# Generating Threat Intelligence by Classification and Association of Security Events
## (research statement)

Cláudio Martins
LASIGE, Faculdade de Ciências,
Universidade de Lisboa - Portugal
claudio.dnm@gmail.com

Ibéria Medeiros
LASIGE, Faculdade de Ciências,
Universidade de Lisboa - Portugal
imedeiros@di.fc.ul.pt

## I. Introduction

In today's world, most of organizations are digital, operating with technologies and processes of the Internet era. The changes in IT infrastructure and usage models, including mobility, cloud computing, and virtualization have dissolved traditional enterprise security perimeters, creating a huge attack surface for hackers and other threat actors [1]. Not only the digital landscape has evolved, but there has also been a significant evolution in cyber threat, as adversaries have advanced their knowledge. They have deployed increasingly sophisticated means of circumventing individual controls within users' local environments and probed further into their systems to execute well-planned and orchestrated attacks [2].

One domain that has emerged during the past years is cyber threat intelligence (CTI or TI simply). This domain combines key aspects from incident response and traditional intelligence and can be defined as "the process and product resulting from the interpretation of raw data into information that meets a requirement as it relates to the adversaries that have the intent, opportunity and capability to do harm". However, compared to other cyber domains, such as incident response and security operations, CTI is still in the early adoption phase, limited by the lack of suitable technologies, known as threat intelligence platforms (TIPs) [1]. Despite organizations recognize the potentiality of CTI, the need for tools that would help them manage the collected information and convert it to actions and knowledge is preventing a mass adoption for this kind of solutions.

CTI information, under a form of open source intelligence (OSINT), can provide the knowledge required by security information and event management systems (SIEMs), which can be collected from many sources and by using TIPs. However, as SIEMs, TIPs receives thousands of security events, being hard to analyse them in order to extract relevant data about threats. According to recent surveys, the volume and quality of data are the most common barriers to effective information exchange [3], [4]. Also, most organizations cannot make valuable the use of their threat data because there is too much, approximately 250 to millions of indicators of compromise (IoCs) per day [5].

This work introduces a platform capable of classifying automatically OSINT provided from diverse sources and several threat categories and create relevant threat intelligence of quality. The goal is to provide a platform that uses machine learning classifiers to automatically recognize security events from several threat categories and classify them into the correct category. As a second goal the platform leverage from cluster and association algorithms to collect and generate new information representative of threats, even the new and sophisticate attacks. These new data can be integrated with SIEMs assisting them in finding ways to benefit from OSINT in order to increase their detection capabilities and reducing the number of false positives and false negatives.

## II. Threat Intelligence and Platforms

Threat intelligence (TI) can be defined as "evidence-based knowledge, including context, mechanisms, indicators (...) about an existing or emerging advice (...) that can be used to inform decisions" [6]. TI is useful to an organization only if it is actionable [7]. Its usage is growing in popularity and use amongst organizations according [8]. This trend followed the evolution of targeted attacks as they require a different level of response that is more specific to the organization [9].

TI is collected as a form of OSINT and processing this information currently is becoming an important task for obtaining cybersecurity threat information and protect organizations against such threats [10], [11].

Threat intelligence platforms (TIPs) were introduced with the purpose of collecting information about malicious threats and filling the industry standard gap in threat intelligence sharing. TIPs usually vary in objective, in scope of their action, and in their capacities. Despite their differences, TIPs have multiple advantages that enable organizations to easily bootstrap the core processes of collecting, normalizing, enriching, correlating, analyzing, disseminating and sharing of threat related information [12]. However, current solutions have some limitations that prevents theirs mass adoption. Such limitations are among of the processing of the quantity of TI that is collected, the technologies used for their processing and the resulting information, which lacks of quality [3], [4]. The missing of quality turns hard the work of the security

analysts since they need to analyze TI manually and try to extract useful information, which such tasks take some time. Some of these limitations that we consider for our work are presented next [3], [4].

- *Limited advanced analytics capabilities and tasks automation.* Most TIPs have limited capabilities related to aggregation and composition of security events, as well as the capability of de-duplicate, and automatically tag and classify data.
- *Shared threat information is too voluminous.* One of the problems is the overload of threat information shared via open source, commercial sources and communities. Combining shared threat information from different sources makes the relevant intelligence hard to find and makes it difficult to generate value out of it.
- *Limited technology enablement in threat triage.* There is limited technology enablement to facilitate the relevancy determination process. Currently, this process is done manually, in a complex way and dependent on the analyst.
- *Limited analysis capabilities.* Most TIPs have limited capabilities related to browsing, attribute-based filtering, advanced searched information, pivoting, exploration and visualization. Moreover, few platforms provide integration with third party tools that could help addressing these limitations.

## III. Generating Quality Threat Intelligence

We propose a platform to automatically classify events into threat sub- and categories, next associate such events by threat category, and then generate summarized data representative of malicious threats, i.e., for doing classification and association of events and represent them as a new data which characterize and identify attacks that are not possible by analyzing security events individually. The architecture of the platform is represented in Fig. 1. The main components are the following:

- *TIP:* A TIP to collect and normalize OSINT data provided from different and several source feeds, such as malware domains, vulnerabilities, blacklists.
- *Event Enricher:* The goal is to enrich the OSINT data with external data that does not come with OSINT, but can better characterize the treats involved in security events.
- *Attribute Analyser:* Each received event will be processed to identify key information to facilitate their analysis. The parsed information will be analysed to identify the type of attribute.
- *Attribute Classifier:* This component will classify each event attribute according to a common taxonomy (e.g., infoleak:automatic-detection="bitcoin-address").
- *Association Ruler:* Based on the classification of each attribute and using an associative ruler machine learning algorithm, will be given to each event its threat category.
- *Event Classifier:* Based on the output of the Association Ruler, this component will classify each event according to a common threat taxonomy (e.g. malware_classification:malware-category="Ransomware"), and then generate new TI by summarizing the group of events of each category in a a single security event.

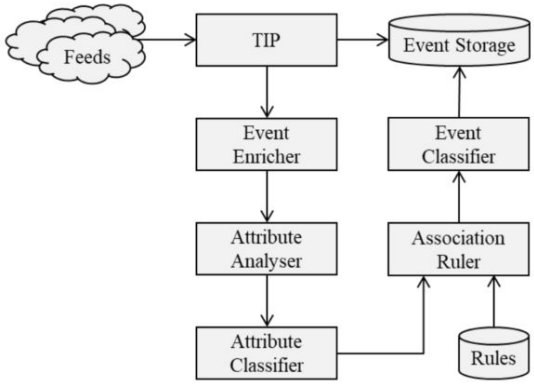


Fig. 1. Event classification and association platform architecture

We are currently developing a prototype of our platform. We use MISP TIP to gather and normalize the security events. To enrich the events (e.g., enrich URL malware domains with VirusTotal information) we developed python scripts that integrate with the core API of MISP. The classification and association of events are being implemented by us based on machine learning algorithms.

*Acknowledgements.* This work was partially supported by the EC through funding of DiSIEM project, ref. project H2020-700692, by national funds through FCT/MCTES (PIDDAC)/FEDER with reference to project AAC-2/SAICT/2017-029058 (SEAL), and LASIGE Research Unit, ref. UID/CEC/00408/2019.

## References

[1] S. W. Headquarters, "Advanced Persistent Threats: A Symantec Perspective. Preparing the Right Defense for the New Threat Landscape," 2011.
[2] SWIFT, "The evolving cyber threat," 2017.
[3] ENISA, "Exploring the opportunities and limitations of current Threat Intelligence Platforms," 2017.
[4] C. P., D. L., and H. C., "A Study on Advanced Persistent Threats." In Proceedings of 15th IFIP TC 6/TC 11 International Conference, Sept. 2014, pp. 63 – 72.
[5] W. Tounsi and H. Rais, "A survey on technical threat intelligence in the age of sophisticated cyber attacks," *Computer Security*, vol. 72, no. C, pp. 212–233, Jan. 2018.
[6] Q. Eijkman and D. Weggemans, "Open Source Intelligence and Privacy Dilemmas: Is it Time to Reassess State Accountability?" 2013.
[7] Webroot, "Threat Intelligence: What is it, and How Can it Protect You from Today s Advanced Cyber-Attacks," 2014.
[8] SANS, "CTI in Security Operations: SANS 2018 Cyber Threat Intelligence Survey," 2018.
[9] M. Bromiley, "Threat Intelligence: What It Is, and How to Use It Effectively," 2016.
[10] X. Liao, K. Yuan, X. Wang, Z. Li, L. Xing, and R. Beyah., "Acing the ioc game: Toward automatic discovery and analysis of open-source cyber threat intelligence," in *ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 755–766.
[11] C. Sabottke, O. Suciu, and T. Dumitras, "Vulnerability disclosure in the age of social media: exploiting twitter for predicting real-world exploits," in *In 24th USENIX Security Symposium*, 2015, pp. 1041–1056.
[12] S. Barnum, "Standardizing Cyber Threat Intelligence Information with the Structured Threat Information eXpression (STIX)," 2014.