Enhancing Information Sharing and Visualization Capabilities in Security Data Analytic Platforms

Gustavo Gonzalez-Granadillo[1] Mario Faiella[1] Ibéria Medeiros[2] Rui Azevedo[2] Susana Gonzalez-Zarzosa[1]
[1]Atos Research & Innovation, Cybersecurity Laboratory, Spain
{gustavo.gonzalez, mario-ferdinando.faiella, susana.gzarzosa}@atos.net,
[2]LASIGE, Faculty of Sciences, University of Lisboa, Portugal
imedeiros@di.fc.ul.pt, razevedo@lasige.di.fc.ul.pt

***Abstract*—Collecting and processing Open Source Intelligence (OSINT) data is becoming a fundamental approach for obtaining cybersecurity threat information and awareness. Different types of useful information and Indicators of Compromise (IoCs) are obtained from OSINT sources, which keep security analysts updated about new and possible threats against the IT infrastructures they protect. However, skimming through various news feeds is a time consuming process and a source of all kinds of information (sometimes unuseful and not related to the monitored infrastructure) for any security analyst. Based on these shortcomings, we propose a Context-Aware OSINT Platform as a tool for enhancing visualization and information sharing capabilities in security data analytic platforms. The tool is not only able to collect OSINT data, but also to process it and filter only the relevant parts, thus enriching the attributes of the detected data, and consequently, decreasing the amount of information and the time required to analyze and act upon.**

I. Introduction

In this technological computer connected world, systems that guarantee security in computer transactions and technological environments play a fundamental role. This is the main goal of the Security Data Analytic Platforms, which are also commonly referred to as Security Event and Information Management (SIEM). They provide real time analysis of security events generated by network and system devices. Such platforms are fed with high volumes of data (e.g., logs, alerts, events) coming from multiple and heterogeneous sources and process them on the fly. They normally have the capacity to collect, store and correlate events generated by a managed infrastructure.

Traditionally, these platforms are composed of several modules (e.g., source device, data collection, data normalization, rule engine, data storage, reporting), each performing a specific task that needs to work properly with the others, otherwise the entire system will fail [1]. Despite the benefits of all security data analytic platforms, current solutions are limited in terms of correlation, storage, visualization, and performance [2], [3].

In addition, these platforms generally use a static approach for threat identification, and this is not sufficient anymore to face new threats [4]. Integration among external sources and internal monitoring systems is essential for improving detection and reaction capabilities against these new types of cyber threats [5]. This consideration is valid not only when data is gathered from external sources, but also when internal knowledge has to be shared with other entities (e.g., trusted partners, public or private shared repositories).

We propose, in this paper a Context-aware Open Source Intelligence (OSINT) Platform, as an approach for enhancing visualization and information sharing capabilities in security data analytic platforms. The approach considers timely access to relevant, accurate and sensitive information about cyber-attacks and the possible incidents that may occur in the infrastructures we are protecting.

The remainder of the paper is structured as follows: Section II presents the proposed improved capabilities of security data analytic platforms. Section III introduces our context-aware OSINT tool and details its main components. Section IV presents a case study to illustrate the applicability of our approach. Related work is presented in Section V. Finally, conclusions and perspectives for future work are presented in Section VI.

II. Improved Capabilities of Security Data Analytic Platforms

Although current security data analytic platforms provide powerful features in terms of correlation, storage, and performance, most of them are very limited when it comes to collect, process, visualize and share Open Source Intelligence (OSINT) data [6]. This section details these limitations and provides some perspectives for enhancements.

A. Enriched Information Sharing Capabilities

In order to cope with current threats, it is important to have timely access to relevant, sensitive threat intelligence information. Sharing this information is not trivial. Threat intelligence must be expressed and, then, shared using specific standards, allowing involved parties to speed up processing and analysis phases of received information, which in turn will achieve interoperability among them. Many standards have been considered in the sharing process, however, the most used, and also the most promising ones are the Structured Threat Information eXpression (STIX[1]), for describing cyber threat information, and the Trusted Automated eXchange of Indicator Information (TAXII[2]), for sharing it in an automated and secure way [7]. Moreover, organizations starting to rely on so-called Threat Intelligence Platforms (TIPs), aiming at overcoming specific limitations of actual detection and monitoring

[1]https://oasis-open.github.io/cti-documentation/stix/intro
[2]https://oasis-open.github.io/cti-documentation/taxii/intro

tools, such as import and information sharing capabilities [5]. An interesting overview of these platforms, together with an exhaustive comparison, is described in [4], where a huge emphasis has been given to the Malware Information Sharing Platform (MISP)[3], especially for dealing with issues related to the establishment of collaborative environments, or simply for sharing data with internal security tools for further correlation.

Modern security data analytic platforms lack the ability to integrate relevant security data coming from public sources with data gathered from the infrastructure by sensors and other security devices. In this context, one potential enhancement will be the development of a component that considers all events occurring in the monitored infrastructure to provide a threat assessment for incoming OSINT data indicating its relevance and priority. This assessment will complement the usage of static information about the monitored infrastructure with dynamic and real-time threat intelligence data reported from inside the own monitored infrastructure in the way of Indicators of Compromise (IoCs).

Furthermore, the use of natural language processing techniques to identify threats from the use of keywords that typically indicate a threat in major languages; such as ddos, security breach, leak and more [8]. This information can be used to tag OSINT data as relevant or irrelevant. In addition to the type of threat, other information from the OSINT sources such as location and entities involved could also be extracted to provide a more comprehensive description of the threat. The prediction confidence of the classifier can be included in the data sent to SIEMs, which will help to avoid the issue of false alarms.

In addition, as information coming from OSINT sources can be duplicated and not correlated, an improvement for current security platforms could be the use of functions (or modules) to discard redundant information and correlate only those common data to facilitate the analysis and evaluation process. More specifically, before sharing IoCs, it is necessary to filter them (deduplication), correlate them (aggregation) and enrich them with useful information (e.g., threat assessment).

B. Enhanced Visualization Capabilities

During the reviewing of the state-of-the-art of existing security data analytic platforms, we observed that the reporting and data visualization capabilities are limited in terms of supporting effective extraction of actionable insights from the huge amount of data being collected by the systems. Although all platforms offer data visualization capacities, most of the visual representations are generic, not designed with particular user needs in mind, or too highly rudimentary to have any significant effect on how the data could be used or interpreted [4], [7].

In addition, existing systems do not have the capacity to use diverse data modes, e.g., statistical modeling outputs, OSINT data collections, or comprehensive models of user behavior. These novel data facets, when combined with the data already being gathered, offer challenges and opportunities for such security platforms.

[3]https://www.misp-project.org/

To enhance the visualization capability of current platforms, they must focus on flexible modules able to work with several data sources that carry heterogeneous characteristics, and with data that is under constant change, i.e., real-time streaming data. In addition, visualization must enable security analysts to better profile the system with novel representations that communicate the provenance of an attack, ongoing activities, vulnerabilities, and the characterization of sessions/users [9], [10].

To enhance the visualization capability of existing security data analytic platforms, we identify the following improvements:

- Design and develop a rich set of specialized visualization models that handle diverse types of data e.g., high-dimensional, temporal, textual, relational, spatial;
- Provide effective overviews, interactive capabilities to focus on details, and mechanisms to compare individual and/or groups of data instances;
- Design and develop visualization models capable of handling the dynamic nature of the data (e.g., streaming system activity logs, OSINT data, etc.) to support real-time analysis and decision-making; and
- Develop a visual summary of user activities that reveals common/abnormal patterns in a large set of user sessions, compares multiple sessions of interest, and investigates in depth of individual sessions [11], [12].

III. Context-Aware OSINT Platform

The Context-Aware Open Source Intelligence (OSINT) platform is able to correlate static and real-time information (e.g., Indicators of Compromise), related to the monitored infrastructure, with data coming from external OSINT sources through data fusion and analysis tools that check the relevance and accuracy of the data. The platform is composed of three main modules: (i) the input module, including the IoC generators, as well as infrastructure tools and devices that aggregate threat-related data; (ii) the operational module, composed of one MISP instance and a heuristic analysis process; and the output module, including the tool dashboard and connections to Security data analytic platforms (as seen in Fig. 1). As a result, this platform generates the following Indicators of compromise (IoC):

- **Composed IoCs (cIoCs):** is the result of the aggregation and normalization of OSINT data, retrieved from various feeds, expressed in different formats (e.g., plaintext, csv).

- **Enriched IoCs (eIoCs):** is the enriched version of a cIoC, obtained after the correlation of the latter with static and real-time information associated to the monitored infrastructure. The result of this process is a threat score (detailed later) that will be added as a new attribute. For this reason the word ”enriched” has been used.

- **Reduced IoCs (rIoCs):** is the reduced version of the corresponding enriched one. The latter could potentially contain a huge amount of information, not worthy to be visualized, but still useful for future analysis and correlation tasks. Threfore, only the rIoC, with just the

most relevant information from the monitored infrastructure point of view, will be sent to the dashboard, while the eIoC will be stored locally, or shared with external entities.

A. Input Module

This module is composed of two elements in charge of collecting data coming from OSINT sources and the infrastructure. Its main objective is to collect, clean, and aggregate data to feed the operational module with composed Indicators of Compromise (cIoCs) and other threat related data. The remainder of this section details the components of this module.

1) OSINT Data Collector: This component aims to generate cIoCs based on the aggregation of OSINT data. To do so, firstly, the component is configured with different types of OSINT feeds (e.g., malware domains, vulnerability exploitation) provided by several sources, such as free and collaborative organizations. Next, it collects data from these OSINT feeds, known as being events of security, and normalizes them. Normalization is required since OSINT data comes in various formats, such as plaintext and csv. Therefore, to process correctly the security events received, it is necessary that they should be in a common format (e.g., MISP format[4], or STIX), which is recognized by the component. However, distinct feeds can provide the same data, meaning that the component can receive duplicated data. To circumvent and avoid getting duplicate data, the component resorts of a deduplicator mechanism that compares the data received with the data already stored in the database, looking for security events equals to the received ones, and erases the duplicated ones. Afterwards, the component aggregates the security events by threat category, resulting in sets of events regarding a same category. In addition, within each set it looks for interconnections between events, correlating them by the establishment of connections of pair of events. The result of this correlation is sub-sets of events. Lastly, from these sub-sets are generated cIoCs, in which a single (composed) IoC is created from the correlated events.

2) Infrastructure Data Collector: Unlike the OSINT data collector, this component obtains information related to the monitored infrastructure that could lead to internal indicators of compromise (e.g., hashes, signatures, IPs, domains, URLs, etc.). In addition, this component will gather information of internal monitoring devices and operations from the infrastructure (e.g., installed applications, operating systems, threat actors, intrusion tools, vulnerabilities, etc.) that will be contrasted with the data coming form external sources in order to assess their corresponding risk level. This correlation process, between information received from external sources and cybersecurity related data detected with internal security tools, has been defined as a critical activity for obtaining relevant and actionable threat intelligence [13].

Both OSINT and infrastructure data feed the operational module for further processing and analysis.

B. Operational Module

This module is composed of two main components: (i) a MISP instance, in charge of gathering data from both OSINT-based sources and the monitored infrastructures, as well as sending the *enriched IoCs (eIoC)* to internal components, systems and tools or sharing them with trusted organizations; and (ii) a Heuristic component, in charge of performing the heuristic analysis, with the final aim of computing a Threat Score, enriching the cIoCs coming from the OSINT data collector, and obtaining the associated eIoCs. The reminder of this section details these components.

1) MISP Instance: The Malware Information Sharing Platform (MISP) is a free and open source threat intelligence platform used for gathering, sharing, storing and correlating IoCs of targeted attacks, threat intelligence, financial fraud information, vulnerability or even counter-terrorism information.

The MISP instance constituting the Operational Module of the Context-Aware OSINT Platform is composed of a collector entity (for both OSINT and infrastructure data), and a relational database to store locally information about IoCs and the monitored infrastructure. Relying on MISP, all incoming cIoCs will be automatically converted into their MISP format representation for being stored correctly. Then, thanks to specific export modules, they can be retrieved in various formats (e.g., MISP JSON, STIX 1.x and STIX 2.x).

The MISP instance interacts with the Heuristic component by sending the received cIoCs to be further analyzed. During the analysis process, a threat score is computed and added to the previously stored cIoC, converting it to a eIoC, and, optionally, shared with external entities. Considering that a huge amount of information could be included in an eIoC, just a filtered subset of information, identified as the most relevant from the monitored infrastructure point of view, could be sent to internal data analytic platforms, depending on the task that should be performed. This new IoC is called *reduced IoC (rIoC)*, and it also includes the threat score of the associated eIoC. Also data received from the monitored infrastructures, could be stored in the MISP database, in order to perform basic automated correlation steps, when some cIoCs are received, before performing the heuristic analysis. More details on the heuristic analysis is given in the next section.

2) Heuristic Component: This component has a heuristic engine that evaluates a set of attributes or features to produce a threat score that indicates the importance and priority to be given to the analyzed heuristic. Our tool is able to support different standards, thanks to the adoption of MISP. The set of heuristics will be selected depending on what standard is used for representing cybersecurity events.

The proposed Threat Score (TS) function is defined as the sum of all individual heuristic values (X_i) times its corresponding weight factor (P_i). This is based on multiple criteria (e.g., relevance, accuracy, timeliness, variety). The sum is then multiplied by the completeness parameter (C_p), as shown in Equation 1.

$$TS = C_p \times \left(\sum_{i=1}^{t} X_i \times P_i \right) \tag{1}$$

Where

C_p= Completeness criterion: $\frac{Non\_Empty\_Features}{Total\_Features}$

[4]https://www.misp-project.org/datamodels/

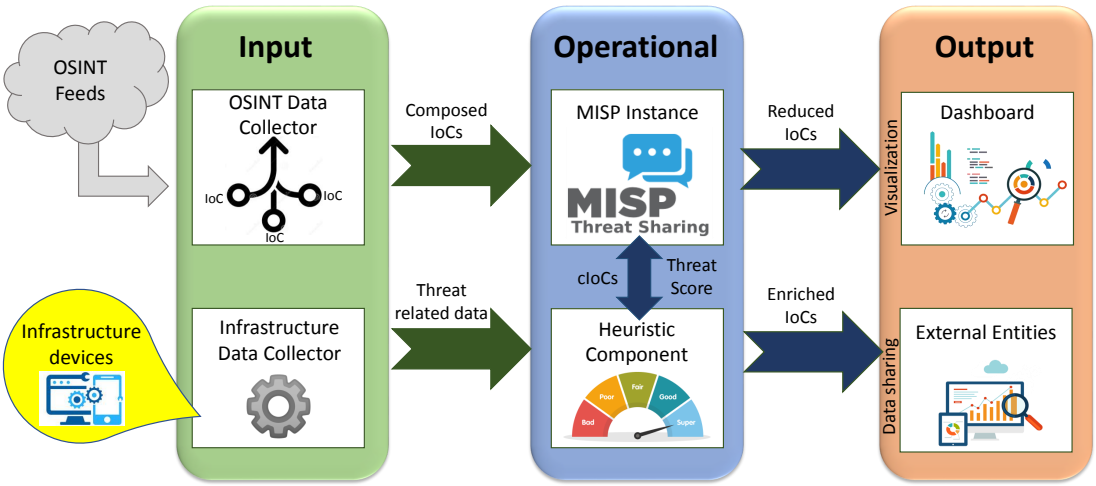


Fig. 1: Context-Aware OSINT Platform Architecture

X_i = Value assigned to a given heuristic's feature based on the information obtained from the IoC during the evaluation

P_i = Weighting Criteria

Considering for instance, three heuristics to be evaluated (i.e., H_1, H_2, H_3), each of which has information of five features (e.g., X_1, X_2, X_3, X_4, X_5), and considering that the features will be weighted as follows (P_1=0,10; P_2=0,25; P_3=0,40; P_4=0,15; P_5=0,10), the resulting threat score (TS) is summarized in Table I.

TABLE I: Example of a Threat Score Computation

	Features					
Heuristics	X_1	X_2	X_3	X_4	X_5	**Threat Score**
H_1	3	4	3	1	5	3,15
H_2	5	2	2	4	0	1,92
H_3	1	1	2	3	3	1,90

As depicted in Table I, features' values are always positive. Such values measure the reliability of the heuristic in identifying a given threat. The resulting TS ranges from zero to five ($0 \leq TS \leq 5$), the higher the TS value, the more reliable the IoC. Thus, as the TS value approaches to zero, the IoC can be considered as poor, incomplete and/or not reliable with a very low priority level.

a) Identification of Heuristics: Regarding heuristics identification, we considered the STIX 2.0 standard, defined as the de-facto standard for describing threat intelligence [7]. It defines twelve STIX Domain Objects (SDOs[5]) to represent each piece of information with specific attributes that are interrelated for a better understanding and more accurate details on the specific event they represent. Among the list of SDOs we have selected the following as the main heuristics to represent cyber threat information in our platform:

- Attack Pattern: type of tactics, techniques, and/or procedures describing ways threat actors attempt to compromise targets;
- Identity: individuals, organizations, or groups, as well as classes of them that could be involved in a security event;
- Indicator: contains patterns used to detect suspicious or malicious cyber activity;
- Malware: malicious code or software used to compromise the confidentiality, integrity, or availability of a victim data or system;
- Tool: legitimate software that can be used by threat actors to perform attacks;
- Vulnerability: mistakes in software that can be directly used by a hacker to gain access to a system or network.

For each heuristic, we identified a set of features that indicate valuable information on the identification of a threat. Examples of these features are provided in Table II.

TABLE II: Example of Heuristic's Features

Heuristics	**Features**
Attack pattern	attack_type, detection_tool, modified, created, valid_from, external reference, kill_chain_phases, osint_source, source_type
Identity	identity_class, name, sectors, modified, created, valid_from, location, osint_source, source_type
Indicator	indicator_type, modified, created, valid_from, external reference, kill_chain_phases, pattern, osint_source, source_type
Malware	category, status, operating_system, modified, created, valid_from, external reference, kill_chain_phases, osint_source, source_type
Tool	tool_type, name, modified, created, valid_from, kill_chain_phases, osint_source, source_type
Vulnerability	operating_system, source_diversity, application, vuln_app_in_alarm, modified, created, valid_from, valid_until, external reference, cve

b) Identification of the Weighting Criteria: considering that threat intelligence information must be relevant, actionable and valuable [14], we identified the following criteria to meet the aforementioned requirements:

- Relevance: evaluates if the information associated to a given feature is useful to identify a threat (e.g., no_info, optional, required);
- Accuracy: OSINT data will be compared to the information coming from the infrastructure to identify if there is a

[5]https://oasis-open.github.io/cti-documentation/stix/intro.html

match with one or more features (e.g., no_info, no_match, partial_match, full_match);
- Timeliness: it evaluates if a detected event is related to an already detected one, by the infrastructure or by the OSINT-based components, and if for instance, such events refer to the same threat, but with a different level of intrusion (e.g., no_info, unseen, unchanged, changed);
- Variety: evaluates the sources (infrastructure, OSINT) from where the information is originated or detected (e.g., no_info, single_source, multi_source, all_sources);
- Completeness: it is measured as the number of features with information over the total number of features from a particular heuristic.

In order to perform the heuristic analysis, a value must be assigned to each feature (e.g., from 0 to 5) based on expert knowledge and the usefulness of the criteria in identifying possible threats, malfunctions or anomalies in the monitored infrastructure.

C. Output Module

The Output module of our platform is mainly in charge of representing graphically the most relevant information contained in the eIoCs, that is the reduced IoCs produced in the operational module. Enriched IoCs can contain a great number of information that can reduce efficacy of the visualization process. In order to avoid such limitations, a reduced IoC (rIoC), composed of information related to the infrastructure, is used for this purpose. The eIoC could be, instead, shared with other external entities, which could be both internal security tools or trusted organizations. The reminder of this section details each component of the output module.

1) Dashboard: it provides a graphical representation of the infrastructure topology by highlighting the alarms and rIoCs associated to each node composing the infrastructure's network, as depicted in Figure 2.

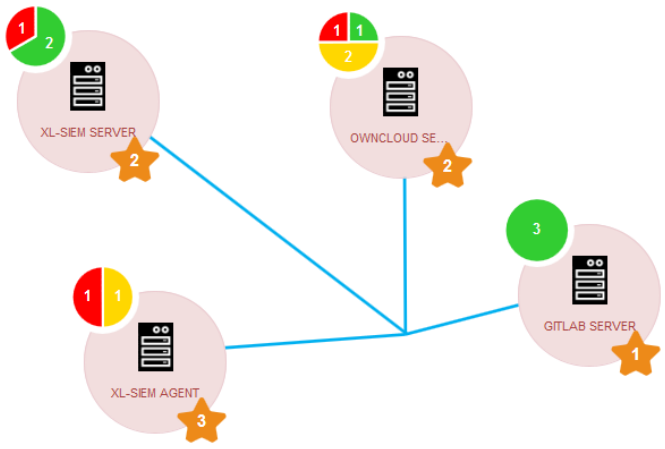


Fig. 2: Platform's Dashboard

Each node will have in its upper left side a circle indicating the number and severity of the alarms (in colors green, yellow and red), and in its lower right side, a star indicating the number of rIoCs related to that particular node.

Alarms will indicate the number of issues, IP source and destination, as well as a brief description of the issue. rIoCs will indicate the number of detected vulnerabilities, the CVE, the associated threat score, a brief description of the vulnerability and the affected application. A system inventory containing the nodes, and their installed applications is required to perform the match.

In addition, the dashboard provides, in a separate tab, information about the type of node (e.g., Server, Workstation); the IP addresses (known, unknown, source, destination); the operating system (e.g., Linux, Windows); and the connected networks (e.g., LAN, WAN).

2) External Entities: The exchange of eIoCs is performed through MISP, which automatically converts all the received information into the MISP JSON format and stores it in the MISP relational database. The JSON format is always used whenever two or more MISP instances are exchanging intelligence among them. However, when sharing with external entities that do not use MISP, as well as systems which are not able to directly handle the MISP format, the usage of other standards is preferable, also for describing a wider set of TI. From this point of view, STIX 2.0 represents a good choice, being the most used in TI domain [4]. MISP comes out with the possibility of exporting internal stored information using this specific standard. Moreover, the modules in charge to perform the conversion are extensible and can be adapted and improved depending on the organization needs, in particular if they need to develop their own custom export module, and add it to MISP.

After these considerations, the idea behind the Context-aware OSINT Platform is to rely on the MISP JSON format to store incoming events, due to the adoption of MISP. This information is then converted into STIX 2.0, if necessary for the analysis, and exported to the Heuristic Component. This last standard will be considered, starting from the heuristic features identification until the evaluation of the Threat Score, which, once computed, will be added to the original cIoC as a custom attribute. To improve the overall quality of the generated eIoCs, additional information associated to the criteria considered in the score evaluation could be used for the enrichment. Finally, the enriched indicator could be exported using both MISP format or STIX 2.0, depending on the receiver's needs.

IV. Use Case: Remote Code Execution

For this case study, we have defined an inventory of the infrastructure's network with nodes and the applications already installed. Every eIoC is checked against this information and, if there is a match, the rIoC is generated, associated to a specific node, and, finally, sent to the Output Module. If there is no match, the rIoC is not generated, while, if the match is with a common keyword (e.g., Linux), the new rIoC is associated with all nodes. Table III summarizes this information.

An Indicator of Compromise associated to a specific vulnerability (CVE-2017-9805) has been received, with information about a critical remote code execution in Apache Struts, which allows the attackers to execute arbitrary code into vulnerable field of POST request body. The severity

TABLE III: Infrastructure Inventory

Nodes	Names	Applications
Node 1	OwnCloud Server	ubuntu, owncloud, ossec, snort, suricata, nids, hids
Node 2	GitLab Server	ubuntu, gitlab, ossec, snort, suricata, nids, hids
Node 3	XL-SIEM Agent	ubuntu, snort, suricata, nids, php
Node 4	XL-SIEM Server	debian, apache, apache storm, apache zookeeper, apache struts, mysql, nessus, opennvas
All Nodes	-	linux

vulnerability is assessed as high, with CVSS[6] v3.0 equals to 8.1.

Since the STIX 2.0 IoC refers to the vulnerability type, we will analyze all possible features from this heuristic. Table IV summarizes all possible features, attributes, and scores that could be obtained from an IoC of type vulnerability.

TABLE IV: Features, attributes and scores associated to an IoC of type vulnerability

Feature	Description	Attributes and Scores
operating_ system	information about the affected operating system	windows (5), centOS, debian (3), others (1), unknown (0).
source_ diversity	IoC has been previously reported by OSINT or different sources	OSINT_source (1); No_OSINT_source (2); infrastructure_source (3).
application	Affected application identified in the IoC	browser (5), office (4), android (3), web (2), other (1).
vuln_app_ in_alarm	Check if incidents/alarms are related to specific applications	present(2), not_present (1).
modified/ created	Timestamp related to object creation/last modification	last_24h (5), last_week (4), last_month (3), last_year (2), other (1).
valid_from	From when the IoC can be considered valid	last_week (3), last_month (2), last_year (1), other (0).
valid_until	Until when the IoC can be considered valid	less_or_equal_to_current_date (0); greater_than_current_date (5).
external_ refer- ences	External references checked against a local inventory	multi_known_ref (5); single_known_ref (3); unknown_ref (1); no_ref (0).
cve	Check if CVE is found in the information provided by the IoC, and if so, check the CVSS	No CVE (0), CVE with no CVSS (1), CVE with low CVSS (2), CVE with medium CVSS (3), CVE with high CVSS (4), CVE with critical CVSS (5).

A. Information Sharing Process

In this subsection, the information sharing process is summarized from a technical point of view. Both OSINT data and Infrastructure Data Collectors send IoCs to the MISP instance of the Operational Module through a set of API provided by the latter. A specific open source library, written in Python, called PyMISP[7], exits to interact directly with the MISP platform.

An event coming from the Infrastructure Data Collector is simply stored internally and used later during the heuristic analysis for the threat score evaluation. The other events, instead, which come from the OSINT Data Collector, trigger a built-in automated, and real-time, sharing mechanism, based on the asynchronous messaging library zeromq[8], allowing the Heuristic Component to receive them, in STIX 2.0 format, and start the correlation with the stored infrastructure data. Once the threat score is computed, the eIoC is generated enriching the MISP JSON version of the cIoC, stored in the MISP database, adding the threat score as a new MISP attribute.

When performing the heuristic analysis, if a vulnerability associated to at least one asset of our monitored infrastructure is detected, this related information is extracted and used to build the rIoC, which will be sent directly to the Dashboard through specific web sockets, developed relying on the socket.io library[9], keeping the link to the associated eIoC stored in the MISP database.

The Dashboard is aware of the topology of the monitored network, and it can associate each rIoC to the potentially affected nodes, to visualize correctly the received data.

B. Preliminary Results

By contrasting the information of the IoC with the list of features presented in Table IV, we identified that the reported incident affects debian OS, it was first reported from OSINT, it affects the Apache Struts web application framework, there are no alerts from the infrastructure related to this application, the IoC was created and last modified on 2017-09-13, it is valid for one year, there exists external references from the Common Attack Pattern Enumeration and Classification (CAPEC[10]) and the Common Vulnerabilities and Exposures (CVE[11]).

Table V summarizes the assessment results associated to the IoC from a remote code execution. Please note that information about the date at which the IoC will be valid is missing, therefore, the valid_until feature is empty and discarded from our analysis. Each feature is assigned a heuristic value (X_i) that corresponds to the detected attributes and scores from Table IV. For instance, according to the description of the incident, the affected operating system is debian, therefore X_i= 3 for this feature, the CVE has a high CVSS, therefore X_i= 4 for this feature.

TABLE V: Threat Score Results

Feature	X_i	R	A	T	V	Total	P_i
operating_system	3	5	1	1	1	8	0.0952
source_diversity	1	5	1	1	1	8	0.0952
application	2	5	5	1	1	12	0.1429
vuln_app_in_ alarm	1	5	1	1	1	8	0.0952
modified/created	2	1	1	1	1	4	0.0476
valid_from	1	1	1	1	1	4	0.0476
valid_until	0	0	0	0	0	0	0.0000
external_references	5	7	10	1	5	23	0.2738
CVE	4	10	5	1	1	17	0.2024

In addition, each feature is affected by a weighting factor (P_i) composed of four criteria: Relevance (R), Accuracy (A), Timeliness (T), and Variety (V), described in Section III-B2b. The weighing factor criteria has been assessed based on expert knowledge, which determines the P_i vale as the total number of points associated to a given feature over the total number of points of all features.

[6]https://nvd.nist.gov/vuln/detail/CVE-2017-9805
[7]https://github.com/MISP/PyMISP
[8]http://zeromq.org/
[9]https://socket.io/
[10]https://capec.mitre.org/
[11]https://cve.mitre.org/

Following Equation 1, and using the scores values presented in Table V, we compute the threat score for the Remote Code Execution (RCE) IoC as: $TS_{(RCE)} = \frac{8}{9} \times \left(\sum_{i=1}^{t} X_i \times P_i\right) = 2.7406$

C. Analysis of Results

The Threat Score presented in Section IV-B is composed of two types of weights: (i) individual weights assigned to every attribute; and (ii) global weight (i.e., completeness criterion) assigned to the heuristic. The threat score is a positive value that considers static and dynamic parameters. The score ranges from zero to five ($0 \leq TS \leq 5$), the higher the TS value, the more reliable the IoC. Thus, as the TS value approaches to zero, the IoC can be considered as poor, incomplete and/or not reliable.

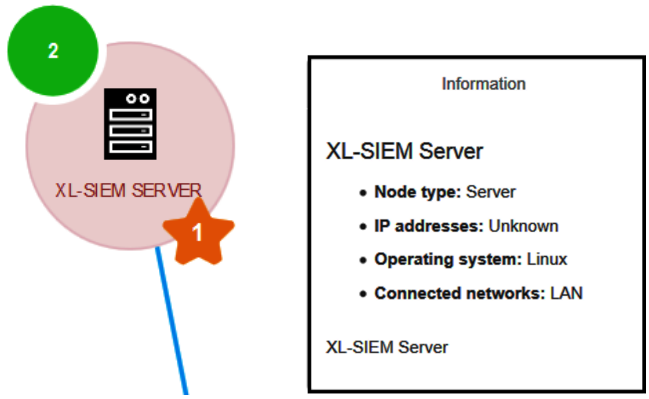
For the Remote Code Execution example, the resulting TS assigned to the IoC is equivalent to 2.7406, which places the relevance of this IoC in the average position. Priority for treatment of the risks associated to the detected risks should be given based on the TS value.

The TS will provide detailed information about relevance, timeliness, accuracy, completeness and variety properties of the evaluated IoC. This information is used by (i) SIEMs, as an input to develop new correlation rules in order to improve incident detection and response; and (ii) Security analysts, as the input to start rapidly and more accurately the reaction process based on the actionable threat intelligence data provided by the context aware intelligence sharing platform.

Please visit https://caisplatform.wixsite.com/english for more examples of Threat Score computation and details on the use case analysis.

D. Visualization of Results

Figure 3 depicts the visualization data related to the affected node in our proposed platform. Please note that the visualization data presented by the tool are related to the network topology, the assets that belong to the infrastructure, and the security data related to each asset (e.g., alarms, vulnerabilities, threat score data).



(a) Graphical Representation (b) Node Details

Fig. 3: Node Visualization Data

Figure 3(a) shows the affected node (i.e., XL-SIEM SERVER) with two associated alarms, both with low severity (green); as well as one enriched IoC. Figure 3(b) provides more details about the node (e.g., node type, IP address, operating system). This information is obtaining by clicking on top of the node graphical representation.

In addition to the information about the affected node, the platform's dashboard provides detailed information about the security issues affecting such node. Figure 4 details the two alarms and the CVE associated to Node 4 (XL-SIEM Server). The associated vulnerability is about a remote execution code, with a Threat Score of 2.7406 points.

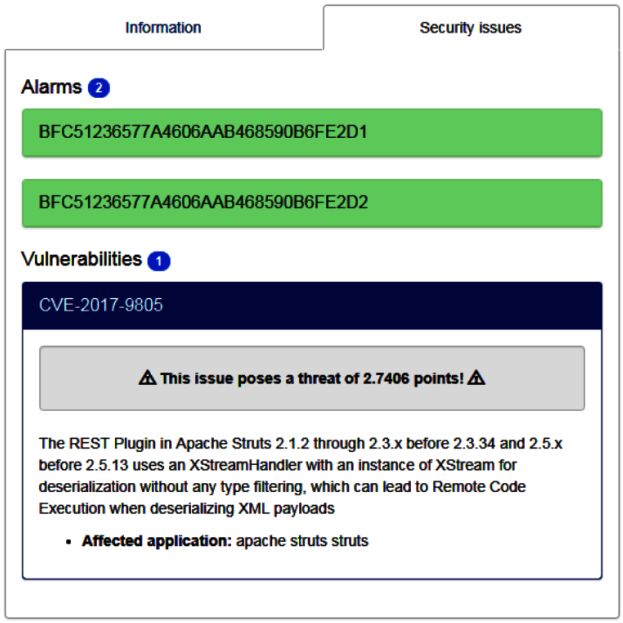


Fig. 4: Security Issues Detailed Information

As shown in Figure 4 a reduced IoC (rIoC) is generated by the tool, where information about the vulnerability (i.e., CVE, description, and the affected infrastructure) is provided. The threat score indicates the associated severity of this vulnerability which helps security analysts in their prioritization.

V. Related Work

Several companies have developed security data analytic platforms (e.g., SIEMs) aiming at detecting network attacks and anomalies in an IT system. Gartner reports [15] analyze the tools available in the market and provides a list of the top leading vendors (e.g., IBM[12], LogRhythm[13], Splunk [14], HPE[15], and Intel Security[16]) that provide products with good requirements behavior and have the foresight for future requirements.

[12] http://www-01.ibm.com/support/docview.wss?uid=swg27047107
[13] https://logrhythm.com/
[14] https://www.splunk.com/pdfs/technical-briefs/splunk-as-a-siem-tech-brief.pdf
[15] http://www.cybersecurity.my/mycc/document/mycpr/C076/HPE%20ArcSight%20ESM%20ST.pdf
[16] http://bluekarmasecurity.net/wp-content/uploads/2014/01/McAfee-WhitePaper-SIEM.pdf

Besides the great variety of commercial and open-source security data analytic platforms, current tools are unable to cope with the new and complex attack patterns. Most of them lack of capabilities to collect, process, store and use Open Source Intelligence (OSINT) data to identify, visualize and prioritize threats [3], [15]–[17].

Few tools have been proposed as a viable solution. Owen [18], for instance, proposes Moat, a powerful tool that covers known bad actors and consume data from multiple sources such as vulnerability systems and port scanners. Moat has been integrated with SIEMs using Structured Threat Information eXpression (STIX) and Extensible Markup Language (XML) formats for sharing purposes but it is not yet defined for other well-known standards such as the Trusted Automated eXchange of Indicator Information (TAXII).

In addition, some commercial SIEMs (e.g., LogRhythm) have added security intelligence to its SIEMs and analytic platforms. Their approach uses rich context enabled by threat intelligence from STIX/TAXII-compliant providers, commercial and open-source feeds, as well as internal honeypots. However, although the platform uses these data to reduce false-positives, detect hidden threats, and prioritize concerning alarms, more research is needed about threat intelligence sharing platforms, and their integration with other security tools

Our research differs from previous work as it introduces a platform that integrates OSINT data feeds with threat intelligence platforms aiming at enriching visualization and information sharing capabilities with aggregated and useful data that helps in the analysis and prioritization of threats and malicious events.

VI. Conclusions and Future Work

In this paper we present a Context-Aware OSINT platform as an extended visualization and information sharing capabilities in current security data analytic platforms. The proposed solution is composed of three main modules: (i) Input Module for the collection, normalization, and aggregation of IoCs and other threat related data coming from OSINT sources and the monitored infrastructure; (ii) Operational Module, able to compute a threat score that will enrich the evaluated IoC, and, in turn, will generate reduced and enriched IoCs, relying on a MISP instance; and (iii) Output Module for results visualization and sharing with internal and external entities.

Future work will concentrate in developing new features to enrich the threat score analysis, improving the quality of the refined threat intelligence to be shared, providing not only the final threat score, but also detailed information about each single criterion used in the evaluation of the score itself, which will be properly displayed through the dashboard, for improving the quality of the information available to the security analyst. Furthermore, visualization capabilities will be further enhanced to deal with the representation of a huge amount of alarms and rIoCs. In addition, the obtained results will be compared with other existing tools in terms of detection, false positive and false negative rates.

Acknowledgment

The research in this paper has received funding from the EC through funding of DiSIEM project, ref. project H2020-700692, NeCS project, ref. project H2020-675320 and LASIGE Research Unit, ref. UID/CEC/00408/2019.

References

[1] D. Miller, S. Harris, A. Harper, S. Van Dyke, C. Blask.: *Security Information and Event Management (SIEM) Implementation*, Mc Graw Hill, (2010).

[2] A. Barros.: *SIEM Correlation is Overrated*, Gartner Blog Network, (2017).

[3] K. Scarfone.: *Comparing the best SIEM systems on the market*, Online Research available at: http://searchsecurity.techtarget.com/feature/Comparing-the-best-SIEM-systems-on-the-market, (2015).

[4] W. Tounsi, H. Rais.: *A survey on technical threat intelligence in the age of sophisticated cyber attacks*, Computers and Security, vol. 72, pp. 212–233, (2018).

[5] Threat Connect.: *THREAT INTELLIGENCE PLATFORMS: Everything Youve Ever Wanted to Know But Didnt Know to Ask*, Technical Report available at https://www.threatconnect.com/wp-content/uploads/ThreatConnect-Threat-Intel-Platform-ebook.pdf, (2018).

[6] G. Gonzalez-Granadillo, S. Gonzalez-Zarzosa, M. Faiella.: *Towards an Enhanced Security Data Analytic Platform*, 15th International Conference on Security and Cryptography (SECRYPT), (2018).

[7] C. Sauerwein, C. Sillaber, A. Mussmann, R. Breu.: *Threat Intelligence Sharing Platforms: An Exploratory Study of Software Vendors and Research Perspectives*, International Conference on Wirtschaftsinformatik, (2017).

[8] E. Nunes, A. Diab, A. Gunn, E. Marin, V. Mishra, V. Paliath, P. Shakarian.: *Darknet and Deepnet Mining for Proactive Cybersecurity Threat Intelligence*, pp. 16, (2016).

[9] Esri.: *The Geospatial Approach to Cyber security: An Executive Overview*, White paper, https://www.esri.com/ /media/Files/Pdfs/library/whitepapers/pdfs/ geospatial-approach-cybersecurity.pdf, (2014).

[10] A. Slingsby, J. Dykes, J. Wood.: *Exploring uncertainty in geodemographics with interactive graphics*, Transactions on Visualization and Computer Graphics, 17(12), 2545-2554, (2011).

[11] P. H. Nguyen, C. Turkay, G. Andrienko, N. Andrienko, O. Thonnard, J. Zouaoui.: *Understanding User Behaviour through Action Sequences: from the Usual to the Unusual*, IEEE Transactions on Visualization and Computer Graphics, (2018).

[12] S. Chen, S. Chen, N. Andrienko, G. Andrienko, P. H. Nguyen, C. Turkay, O. Thonnard, X. Yuan.: *User Behavior Map: Visual Exploration for Cyber Security Session Data*, IEEE Symposium on Visualization for Cyber Security VizSec18, Germany, (2018).

[13] F. Skopik, G. Settanni, R. Fiedler.: *A problem shared is a problem halved: A survey on the dimensions of collective cyber defense through security information sharing*, Computers and Security, vol. 60, pp. 154–176, (2016).

[14] H. Dalziel.: *How to define and build an effective cyber threat intelligence capability*, In Syngress, eBook ISBN: 9780128027523, (2014).

[15] K. M. Kavanagh, O. Rochford, T. Bussa.: *2016 Magic Quadrant for SIEM*, Gartner Technical Report G00290113, (2016).

[16] K. Sheridan.: *Future of the SIEM*, Dark Reading, threat intelligence article, (2017).

[17] R. Caccia, O. Cassetto, B. Shteiman.: *The Future of SIEM*, International Information Systems Security Certification Consortium (ISC^2) Webminar available at https://www.brighttalk.com/, (2017).

[18] T. Owen.: *Threat Intelligence & SIEM*, Masters Research Project, Lewis University, (2015).