Compliance to data protection and purpose control using process mining technique

Azadeh Sadat Mozafari Mehr

Eindhoven University of Technology, the Netherlands
email a.s.mozafari.mehr@tue.nl

Abstract. The business processes of an organisation are executed in certain boundaries. Some of the restrictions are raised from the environment of the organisations such as regulatory and supervisory constraints. One of the regulations that is imposed on organisations is the European General Data Protection Regulation (GDPR). The most important aspect of the GDPR rules is how organisations handle personal data of their customers. In this research, we focus on this aspect of the GDPR. Our goal is to develop a solution that enables organisations to deal with the challenges of becoming compliant with GDPR. We plan to use and improve process mining techniques to tackle the problems such as discovering data-flow and control-flow of business processes that have interaction with personal data of customers. Our approach consists of four phases: (1) discover process model based on purpose, (2) translate regulatory rules to technical rules, (3) develop privacy policy model base on the GDPR, (4) conformance analysis.

Keywords: Process Mining· GDPR· Compliance Checking · Rule Translation

1 Introduction

Generally, business processes of an organisation are executed in certain boundaries. These restrictions are defined by a set of business rules. Some of these rules are regulatory and supervisory constraints, some are based on the domain standards, some are implemented according to the internal regulations of each organisation and many are defined by the trading partners [17]. The complexity of the organisation processes, thus, has increased significantly and there is need for constant monitoring of the implementation and execution of these complex processes and assessment of their conformance with business rules. Given the issues raised, organisations face major challenges. They must manage complex organisational and cross-organisational processes, monitor running processes and examine conformance with all business rules defined by managers, governments and stakeholders [14]. One of the regulations that is imposed on organisations is the European General Data Protection Regulation (GDPR) [9]. The most important aspect of the GDPR rules is how organisations handle personal data of their customers, in which “the purpose of using data by organisation” and “the

consent of the customer" have more prominent roles. New and improved rights for the customer, such as "the right to be forgotten", impacts companies because such rights need to be accommodated in their internal processes. In this work we focus on these important aspects of GDPR and address the most important challenges that companies may face to become compliant:

• Most organisations rarely have a global picture and knowledge about the data-flow between their business processes. Therefore, they cannot find the points of the whole process that need improvement to become compliant with the GDPR [2].

• For auditing or internal assessment of their processes, organisations need to identify which process activities interact with personal data of the customers. At the next step, they require a mechanism to distinguish between personal data and other information [7]. Furthermore, they should investigate whether the data is used for the intended purpose [13].

• Organisations should handle their processes in such way to allow users to consent to some, but not all processes [7]. To this end, they should clearly state their privacy policy, the purpose of collecting and using the data. This can only be done when the organisation itself has a clear view of data-flow and control-flow of its processes in reality.

• Organisations should track the consent of their customers. Revoking the consent has impacts on the execution of the activities that use personal data. On the one hand, current process cannot use the data and on the other hand, processes that may be run in the future might not be able to use the data of those group of customers. For this reason, organisations require to identify the processes that are associated with personal data at the runtime. We are investigating the challenges and we aim to concentrate on developing solutions to deal with the mentioned challenges. We plan to use and improve process mining techniques to tackle the problems such as discover the data-flow and control-flow of business processes that have interaction with personal data of customers. Our goal is to develop a mechanism to translate regulatory rules(in this work GDPR rule set) into technical rules and model GDPR compliance rule set as patterns. Finally, we plan to use conformance techniques to assess whether the discovered process model is complied with GDPR compliance rule.

2 Approach and related work

Our overall goal is to provide a framework for checking compliance with data protection and purpose control using process mining techniques. To implement the solution, our approach consists of four main steps(Fig.1):

1. Discover process model based on purpose: Usually, process mining is performed to discover, monitor and improve real processes (not assumed or modeled processes) and is based on the knowledge extracted from the event logs of information systems [12, 4]. Typically, there is a gap between a modeled process and the trend expected to occur in the process run and what occurs in reality. The goal of process mining is to identify and decrease problems caused

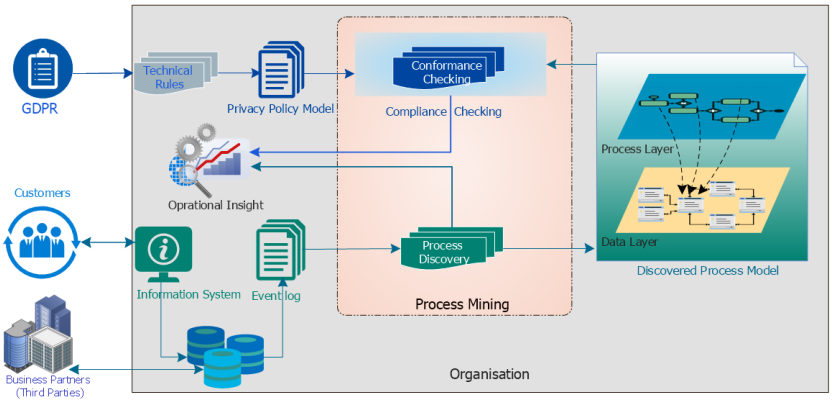

Fig. 1: Overview of the Proposed Approach

by these deviations. Process mining techniques and algorithms are categorized as three main types including discovery, conformance, and extension [4]. Process discovery is the most common process mining technique and can automatically detect and generate a process model based on the event log and reality [3]. We have a plan to continue and improve the discovery technique of "Object-Centric Process Mining" [10]. In[10], a modeling language which combines data models with declarative models to discover an object-centric behavioral constraint model was proposed. The goal of this study was to discover a process model from the log where the scope is data/object-centric processes supported by CRM and ERP systems. Since customer is a basic concept in designing such systems,initially by considering structured data model, we aim to use their approach to find in which parts of the process execution there are interactions with customer data. In the next step, we intend to extend their algorithm and add purpose specification to the core part of it. In [10], although the approach is capable to find the intersection points between the process layer and data layer, it cannot discover the purpose of this association. In this research, we aim to cover this gap.

2. Translate regulatory rules to technical rules: A fundamental step towards the solution is regulatory analysis to develop a concrete understanding of the underlying business needs. This analysis must be done from two points of view: the regulations on data protection, particularly the GDPR, and the business needs for facilitating compliance thereof. Regarding to our goals, policies that should be formalized as a rule include concepts such as the data ground under which personal data falls, roles of the entities requesting and processing personal data, operations and services performed over personal data, attributes of all the involved entities, purposes of requesting/processing data. Therefore, having the regulatory analysis as the starting point, generalization and creation of an abstract model of privacy policy base on the GDPR comes next.

3. Develop privacy policy model base on the GDPR: To check compliance of the log with GDPR, at first we should translate the GDPR rule set to patterns. There are two basic types of compliance checking: forward and

backward compliance checking [14]. At this phase we plan to focus on backward compliance checking based on event data. Backward compliance checking assesses compliance between process executions and all compliance rules, the result will show when and where a particular rule was violated. A variety of conformance checking techniques have been proposed based on an event log and process model (e.g., Petri-net) [5, 6, 8, 16]. [1, 11] proposed approaches based on temporal logic. In [15], the authors provide Petri-net patterns modeling typical compliance rules. These rules can be instantiated for a particular process, i.e., the abstract activities in the pattern are replaced by concrete activities recorded in the event log. The log complies to the rule if each log trace is described by the Petrinet pattern. In case a trace is not described, they locate where the trace deviates from the pattern. In this work, they focus on compliance with control-flow. As our final goal is checking compliance of discovered models with GDPR, to this end, similarly to the work [15], we intend to model and generalize the mentioned aspect of the GDPR as an abstract process model. We realized that this model should include and concern two concepts of "the purpose of each activity" -that can be one of the followings: (DC) Data collection,(DU) data usage,(NA) none of them or (DC/DU) both of them- and "the consent of the customer". Since the OCBC model is based on the declarative process model, we cannot use their idea to generalize and model the GDPR rule set as Petri net models. In their work, they focus on compliance to control-flow whereas in our work we require to concentrate on compliance to data-flow more than control-flow.

4. Conformance analysis: In conformance checking, an existing model is compared with the model discovered from the event log. The goal of this technique is to assess whether a process model discovered from the log conforms to the assumed and predetermined process. This technique can be applied to different aspects of the process, such as for an ideal process model, organisational vision and business rules and policies [16]. In this phase, we will use this technique to check the compliance of business process with the GDPR.

3 Conclusion

In this paper, we mentioned the most important challenges that organisations meet to become compliant with the GDPR. We defined our goals and the scope that we plan to concentrate on. We outlined our research plan to provide a framework for checking compliance to data protection and purpose control using process mining techniques. Besides explaining each phase of our approach, we summarized the related works and techniques that we intend to use or extend at our future plan for the implementation of each phase.

Acknowledgment

The author has 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., de Beer, H.T., van Dongen, B.F.: Process mining and verification of properties: An approach based on temporal logic. In: Meersman, R., Tari, Z. (eds.) On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE. pp. 130–147. Springer Berlin Heidelberg, Berlin, Heidelberg (2005)
2. van der Aalst, W.M.P., Schmiedel, T.: Extracting Event Data from Databases to Unleash Process Mining, pp. 105–128. Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-14430-6_8, https://doi.org/10.1007/978-3-319-14430-6_8
3. Van der Aalst, W.: Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer, Verlag Berlin Heidelberg (2011). https://doi.org/10.1007/978-3-662-49851-4, https://doi.org/10.1007/978-3-662-49851-4
4. Van der Aalst, W.: Data Science in Action. Springer, Verlag Berlin Heidelberg (2016). https://doi.org/10.1007/978-3-642-19345-3, https://doi.org/10.1007/978-3-642-19345-3
5. Aalst, van der, W., Adriansyah, A., Dongen, van, B.: Replaying history on process models for conformance checking and performance analysis. WIREs Data Mining and Knowledge Discovery **2**(2), 182–192 (2012). https://doi.org/10.1002/widm.1045
6. Adriansyah, A., Dongen, van, B., Aalst, van der, W.: Conformance checking using cost-based fitness analysis. In: Chi, C., Johnson, P. (eds.) Proceedings of the 15th IEEE International Conference on Enterprise Distributed Object Computing (EDOC 2011, Helsinki, Finland, August 29-September 2, 2011). pp. 55–64. Institute of Electrical and Electronics Engineers (IEEE), United States (2011). https://doi.org/10.1109/EDOC.2011.12
7. Basin, D., Debois, S., Hildebrandt, T.: On purpose and by necessity: compliance under the gdpr. In: Business Process Management. Financial Cryptography and Data Security, Nieuwport, Curaçao (2018)
8. Calders, T., Günther, C.W., Pechenizkiy, M., Rozinat, A.: Using minimum description length for process mining. In: Proceedings of the 2009 ACM symposium on Applied Computing (SAC'09). pp. 1451–1455. ACM Press (2009). https://doi.org/http://doi.acm.org/10.1145/1529282.1529606
9. of the European Union, C., Parliament, E.: 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 (general data protection regulation) (2016)
10. Li, G., de Carvalho, R., van der Aalst, W.: Automatic discovery of object-centric behavioral constraint models. In: Abramowicz, W. (ed.) Business Information Systems. pp. 43–58. Lecture Notes in Business Information Processing, Springer, Germany (6 2017). https://doi.org/10.1007/978-3-319-59336-4_4
11. Montali, M., Pesic, M., Aalst, W.M.P.v.d., Chesani, F., Mello, P., Storari, S.: Declarative specification and verification of service choreographiess. ACM Trans. Web **4**(1), 3:1–3:62 (Jan 2010). https://doi.org/10.1145/1658373.1658376, http://doi.acm.org/10.1145/1658373.1658376
12. de Murillas, E.G.L., van der Aalst, W.M.P., Reijers, H.A.: Process mining on databases: Unearthing historical data from redo logs. In: Motahari-Nezhad, H.R., Recker, J., Weidlich, M. (eds.) Business Process Management. pp. 367–385. Springer International Publishing, Cham (2015)

13. Petković, M., Prandi, D., Zannone, N.: Purpose control: Did you process the data for the intended purpose? In: Jonker, W., Petković, M. (eds.) Secure Data Management. pp. 145–168. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)
14. Ramezani, E., Fahland, D., van der Aalst, W.M.P.: Where did i misbehave? diagnostic information in compliance checking. In: Barros, A., Gal, A., Kindler, E. (eds.) Business Process Management. pp. 262–278. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
15. Ramezani Taghiabadi, E., Fahland, D., van Dongen, B.F., van der Aalst, W.M.P.: Diagnostic information for compliance checking of temporal compliance requirements. In: Salinesi, C., Norrie, M.C., Pastor, Ó. (eds.) Advanced Information Systems Engineering. pp. 304–320. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
16. Rozinat, A., van der Aalst, W.M.P.: Conformance checking of processes based on monitoring real behavior. Inf. Syst. **33**(1), 64–95 (Mar 2008). https://doi.org/10.1016/j.is.2007.07.001, http://dx.doi.org/10.1016/j.is.2007.07.001
17. Skopik, F., Schall, D., Dustdar, S.: Modeling and Mining of Dynamic Trust in Complex Service-Oriented Systems, pp. 29–75. Springer Vienna, Vienna (2011), https://doi.org/10.1007/978-3-7091-0813-0_3