TECHNICAL AND LEGAL ASPECTS OF DATABASE'S SECURITY IN THE LIGHT OF IMPLEMENTATION OF GENERAL DATA PROTECTION REGULATION

Paweł Drąg,¹ Mateusz Szymura²

Abstract: In the modern era, information is not only a valuable commodity, but also a potential source of threat, especially when it comes to personal data. The implementation of the General Data Protection Regulation seeks to unify regulations and safeguards in a same manner across the EU. The following paper surveys how the legal aspects of GDPR influence the existing technical framework of databases containing personal data. In this research we want to show if the already existing technical infrastructure and safeguards implemented in databases containing personal data are sufficient and if not, if implementing new ways of protecting of data will require creating entire new system of databases or only changing of existing framework. Therefore, we combine an analysis of legal texts with a technical analysis of existing and newly implemented safeguards. While the GDPR doesn't answer what safeguards should be implemented (in the spirit of technological neutrality), the notion of pseudonymisation of the data is strongly advocated through the Regulation. In our paper we tried to show the algorithm, which create a pseudonymisation function that can change personal data into generic data with the possibility to reverse that process and utilise data after de-pseudonymisation. Implementing safeguards based on the following function create a more safe environment for data safekeeping, while give nearly immediate access to data for authorised person, who can reverse pseudonymisation and transform generic data once more into personal data.

UDC Classification: 004.6; **DOI:** <http://dx.doi.org/10.12955/cbup.v6.1294>

Keywords: Data Protection Regulation, Database, Security

Introduction

In the modern era, information is not only a valuable commodity, but also a potential source of threat, especially when it comes to personal data. The replacement of the current EU data protection legislation, Data Protection Directive 95/46/EC (DPD) and the implementation of the General Data Protection Regulation (GDPR) seeks to unify regulations and safeguards in a same manner across the EU. This will be possible due to the legal character of a Regulation, which is binding and directly applicable in all Member States of the European Union.

The idea of data protection laws shows us conflict of two values of the information-based society - the desire to protect personal data and the necessity of effective processing of data. GDPR tries to resolve this conflict by imposing minimum standards, which should guarantee both.

But not all of the scholars agree that the GDPR in fact ensures any of these goals. Koops suggests that the current proposal of data protection based on the GDPR will fail based on three fallacies – the delusion of protection one's data, the simplification of the law, which actually starts to be even more complex than before and the whole idea of comprehensive protection that it is more regulatory than functional (Koops, 2014).

Meanwhile, aside from legal interpretation of the provided safeguards there is also the question of practical applicability of already existing safeguards. Duncan and Whittington emphasize significantly increasing challenges in the protection of databases in operating systems, especially in those that use cloud computing, in which case a corporate firewall doesn't protect external services and potential consequences of attack on the database could be much more severe and untraceable due to the destruction of system log data by the attacker (Duncan & Whittington, 2017).

Data Protection Framework in the European Union

To begin with the description of the current EU legal framework of data protection we need to make a few introductory remarks about the data protected by the GDPR.

Definition of personal data can be found in article 4(1) of the GDPR which states that personal data is *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* (GDPR, article 4(1)).

¹ Wrocław University of Science and Technology, Department of Control Systems and Mechatronics, Poland. E-mail: pawel.drag@pwr.edu.pl.

² University of Wrocław, Faculty of Law, Administration and Economics, Poland. E-mail: mateusz.szymura@uwr.edu.pl

There are also three categories of sensitive personal data which will demand a higher level of protection. The first type of this data is genetic data, which is personal data related to *the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person*. Another type of sensitive data is biometric data, which is a result of processing data connected to physical, psychological and behavioral characteristics of the natural person. The last part of this group is medical data, which is a data related not only to the physical and mental health of the natural person, but also to the provision of health care services (GDPR, article 4 (13), (14) & (15)).

This data, alongside data revealing the racial or ethnic origin of a person, as well as their political, religious and philosophical beliefs, trade union membership, and/or their sex life or sexual orientation are protected to the extent that the GDPR provides general rules prohibiting processing of those types of data. Exceptions to this rule can be made only under specific circumstances. Processing could be done only on one of the following bases:

- a) explicit consent from the natural person,
- b) when it is necessary for the purposes of carrying out the obligations and exercising specific rights of the controller in the fields of employment, social security and social protection law,
- c) when it is necessary to protect the vital interests of the data subject or other natural person,
- d) when non-profit organizations (foundation, association, trade union) process data of its members in legitimate activities and with appropriate safeguards,
- e) in the case of legal claims or by the courts acting in a judicial capacity,
- f) when data is already made public by the data subject,
- g) when it is necessary for reasons on substantial public interest,
- h) for the purpose of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services,
- i) when it is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes.

While these reasons allow to process special categories of personal data, Members State of the EU can introduce further conditions and limitations in order to protect their citizens (GDPR, article 9).

All data processing must be done in compliance of two pillars: data protection principles and personal rights of the natural person. Data protection principles are shown in article 5 of the GDPR. All data should be processed lawfully, fairly and in a transparent manner. The purpose of processing needs to be at the same time specified, explicit and legitimate, with important exemption of processing data for public interest and scientific research which are deemed compatible with the initial purpose of data processing. The processing itself need to be limited only to purposes stated initially. In order to protect the data, it needs to be kept in a form that allows the identification of the subject of the data only for the necessary period and in a manner that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures (GDPR, article 5).

The GDPR states, that the controller should process personal data that could be used to identified natural person only when it's necessary. Legislators clearly favor processing of data in forms in which the data controller is able to demonstrate, that it is not in a position to identify the subject of the data (GDPR, article 11).

The second of the aforementioned group of compliance requirements are the personal rights of the data subject. The GDPR provides the following rights:

- a) The right to be informed - which includes the right to be informed about collection of data, as well as its usage. This particular right, as stated in articles 13 and 14 of GDPR creates an obligation of transparency on the side of the controller and processor of the data, as well as from a third party, which lawfully received the data from one of them. The individual need to be informed about the purpose for the processing of data, periods of retention of the data with whom it could be shared. The following information needs to be provided in a way that it is easy to access and understand.

- b) The right of access - under article 15 of the GDPR, an individual has not only access to their personal data but also information mentioned above, as well as to information about the procedure of rectification or erasure of the personal data, information about the right to lodge a complaint to the supervisory authority and information about the existence of automated decision-making in the processing of the personal data.
- c) The right to rectification - in case of inaccurate personal data, the GDPR in article 16 enacts the right to demand rectification of the data. While the GDPR doesn't give a definition of inaccurate data it should be considered both wrong and missing data.
- d) The right to erasure - also known as a right to be forgotten, which is granted to an individual in cases, when there is no longer necessary to process their personal data, when the individual withdraws their consent or the data was processed unlawfully. This particular right isn't absolute - article 17(3) of GDPR enacts limitations for this privileged of the individual - especially in the fields of public interest, scientific research and in case of exercising one's right for the freedom of expression.
- e) The right to restrict processing - which isn't as absolute as the right to erasure, but limits processing only to a storage capacity, with exception when data needs to be used in order to exercise or defend legal claims or for the protection of the rights of another natural or legal person or for reasons of important public interest (GDPR, article 18).
- f) The right to data portability - this right allows natural person to obtain data from the controller, which needs to assure that data can be moved, copied and transferred in secure way from one processor to another (GDPR, article 20).
- g) The right to object to the processing of personal data - which varies depending on the purpose of the data processing. In the case of direct marketing purposes, the individual always enjoys this right. Contrary in other situations, the controller can demonstrate compelling legitimate grounds for the processing, which override the interests, rights and freedoms of the individual, which in effect allow the controller to process the data despite an individual's objection (GDPR, article 21).

The GDPR contains two basic obligations when it comes to database security - to secure the data and to inform certain subjects in case of a breach. Personal data breach is defined in the GDPR as a *breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to, personal data transmitted, stored or otherwise processed* (GDPR, article 4(12)). The general obligation to implement security measures is regulated in article 32 of the GDPR. Controllers and processors of data need to implement safeguards which can vary based on several reasons: technical (*state of the art*), economical (*cost of implementation*) and likelihood and severity of potential data breach. Organizations need to assess the level of security in the processing of data to protect them from *accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to personal data transmitted, stored or otherwise processed*. The following effects could be achieved by pseudonymisation and encryption of personal data, ensuring the protection of processing services, retaining the ability to restore data in cases of accidents as well as regular testing and assessing effectiveness of technical and organizational safeguards (GDPR article 32).

One of the important elements that need to be considered in connection to databases security is the form of data storage. Hintze suggests, that all of the risk mentioned in article 32 of the GDPR could be significantly lowered with a procedure of de-identification, and the stronger the levels of de-identification, the lower the risks of potential misuse of data gained from a breach are (Hintze, 2016).

In the case of a personal data breach, the controller is obliged to inform not only the data subject, but also the proper supervisory authority. These notifications should be done without undue delay, but in the case of supervisory authority, the GDPR creates maximal period of 72 hours for notification. There are also different basis for notification - while the supervisory authority needs to be notified in any case of personal data breach, only those data breaches that potentially possess a higher risk to the rights and freedoms of natural persons are reasons for the notification made directly to the data subjects (GDPR, article 33(1) and 34 (1)). Informational duties have very similar character in both situations. While information transferred to a supervisory authority is broader and describes the overall character of a data breach and the nature of the personal data breach, with categories and approximate numbers of data subject concerned, consequences of the breach for personal data and the

proposed and taken measures, the form of information for natural persons emphasize a plain form of explanation about the nature of breach (GDPR article 33 (2 a-d) and article 34(2)).

Technological Aspects of the GDPR

When it comes to translate law into technical aspects, most important is the suggestion containing Recital 15 of the GDPR, which states, that the protection of a natural person should be technologically neutral. This means that regardless if the data is processed by a state-of-the-art technology automatically or it is contained in a filing system and processed manually it should serve the same purpose - the protection of personal data. Only the controller and the processor decide how they will process personal data, but at the same time they are responsible for their safety, but also for their integrity and confidentiality, which article 5 f of the GDPR clearly states (GDPR Recital 15 & article 5f).

While the GDPR states a general rule of technological neutrality, it indicates importance of a few technical principles: the pseudonymisation of data, its integrity, availability and confidentiality. The process of data as described in article 4(5) of the GDPR as processing of personal data in a way that the data cannot be attributed to a specific data subject without a certain key or additional information, which is protected and stored separately and could be used to decrypt it. In fact, pseudonymisation is a process of anonymization of the data, that under certain circumstances could be reversed, which is in the same means of protection of the data and the way of its processing, since irreversible anonymization would not serve both those purposes. Recital 26 of the GDPR clearly states that the principles of data protection don't apply to anonymous information, which are described as information which does not relate to an identified or identifiable natural person, or personal data rendered anonymous in such manner that the data subject is no longer identifiable (GDPR, Recital 26).

At first glance the GDPR doesn't state the duty of pseudonymisation of data, but rather a soft approach to persuade to use that a partial method of data protection can be found in Recitals 28 and 29. The first of the cited recitals states, that the application of pseudonymisation of personal data can reduce the risks to data and help controllers and processors to meet their data-protection obligations. The second recital states that the controller that allows the pseudonymisation of data should offer that protection to other entities he controls in order to create incentives for the application of pseudonymisation (GDPR article 4(5), Recitals 28 and 29).

Some authors point out that the procedure of pseudonymisation is still vulnerable to re-identification risks, which means that a potential attacker still can identify the data subject with the datasets. They pinpoint three potential risks - singling out, link-ability and inference. The first of these risks is the ability to isolate records about the same subject and to single them out from a database. The second one is the possibility to link records concerning two data subjects in the same category of data. The third one is the possibility to deduce the value of the attribute of personal data, from the values of other attributes. As those risk couldn't be avoided, but only mitigated, the authors suggest using data processing techniques that lead to the randomization and generalization of markers, as well as to masking direct identifiers (Hu et al., 2017). The second technical principle of processing and protection of personal data is its integrity as shown in the already cited articles 5f and 32.1 b of the GDPR. While the defragmentation of data could be perceived as one way to protect data, the GDPR favors pseudonymisation as a way to divide information. Defragmentation of data is a potential threat to its completeness, accuracy and consistency and should be avoid in general.

The principle of availability states that the obligation is connected not only to the data, but to the whole processing systems and services. Processing data systems should be designed to resist, at a given level of confidence, accidental events or unlawful or malicious actions that compromise the availability, authenticity, integrity and confidentiality of stored or transmitted personal data, as stated in Recital 49 of the GDPR (GDPR, Recital 49). Securing confidentiality in the processing of personal data is at the same moment the most important and the hardest to implement. It requires not only technical safeguards to protect personal data from external attacks, but also organizational ones to limit the possibility of data leakage from internal misuse of data. The fact that three Recitals emphasizes the obligation to ensure confidentiality of data shows how important this particular principle is for European legislators (GDPR Recital 39 and 83)

We came to the conclusion that pseudonymisation is the most suggested method of data protection based on the GDPR. But there is also the question of how this can be achieved.

The new data protection approach

The appropriate treatment of personal data is one of the most important challenges connected with the application of information technologies. The variety of used devices, different communication protocols, as well as generic data kinds, justify a question about the safeguard of privacy according to all the procedures connected with data storage, communication and processing. The privacy protection aspects reflect the observed development of the technology, like data-bases of public health services or even personal genetic information (Anisetti et al., 2018). Moreover, the amount of information resultant from daily life activity, requires Big Data Analytics tools for analysis. Therefore, the question of how software methods should reflect the GDPR legislation is highly motivated.

In general, it is important to indicate that some disconnections between programming procedures and GDPR regulations can be observed. To solve these difficulties, a new data protection approach procedure, based on the Privacy-Aware Data Flow Diagrams PA-DFDs (Antignac et al., 2016), was designed. The main parts of the new data protection approach are the definitions of generic data, personal data, as well as the reversible pseudonymisation function.

Definition 1

Generic Data (gdata): all data that cannot be assigned to a specific person.

As an opposition to generic data, the personal data can be defined

Definition 2

Personal Data (pdata): all data that can be assigned to a specific person.

Moreover, the reversible pseudonymisation function f_{PA} is considered.

Definition 3

Let us considered the function f_{PA} such that

$$f_{PA}(pdata) = gdata \quad (2)$$

and

$$f_{PA}^{-1}(f_{PA}(pdata)) = pdata. \quad (1)$$

The application of the reversible pseudonymisation function (1) enables us to translate the Personal Data into Generic Data. Finally, the function f_{PA} can be used to transform the obtained Generic Data also in the opposite direction. The diagram of the proposed methodology was presented in Figure 1.

The function can be represented by an effective coding/decoding procedure (Demir et al., 2018; Wallace, 2016; Bauer et al., 2016)

Conclusion

The new legal framework of European data protection laws tries to create a unified legal environment of data protection within the EU. While the GDPR introduces principles of technical neutrality, it also creates several requirements for data processing, that each system chosen by a controller needs to be in compliance with. From those two most fundamental groups of requirements, there are contained the data protection principles and personal rights of the natural person. In order to create effective safeguards a few technical principles need to be implemented as well, which are the pseudonymisation of data, its integrity availability and confidentiality. All systems should create conditions that allow the fulfillment of these technical principles. From these principles, the most important one is the pseudonymisation of data, which is not only a way to protect data from a breach, but also to render the

data useless in the case of a potential breach. Finally, the new data protection approach based on the reversible pseudonymisation function f_{PA} was proposed.

Acknowledgments

The work of Paweł Drąg has been supported by the grant 0401\0146\17 at Wrocław University of Science and Technology, Department of Control Systems and Mechatronics.

References

- Anisetti, M., Ardagna, C., Bellandi, V., Cremonini, M., Frati, F., Damiani, E. (2018). Privacy-aware Big Data Analytics as a service for public health policies in smart cities. *Sustainable Cities and Society*, vol. 39, pp. 68-77. <https://doi.org/10.1016/j.scs.2017.12.019>.
- Antignac, T., Scandariato, R., Schneider, G. (2016). A privacy-aware conceptual model for handling personal data. In: Margaria T., Steffen B. (eds.), *International Symposium on Leveraging Applications of Formal Methods*, pp. 942-957, Springer. https://doi.org/10.1007/978-3-319-47166-2_65.
- Bauer, C.R.K.D., Ganslandt, T., Baum, B., Christoph, J., Engel, I., Löbe, M., Mate, S., Stäubert, S., Drepper, J., Prokosch, H.-U., Winter, A., Sax, U. (2016). Integrated Data Repository Toolkit (IDRT). A Suite of Programs to Facilitate Health Analytics on Heterogeneous Medical Data. *Methods of Information in Medicine*, vol. 55, np 2, pp 125-135. <https://doi.org/10.3414/ME15-01-0082>.
- Demir, L., Kumar, A., Cunche, M., Lauradoux, C. (2017). The Pitfalls of Hashing for Privacy. *IEEE Communications Surveys and Tutorials*, vol. 20, pp 551 - 565. . <https://doi.org/10.1109/COMST.2017.2747598>.
- Duncan, B., Whittington, M. (2017). Creating and Configuring an Immutable Database for Secure Cloud Audit Trail and System Logging. *International Journal On Advances in Security*, vol.10, no. 3-4, pp. 155-166.
- General Data Protection Regulation, GDPR. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC.
- Hintze, M. (2016). Viewing the GDPR through a de-identification lens: a tool for compliance, clarification, and consistency. *International Data Privacy Law*. <https://doi.org/10.1093/idpl/ix020>.
- Hu, R., Stalla-Bourdillon, S., Yang, M., Schiavo, V., Sassone, V. (2017). Bridging Policy, Regulation, and Practice? A Techno-Legal Analysis of Three Types of Data in the GDPR. Hu, Runshan and Stalla-Bourdillon, Sophie and Yang, Mu and Schiavo, Valeria and Sassone, Vladimiro, *Bridging Policy, Regulation, and Practice? A Techno-Legal Analysis of Three Types of Data in the GDPR* (September 1, 2017). In: van Brakel, R.L.R., Gutwirth, S., De Hert P. (eds.) *Data Protection and Privacy: The Age of Intelligent Machines.*, Hart Publishing, 2017. <https://ssrn.com/abstract=3034261>.
- Koops, B.-J. (2014). The trouble with European data protection law. *International Data Privacy Law*, 2014, vol. 4, no. 4, pp. 250-261. <https://doi.org/10.1093/idpl/ipu023>.
- Wallace, S.E. (2016). What Does Anonymization Mean? DataSHIELD and the Need for Consensus on Anonymization Terminology. *Biopreservation and biobanking*, vol.14, no.3, pp. 224-230. <https://doi.org/10.1089/bio.2015.0119>