AIL information leaks analysis and the GDPR in the context of collection, analysis and sharing information leaks

CIRCL Computer Incident Response Center Luxembourg

2018-06-16

Contents

AIL information leaks analysis and the GDPR in the context of collection, analysis and sharing information leaks

Introduction

The General Data Protection Regulation (GDPR) aims to reduce legal uncertainty and limit the interpretations by setting out clear conditions for the processing and sharing of personal data. It is applicable since May 25th, 2018 and applies regardless of whether the processing takes place in the European Union (Article 3). The GDPR distinguishes the roles and obligations of data controllers and data processors, provides definitions of personal data and establishes the conditions under which information can be processed and shared.

The Analysis for Information Leaks (AIL) is a modular framework to analyse potential information leaks from unstructured data sources such as Pastebin and other publicly available information sources. AIL is composed of modules which allow filtering the information leaks related to specific data categories. AIL has a set of default modules and custom modules can be developed. The output provided by AIL can be imported into MISP through ail-leak object and then shared with the MISP community. CSIRTs may also choose to publish information on the organisations affected by the information leaks in order to notify victims and raise awareness not only among other CSIRTs, but organisations, service providers and individuals as well. For example, CIRCL released and regularly update its TR-46 article in aiming at suggesting appropriate reactions of users of the service that leaked the information.

When using AIL, a user usually perform the three steps below:

1. collection (by crawling public websites and collecting unstructured data from sources such as Pastebin)
2. analysis (analysis of information collected via AIL framework) and
3. sharing (sharing information leaks with the ail-leak object in MISP).

This process is illustrated in the figure below:

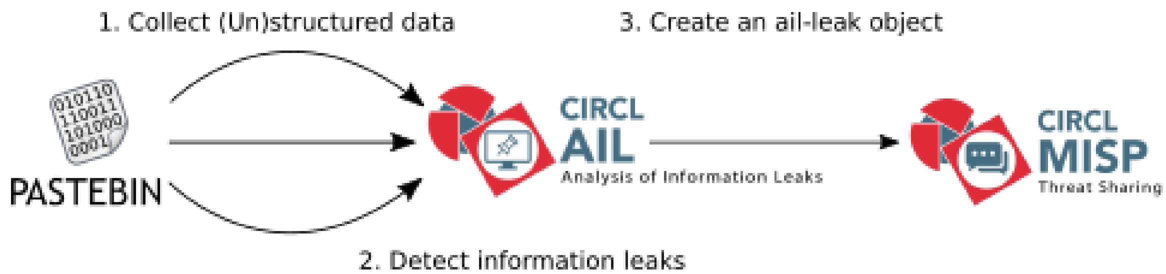

Figure 1: AIL processes of information leaks

FIGURE 1: STEPS OF THE PROCESSING OF INFORMATION LEAKS

What information processed by AIL is personal data?

Personal data is defined as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person" (Art. 4(1), GDPR).

Because AIL's main purpose is to collect and analyse information leaks, the data obtained through the tool is likely to contain personal data. Furthermore, although Pastebin is used by programmers and legitimate users, it is also commonly used by threat actors to disclose compromised sites, database dumps (e.g. user data, credentials and credit card details) and other personal information. Thus, the categories of personal data exchanged vary depending on the type of actors involved. The types of data collected, analysed and shared via AIL and via the MISP ail-leak object are not limited to the AIL default modules. AIL default modules may contain the following types of personal data (non-exhaustive list):

- Credentials (user names, passwords, etc.);
- Credit cards information;
- Email addresses;
- Phone numbers;
- API Keys (can be linked to an identity or account).

In addition, users of AIL have the possibility to create new modules and choose data sources analysed to identify potential information leaks. It is not impossible that these data sources contain personal data.

As highlighted in related work on the GDPR and information sharing, IP addresses can be considered as personal data, including dynamic IP addresses in certain circumstances. Hence, it is important to note that AIL may process IP addresses as they can be part of an information leak.

As previously explained, information leaks are shared through MISP via the ail-leak object. This object contains advanced combinations of attributes describing an information leak. Some of the attributes of ail-leak may contain personal data such as the following (non-exhaustive):

- "Raw data": the raw data as collected from the original source (e.g. Pastebin);
- "Text": contains a description of the leak, including potential victims' information.

It is to be noted that AIL may process special categories of data, as defined in the GDPR (Art. 9), i.e. data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation. For example, information posted on Pastebin may relate to political, religious or philosophical beliefs (i.e. from social media) and could be collected via AIL by users in their own custom modules. In such instances, the GDPR (Article 9) prohibits the processing of special categories except when specific conditions apply (e.g. if the processing constitutes a legitimate interest of the data controller provided that appropriate safeguards are applied to protect the data or for the reasons of a "substantial public interest, on the basis of Union or Member State law" etc.).

Who is the Controller when collecting, analysing and sharing information leaks via an AIL instance and an ail-leak MISP object?

The GDPR clarifies the differences in the roles and responsibilities of data controllers and data processors. According to Art. 4(7), the data controller "determines the purposes and means of the processing of personal data", either alone or in partnership with other data controllers ("joint controllers"). As highlighted in previous work, "When the peers decide to process the shared information (e.g. store, update, and integrate it in other systems), they become the data controller of the separate processing. Any peer having access to a piece of information is responsible for determining the purposes of processing activities which can include whether to share it or not to share."

Considering this interpretation, it is important to highlight that AIL is a framework used internally within organisations and is therefore not meant to be exposed or accessible to other entities. Hence, in most cases, the data controller is the operator of an AIL instance, i.e. the one who benefits from the AIL analysis, as it determines the purpose for which the information leaks analysis will be used and the means. These purposes should be determined based on legal grounds for processing personal data (which is further explored in the next section).

What are the grounds for processing personal data for the processing activities related to collecting, analysing and sharing information leaks?

The following paragraphs detail the legal grounds under which personal data can be processed for the purpose of information leaks collection, analysis and sharing.

Art. 6(1)(a) – Consent

Consent of the data subject is in many cases not feasible in practice and often impossible or illogical to obtain. The consent must be given by a statement or a clear affirmative action. Nonetheless, the data controller should be able to demonstrate it (Art. 7 Conditions for consent). In the context of information leaks, it is difficult:

- To identify the data subject victim of the leak. Even when the leak concerns email addresses or phones, usually the suspicion that the actual warning or consent email could be a phishing is too high for the data subject, and hence is ignored;

- To ask for the consent of each data subject targeted by information leaks as many people can be involved (e.g. 117 millions LinkedIn accounts leaks from a 2012 hack);

For some cases, data subjects that are victims of a data breach may be asked for consent by the entity that has detected the information leak prior to sharing their personal data with ail-leak object in MISP.

Art. 6(1)(c) – Legal obligation

This legal ground applies mostly to CSIRTs, in accordance with the powers and responsibilities set out in CSIRTs mandate and with their constituency, as they may have the legal obligation to collect, analyse and share information leaks without having a prior consent of the data subject. Some CSIRTs, depending on their contractual agreement may only be able to collect and detect information leaks and will not be able to share them with MISP.

Art. 6(1)(d) – Vital interest

The purpose of monitoring information leaks can hardly justify a vital interest. It is still possible to identify rare exceptions of information leaks that can threaten the life of individual (e.g. IP address list publicly available of critical machines vulnerable of remote code execution into a hospital). For more information or discussion on the use of vital interest for CSIRT, please refer to CIRCL GDPR workshops FAQ, specifically question 2 (Q2).

Art. 6(1)(e) - Public interest

Entities such as public CSIRTs can process personal data if acting under a specific mandate or delegation from an official authority to detect information leaks that could affect public interest.

Art. 6(1)(f) - Legitimate interest

Recital 49 explicitly refers to CSIRTs' right to process personal data provided that they have a legitimate interest and that such interests are not overridden by the fundamental rights and freedoms of data subjects. Collecting, processing and sharing information leaks constitutes a legitimate interest specifically for CSIRTs, as it is aligned with the purpose and scope of most CSIRTs mandates. Indeed, collecting, detecting and sharing information leaks will enable CSIRTs to better prevent and mitigate attacks by, for example, identifying credentials leaks that can be used to connect to critical system.

Does the GDPR allow CSIRTs to collect, analyse and share information leaks through AIL and MISP ail-leak object?

As noted above, the GDPR enables information exchange of personal data as long as it is performed for the purposes of ensuring network and information security or if it constitutes the legitimate interest of the data controller (e.g. preventing unauthorised access to sensible machine after credential leaks) (Recital 49). A processing activity should comply with the six principles in Art. 5, which could be summarized as: "lawfulness, fairness and transparency", "purpose limitation", "data minimisation", "accuracy", "storage limitation".

In most cases when collecting, analysing and sharing information leaks, the information leaks come from stolen data, hence the processed personal data has not been obtained from the data subject. In this case, Art. 14 triggers the application of the transparency principle. This article requires that specific information, such as identity and contact details of the controller, is provided to the data subject. However, it can be difficult to provide such information to the information leak victims as the list can be very long. The GDPR has foreseen such use cases where consent may be difficult or impossible to obtain and provides some exceptions. Article 14(5)(b) is of particular relevance to AIL usage cases, as it states that Art. 14(1) to (4) shall not apply if "the obligation referred to in paragraph 1 of this Article is likely to render impossible or seriously impair the achievement of the objectives of that processing". This restriction needs however to be balanced with "appropriate measures" such as "making the information publicly available". For example, CSIRTs could make information about their processing activities publicly available in line with RFC 2350 and GDPR Art. 14(1) and (2).

As AIL is a modular platform, processing modules can be removed, modified or added into the platform. For instance, removing the Credit Cards module deactivate the detection of credit card information leaks. In addition, when creating an ail-leak object, it is possible to select only necessary personal data leaked to share in order to provide data minimisation and purpose limitation.

By default AIL will keep all information received which can contain personal data. However, the AIL project includes a script to remove all of the information inside the platform. This script needs to be launch manually but can be automated and enable storage limitation. Furthermore when sharing information leaks with ail-leak object, MISP provides features to keep the accuracy of the attribute. By creating a "proposal notification", it is possible to propose a modification for each attribute into MISP. A rating system called "sightings" is also implemented into MISP to confirm whether an attribute is valid or not.

Information stored into AIL cannot be modified through the user interface once received, meaning that its integrity cannot be altered. In addition, MISP contains user roles and an approval feature to restrict modifications of an ail-leak object further ensuring the integrity of personal data. Information present in the AIL user interface is publicly accessible. However, it is possible to restrict the access with network configuration (e.g. IP address filtering) for the information confidentiality. An authentication system into MISP is also implemented to filter access to the list of events that can contain ail-leak object.

Conclusion

The benefits of the AIL project and the collection, analysis and sharing of information leaks via the ail-leak object in MISP are in support of threat intelligence. On one hand, discovering additional leaks in a timely manner allows CSIRTs to improve their incident response time. On the other hand, an early warning of an information leak incident to the affected parties and constituents can help them take action sooner to mitigate the impact of the information leaks. Furthermore, the GDPR does not prevent the collection, analysis and sharing of information leaks with AIL and ail-leak object in MISP as long as the processing of the personal data is aligned with the Regulation principles and is based on a lawful ground.

Reference

1. Information sharing and cooperation enabled by GDPR
2. CIRCL(2018) “AIL Framework for Analysis of Information Leaks: From a CSIRT use-case towards a generic analysis open source software”
3. List compiled based on MISP Objects Guide
4. CIRCL, “TR-46 – Information Leaks Affecting Luxembourg and Recommendations”
5. AIL framework - Analysis Information Leak framework

Acknowledgment

This document was partially funded by CEF (Connecting Europe Facility) funding under CEF-TC-2016-3 - Cyber Security *Improving MISP as building blocks for next-generation information sharing*.

Co-financed by the European Union
Connecting Europe Facility

Contact and Collaboration

If you have any question or suggestion about this topic, feel free to contact us. This document is a collaborative effort where external contributors can propose changes and improvement the document.