GROOT: A GDPR-based Combinatorial Testing Approach

Said Daoudagh[0000−0002−3073−6217] and Eda Marchetti[0000−0003−4223−8036]

ISTI-CNR, Pisa, Italy
{said.daoudagh, eda.marchetti}@isti.cnr.it

Abstract. For replying to the strict exigencies and rules imposed by the General Data Protection Regulation (GDPR), ICT systems are currently adopting different means for managing personal data. However, due to their critical and crucial role, effective and efficient validation methods should be applied, taking into account the peculiarity of the reference legal framework (i.e., the GDPR). In this paper, we present GROOT, a generic combinatorial testing methodology specifically conceived for assessing the GDPR compliance and its contextualization in the context of access control domain.

Keywords: Combinatorial Testing · Data Protection · GDPR.

1 Introduction

Nowadays, quality of Information and Communication Technology (ICT) systems and modern applications is strictly tied with the security and privacy. However, most the times, due to the peculiarity of the General Data Protection Regulation (GDPR) [8], effective and efficient validation methods have to be applied for avoiding possible violations. In this paper, we present GROOT, a combinatorial testing methodology specifically conceived for assessing the GDPR compliance of ICT systems in processing Personal Data. We specifically contextualize GROOT into the Access Control (AC) domains, because they are the most promising approach for taking in consideration the peculiarities of the GDPR [6, 5]. Indeed, Access Control Systems (ACSs) aim to ensure that only the intended subjects (e.g., Data Subject, Controller and Processor) can access the protected data (e.g., Personal Data or special Categories of Personal Data) and get the permission levels required to accomplish their tasks and no much more.

The testing of ACSs represents a key activity to guarantee the trustworthiness of (personal or sensitive) data and protect information technology systems against inappropriate or undesired user access [4]. However, testing is still a time consuming, error prone activity and a critical step of the development process. Bad choices in each stage of the testing phase may compromise the entire process, with the risk of releasing inadequate security and privacy solutions that allow unauthorized access (*security perspective*) or unlawful processing (*legal perspective*).

Indeed, several strategies for the generation of test cases (i.e., access requests) for access control systems have been defined in scientific literature. They leverage the application of combinatorial approaches to access control policies values for generating test inputs [2]; or exploit data flow for test cases generation starting from policies specification [17]; or are based on the representation of policy implied behavior by means of models [9, 1]. However, to the best of our knowledge, there are few proposals for assessing the compliance with the GDPR [10, 7], and none targeting the testing access control systems in the context of the GDPR. Therefore, our work aims at advancing the state-of-the-art by providing, for the first time, the GdpR-based cOmbinatOrial Testing (GROOT) strategy, i.e., a general combinatorial strategy for testing systems managing GDPR's concepts (e.g., Data Subject, Personal Data or Controller). To better illustrate the GROOT procedural steps, an application example is also provided.

Outline: Section 2 provides an overview of the main concepts, Section 3 illustrates the GROOT methodology and its application. Finally, Section 4 concludes the paper and depicts future works.

2 Background

GDPR Concepts. The GDPR is the currently European Regulation for the protection of *Personal Data*. In its Art. 4, the GDPR defines *Personal Data* as "any information relating to an identified or identifiable natural person ('data subject')", whose data are managed by a *Controller*. The *purpose* of the *processing* of Personal Data is determined by the controller, and this "processing shall be lawful only if and to the extent that at least one of the" six legal bases "applies" (Art. 6). In particular, one of those legal bases is the consent given by the data subject "to the processing of his or her Personal Data for one or more specific purposes" (Art. 6.1(a)). The GDPR also sets other fundamental rights of the data subject, such as the right of access (Art. 15) and the right to data portability (Art. 20).

Access Control. *Access Control* (AC), implemented through *Access Control Mechanism* (ACM), provides a decision to an authorization request, typically based on predefined *Access Control Policy* (ACP). This is a specific statement of what is and is not allowed on the basis of a set of rules. For instance, a policy contains a set of rules that specify who (e.g., Controller, Processor or Data Subject) has access to which resources (e.g., Personal Data) and under which circumstances (e.g., based on the Consent and Purpose) [15].

Representing the GDPR. Implementing the GDPR's requirements is a challenging task, and a standardized solution is still missing. The most promising approaches can be divided into: using Semantic Web technologies, i.e., ontologies, using UML representation and using access control policies specification. Concerning the first group, recent proposals are [14], which models the legal concepts through the Privacy Ontology (PrOnto), and the GDPR text extensions [13] where the GDPR is represented as inked data resource. Works in

the second group use the UML notation for representing the GDPR's concepts. Among them we refer to [16] where the authors use the UML model for designing automated methods for checking the GDPR compliance, and [11] where the authors use an educational e-platform paradigm for combining the regulation, information privacy and best practices. The third group represents the legal concepts through access control policies. In particular, authors in [6] propose a semantic model to represent the GDPR consent customized for the XACML reference access control architecture, whereas in [3] authors provide a life cycle for the development of access control policies and mechanisms in reference to the GDPR's demands.

Our proposal requires (and exploits) the possibility of having a structured and machine readable specification of the legal concepts. The aim is therefore to provide a methodology independent from any GDPR representation. The adaptation of the methodology to the different GDPR's representations is left and handled during the development stage of the GROOT proposal.

3 GROOT

GROOT is a general combinatorial testing approach, for validating systems managing GDPR's concepts (e.g., Data Subject, Personal Data or Controller). In the following, we first illustrate the GROOT methodology, and then we show its usage in the context of access control.

3.1 GROOT Methodology

In illustrating the GROOT methodology, we use the following definitions:

Definition 1 (GDPR-based SUT Model) *A GDPR-based SUT Model is a tuple* $Model_{GDPR}(PAR, V)$*, where:*

- $PAR \subseteq \{DS, PD, DC, DP, C, P, PA, TP\}$ *is the set of parameters that affect the GDPR-based SUT, where DS = Data Subject, PD = Personal Data, DC = Controller, DP = Processor, C = Consent, P = Purpose, PA = Processing Activity, TP = Third Party, and*
- $V = \{V_i \mid \ i \in PAR$ *and* V_i *is the set of values for the parameter* $i\ \}$ *is the set of sets of the values that can be selected for each parameter.*

Definition 2 (GDPR-based Test Case) *Given a GDPR-based SUT Model* $Model_{GDPR}(PAR, V)$*, a GDPR-based Test Case is a tuple* $TC_{GDPR}(ATT)$ *where:* $ATT = \{ATT_i \mid ATT_i \subseteq V_i\ ,\ i \in PAR$ *and* $\ V_i \in V\}$*.*

The GROOT methodology takes as an input a GDPR-based implementation, that is a representation of the GDPR in terms of a specification language. As detailed in Section 2, currently, different proposals are available and can be used for the purpose. Under this hypothesis GROOT is composed of three main steps (see Figure 1): GDPR-based Model Derivation; Test Cases Generation; and Test Cases Translation.

GDPR-based Model Derivation (Step ①). In line with Definition 1, the GDPR-based SUT Model of the GDPR-based implementation is then derived. For this, the GDPR-based implementation is parsed in order to identify the set of parameters P, and the associated set of sets V. More precisely, for each parameter i, the subset V_i, containing the values used in the GDPR-based implementation, is derived.

Test Cases Generation (Step ②). In this step, the combinatorial testing is performed. Based on the derived parameters' values sets, different combinatorial strategies can be adopted such as: all-combinations, pairwise combinations or t-wise combinations. For instance, in the all-combinations test strategy according to the Definition 2, for each parameter i and its set of value V_i, the power set of V_i ($P(V_i)$) is derived, i.e., all possible subsets of V_i. Then, the obtained power sets $P(V_i)$ are combined so as to derive the test cases i.e., the $TC_{GDPR}(ATT)$ tuples. Because combinatorial testing is a costly activity, the selection of the best combinatorial strategy, that could be adopted, may depend on different testing objectives such as: coverage, effectiveness, reduction or prioritization.

Test Cases Translation (Step ③). According to the domain specific language, each of the obtained $TC_{GDPR}(ATT)$ tuples in Step ② is translated into specific executable test case. In the context of AC, a test case is represented through an AC request that can be evaluated by the ACM.

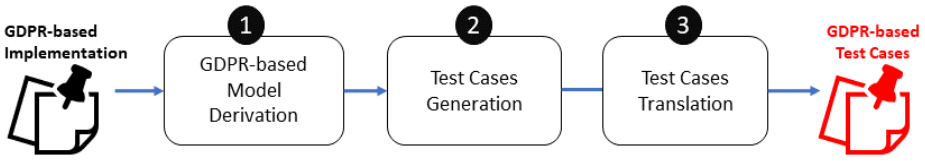

Fig. 1. GROOT Methodology.

3.2 Using GROOT

In this section, we illustrate the application of GROOT through a use case scenario based on a realistic fitness environment. More precisely, we consider Alice, a Data Subject, who wants to use a smart fitness application to monitor her daily activities to achieve a predefined training objective. In this case, we suppose that a customized (mobile) application is provided by a generic myFitness company (Controller). To meet Alice's needs, myFitness has so far defined two purposes (MyCholesterol and Untargeted Marketing), each related to a specific data set of Personal Data and achieved by allowing access to perform a specific set of Actions. Specifically, the MyCholesterol purpose is achieved by performing AGGREGATE, DERIVE, and QUERY actions; whereas the Untargeted Marketing purpose is achieved by performing COLLECT, QUERY, and SEND actions. At the time of subscribing to the myFitness application, Alice provided her Personal

Data (i.e., e-mail, Age, Gender, and Blood Cholesterol) and gave her consent to process her e-mail and Age for Untargeted Marketing purpose, and her Blood Cholesterol for MyCholesterol purpose. In turn, myFitness gave Alice controller's contacts that include: orgName, address, e-mail, and phone number.

GDPR-based Implementation. In this application example, the GDPR-based implementation refers to the Art. 6.1(a) of the GDPR. In the context of AC, considering for instance the GDPR formalization proposed by [6], the article is represented through the access control policy (called Alice's policy) reported in the listing below. The policy allows a lawfulness of processing of Personal Data related to Alice and it is composed of two rules (**R1** and **R2**):

Alice's Policy:

R1: permission(data_controller=myFitness, data_subject=Alice, personal_data={Blood Cholesterol, Age, Gender}, purpose=MyCholesterol, action={DERIVE, AGGREGATE, QUERY}, consent=TRUE)

R2: permission(data_controller={myFitness, address}, data_subject=Alice, personal_data=Email, purpose=UntargetedMarketing, action=SEND, consent=TRUE)

For instance, R1 allows `data_controller` (who) to process `personal_data` (which resources) because of the `consent` (under which circumstances).

GDPR-based Model Derivation (Step ①). According to the GROOT methodology presented in the previous section, the GDPR-based Model is parsed to derive the PAR, and the associated values of the parameters. In the case of Alice's policy, the identified set of parameters derived from the policy elements is $PAR \subseteq \{DS, PD, DC, C, P, PA\}$. For instance, the values associated with parameter P is $V_P = \{MyCholesterol, UntargetedMarketing\}$. In line with Definition 1, the result of this step is represented in tabular form in Table 1. The first column (labeled PE) reports the related Alice's policy elements, the second column (labeled PAR) reports the derived parameters, and the last column (labeled V_{PAR}) lists the related values.

Test Cases Generation (Step ②). The combination of the parameters' values of Table 1 is computed in order to derive the set of test cases. Different strategies can be adopted in this step. By considering the all-combination, for each parameter $j \in PAR$, the power set of the associated values is derived. For instance, the power set associated with parameter P (i.e., Purpose) is P_{V_P} = {{}, {UntargetedMarketing}, {MyCholesterol}, {UntargetedMarketing, MyCholesterol} }. Possible test cases are $TC_{GDPR}(ATT\_1)$ and $TC_{GDPR}\text{-}(ATT\_2)$ where ATT_1 ={DC=myFitness, DS=Alice PD={Blood Cholesterol}, P= MyCholesterol, PA=DERIVE, C=TRUE} and ATT_2 ={C=myFitness, DS=Alice, PD={Email,Age}, P=UntargetedMarketing, PA=SEND}.

For all-combination the cardinality of the derived test suite is 16.384, because the number of test cases follows exponential growth with the numbers of values' parameters. The number of generated test cases can be reduced by considering different approaches. For instance, by applying the pairwise technique the

PE	PAR	V_{PAR}
data_subject	DS = Data Subject	Alice
personal_data	PD = Personal Data	Blood Cholesterol, Age, Gender, Email
data_controller	DC = Controller	myFitness, Address
consent	C = Consent	TRUE
purpose	P = Purpose	UntargetedMarketing, MyCholesterol
action	PA = Processing Activity	DERIVE, AGGREGATE, QUERY, SEND

Table 1. GDPR-based SUT Model Associated of Alice's policy.

cardinality of test suite has been reduced to 259 covering the 16.384 variants. However, it is out of the scope of this paper discussing solutions for managing the explosion problem of combinatorial testing. For more details, we refer to [12].
Test Cases Translation (Step ③). Finally, each of the obtained test cases is translated into an executable one. In the context of AC, possible AC requests, associated with $TC_{GDPR_{ATT\_1}}$ and $TC_{GDPR_{ATT_2}}$ respectively, are reported below. For instance, **Req1** states that *myFitness* (who) wants to process *Blood Cholesterol* (which resources) for *MyCholesterol* purpose (under which circumstances).

Example of Access Control Requests using GROOT:

Req1: request(DC=myFitness, DS=Alice PD=Blood Cholesterol, P=MyCholesterol, PA=DERIVE, C=TRUE)
Req2: request(C=myFitness, DS=Alice, PD={Email,Age}, P=UntargetedMarketing, PA=SEND)

4 Conclusions and Future Work

In this paper, we presented GROOT, a combinatorial testing strategy specifically conceived for assessing the compliance with the GDPR of systems managing personal data. We have firstly presented the conceived methodology, which consists of three main steps, then we have exemplified its application by considering a realistic use case scenario coming from fitness environment. In particular, we illustrated how to apply GROOT for testing GDPR-based access control policies. It is part of our work-in-progress the assessment of the GROOT approach by considering real case studies as well as the use of mutation approaches for evaluating its test effectiveness. We are also working on the GROOT implementation in order to automatize the overall proposed process. As a future work, we will customize GROOT approach by considering other technologies such as consent management systems.

Acknowledgement

This work is partially supported by the project BIECO H2020 Grant Agreement No. 952702, and by CyberSec4Europe H2020 Grant Agreement No. 830929.

References

1. Abassi, R., El Fatmi, S.G.: Security policies a formal environment for a test cases generation. In: Artificial Intelligence and Security Challenges in Emerging Networks, pp. 237–264. IGI Global (2019)
2. Daoudagh, S., Lonetti, F., Marchetti, E.: XACMET: XACML testing & modeling. Softw. Qual. J. **28**(1), 249–282 (2020)
3. Daoudagh, S., Marchetti, E.: A life cycle for authorization systems development in the GDPR perspective. In: Proc. of the 4th Italian Conference on Cyber Security, Ancona, Italy, February 4-7, 2020. CEUR, vol. 2597, pp. 128–140 (2020)
4. Daoudagh, S., Marchetti, E.: Graduation: A gdpr-based mutation methodology. In: Quality of Information and Communications Technology. pp. 311–324. Springer International Publishing, Cham (2021)
5. Daoudagh, S., Marchetti, E., Savarino, V., Bernardo, R.D., Alessi, M.: How to improve the GDPR compliance through consent management and access control. In: Proc. of the 7th International Conference on Information Systems Security and Privacy, ICISSP 2021, February 11-13, 2021. pp. 534–541. SCITEPRESS (2021)
6. Davari, M., Bertino, E.: Access control model extensions to support data privacy protection based on GDPR. In: IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, December 9-12, 2019. pp. 4017–4024. IEEE (2019)
7. Drozdowicz, M., Ganzha, M., Paprzycki, M.: Semantic access control for privacy management of personal sensing in smart cities. IEEE Transactions on Emerging Topics in Computing pp. 1–1 (2020)
8. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation). Official Journal of the European Union **L119**, 1–88 (May 2016), http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC
9. Khamaiseh, S., Chapman, P., Xu, D.: Model-based testing of obligatory ABAC systems. In: 2018 IEEE International Conference on QRS 2018, Lisbon, Portugal, July 16-20, 2018. pp. 405–413. IEEE (2018)
10. Mahindrakar, A., Joshi, K.P.: Automating gdpr compliance using policy integrated blockchain. In: 2020 IEEE 6th Intl BigDataSecurity, IEEE Intl HPSC and IEEE Intl IDS. pp. 86–93 (2020)
11. Mougiakou, E., Virvou, M.: Based on gdpr privacy in uml: Case of e-learning program. In: 2017 8th International Conference on Information, Intelligence, Systems Applications (IISA). pp. 1–8 (2017)
12. Nie, C., Leung, H.: A survey of combinatorial testing. ACM Computing Surveys (CSUR) **43**(2), 1–29 (2011)
13. Pandit, H.J., Fatema, K., O'Sullivan, D., Lewis, D.: Gdprtext - gdpr as a linked data resource. In: The Semantic Web. pp. 481–495. Springer, Cham (2018)
14. Robaldo, L., Bartolini, C., Palmirani, M., Rossi, A., Martoni, M., Lenzini, G.: Formalizing gdpr provisions in reified i/o logic: the dapreco knowledge base. Journal of Logic, Language and Information **29**(4), 401–449 (2020)
15. Sandhu, R.S., Samarati, P.: Access control: principle and practice. IEEE Communications Magazine **32**(9), 40–48 (Sep 1994)
16. Torre, D., Soltana, G., Sabetzadeh, M., Briand, L.C., Auffinger, Y., Goes, P.: Using models to enable compliance checking against the gdpr: an experience report. In: 2019 ACM/IEEE 22nd International Conference, MODELS. pp. 1–11. IEEE (2019)
17. Zhang, Y., Zhang, B.: A new testing method for xacml 3.0 policy based on abac and data flow. In: 2017 13th IEEE International Conference on Control Automation (ICCA). pp. 160–164 (2017)