FORESIGHT 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 FORESIGHT responsible for 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

FORESIGHT (EC Horizon 2020 Project FORESIGHT, GA Nr. 833673) main objective is to develop a federated cyber-range solution that will allow the generation of realistic and dynamic scenarios that are based on identified and forecasted trends of cyber-attacks and vulnerabilities extracted from Cyber Threat Intelligence (CTI) gathered from a variety of internal and external (online) sources, aiming in that way enhancing at the preparedness of cybersecurity professionals at all levels and advance their skills towards preventing, detecting, reacting and mitigating sophisticated cyber-attacks. This is achieved by delivering an ecosystem of networked realistic training and simulation platforms that collaboratively bring unique cyber-security aspects from the aviation, smart grid, and naval domains. The proposed platform will extend the capabilities of existing cyber-ranges and will allow the creation of complex cross-domain/hybrid scenarios to be built jointly with the IoT domain. Emphasis is given on the design and implementation of realistic and dynamic scenarios that are based on identified and forecasted trends of cyberattacks and vulnerabilities extracted from cyber-threat intelligence gathered from the dark web; this will enable cybersecurity professionals to rapidly adapt to an evolving threat landscape. The development of advanced risk analysis and econometric models will prove to be valuable in estimating the impact of

[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.

cyber-risks, selecting the most appropriate and affordable security measures, and minimising the cost and time to recover from cyber-attacks. Innovative training curricula, guiding cyber-security professionals to implement and combine security measures using new technologies and established learning methodologies, will be created, and employed for training needs; they will be linked to professional certification programs and be supported by learning platforms. Aside from the development of skills, the project aims at a holistic approach to cyber-threat management with the goal of cultivating a strong security culture. As such, the project puts considerable emphasis on research and development (i.e., research on cyber-threats, development of novel ideas, etc.) as the key to increasing training dynamics and awareness methods for exceeding the rate of evolution of cyber-attackers.

2. Data Cotroller

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

3. Data Processing

The purpose of data collection in this project is to identify new tools utilised for each cyberattack, (attack vectors, Indicators of Compromise (IoC), IP addresses, malware hashes, etc,), thus expanding the existing knowledge on new cybersecurity attacks and tactics, techniques, and procedures used by adversaries, on vulnerabilities, on data breaches, etc. The collected and processed data will be used for the generation of realistic and dynamic scenarios that are based on identified and forecasted trends of cyber-attacks and vulnerabilities. The collection and processing of data coming from various sources is vital for the extraction of useful cyber-threat intelligence to be used for the above purposes. The legal basis under which the data shall be processed is article 6(1)(f) GDPR. The legitimate interest of the controllers lies with the pursuit of scientific research purposes confirmed by the European Commission. Such data will be collected by the honeypots as well as the web and social media crawling that will be tested in the context of the FORESIGHT pilots.

What personal data is being processed?

The following categories of personal data publicly available on social media and/or the surface and dark web will be collected and processed:

- IP addresses
- E-mail addresses
- Social media and forum posts including the language, textual content, hashtags, images/videos, whether the post is a reply to another post, as well as the number of retweets (in case of twitter) and the number of likes;
- Social media and forum account information, including the names and surnames, birth dates, birth places, marital status, addresses, tax information, and phone numbers, username, description (if any from the user), location, as well as the number of friends, followers and favourites;
- Social media account interactions, including user mentions;

No special categories of personal data (art. 9(1) GDPR) will be collected, 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, social media platforms/Web forum communities including the ones referred explicitly to the terms of use on behalf of minors, thus in accordance with user expectation of privacy, and will be properly will be anonymised and/or pseudonymised. The collected data will mostly include non-personal information. In accordance with the data minimisation principle, only personal data (i.e., IPs) that are deemed useful for the project will be processed subject to a privacy-by-design technique, while the majority will be deleted immediately. IPs can be used to correlate different attacks and result in enriching the collected cyber threat intelligence. Further, details are provided in the "What is the purpose of the processing" section.

What is the purpose of the processing?

The use of personal data included in the collected data (i.e., IPs) enables the data controller to correlate the gathered data to extract more advanced intelligence about cyber threats. The purpose of the data controller is not to collect personal data, but to use this data to enrich the collected intelligence. This will facilitate the generation of more realistic scenarios that are based on identified and forecasted trends of cyber-attacks and vulnerabilities which will result in the better education of the users of the FORESIGHT platform against cyber-attacks. The main sources of interest to be processed should primarily include information about cyber threats and their purpose should be to educate people and organisations against cyber threats. This is in line with the aim of the FORESIGHT project, which is to develop a federated cyber-range. Some of the sources that will be monitored (e.g., specific web pages from the Dark Web) can have different uses in the framework of the scientific research (FORESIGHT is a research program) for collecting or sharing data, such as Dark Web forums that sell new vulnerabilities (i.e., 0-days). This information can be used by the FORESIGHT project to inform and educate its users about those new vulnerabilities to increase their security. The project is interested neither in the collection nor in any other type of personal data processing made by this tool and considers it a by-product. The IPs and usernames that will be collected will be encrypted by design.

Data security

The FORESIGHT 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. Examples include rule-based anonymisation techniques and machine learning based techniques that will be applied to the recognised names that may correspond to personal data. Personal data considered useful for the project will be either pseudonymised and deleted or encrypted and stored in a password-protected database. IPs and email addresses will be pseudonymised on the fly (IPs are replaced with the expression "rectified_IP'' while emails are replaced with the expression "rectified_email"). The retrieved text will be stored encrypted using the AES256 encryption algorithm. The key will be securely stored in the machine that hosts the information gathering module. All the personal identifiers (if any) from all the other data sources (surface web) will be removed (e.g., Photos blurring etc.) before being further processed, so that access to raw data with further links to personal data will not be possible. For any interaction with the server (e.g., remote control or data transfer), secure protocols such as ssh/stfp will be used. Additionally, only authorised personnel will have access to this machine. 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, 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 anonymised/pseudonymised data needed to fulfil their tasks. In addition to that, a specific Data Protection Impact Assessment has been clearly carried out prior to any type of processing. After the end of the project (March 2023), secure deletion tools will be applied by the members of the FORESIGHT Joint Controllership agreement (see section "Will the collected data be shared?" for more details), that will make the data irretrievable.

Will the collected data be shared and who will be responsible for all the data when this study is over?

The collected personal data may be disclosed: (1) to all partners of the Consortium, through a password protected system; and (2) if this is required to third parties (including data processors if exist) 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. It is also highlighted that no personal data will be transferred outside the European Union (EU) or the European Economic Area (EEA).
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 anonymised or pseudonymised form will be the duration of the project plus five (5) years after the end of the project [i.e., March 2023], to be available for demonstration in case of an inspection or an audit, as long as required to achieve the above purposes of 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 FORESIGHT (*see section 3: What personal data is being processed?*) 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?

Your data will be used for (i) for 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 FORESIGHT 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) or other applicable 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