Process Mining Meets GDPR Compliance: The Right to be Forgotten as a Use Case

Rashid Zaman and Marwan Hassani

Process Analytics Group, Faculty of Mathematics and Computer Sceince,
Eindhoven University of Technology, The Netherlands
{r.zaman,m.hassani}@tue.nl

Abstract. In a bid to ensure privacy of personal data of data subjects, the General Data Protection Regulation(GDPR) entails stringent obligations on organizations and businesses qualifying as data controllers and data processors. The regulation additionally bestow data subjects certain rights over their personal data, *right to be forgotten* generally being perceived the landmark. Fulfilling the GDPR's obligatory requirements and provisioning of the data subject's rights implicates considerable changes to the existing (pre-GDPR era) business and organizational processes. Being a non-trivial task, several technical as well as procedural challenges are associated. The case for organizations having intertwined or cascaded business processes and business processes stretched over multiple organizations is even more complicated. Process mining discipline has been found highly effective in automatically discovering, conformance/compliance analysis, and enhancement of business processes, organizational workflows, healthcare procedures/guidelines to name a few. Process mining techniques therefore have a great potential to assist and guide the transformation of pre-GDPR era (presumably GDPR non-compliant) business or organizational processes into GDPR-compliant processes, and afterwards monitor the compliance during execution. In addition to the current state of the art offline process mining techniques, stable online conformance checking and online model repair techniques needs to be developed for ensuring compliance to the regulation. We are highlighting the challenges associated with implementation of the right to be forgotten, and the GDPR in general.

Keywords: GDPR · Business processes · Compliance · Conformance · Right to be Forgotten.

1 Introduction

With increasing dependence and usage of internet by people and integration of software systems to interface and facilitate these users, either implicitly or

Published in the proceedings of the Doctoral Consortium of ICPM 2019.

explicitly trails are left behind. These trails contains usage behavior information as well as personal data of the users. Traditionally, organizations have been processing this data for discovering patterns and useful insights to support their business decision-making. Apart from in-house processing, data has also been outsourced to third parties for processing on behalf of the outsourcing body or for purposes specific to the outsourced body, probably not in line with the primary logging purpose. Processing of data in this manner and some data breaches have raised privacy concerns at the customers side.

To cope with the privacy issues arising from storage, access and processing of the (personal) data of users, the European Union adopted the General Data Protection Regulation(GDPR)[1] in 2018. Calling for considering "privacy by design" and "privacy by default", the GDPR impose strict measures on organizations and businesses regarding the processing to ensure privacy of users data. Another set of regulatory obligations of the GDPR, centered around the data subjects, includes but not limited to right to be informed, access to personal data, rectification, portability, restrict processing, and most importantly erasure or forgotten.

Organizations are facing challenges in the GDPR implementation. In some cases, implementing the GDPR requirements upto certain extent prove disadvantageous to the pre-GDPR era working mechanism of businesses. To cope with the challenges associated with transition of business processes from GDPR non-compliance to GDPR compliance and to ensure the compliance prevail throughout process life, Business Process Re-engineering and functional toolkit for GDPR compliance (BPR4GDPR) project[2] has been initiated. The resulting framework of tools and engines will provide support in implementing the major GDPR provisions and will be applicable to broad spectrum of business domains and processes.

BPR4GDPR lifecycle, refer Figure 1, starts with the process identification phase, proceeding with adaptation of business processes to the GDPR compliant version. The adaptation phase is followed by continuous monitoring to detect any execution deviations along the process life. Process mining discovery and conformance/compliance techniques are the most suitable candidates for complementing all the mentioned phases. Therefore, a privacy-aware process miner will be developed as part of the BPR4GDPR holistic framework.

This paper highlights the challenges associated with the GDPR implementation and compliance in general, specifically the challenges associated with granting the right to be forgotten (referred to as RTBF in rest of the text) to the data subjects. Process mining perspective to the problem is presented as well. The scenario is elaborated in the light of an automotive lead generation use case.

The remainder of this paper is structured as follows. Section 2 provides an overview of the most relevant process mining techniques and work done in connection with implementing the GDPR in business process landscape. Section 3 briefly discuss the RTBF and its impact on business processes. Section 4 provides

[1] https://gdpr-info.eu

[2] http://www.bpr4gdpr.eu

problem definition. In section 5 we present the use case for the scenario while in Section 6 we present the challenges associated with implementing the GDPR in general and RTBF specifically. Section 7 provides overview of the envisaged privacy-aware process mining.

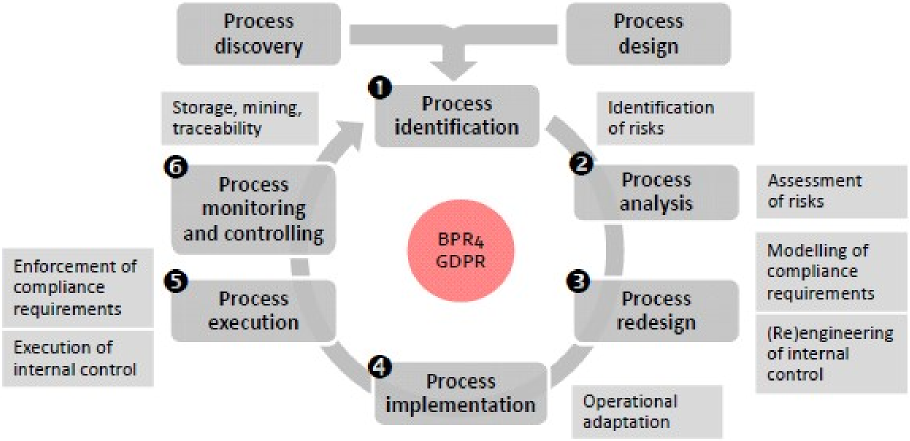

Fig. 1. BPR4GDPR lifecycle

2 Related Work

Process mining [1], though a relatively young discipline, has strategically positioned itself in the business processes landscape. The three main process mining techniques namely Process Discovery, Process Conformance Checking, and Process Enhancement have entered the state of maturity, especially on static event logs of reasonable size. Realization of these techniques for streaming event data [2] and data of massive scale is evolving. Process conformance and compliance checking are of prime importance for the GDPR compliance therefore we will briefly discuss the work done in these relevant areas.

Process conformance checking techniques [3, 4] confronts event log (observed behavior) with business process model (desired behavior) to detect deviations and accordingly assess the inter-harmony level. Process compliance checking [5–8] aims at analyzing the execution of observed behavior in accordance with requisite business rules, desired practices, regulation etc. Compliance checking, in contrast to control-flow orientation of conformance checking, takes into consideration the various characteristics of the activities, like, their order of execution, cardinality of their execution, their predecessor and successor activities, and the associated data or resource attributes.

Some rear work specifically on the GDPR obligatory requirements or similar lines in the context of business processes exist. [9] match audit trails with legitimate execution sequences as per model. Sequences deviating from the legitimate execution sequences are considered as infringements from the specified processing purpose(s). [10] apart from algorithmically bridging process collection to privacy policy, naïvely classifies unused data classes as exception to data minimisation.

3 GDPR Right to be Forgotten

In this section we are introducing RTBF and explaining the impact of RTBF on existing business processes.

3.1 RTBF

Article 17 of the General Data Protection Regulation titled Right to Erasure (mostly known as Right to be Forgotten) entitles data subjects to ask data controller(s) for erasure of their personal data and accordingly binds the controller(s) to erase the requested personal data without undue delay. Although in some exceptional cases the right can be denied, in majority of situations and business cases it remains unavoidable.

3.2 Impact of RTBF on Businesses

RTBF presents major and far-reaching impact on majority of the business processes starting from implementation. Conventionally, having no such obligation in place, organizations and businesses have been processing personal data of people, in cases without explicit consent, upto the extent required for their businesses, retaining it for longer periods, and even subletting it to other businesses to be used for their own processing purposes.

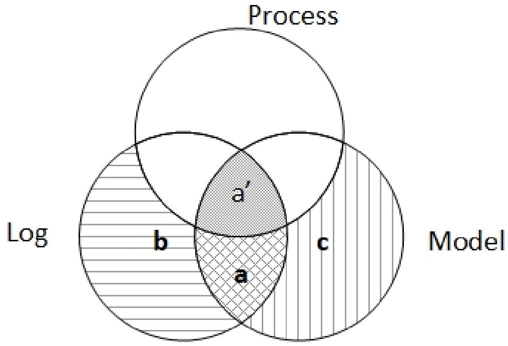

Fig. 2. Model, Log and Process

In process-model-log nomenclature, referring to Figure 2, usually only a portion (for example "a" and "a$'$" in the figure) of the log "L" and model "$\mathcal{M}$" are in conformance while remaining portions (for example "b" of the log "L" and "c" of the model "$\mathcal{M}$") are usually non-conforming. Inclusion of the GDPR obligations "G" to the existing process and consequently model "$\mathcal{M}$" shall further push away the log "L" and model "$\mathcal{M}$" thus resulting in increased non-conformance. Due to severe financial consequences of the GDPR irregularities, such non-conformance is more critical in nature in comparison to conventional non-conformance therefore necessitating forward compliance through model adaptation and backward compliance through monitoring.

4 Problem Definition

A pre-GDPR era process model, supposedly the GDPR non-compliant, is a Petri net, $N = (P, T, F, A)$ where P in the tuple represents a finite set of places, T is the set of transitions, F is the set of arcs connecting places to transitions and transitions to places, and A is the set of activities.

In post-GDPR era, processes need to have an erasure mechanism in place for data subjects D_{s} to be able to exercise right to be forgotten. The erasure mechanism itself can be conceived as a (sub)process and therefore a Petri net, $N_{\mathrm{Grtbf}} = (P_{\mathrm{Grtbf}}, T_{\mathrm{Grtbf}}, F_{\mathrm{Grtbf}}, A_{\mathrm{Grtbf}})$ where $P_{\mathrm{Grtbf}}, T_{\mathrm{Grtbf}}, F_{\mathrm{Grtbf}}, A_{\mathrm{Grtbf}}$ represents the places, transitions, transitions and places inter-relational arcs, and activity labels in the GDPR erasure (sub)process model respectively.

Business process model N and erasure process N_{Grtbf} are disjoint in nature i.e., $T \cap T_{\mathrm{Grtbf}} = \emptyset$. Therefore, a blanket GDPR-compliant version of N can be obtained by overlaying erasure process N_{Grtbf} over the business process model N i.e., $N + N_{\mathrm{Grtbf}}$ or $N \oplus N_{\mathrm{Grtbf}}$. Due to the disjointedness of the N and N_{Grtbf}, $N \oplus N_{\mathrm{Grtbf}}$ distinguishes from the conventional model merging [11] and model repair [12] techniques.

From behavioral point of view, in case of an erasure request, the N_{Grtbf} shall seize further processing of the relevant data, in other words reset the N. Therefore, $N \oplus N_{\mathrm{Grtbf}}$ becomes a *Reset net* (P, T, F, R) where R is the Reset Arcs defining function. An erasure request in Marking M shall result in Marking M' such that $M' = \mathit{\Pi}_{\mathrm{P} \oplus P_{Grtbf} \setminus R(t)\, +\, P}$ without undue delay.

5 Use Case

Conventionally, automotive car dealerships acquire data from multiple sources associated with automotive industry, for instance, repair/maintenance services providers, automotive showcase events organizers, automobiles manufacturers. Referring to the non-shaded left-top part of the Figure 3, these sources generate/store data their customers data through transactions and interactions, for instance, people recording interest in vehicles at showcase events.

Data is acquired from these multiple sources by car dealerships and processed for identifying quality leads. Identified leads are accordingly contacted for the

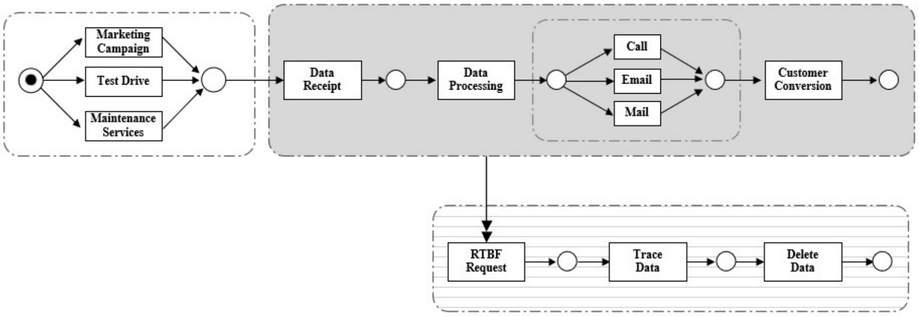

Fig. 3. Lead Generation process model with Blanket GDPR RTBF

purpose of conversion to customer, up-sell or cross-sell, refer right-top shaded part of Figure 3. In order to be GDPR RTBF compliant, RTBF process N_{Grtbf}, refer to the pattern-filled lower part of Figure 3, is annexed to the process model *N*, resulting in a blanket GDPR RTBF compliant version *(P,T,F,R)*.

6 Challenges

The challenges associated with the GDPR implementation in general and blanket RTBF GDPR adaptation specifically are presented in the light of the automotive lead generation use case.

6.1 Model Adaptation

Blanket GDPR RTBF compliant models *(P,T,F,R)* looks trivial to realize but in essence poses a threat to the working model of many businesses. Devising process models protecting business interests and ensuring the GDPR compliance at the same time are a challenge to the community. Example of such threats in automotive lead generation are customers (promptly) exercising RTBF after violating traffic rules during test drive with the car dealers. This and many other possible threats, resulting from the blanket GDPR adaptation, potentially leads the system to in-consistent state. Safe models, being compliant with the GDPR and caring for the intricacies of the incumbent business as well, are intended.

6.2 Monitoring

Conventional conformance/compliance techniques are mainly oriented towards control flow and associated characteristics of activities. The GDPR is concerned with the privacy of the data being accessed and processed during execution of business process activities. In the case of automotive lead generation process, identification of quality leads involves processing the personal and financial

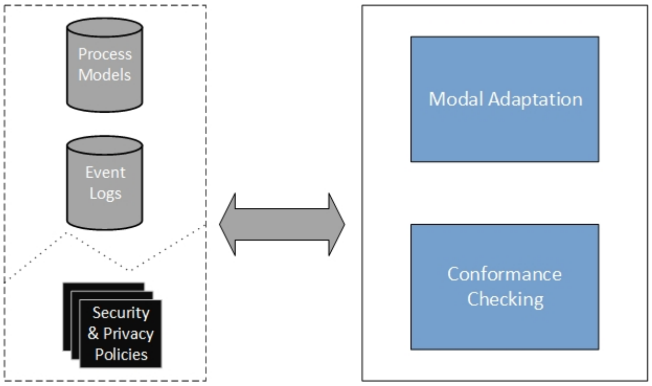

Fig. 4. Privacy-aware process miner

particulars of the data subjects. Therefore, conformance/compliance techniques in this perspective needs to take into account the access policies, processing policies, data anonymisation and encryption policies etc. and perform conformance/compliance checking in the light of these policies. In our use case, the unit responsible for correspondence with the leads shall have limited access to the leads data necessary for job in hand like correspondence particulars but not the financial details of the leads.

6.3 Online Conformance Checking

Non-conformance to the GDPR has severe consequences, particularly financial ones. Therefore, backward conformance/compliance checking shall not be relied upon completely. Online conformance checking techniques, having the potential to detect deviations at the point in time when they happen, needs to be devised to avoid or mitigate non-conformance. Scalability is going to be a major issue for online conformance checking techniques.

6.4 Online Model Adaptation

It is not a viable solution to take processes offline for adaptation(s) in light of non-conformance, unless complete re-designing is unavoidable. Therefore, online model adaptation techniques needs to be developed which shall automatically detect any changes in business working and update the process model accordingly at run-time in light of the changes detected [13]. A challenge for the approach will be to distinguish noise from real changes.

7 Privacy-aware process mining

Process mining is conventionally oriented towards control, data and resource perspectives of processes. Privacy perspective for process mining shall additionally take into consideration the privacy of the data and data artifacts manipulated by the process activities during execution. The GDPR has different requirements for different states of data, for instance mandatory consent acquisition from data subjects at collection of (minimal) data, processing of data inline with the acquired consent, encryption of data during lawful transit to third parties or countries, and anonymisation while resting the data. Refer to Figure 4, the two main constituents of our envisaged privacy-aware process miner are model adaptation and conformance checking. In addition to process model(s) and process event logs, security and privacy policies in an adequate formalism is the third required input for the miner.

Model adaptation module shall take into consideration the relevant GDPR requirements and adapt the process in hand such that it becomes fully GDPR compliant, while syntactically remaining as close as possible to the existing version. Considerable changes in the underlying process model can be quantified using one of the distance measures introduced in [13] to detect concept drifts. Conformance checking in the privacy-aware process miner kit shall check for the conformance of the event log with process model, taking into consideration the relevant security and privacy policies as well.

Acknowledgments

The authors have received funding within the BPR4GDPR project from the European Union's Horizon 2020 research and innovation programme under grant agreement No 787149.

References

1. van der Aalst, W.M.P.: Process Mining: Data Science in Action, 2nd edn. Springer, Berlin (2016). https://doi.org/10.1007/ 978-3-662-49851-4
2. van Zelst, Sebastiaan J., Alfredo Bolt, Marwan Hassani, Boudewijn F. van Dongen, and Wil MP van der Aalst."Online conformance checking: relating event streams to process models using prefix-alignments." International Journal of Data Science and Analytics (2017): 1-16.
3. Adriansyah, Arya, Boudewijn F. van Dongen, and Wil MP van der Aalst. "Conformance checking using cost-based fitness analysis." Enterprise Distributed Object Computing Conference (EDOC), 2011 15th IEEE International. IEEE, 2011.
4. Carmona, Josep, Boudewijn van Dongen, Andreas Solti, and Matthias Weidlich. Conformance Checking: Relating Processes and Models. Springer, 2018.
5. Ramezani, Elham, Dirk Fahland, and Wil MP van der Aalst. "Where did i misbehave? diagnostic information in compliance checking." International conference on business process management. Springer, Berlin, Heidelberg, 2012.
6. Ramezani, Elham, Dirk Fahland, and Wil MP van der Aalst. "Supporting domain experts to select and configure precise compliance rules." International Conference on Business Process Management. Springer, Cham, 2013.
7. Taghiabadi, Elham Ramezani, et al. "Compliance checking of data-aware and resource-aware compliance requirements. " OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". Springer, Berlin, Heidelberg, 2014.

8. De Leoni, Massimiliano, and Wil MP van der Aalst. “Aligning event logs and process models for multi-perspective conformance checking: An approach based on integer linear programming.” Business Process Management. Springer, Berlin, Heidelberg, 2013. 113-129.
9. Petković, Milan, Davide Prandi, and Nicola Zannone. “Purpose control: Did you process the data for the intended purpose?.” Workshop on Secure Data Management. Springer, Berlin, Heidelberg, 2011.
10. Basin, David, Søren Debois, and Thomas Hildebrandt. “On purpose and by necessity: compliance under the GDPR.” FC. Springer, Berlin Heidelberg (2018).
11. La Rosa, Marcello, et al.“Merging business process models.” OTM Confederated International Conferences” On the Move to Meaningful Internet Systems”. Springer, Berlin, Heidelberg, 2010.
12. Fahland, Dirk, and Wil MP van der Aalst. “Model repair—aligning process models to reality.” Information Systems 47 (2015): 220-243.
13. Hassani, Marwan, “Concept Drift Detection Of Event Streams Using An Adaptive Window.” International ECMS Conference On Modelling And Simulation (to appear).