GDPR Compliance Through Authorization Systems

Said Daoudagh*
said.daoudagh@isti.cnr.it,di.unipi.it
ISTI-CNR and University of Pisa
Pisa, Italy

KEYWORDS

Access Control, GDPR, Privacy, Testing

1 INTRODUCTION & PROBLEM OF INTEREST

The General Data Protection Regulation (GDPR) is the new EU legal framework for the protection of Personal Data of European citizens. It aims to harmonize data protection law in Europe and to strengthen the rights of individuals. The GDPR imposes several duties, and consequently a system of fines, for the Controller and the Processor, i.e., the data managers for the processing of Personal Data. Personal Data is any information about a Data Subject, i.e., an identified or identifiable natural person. In order to avoid severe penalties, the controller and processor need to: i) demonstrate the compliance with the GDPR (required by the "Accountability principle" (Art. 5.2)); ii) demonstrate the appropriate technical security level ("integrity and confidentiality" (Art. 5.1(f))); iii) adapt and rethink their data practices so as to be aligned with the "data protection by design and by default" approach (Art. 25).

In this scenario, if on one side big organizations are currently successfully investing large amount of money both in technologies and legal consulting to assure their compliance with the GDPR, on the other Small and Medium-sized Enterprises (SMEs) have not the same economic power. SMEs are currently looking for low cost, easy to use solutions for being aware about the GDPR requirements and prepared to comply with its provisions. For the SMEs one of the major difficulties comes from the GDPR technical interpretation: the technology neutral nature of the GDPR poses important application challenges for software architects, developers and security experts having no sufficient legal expertise for translating GDPR provisions into technical requirements. Indeed, for all organizations, and specifically for SMEs, being (by-design) compliant with the GDPR means having solutions that: (1) are general purpose; (2) must take in consideration the regulation by-design; (3) must be easily integrated with the existing business process; and finally (4) must be rooted in the GDPR principles dictated in Art. 5.

At state of the practice there is no a comprehensive ready-to-apply solution for the above mentioned challenges, therefore the underlining idea of the PhD projet: leverage the Access Control (AC) systems, de facto mechanism used to restrict data access, as a technical solution for protecting "personal data by-design", and gaining legal compliance with the GDPR. The choice of AC system has two important strengths: their structure and their applicability. Structurally, AC are based on Access Control Policies (ACPs), i.e., a set of rules that specify who has access to which resources and under which circumstances.Consequently, AC systems satisfy by construction the principle of *Integrity and Confidentiality*. The idea of this PhD research of enriching them with policies elicited from the GDPR's provisions, leverages the AC systems to realize the compliance by-design to the GDPR's obligations. Considering the applicability, AC are: i) general-purpose models supported by the eXtensible Access Control Markup Language (XACML) standard and a reference architecture; ii) easily integrable within the existing business process as externalized authorization service, so as to decouple the business logic and authorization.

From a technical point of view, having ACPs expressing requirements aligned with GDPR's provisions involves two conceptual mappings: i) a trivial association with Resources to Personal Data and with Controller, Processor, or Data Subject to whom requesting access to them; ii) a more challenge association to *identify*, to *extract*, to *translate* and to *encode* the GDPR's provisions into enforceable ACPs [12]. This is mainly because the GDPR's provisions can be ambiguous, can include implicit information, are unstructured and not straightforwardly expressible in a formal policy.

To this purpose, an outcome of the PhD research is a systematic development process for implementing AC systems and their ACPs compliant-by-design with the GDPR. This assures that the leveraged AC systems can protect personal data (*security perspective*) and process them lawfully (*legal perspective*).

2 LIMITS OF THE EXISTING SOLUTIONS

In literature there are different works trying to give an answer on how to comply with the GDPR, but most of them are either in a early stage, or lack of automation and tools supporting their proposal in real scenarios. Among them seminal works are: [11], where an approach to extract ACPs from the Data Protection Directive (Directive 95/46/EC in force before the GDPR) is presented; [7] where a systematic approach for implementing AC policies in an industrial setting is discussed, but without taking in consideration any legal framework. Inspired by the principle of *Data Protection by-design*, the present dissertation combines those proposals and leverage them so as to provide a unified, inexpensive, adaptable, efficient and effective framework able to systematically develop ACPs in reference to the legal framework of the GDPR.

3 RESEARCH STRATEGY, NOVELTY & EXPECTED OUTCOMES

The PhD project has been conceptualized around three axes of security and data (privacy) protection: Access Control, Data Protection by-Design and Access Control Testing & Monitoring. These axes are all related to the GDPR's demands that Controllers shall obey for being GDPR compliant-by-design. Additionally, in order to provide a ready-to-use, low-cost and effective solution for the SMEs, a set of supporting tools, methodologies and guidelines for developing Access Control Systems (ACSs) and ACPs compliant-by-design

*Supervisor: Dr. Eda Marchetti (ISTI-CNR). Co-Supervisor: Dr. Anna Monreale (University of Pisa).

with the GDPR are provided. Hence, the main outcomes of the PhD project are: an Agile GDPR-based Authorization Development Life Cycle (ADLC) [10] and its supporting automation. The novelty is that these outcomes are (by-design) conceived *[t]aking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing [...], which are designed to implement data-protection principles*, as in Art. 25.

Considering the ADLC, the proposal of this PhD is an improvement of an existing approach for implementing authorization systems within enterprises [7]. The result is an Agile ADLC, which is profoundly rooted in the GDPR's "Data Protection by Design" approach (Art. 25) and the "Confidentiality and Integrity" principle defined in Art. 5.1(f), composed of eight phases: (1) Define GDPR-based use case (①); (2) Gather authorization requirements (②); (3) Identify required attributes (③); (4) Author the authorization policies (④); (5) Test ACPs & AC mechanisms (⑤); (6) Deploy the architecture (⑥); (7) Deploy the policies (⑦); and (8) Run access reviews (⑧). Considering the current advancement of the PhD project steps from ① to ④ have been developed and presented in [2, 3]; step ⑤ is currently under development [5, 9]; whereas, steps from ⑥ to ⑧ will be part of the next months activity, and some results are already obtained[8].

For the ADLC supporting automation, this proposal is the first attempt to integrate, in a unique automated environment, different available solutions for extracting, implementing and testing the data protection regulation. As in Figure 1, the ADLC implementation is composed of three main modules: (1) GDPR-Based Access Control Policies Management (module Ⓐ); (2) Access Control System (module Ⓑ); and (3) GDPR Analytics (module Ⓒ).

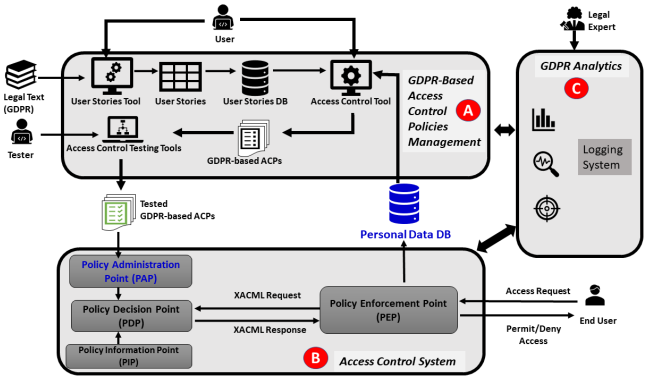

Figure 1: The Proposed GDPR-based Environment.

Very briefly, module Ⓐ provides facilities to perform steps from ① to ⑤ of the Agile ADLC. In particular, it takes as input a Legal Text (in our case the GDPR), analyses the articles related to ACs and creates a *Data Protection Backlog* [2] containing a set of User Stories organized in Epics and Theme. The User Stories are then translated into enforceable ACPs written in XACML standard encoding the GDPR principles, by taking into account real attributes contained in `Personal Data DB` [3]. Module Ⓐ integrated also ACPs and the AC mechanisms testing tools that have the following main features: (1) *Test Case Generation* [6, 9]; (2) *Mutation Generation* [4]; (3) *Test Cases Execution & Result Analyzer*; (4) *Testing Strategy Enhancement*; and finally, (5) *Oracle Derivation* [5, 9].

The `Personal Data DB` component of Figure 1 contains Personal Data, whose access is regulated by the module Ⓑ, i.e., an adapted and extended version of the XACML reference architecture with new features such as logging functionalities.

Finally, facilities for collecting and managing information for the GDPR compliance and audit purposes [1, 8] are included in module Ⓒ, which is currently under development. The module contains logging systems, monitoring capabilities, and reporting functionalities of the proposed environment so that data mining and machine learning techniques can be adopted to construct behavioral models to discover and notify unwanted behaviors.

4 OUTCOMES EVALUATION

The evaluation of the PhD's outcomes spreads to different directions: i) involving legal, security and data protection compliance experts, as well as practitioners and scholars specialized in the GDPR interpretation and realization; ii) asking co-partners of CyberSec4Europe project[1], and private companies to provide feedbacks for the proposed methodology and to assess the associated technological solutions; iii) developing a complete real case study inside the ISTI-CNR so as to evaluate the PhD proposal in a real environment. Additionally, the PhD research outcomes are continuously submitted to international conferences and journals so as to evaluate their novelty and research impact.

REFERENCES

[1] Cesare Bartolini, Antonello Calabrò, and Eda Marchetti. 2019. GDPR and business processes: an effective solution. In *Proc. of APPIS 2019, Las Palmas de Gran Canaria, Spain, January 07-09, 2019.* 7:1–7:5.
[2] Cesare Bartolini, Said Daoudagh, Gabriele Lenzini, and Eda Marchetti. 2019. GDPR-Based User Stories in the Access Control Perspective. In *Proc of QUATIC 2019, Ciudad Real, Spain, September 11-13, 2019.* 3–17.
[3] Cesare Bartolini, Said Daoudagh, Gabriele Lenzini, and Eda Marchetti. 2019. Towards a Lawful Authorized Access: A Preliminary GDPR-based Authorized Access. In *Proc. of ICSOFT 2019, Prague, Czech Republic, July 26-28, 2019.* 331–338.
[4] A. Bertolino, S. Daoudagh, F. Lonetti, and E. Marchetti. 2013. XACMUT: XACML 2.0 Mutants Generator. In *Proc. of Mutation 2013.* 28–33.
[5] Antonia Bertolino, Said Daoudagh, Francesca Lonetti, and Eda Marchetti. 2018. An Automated Model-based Test Oracle for Access Control Systems *(AST '18).* ACM, New York, NY, USA, 2–8.
[6] Antonia Bertolino, Said Daoudagh, Francesca Lonetti, Eda Marchetti, and Louis Schilders. 2013. Automated testing of eXtensible Access Control Markup Language-based access control systems. *IET Software* 7, 4 (2013), 203–212.
[7] David Brossard, Gerry Gebel, and Mark Berg. 2017. A Systematic Approach to Implementing ABAC. In *Proc. of the 2Nd ACM ABAC '17.* ACM, New York, NY, USA, 53–59.
[8] Antonello Calabrò, Said Daoudagh, and Eda Marchetti. 2019. Integrating Access Control and Business Process for GDPR Compliance: A Preliminary Study. In *Proc. of ITASEC 2019, Pisa, Italy, February 13-15, 2019.*
[9] S. Daoudagh, F. Lonetti, and E. Marchetti. 2019. XACMET: XACML Modeling & Testing. *Software Quality Journal* (2019). To appear.
[10] Said Daoudagh and Eda Marchetti. 2020. A Life Cycle for Authorization Systems Development in the GDPR Perspective. In *Sumitted to ITASEC 2020, Ancona, Italy, February 4-7, 2020.*
[11] Kaniz Fatema, Christophe Debruyne, Dave Lewis, Declan O'Sullivan, John P. Morrison, and Abdullah-Al Mazed. 2016. A Semi-Automated Methodology for Extracting Access Control Rules from the European Data Protection Directive. In *2016 IEEE SP Workshops 2016, San Jose, CA, USA, May 22-26, 2016.* 25–32.
[12] Xusheng Xiao, Amit Paradkar, Suresh Thummalapenta, and Tao Xie. 2012. Automated Extraction of Security Policies from Natural-language Software Documents. In *Proc. of the ACM SIGSOFT FSE '12.* New York, NY, USA, Article 12, 11 pages.

[1] It is an ongoing pilot for a future European Cybersecurity Competence Network.