Data Sanitisation and Redaction for Cyber Threat Intelligence Sharing Platforms

Cagatay Yucel, Ioannis Chalkias, Dimitrios Mallis, Deniz Cetinkaya, Jane Henriksen-Bulmer, Alice Cooper
Bournemouth University,
Fern Barrow, Poole, Dorset,
BH12 5BB,
United Kingdom
Email: {cyucel, ichalkias, dmallis, dcetinkaya, jhenriksenbulmer, s4918533}@bournemouth.ac.uk

***Abstract*—The recent technological advances and the recent changes in the daily human activities increased the production and sharing of data. In the ecosystem of interconnected systems, data can be circulated among systems for various reasons. This could lead to exchange of private or sensitive information between entities. Data Sanitisation involves processes and practices that remove sensitive and private information from documents before sharing them with entities that should not be exposed to the removed information. This paper presents the design and development of a data sanitisation and redaction solution for a Cyber Threat Intelligence sharing platform. The Data Sanitisation and Redaction Plugin has been designed with the purpose of operating as a plugin for the ECHO Project's Early Warning System platform and enhancing its operative capabilities during information sharing. This plugin aims to provide automated security and privacy-based controls to the concept of CTI sharing over a ticketing system. The plugin has been successfully tested and the results are presented in this paper.**

I. Introduction

The advances in networks and other technological advances (e.g. advances which have led to more devices connecting to the internet) as well as the radical changes that society is experiencing (e.g. increased use of social media and increased remote working due to the COVID-19 pandemic) have resulted to the increased production and sharing of data. In the ecosystem of interconnected systems that constitutes the cyber realm, data can be circulated among systems with different considerations and policies on the issues of information sharing, processing and privacy. This could lead to exchange of private or sensitive information between entities; with the information reaching entities that do not (or should not) need to obtain it.

Data Sanitisation involves processes and practices that remove sensitive and private information from documents, before sharing them with entities that should not be exposed to the redacted content. At its early stages, when the data were displayed on paper, early sanitisation techniques started as printing a black opaque ink tape over the portions of texts during the declassification and dissemination of classified documents. The main concept has been preserved, even though it is also now being applied to digital information. In the context of cybersecurity, information may include critical and sensitive information regarding an organisation. The information of vulnerabilities, weaknesses and even enumeration of related cybersecurity incidents can be critical in terms of open-source intelligence. Moreover, sharing of cybersecurity incidents can be proven dangerous when including malicious data inside a Cyber Threat Intelligence (CTI) token. Uploaded and included documents must be scanned thoroughly before being shared with other organisations. As CTI sharing platforms are used by security experts, and ideally they are assumed to be trusted environments, these platforms can also be exploited for the dissemination of malicious data.

This paper presents the design and development of a data sanitisation and redaction solution for a CTI sharing platform. The Data Sanitisation and Redaction Plugin (DSR) has been designed with the purpose of operating as a plugin for the ECHO (European network of Cybersecurity centres and competence Hub for innovation and Operations) Project's Early Warning System (EWS) platform, and enhancing its operative capabilities during information sharing . The EWS is a platform for secured collaborative information sharing of cyber-relevant information [1]. This plugin aims to provide automated security and privacy-based controls to the concept of CTI sharing over the ticketing system. A ticket in the EWS encompasses CTI information with the creation of attributes and facets. These attributes can contain scripts, IP addresses, user credentials and malicious files and indicators. Before sharing a ticket on the EWS, the user of the system must ensure that private information and attributes are not included in the ticket. In addition, harmful scripts, executables, unauthorised downloads and injections can easily be overlooked. Hence, data sanitisation and redaction steps are required in the process, before sharing the tickets. Our proposed solution can successfully identify and isolate aforementioned threats and remove sensitive data.

The remainder of this paper is structured as follows. Next section presents the related work and background research. Section 3 introduces the DSR plugin and describes the main components. Section 4 includes the test results and evaluation of the study. Finally, conclusions and future work are given in the last section.

II. Related Work

Sanitisation and redaction of data have been under the microscope of the research community for more than two decades. Initially introduced as an answer to security and privacy risks associated to data mining techniques and machine learning algorithms that were used to extract sensitive knowledge from publicly available information [2]. Early research on data sanitisation discusses distributed transactional databases and addresses the problem of maximising the utilisation of data with the use of sensitive association rules where sensitive and private data are being redacted [3]. In the relevant literature, various privacy preserving models are presented towards data and/or knowledge protection through sanitisation [4], [5]. The existence of multiple sanitisation solutions (and the lack of commonly agreed practises can be considered another added weight to the fact that there is not a common solution for sharing the cyber threat data yet [6].

Cyber security solutions like SIEM tools, Firewalls, IDSs and IPSs have been significantly improved over the years offering now advanced endpoint, cloud control and protection. However, it has been proven that they can further upgrade their performance by taking advantage of free or paid Threat Intelligent Feeds [7]. Those are up-to-date streams of data related to current or potential security threats. Hence, like in any other exchange of sensitive data, the security and privacy aspects of sharing CTI have been also a concern. Related extensive research has taken place since 2015, especially with the development of Structured Threat Information Expression (STIX) [8]. The models that are integrating with STIX are developed and utilised in the research community as it is the de-facto standard for sharing CTI [9], [10].

In [11] and [12], security analysis of CTI sharing with third party analysis platforms over cloud are presented. Both studies address the studies conducted towards this aim in the EU project of Collaborative and Confidential Information Sharing and Analysis for Cyber Protection (C3ISP). Based on the C3ISP framework, where four pilot projects offered a wide range of use cases to test security and trust requirements prior to CTI sharing, [13] proposed a five-level trust model for a cloud-edge based data sharing infrastructure. The results, although slightly premature, showed that both data analytics and data sanitisation can be performed while testing several deployment models. Moreover, a blockchain-based network is prototyped and shown in [14] as a solution for CTI sharing platforms suffering from poorly defined trust barriers and data privacy issues between the partners.

A. Data Sanitisation under the scope of the GDPR

In May 2018, the General Data Protection Regulation (GDPR) [15] was enacted across the EU and with this, came an obligation on organisations to process data securely and to safeguard any private date, called "Personal Data" (PI) (Article 4(1), GDPR) by restricting processing and putting in place strong security protection around any sharing of data. GDPR centres around 7 data protection principles; data processing must be (1) fair, transparent and lawful; (2) data should be processed for a specific, explicit purpose only; (3) minimisation, only relevant data should be collected; (4) accurate; (5) storage limitation, stored only for period needed, ideally in a anonymised or sanitised format from the start; (6) Secure, ensuring the integrity and confidentiality of the data is maintained and safeguarded; and (7) Accountability, an obligation to integrate measures that facilitates being able to demonstrate compliance. These principles were devised to ensure only necessary and relevant data is processed and to safeguard the individual's whose data is being processed. This, in terms of the CTI sharing platform, will be achieved by ensuring that the sanitisation process will also remove or redact any PI from the collection of CTI tickets and warnings to ensure compliance with GDPR before this is uploaded to the EWS.

III. Description of the DSR Plugin

This section presents the high level overview of the proposed plugin. There are three main components of the DSR plugin.

- Sanisation Component
- Redaction Component
- EWS Core Communicator

The high level architecture of the plugin is shown in Figure 1. Further analysis produces eight components that describe the operation of the plugin in detail. These are explained below:

- **Core** – The core of the plugin is responsible for sending and receiving updates to and from the classes that the plugin is implementing. It is the main orchestrator of the whole software.
- **Parser** – With this class, the received ticket is formed into a Data Transfer Object (DTO); the parser supports functionality for the formats of JSON, STIX and IOC. The parser transforms the ticket into data that can be processed by the functions of the plugin.
- **DTO** – The DTO is the data structure that is transferred between classes. The DTO is formed by the parser when the Web hook is captured from the E-EWS core. Once formed, it is transferred to the sanitisation component. After the conclusion of the sanitisation process, the DTO is requested by the core and is delivered to the redaction stage where information will be redacted, if needed.
- **Sanitiser** – Once the DTO is sent from the core to the sanitisation component, the processes of the Sanitiser check and remove certain properties of a malicious object. The implemented operations of this sanitiser are a script checker and a downloader detector, while the functionality is complemented by the use of an external antivirus interface, which in that case is the Virustotal API.
- **Redactor** – This class applies the redaction/anonymisation techniques to the data. The Redactor receives the DTO and with a use of regular expressions it offers capabilities in removing private information (e.g. SSNs, emails, phone numbers, credit cards etc.),

according to the user's policies and decisions on redacting information.

- **Configuration Checker** – This class checks the configuration that the user chose before proceeding to redacting information.
- **Report Generator** – After the processes of sanitisation and redaction have been concluded, the core feeds the data to generate a report for the user to evaluate the redaction process, and also to feed vector data into the machine learning engine.
- **Machine Learning Engine** – The machine learning engine is utilising the data of the generated report in order to train the redaction and sanitisation process after every time that a ticket has been received and processed.

The communication between the classes of the DSR pulgin for the sanitisation and redaction processes are given in Figures 2 and 3 respectively. For the sanitisation process, the Plugin Core receives the message and exchanges it with the Sanitation Class in order to provide the necessary sanitisation process, as previously described. The redaction process requires the exchange of the DTO between Plugin Core and each one of the classes named after the components, i.e. the Configuration Checker, Redactor, Report Generator and Machine Learning Engine. For each one of the exchanges of the DTO, the classes return and update to the Plugin Core with the actions provided by them, according to their function.

A. DSR Plugin's Main Features

The DSR plugin can successfully identify and isolate/neutralise any of the following:

- Malicious Scripts (e.g. JavaScript, VBScript, HTML, Perl, C++, ActiveX and Flash scripts)in ticket attachments and attributes,
- SQL scripts that are included in tickets,
- Malicious URLs and links in a file that could result to downloading malicious files,
- Tickets that include malicious headers or parts of malicious code as evidence.

For the purpose of privacy protection and for securing sensitive data, DSR can remove information in tickets that could contain the following:

- IP addresses,
- Emails,
- Username-Password pairs,
- URLs,
- Date formats for the removal of information such as the Date of Birth,
- Number formats that is related to credit card numbers, phone numbers, SSNs, etc.

Depending on the needs of the users, different regular expressions can be applied to match the different criteria that analysts use and the requirements and specification brought by different occasions (e.g. the different structure of a Social Number or an address in each countries).

B. Functional Requirements of the DSR Plugin

Based on the aforementioned objectives, DSR has been developed under the scope of the following functional requirements:

FR1: The attributes of the tickets must be investigated for IP addresses to prompt the user.

FR2: The attributes and the ticket description must be searched for email patterns to prompt the user.

FR3: The attributes and the ticket description must be searched for SSN patterns to prompt the user.

FR4: The attributes must be searched for credit card number patters to prompt the user.

FR5: The users must be warned about date information such as Date of Birth to comply with privacy preserving regulations.

FR6: URLs must be avoided in a ticket unless they are a part of the CTI.

FR7: The users must be able to create new regex and patterns depending on their needs of CTI.

FR8: JavaScript, VBScript, HTML, Perl, C++, ActiveX and Flash scripts must be detected automatically and removed.

FR9: SQL Injections must be detected and automatically removed.

FR10: Fileless malware and malware downloaders, hidden in URLs and scripts, must be detected and automatically removed.

FR11: Malwares included in CTI should not be in an executable condition and should be neutralised in a way that the malware remains analysable.

FR12: The results of searches and sanitisations must be available to the users through a status update over tickets.

FR13: A mechanism for escaping scripts and texts should be provided when a script needs to be shared

The sanitisation component of the plugin aims to automatically remove and wipe malicious data from the cyber threat intelligence data, provided as DTO objects which are defined by the core of the warning system. The redaction plugin automatically redacts private data from the document and additionally trains the Machine Learning engine. The EWS Core communicator is an interface to the EWS Core for the Web Hooks and SDK integration.

C. DSR Plugin's Use Cases

As a result of the requirements and the development plans of the DSR, the following use case where designed and used for the evaluation of the plugin:

- A user that can check for stolen information.
- A user that can identify dates and redact these date data.
- A user that can redact according to the privacy preserving policies and regulations.
- A user that can the header of the malware to neutralise.
- A user that can detect a malware downloader.
- A user that can detect and sanitise XSS injections.
- A user that can detect and sanitise SQL injections.
- A user that can detect and identify scripts.

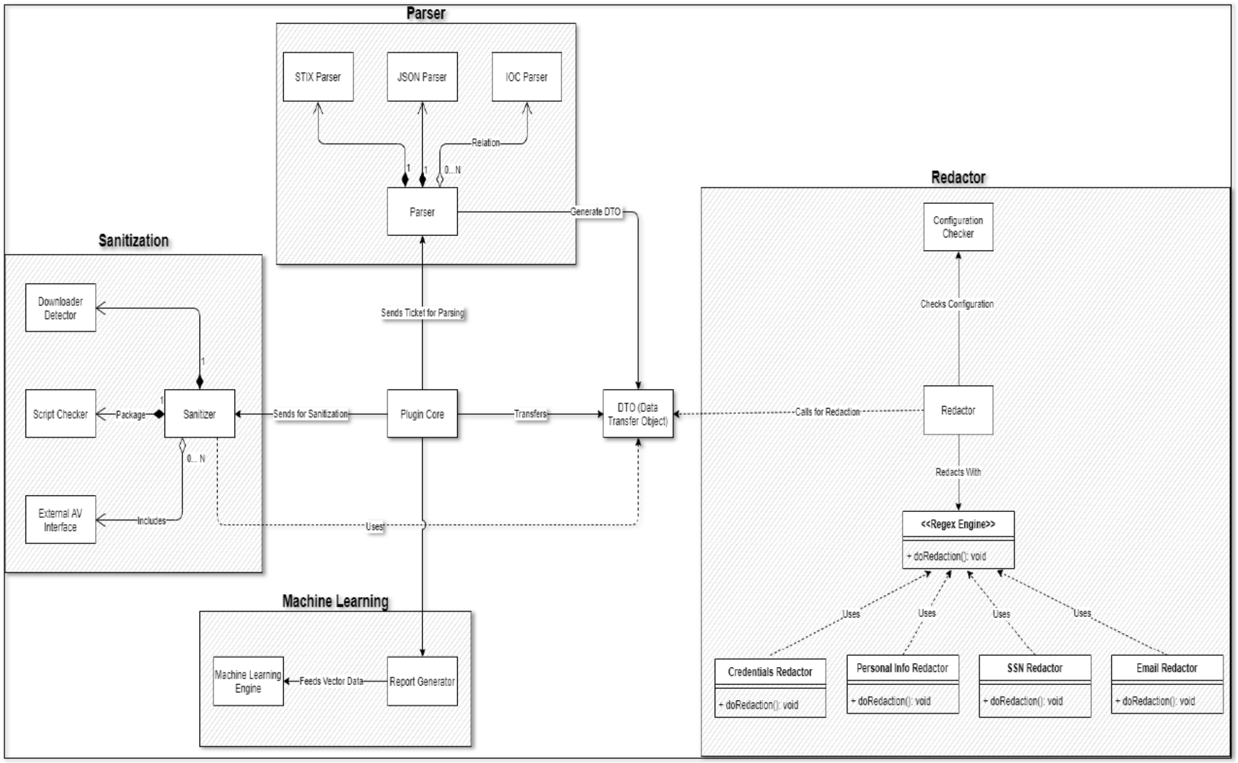

Fig. 1. Class Diagram of DSR

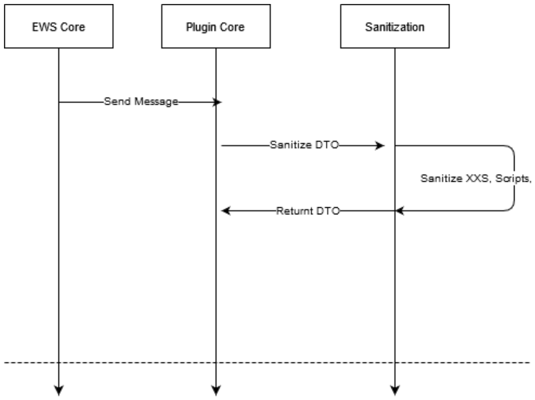

Fig. 2. Sequence Diagram for the sanitisation process of DSR

IV. Testing and Evaluation

For the preliminary testing of the plugin, a CTI dataset of 300 EWS tickets were created. During the creation of this dataset, XSS and SQL injection scripts were randomly selected from the following resources [16], [17]. In addition to this, to create Personally Identifiable information on tickets, the python library Faker version 3.0.1 was utilised [18]. These tickets were fed to the plugin as webhooks coming from the EWS system and each one of them was classified either “malicious” or “privacy threatening tickets”. These results are provided in Table I. Rigorous testing of the plugin which will be enhanced by machine learning classification methods will resume as a future work.

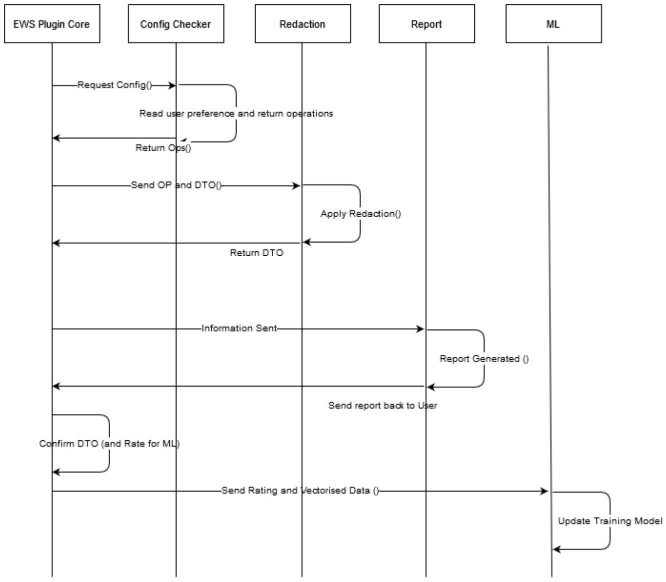

Fig. 3. Sequence Diagram for the redaction process of DSR

TABLE I
Preliminary Test Results for 300 Tickets

	TP	TN	FP	FN	Acc
Privacy Tests	100	188	12	0	0.96
Malicious Scripts Test	195	86	26	5	0.90

V. Conclusion

We presented a data sanitisation and redaction solution for sharing cyber threat intelligence data. Our solution has been implemented as a plugin to the E-EWS CTI ticket sharing platform in the ECHO project. This plugin provides an automated security and privacy preserving controls before sharing the tickets.

Acknowledgement

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

References

[1] E.C.H.O., "Project summary," https://echonetwork.eu/project-summary/, accessed: 2021-04-26.

[2] M. Atallah, E. Bertino, A. Elmagarmid, M. Ibrahim, and V. Verykios, "Disclosure limitation of sensitive rules," in *Proceedings - 1999 Workshop on Knowledge and Data Engineering Exchange, KDEX 1999*, 1999, pp. 45–52.

[3] A. Amiri, "Dare to share: Protecting sensitive knowledge with data sanitization," in *Decision Support Systems*, vol. 43, 2007, pp. 181–191.

[4] I. S. Alwatban and A. Z. Emam, "Comprehensive survey on privacy preserving association rule mining: Models, approaches, techniques and algorithms," *International Journal on Artificial Intelligence Tools*, vol. 23, pp. 1–29, 2014.

[5] I. Vakilinia, D. Tosh, and S. Sengupta, "3-way game model for privacy-preserving cybersecurity information exchange framework," in *Proceedings - IEEE Military Communications Conference MILCOM*, 2017, pp. 829–834.

[6] T. D. Wagner, K. Mahbub, E. Palomar, and A. E. Abdallah, "Cyber threat intelligence sharing: Survey and research directions," *Computers & Security*, vol. 87, p. 101589, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016740481830467X

[7] R. M. Lee, "2020 SANS Cyber Threat Intelligence (CTI) Survey," Tech. Rep. February, 2020.

[8] S. Barnum, "Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX™)," Tech. Rep., 2014.

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

[10] "PRACIS: Privacy-preserving and aggregatable cybersecurity information sharing," in *Computers and Security*, vol. 69, 2017, pp. 127–141.

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

[12] W. Fan, J. Ziembicka, R. De Lemos, D. Chadwick, F. Di Cerbo, A. Sajjad, X. S. Wang, and I. Herwono, "Enabling Privacy-Preserving Sharing of Cyber Threat Information in the Cloud," in *Proceedings - 6th IEEE International Conference on Cyber Security and Cloud Computing, CSCloud 2019 and 5th IEEE International Conference on Edge Computing and Scalable Cloud, EdgeCom 2019*, 2019, pp. 74–80.

[13] D. W. Chadwick, W. Fan, G. Costantino, R. de Lemos, F. Di Cerbo, I. Herwono, M. Manea, P. Mori, A. Sajjad, and X. S. Wang, "A cloud-edge based data security architecture for sharing and analysing cyber threat information," in *Future Generation Computer Systems*, vol. 102, 2020.

[14] D. Homan, I. Shiel, and C. Thorpe, "A new network model for cyber threat intelligence sharing using blockchain technology," in *2019 10th IFIP International Conference on New Technologies, Mobility and Security, NTMS 2019 - Proceedings and Workshop*. IEEE, 2019, pp. 1–6.

[15] European Parliament and the Council of Europe, "General data protection regulation (gdpr)," European Parliament and the Council of Europe, Brussels, Regulation (EU) 2016/679 5419/1/16, April 2016.

[16] Payloadbox, "XSS Payload List." [Online]. Available: https://github.com/payloadbox/xss-payload-list

[17] ——, "SQL Injection Payload." [Online]. Available: https://github.com/omurugur/SQL_Injection_Payload/

[18] Faker, "Faker for Python." [Online]. Available: https://faker.readthedocs.io/