Behavior Verification for Business Processes based on Testing and Anomaly Detection

Kristof Böhmer

University of Vienna, Faculty of Computer Science
kristof.boehmer@univie.ac.at

1 Introduction

Business processes are applied in a wide range of areas, implementing various *mission critical functionalities*. For example, today's processes enable the collaboration of multiple organizations, integrate multiple resources, and handle data from various sources. The latter frequently includes private, sensitive and confidential information (e.g., private user data or medical records). To provide all these functionalities and to meet numerous requirements business processes must be deeply integrated into organizations and related IT systems, cf. [15]. Thus, it is vital to ensure that today's process are as fault and security incident *free* as possible to prevent potential negative effects on process driven organizations.

Current research has already picked up challenges regarding the detection and handling of faults (e.g., based on testing, cf. [10, 2]) and security incidents (e.g., based on anomaly detection, cf. [13, 7]). Nevertheless, the current state of research was found to be limited regarding the analysis of dynamic business *process runtime behavior*. As only runtime behavior represents the real – potentially faulty or insecure – behavior that occurs during the execution of predefined process behavior (e.g., given by process models, cf. [12]) we see this as a major limitation. Moreover, additional limitations in existing work were identified:

First of all, a common understanding, an overview, and requirements for testing approaches – to identify faults in business processes – are currently missing, cf. [2]. Further, it was found that existing process testing approaches frequently require a significant amount of expert knowledge and skills in various areas, such as, lesser-used formal test definition languages, which we assume as being hardly available at process modeling experts. Finally, various process execution scenarios, such as parallel executions are insufficiently supported - rendering a major area in the process domain uncovered.

When analyzing the state of research in the business process runtime security domain (i.e., anomaly detection during business process executions) a different picture emerges, cf. [7]. Existing anomaly detection approaches were found to be severely limited in regard to their robustness and anomaly detection capabilities. For example, anomalies can occur in multiple forms and complexities, cf. [11], but only the most basic ones are currently supported [7]. Further, existing work only focuses on single individual process instances, where the detection of hidden malicious actions requires to aggregate multiple behavior sources (perspectives,

resp.). Moreover, processes handle information from various data sources and formats. Hence, anomaly detection approaches are required that can automatically analyze various data formats for anomalies, to reduce manual efforts.

Finally, we want to point out that business processes have a life cycle that describes their design, execution, monitoring, and optimization. So far the definition of behavior (i.e., process modeling and optimization) and the occurrence of behavior (i.e., process execution and monitoring) have been considered separately. However, both areas are closely connected and must be taken into consideration at once to foster fault free *and* secure processes.

2 Contribution

This thesis concerns on the fault free definition of process behavior *and* the secure anomaly free execution of such behavior, cf. Fig 1, and tackles both areas with multiple publications: [2, 6, 1, 3, 9, 7, 4, 8, 5] Specifically, process testing approaches are defined and analyzed to investigate their applicability, for example, to prevent faults, in the process modeling and optimization phase. Through this the quality of process models is assumed to be increased so that faults are less likely to hamper an organization's performance or to affect the bound of trust between organizations, partners, and customers. For the other two phases, i.e., the execution and monitoring phase, anomaly detection approaches are proposed and applied to prevent security incidents and data breaches. Hereby, anomalous process execution behavior can be identified, detected, and its impact is analyzed to identify abnormal behavior that potentially indicates outside attacks or inside threats. A condensed overview on the contribution is given in the following:

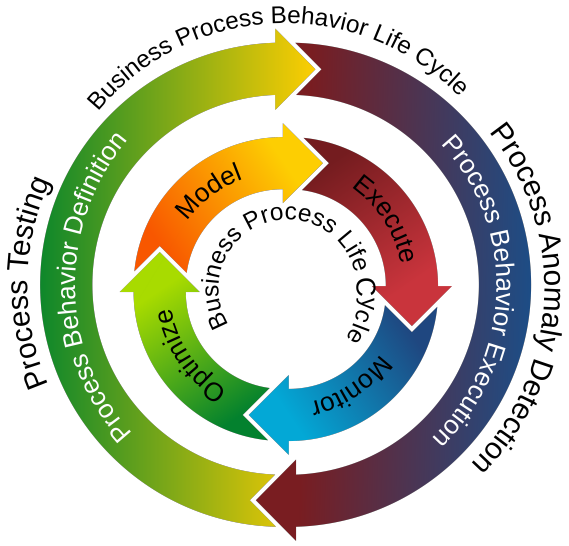

Fig. 1. Classic Business Process Life Cycle and its relation to the Behavior Life Cycle.

- Two *systematic literature reviews* on business process testing and anomaly detection have been conducted [2, 7]. Both provide a detailed investigation along with an analysis of the given support for upcoming challenges, such as, concurrent process behavior – to provide research directions.
- Two approaches are presented to generate, prioritize, and select business process test cases based on historic behavior and machine learning concepts [6, 1, 3]. Hereby, the first approach focuses on the verification of complex interleaving concurrent process model executions and test case prioritization. The latter approach investigates the application of genetic algorithms to select business process test cases. Test case selection and prioritization are a necessity because today's process test suites are frequently auto-generated and through this significantly larger than manually created ones. So executing all test cases would result in escalating test execution times. Hence, reaching a sufficient test execution performance requires to identify, select, and prioritize test cases based on their relevance for given requirements, such as, a chosen minimum process activity coverage.
- A novel testing approach evaluates the applicability of process mining based verification and conformance checking techniques to identify faults [9]. Hereby, it can be exploited that the chosen mining approaches are already deeply rooted in the process domain and include support for a range of characteristics that are unique for that domain. A related experimental evaluation is conducted based on realistic processes from the energy domain.
- A generic unsupervised anomaly detection approach for arbitrary textual data is proposed [4]. The presented approach enables to generate signatures based on historic data exchanges. Furthermore we provide a detailed evaluation with hundreds of thousands of realistic anomalous and non-anomalous data instances to analyze its feasibility and applicability.
- A multi perspective anomaly detection prototype that is capable of detecting point, contextual, and collective anomalies in process executions is presented [5]. It is capable of calculating the likelihood of process execution behavior to differentiate between benign and unlikely anomalous behavior. In addition techniques are proposed to flexibly deal with unexpected and unknown behavior to reduce the number of incorrectly identified anomalies when dealing with evolving, volatile, and changing processes.
- Process models are executed in complex execution scenarios where multiple processes and process instances are executed concurrently and overlap each other. These dynamics are hard to analyze and could, therefore, be utilized by – inside – attackers to hide their malicious intentions. Hence, a multi-instance anomaly detection approach is presented [8]. It transforms recorded historic executions of all processes utilized in an organization into signatures to detect temporal anomalies in concurrently executed large scale process execution scenarios (i.e., it is capable of securing and monitoring all process executions in an organization at once). Its feasibility is shown based on a prototypical implementation and multiple real world data sources.

3 Methodology

This thesis follows the so called *design science* methodology. For this it applies a design science information system research framework that was proposed in [16]. The following list describes how each of the design science research guidelines, as outlined in [16], were taken into account by the thesis:

GL1: Design as an artifact This thesis creates several artifacts that foster the fault-freeness and security of business processes – which is relevant for process driven organizations. For this all the testing and anomaly detection approaches proposed in this work were also prototypically implemented.

GL2: Problem relevance The conducted systematic literature reviews enabled to identify key challenges, gaps, and limitations in the process testing and anomaly detection domain and showed that both research areas are relevant and broadly accepted. This motivated the research questions addressed within this thesis and shows their relevance. Moreover it was found that security is a key concern in the process domain, cf. [14]. This is not surprising, as security incidents can result in a substantial damage to an organization.

GL3: Design evaluation This thesis rigorously evaluates each created artifact. For example, the algorithms, approaches, and concepts presented in this thesis were *all prototypically implemented* and evaluated based on *real life* and/or *realistic synthetic data*. Hereby, the applicability, performance, and functionality of the proposed approaches were shown. In addition identified limitations and future work were pointed out and discussed.

GL4: Research contributions This thesis contributes by creating artifacts that tackle open challenges in the testing and security domain (e.g., algorithms that enable to identify various anomaly types). Moreover methodologies are presented. These are, for example, metrics that enable to measure the fault detection likelihood of test cases or a fault risk for process activities.

GL5: Research rigor This thesis applied rigorous methods and techniques to construct and evaluate the designed artifacts. For this the presented research is evaluated and compared with competing techniques whenever possible. For example, the presented test case selection approach is compared with multiple alternatives. In addition fundamental design decisions are discussed and motivated and formal concepts are applied whenever appropriate.

GL6: Design as a search process This motivated the applied research process. For example, the systematic literature reviews enabled to identify and organize existing knowledge and shortcomings. In addition the iterative nature of the search process, cf. [16], is reflected by our research process. For example, the presented security focused research tackles multiple successive and interrelated research areas – which will, as future work, be joined to a coherent security framework. The intermediate results identified throughout the conducted research and search process have influenced our approaches and are reflected by our research and the constructed artifacts.

GL7: Communication of research The research has been presented on conferences, in proceedings, and journals addressing the BPM, Service, and Database research community: [2, 6, 1, 3, 9, 7, 4, 8, 5]

4 Conclusion

This thesis is focusing on two aspects, i.e., the secure and fault free definition and execution of business processes. Through this the thesis can provide a holistic approach that combines process testing and anomaly detection to ensure not only fault free but also secure process model executions. As the world and our lifes tend to become more and more affected by and oriented towards processes we assume this to be a necessity for organizations, but also for today's society.

References

1. Böhmer, K., Rinderle-Ma, S.: A genetic algorithm for automatic business process test case selection. In: On the Move: CoopIS. pp. 166–184. Springer (2015)
2. Böhmer, K., Rinderle-Ma, S.: A systematic literature review on process model testing: Approaches, challenges, and research directions. CoRR abs/0902.0885 (2015)
3. Böhmer, K., Rinderle-Ma, S.: Automatic business process test case selection: Coverage metrics, algorithms, and performance optimizations. Cooperative Information Systems pp. 174–190 (2016)
4. Böhmer, K., Rinderle-Ma, S.: Automatic signature generation for anomaly detection in business process instance data. In: Enterprise, Business-Process and Information Systems Modeling, pp. 184–199. Springer (2016)
5. Böhmer, K., Rinderle-Ma, S.: Multi-perspective anomaly detection in business process execution events. In: On the Move: CoopIS. pp. 80–98. Springer (2016)
6. Böhmer, K., Rinderle-Ma, S.: A testing approach for hidden concurrencies based on process execution logs. In: Service-Oriented Computing. Springer (2016)
7. Böhmer, K., Rinderle-Ma, S.: Anomaly detection in business process runtime behavior - challenges and limitations. CoRR abs/1705.06659 (2017)
8. Böhmer, K., Rinderle-Ma, S.: Multi instance anomaly detection in business process executions. In: Business Process Management. pp. 80–98. Springer (2017)
9. Böhmer, K., et al.: Application and testing of business processes in the energy domain. Datenbanksysteme für Business, Technologie und Web (2017)
10. Bures, M., Cerny, T., Klima, M.: Prioritized process test: More efficiency in testing of business processes and workflows. In: International Conference on Information Science and Applications. pp. 585–593. Springer (2017)
11. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM Computing Surveys 41(3), 15 (2009)
12. Halfond, W.G., Orso, A.: Combining static analysis and runtime monitoring to counter sql-injection attacks. In: ACM SIGSOFT software engineering notes. vol. 30, pp. 1–7. ACM (2005)
13. Hsu, P.Y., Chuang, Y.C., Lo, Y.C., He, S.C.: Using contextualized activity-level duration to discover irregular process instances in business operations. Information Sciences 391, 80–98 (2016)
14. Leitner, M., Rinderle-Ma, S.: A systematic review on security in process-aware information systems–constitution, challenges, and future directions. Information and Software Technology 56, 273–293 (2014)
15. Niedermann, F., Radeschütz, S., Mitschang, B.: Deep business optimization: A platform for automated process optimization. ISSS/BPSC 2010, 168–180 (2010)
16. Von Alan, R.H., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS quarterly 28(1), 75–105 (2004)