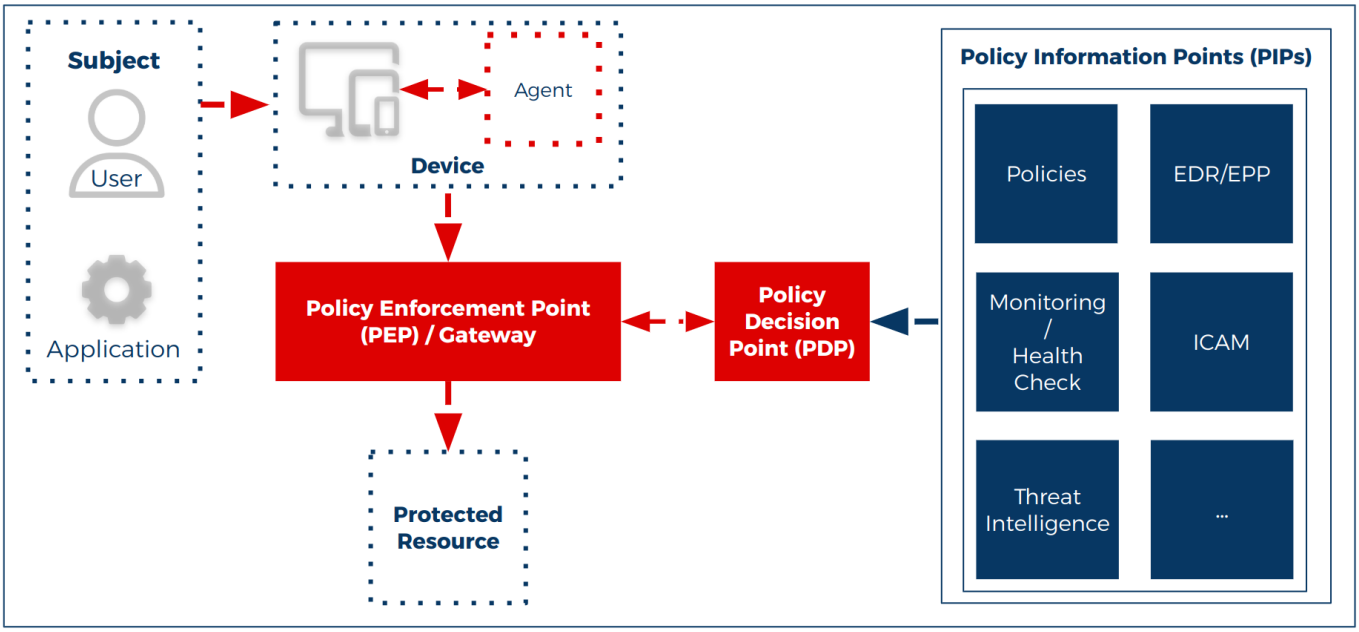

Fig. 1. Zero Trust Architecture

the authors defined a framework that provides the capability of handling unknown values for trust-building parameters or the seamless integration of dimensions. It takes into account various aspects linked to trust in the information, as well as the intrinsic uncertainty of the information itself. Information can be contradictory. Certain dimensions (reliability, competence, plausibility, credibility) may not be known. The information is assessed to yield the final degree of trust that helps make cybersecurity decisions. We are working on implementing their framework as a basis for a first experiment in this field. Section III-C describes the framework we use.

As the integration of trust in ontologies is a complex subject, we have chosen not to consider ontologies as a whole but only the simplest possible IoCs. We take into consideration the information received during the dissemination phase of OODA loop. Our work focuses on the stage of the trust-building process that computes trust in the aggregated information, considering the dimensions related to the trustworthiness of information.

C. Multivalued Logic Framework

The theoretical model of the multivalued logic framework represents the level of activation of each dimension with a single degree of a totally ordered set. Thus, it provides a multivalued logic framework with M truth degrees. For our experiments, we used degrees ranging from 0 to 4 as shown in figure 2. The 5-level scale is chosen according to the annex to STANAG 2022 for NATO forces [9] and to the research papers [6] [7] [8]. The NATO Admiralty code is also used by the platform OpenCTI to note the reliability of the intelligence sources and the credibility of information. Human operators are used to using a trust scale with a maximum of 5 degrees to describe the gradation of trust. The meaning of each degree depends on the dimension, as described in the figure 2. The computation of the trust integrates sequentially the mentioned dimensions as a process for evaluating the rating as illustrated in figure 3. The integration process follows a sequence that moves from the most general dimensions (e.g., the trust one has in the source and its expertise on the topic) to the most specific ones (e.g., compatibility with the agent's existing knowledge and corroboration by other sources). It assesses trust beginning with global factors and refining progressively. The sequential aggregation proceeds from left to right to calculate the trust score for a piece of information. The impact of the direction of the sequential influence of the dimensions on trust is represented by the multiple arrows, while the shaded disks represent the current level of trust. First, framework authors considered the source reliability as an a priori factor influencing the trust-building process. At this stage, the piece of information is assigned the level of trust of its source. Second, they integrated the competence of the source as an additional level of trust. Then, they aggregated it with the source score to update the trust score. Third, they assessed the plausibility of the information by determining whether it contradicts existing knowledge. Then, they aggregated the plausibility dimension to the recently updated score resulting in a revised trust score. We note that the competence and the plausibility lower the overall evaluation as illustrated by the downward-pointing arrows. And so forth, they consolidated their opinion by considering the credibility of the information provided elsewhere. They incorporated the credibility dimension into the recently updated score to obtain the final trust score. Unlike the previous steps that only lower the score, the final step allows for either an increase or decrease in the score

as illustrated by the downward- and upward-pointing arrows. For further details on the mathematical trust-building modeling of this framework, refer to the papers [6] [7] [8].

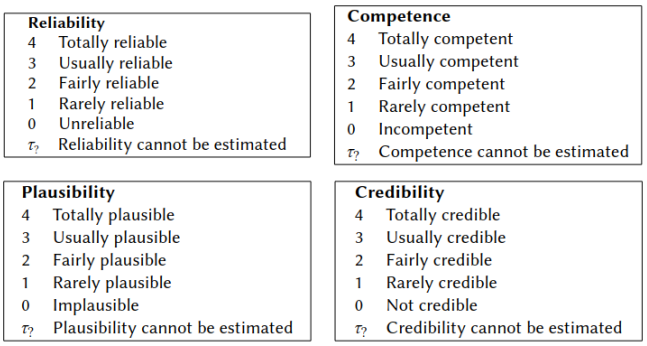

Fig. 2. Examples of degrees for the dimensions

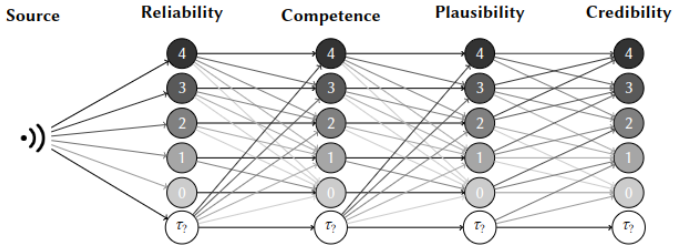

Fig. 3. Examples of degrees and sequential projection of dimensions of trust

IV. Conclusion

In this paper, we have described the ZT security model and the use of CTI information for decision-making in cybersecurity environments. We have outlined that this helps organizations improve their system's security and respond to real and perceived threats. For our initial experiment, we selected a theoretical model based on multivalued logic to formalize trust in the received information. This research aims to explore and understand the mechanism by which trust is built. Our objective is to assess the relevance of the implemented model through case studies and experiments. In future research, we aim to introduce this trust when aggregating more complex data models, such as STIX-type ontologies, and thus introduce this notion into CTI tools. To build on this work, we intend to validate the model by implementing it on real-world devices to assess its validity and performance in practical scenarios and within ZT environment.

Acknowledgment

Sabine Frittella's work is funded by ANR JCJC 2019, project PRELAP (ANR-19-CE48-0006). Mariam Wehbe's work is funded by Projet CyberINSA France 2030 ANR-23-CMAS-0019[1]. This work is part of the MOSAIC project funded by the European Union, Marie Skłodowska-Curie grant No. 101007627.

[1]https://cyberinsa.insa-cvl.fr/

References

[1] Clarke, R. & Knake, R. The Fifth Domain:Defending Our Country, Our Companies, and Ourselves in the Age of Cyber Threats. (Penguin Press,2019,7)

[2] Filigran Open CTI website. (2024), https://github.com/OpenCTI-Platform/opencti

[3] Evan, G. & Doug, B. Zero Trust Networks: Building Secure Systems in Untrusted Networks 1st Edition. *O'Reilly*. (2021)

[4] Johnson, C., Badger, M., Waltermire, D., Snyder, J. & Skorupka, C. Guide to Cyber Threat Information Sharing. (Special Publication (NIST SP), National Institute of Standards,2016)

[5] Cyber Threat Intelligence (CTI): Collection and Processing, *zvelo site internet*, 2020, https://zvelo.com/cti-collection-and-processing/

[6] A. R. d'Allonnes. Chapter 9 An Architecture for the Evolution of Trust: Definition and Impact of the Necessary Dimensions of Opinion Making. , 2017.

[7] A. R. d'Allonnes, M.-J. Lesot. Formalising Information Scoring in a Multivalued Logic Framework. *Information Processing and Management of Uncertainty in Knowledge-Based Systems - 15th International Conference, IPMU*, Part I: 314-324, 2014, doi: 10.1007/978-3-319-08795-5_33 https://github.com/wehbemariam/InformationScoring

[8] A. R. d'Allonnes, M.-J. Lesot. Dynamics of trust building: Models of information cross-checking in a multivalued logic framework. *2015 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE*, 2015, doi: 10.1109/FUZZ-IEEE.2015.7338121

[9] Wikipedia contributors. Admiralty code. *Wikipedia, The Free Encyclopedia*, 2025. [Online; accessed 1-April-2025]. Available: https://en.wikipedia.org/w/index.php?title=Admiralty_code\&oldid=1268492711