ENCRYPT Data Protection Notice

Information for the processing of personal data in accordance with art. 14 GDPR

The purpose of this data protection notice is to inform data subjects about the processing of their personal data. Considering the technical nature of the module and limitations imposed by the research design (i.e., scale), it is considered that informing those data subjects directly would involve a disproportionate effort. For this reason, this information is made publicly available via the project's website in accordance with art. 14 GDPR and with its potentially applicable derogations (art. 14 (5) (b) GDPR[1]), as an effort of enabling the data subjects to be informed about the data processing and to exercise their rights. This notice refers to the specific module of the ENCRYPT responsible for the collection of data from online sources.

Data will be collected from external online sources, as follows:

i. Online public datasets that contain information about data breaches
ii. Social media posts related to cybersecurity
iii. Forum posts (Surface and Dark Web) related to Cyber Threat Intelligence (CTI)
iv. Webpages (Surface and Dark Web) related to CTI

1. The Project

ENCRYPT project's vision it to go beyond-the-state-of-the-art to overcome the limitations of these Privacy Preserving technologies in several aspects. First, it will address the scalability issue by going beyond the single-key FHE paradigm and exploring the application and the practicability of new multi-key and threshold FHE schemes especially in a federated context. Second, to address the drawbacks of each technology in terms of covered threats and performance, ENCRYPT will investigate the combinations of several of these PP methods: TEE with HE, SMPC with HE, DP with HE, etc. Third, ENCRYPT will address the slow computation times associated with the existing solutions for privacy-preserving technologies based on HE or SMPC, by providing hardware acceleration in a user-friendly way, since users will not have to write GPU code for FHE, but rather obtain it from the TornadoVM compiler. Fourth, ENCRYPT will look at the necessary methods to make these advanced PP technologies easier to interact with existing infrastructures and more traditional security mechanisms. It will investigate the use and the application of the transcipher method for the FHE, allowing it to switch from "traditional" symmetric encryption to a homomorphic one, without the need to decrypt the sensitive data. This powerful method will permit not only keeping almost standard symmetric cryptography on the clients' terminals but

[1] Paragraph 5 (b) of this Article provides for an exemption if such information proves impossible or would involve a disproportionate effort, for processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes. In this case, subject to the conditions and safeguards referred to in Article 89(1) GDPR, the controller shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available.

also reducing the bandwidth requirements for exchanging the encrypted data (thus also addressing the scalability drawback).

Since a major impediment in the adoption of these PP technologies is also their "user-friendliness", ENCRYPT will also provide an AI-based recommendation system allowing one to choose one or a combination of them and to configure them in function of the system's requirements and the identified needs in terms of protection of the users' personal data and of performance. Finally, the proposed solutions will be developed and validated in several settings and real-world use cases including the challenging cross-border federated processing of large datasets.

2. Data Controller

The Centre of Research & Technology – Hellas (6th km Harilaou - Thermi, 57001, Thermi-Thessaloniki, Greece) is the Data Controller

3. Data Processing

With respect to the processing of personal data in the Cyber Threat Intelligence (CTI) use case, the relevant activities include the gathering, extraction, correlation, advancement, enrichment and sharing of CTI information. The applicable legal ground for such processing activities carried out through the CTI tool is the legitimate interest of the data controller (CERTH/ITI) pursuant to Article 6(1)(f) GDPR. The processing is in fact necessary for scientific purposes and for the purposes of ensuring network and information security, in accordance with recitals 49 and 50 GDPR.

What personal data is being processed?

The processed data stemming from the social media posts and webpages, with publicly available accounts and with full respect to the terms and conditions of the relevant websites/social media platforms will include:

- IP addresses
- E-mail addresses
- Social media account information, including the username, e-mail, profile picture URL, birth dates, birthplaces, marital status, addresses, tax information, phone numbers, username, description (if any from the user), location, as well as the number of friends, followers, and favourites.
- Social media and forum posts including comments, textual and multimedia content uploaded by social media users, together with relevant metadata, hashtags, and multimedia data (image links, links to articles, posts, etc. found on the surface web)
- Social media account interactions, including user mentions.

No special categories of personal data (art. 9(1) GDPR) are foreseen to be collected (at least not intentionally), nor data relating to criminal convictions (art. 10 GDPR). All data will be collected in accordance with the licences and terms & conditions of the data providers. All data will be

gathered only from public accounts, with the permission defined by the social media platforms and in compliance with the respective terms of use, including the ones referred explicitly to the terms of use on behalf of minors, thus in accordance with user expectation of privacy. All collected data will be pseudonymised, while mentioned users will be left as is, for research purposes. Data minimisation will also be applied, i.e., only data that are necessary for the purposes of the project will be processed. Further, details are provided in the "What is the purpose of the processing" section.

What is the purpose of the processing?

As aforementioned, the data will be used for (i) scientific research purposes, (ii) to facilitate the functionality of other modules of the project, and (iii) for demo purposes.

Data security

The ENCRYPT project implements appropriate technical and organizational measures to ensure an appropriate level of protection against the risks arising from processing, such as accidental or unlawful destruction, loss, alteration, unauthorised disclosure, or access. All data will be collected in accordance with the licences and terms & conditions of the data providers. All data will be gathered only from public accounts, with permission defined by the web/social media platforms and in compliance with the respective terms of use, thus in accordance with user expectation of privacy. In accordance with the data minimisation principle, only the parts of the social media posts that are deemed necessary for the project's objectives will be processed subject to a privacy-by-design technique, while the majority will be deleted immediately. The server hosting this database is accessible only by authorised users through authentication (using passwords of high complexity). A firewall will also be in place to allow only specific (whitelisted) IPs to access the server and to restrict the access of each whitelisted IP only to specific ports/services. Different access privileges to the data are available to ensure that the authorised users will only have access to the stored data on a need-to-know basis, i.e., that the authorised users will have access only to the stored psedonymised data needed to fulfil their tasks. Devices that will store a backup of the data will follow the same security procedures as the main server. For any remote interactions with the server (e.g., remote control or data transfer), secure protocols such as ssh/stfp are used. Any processing of the data is performed on that server. In case processing is needed on other machines, the same security measures of the server will be applied to the respective machine. The metadata of the social media and the webpages will also be stored in a local database that is secured (authentication mechanisms are enabled) and is also IP protected.

Will the collected data be shared?

The collected personal data (in their pseudonymised form) may be disclosed: (1) to all partners of the Consortium, through a password-protected system; and (2) if this is required to third parties for the fulfilment of our legal obligations or is necessary for the fulfilment of the above data processing purposes and is in compliance with the applicable legal framework. The information collected will be also used to contribute towards several journal and conference publications as well as scientific contests, in line with social media/Web Policy. It is also

highlighted that no personal data will be transferred outside the European Union (EU) or the European Economic Area (EEA).

Who will be responsible for all the data when this study is over?

When this study is over, CERTH/ITI will be the only one responsible for the information collected.

How long will data be stored?

The storage duration of the data in their pseudonymised form will be the duration of the project plus five (5) years after the end of the project to be available for demonstration in case of an inspection or an audit, as long as required to achieve the above purposes of the processing, unless a longer retention period is required by law or for the establishment, exercise or defense of legal claims.

Will the collected personal data be used for other purposes?

All personal data collected in ENCRYPT will not be processed for any other purposes outside of those specified in this document.

Will the collected data be processed by automated tools supporting decision-making?

All the relevant collected data will be used for (i) scientific research purposes, (ii) to facilitate the functionality of other modules of the project, and (iii) for demo purposes. Data collected from you will only be used to test the capabilities of the ENCRYPT tools and you will not suffer any consequences of automated processing supporting decision-making. After hashing of your account information, the researchers will not be able to trace back your data back to you.

What are your rights?

Your rights under GDPR are contained within articles 12-23 and 77. Some of your most important rights include:

- *Right to information:* you may request information about whether we hold personal information about you, and, if so, what that information is and why we are holding it. This information shall be provided within a reasonable period after obtaining the personal data, but at the latest within one month of receipt of the request
- *Right to access:* you may request to receive a copy of the personal information we hold about you and to check that we are lawfully processing it.
- *Right to rectification*: you may ask us to rectify the information that we hold about you in case you consider that something is missing or is incorrect.
- *Right to erasure*: you may ask us to erase your personal data at any given moment without a specific reason.
- *Right to object*: you may request to stop processing delete or remove your personal data at any desired moment where there is no good reason for us continuing to process it
- *Right to data portability*: you have the right to request the transfer of your personal data in an electronic and structured form to another party or directly to you. This enables you to take your data from us in an electronically usable format and to be able to transfer your data to another party in an electronically usable format.

- Lodge a complaint with the Hellenic Data Protection Authority (https://www.dpa.gr).

Please note that the aforementioned rights may be restricted in the light of the GDPR (e.g., art. 89 par. 2) and the applicable national data protection legislation.

For the exercise of your rights and for any other data-related information you may contact us at m4d_ethics@iti.gr