[Array 13 (2022) 100120](https://doi.org/10.1016/j.array.2021.100120)

|  |  |  |
| --- | --- | --- |
|  | Contents lists available at [ScienceDirect](http://www.elsevier.com/locate/array) |  |
| Array |
| journal homepage: [www.elsevier.com/locate/array](http://www.elsevier.com/locate/array) |
|  | | |

Process mining usage in cybersecurity and software reliability analysis: A systematic literature review   
Martin Macaka,∗, Lukas Daubnera, Mohammadreza Fani Sanib, Barbora Buhnovaa   
a *Faculty of Informatics, Masaryk University, Botanická 68a, 602 00 Brno, Czechia*   
b *Process and Data Science Chair, RWTH-Aachen University, Aachen, Germany*

|  |  |  |
| --- | --- | --- |
| A R T I C L E | I N F O | A B S T R A C T |
| *Keywords:*  Process mining  Cybersecurity  Software reliability  Systematic literature review | | The digitalization of our society is only possible in the presence of secure and reliable software systems governing ongoing critical processes, so-called critical information infrastructures. The understanding of mutual interdependencies of events and processes is crucial for cybersecurity and software reliability. One of the promising ways to tackle these challenges is process mining, which is a set of techniques that aims to mine essential knowledge from processes, thus providing more perspectives and temporal context to data interpretation and process understanding. However, it is unclear how process mining can help and can be practically used in the context of cybersecurity and reliability.  Therefore, in this work, we investigate the potential of process mining to aid in cybersecurity and software reliability to analyze and support research efforts in these areas. Concretely, we collect existing process mining applications, discuss current trends and promising research directions that can be used to tackle the current cybersecurity and software reliability challenges. To this end, we conduct a systematic literature review covering 35 relevant research approaches to examine how the process mining is currently used for these tasks and what are the research gaps and promising research directions in the area. This work is an extension of our previous work, which focused solely on the cybersecurity area, based on the observation of relative closeness and similar goals of those two fields, in which some approaches tend to overlap. |

**1. Introduction**

The advancement of digitalization in modern society has fueled the role of cybersecurity and software reliability in various domains of critical information infrastructures, such as healthcare or transporta-tion, where potential problems could result in injuries or loss of lives. Nowadays, the key challenge of effective cybersecurity and software reliability assurance is the actual rapid advancement of information technology, with new types of threats and unprecedented discrepancies emerging daily.

While both cybersecurity and software reliability are concerned with different root causes, ultimately, they overlap in their end goal of assuring the availability and integrity of cyber systems [1]. For example, a system can be rendered unavailable by both a successful cyberattack or software bug. Likewise, a system or its data can be inappropriately modified by a malicious insider or accidental miscon-figuration. As such, a holistic approach has to consider the aspects of both fields [2].

Existing cybersecurity and reliability techniques are designed for the discovery and prevention of a specific type of problem, and hence

having difficulties in adapting to new threats [3,4]. Furthermore, the actual security and reliability threats develop over time within complex processes in which minor vulnerabilities (e.g., software bugs, weak separation of authenticated spaces) combine with human/operator er-rors (e.g., credentials leaks) into major problems that are challenging to detect in its early formation stages [5]. Hence, the investigation of security threats is still largely manual [6] or is being addressed with strongly specialized domain-specific techniques to reduce false positives [7,8]. The behavior of entities is often encoded into a mathe-matical model that might be hard to manipulate, abstract, or complex, making it hard to respond to the security threats adequately [9]. Process mining [10] is a set of techniques that could be promis-ing in addressing the aforementioned challenges. In contrast to tradi-tional data-centric approaches (like data mining) and process-centric approaches (such as BPM analysis), process mining involves both data and end-to-end processes in their analysis [11,12] to the benefit of the final results [13,14]. For example, process mining techniques are de-signed to examine when and how a process deviated from the designed

∗ Corresponding au[thor.](mailto:macak@mail.muni.cz)

*E-mail addresses:* [macak@mail.muni.cz](mailto:macak@mail.muni.cz) (M. Macak), [daubner@mail.muni.cz](mailto:daubner@mail.muni.cz) (L. Daubner), [fanisani@pads.rwth-aachen.de](mailto:fanisani@pads.rwth-aachen.de) (M. Fani Sani), [buhnova@mail.muni.cz](mailto:buhnova@mail.muni.cz) [(B. Buhnova).](mailto:macak@mail.muni.cz)

<https://doi.org/10.1016/j.array.2021.100120>  
[Received 26 August 2021; Received in revise](https://doi.org/10.1016/j.array.2021.100120)d form 18 November 2021; Accepted 4 December 2021   
Available online 22 December 2021   
2590-0056/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

*M. Macak et al.*  *Array 13 (2022) 100120*

process model or how a bottleneck activity results in the final delays of a service/product. Moreover, it is designed to discover, monitor, and enhance processes by extracting knowledge from event data (i.e., *event logs*). Process mining has already proven successful in many domains, aiding with challenging tasks such as fraud detection [15], robotic process automation [16], or learning analytics [17]. Furthermore, pro-cess mining became also favored in governing safety-critical processes, such as in healthcare, where it supports critical hospital processes and patient treatment [18]. In [19], it is explained why process mining can be beneficial for advanced analysis and provides several use cases of it in different fields.

The evidence-based benefits and capabilities of process mining in similar domains such as software engineering [20] and confidential-ity [21] make it a promising candidate to address the challenges in cybersecurity and software reliability analysis — using its techniques, we might be able to more effectively detect unexpected behavior [22], identify issues [23], detect deviations [24], or verify whether the system conforms to the designed process [25]. Furthermore, it might have the capability to offer an overview of alerts in the system, detect malware in a system, detect frauds, verify the user behavior, or identify software bugs. This brings new opportunities of process mining to ad-vance the state of research in the cybersecurity and software-reliability situations with uncertainty about the actual processes underlying run-time system behavior, which needs to be reconstructed based on the observed events in the system.

Currently, there is no comprehensive systematic literature review that would help the researchers and practitioners understand where and how exactly the process mining could aid in these specific situations. This paper aims to enhance researchers’ efforts in cybersecurity and software reliability areas via an overview of research approaches using process mining for the purpose of cybersecurity and software reliability in various domains and for various tasks. The papers are grouped by their application domain to give insight into the current progress of process-mining usage in each one of them. We identify the techniques that are used for this purpose, together with their properties, and discuss possible research directions for further progress in this area. It is an enhancement to our previous study [26], which was limited in scope to the cybersecurity area. The rationale behind the inclusion of the software reliability into the systematic literature review is the observation of relative closeness, overlap in techniques, and similar goals of those two fields when conducting the original review. For this reason, we aim to explore the broader scope, as the techniques might support each other based on the overlap. Furthermore, the combined study should provide greater insight into the field for researchers as well.

The remainder of this paper is structured as follows. Section 2 con-tains a background of process mining. Section 3 provides an overview of existing literature reviews on process mining in other areas. Then, in Section 4, we formulate the research questions and describe our methodology. The usage of process mining to ensure cybersecurity is then detailed in Section 5, followed by Section 6, which focuses on the application of process mining for software reliability. The results of the review and the answers to our research questions are in Section 7. Section 8 discusses the threats to the validity of our review. Finally, Section 9 concludes the paper.

**2. Background**

Process mining techniques have proved to be very successful in (1) *process discovery*, which aims to find a descriptive model of the under-lying process from event logs, (2) *conformance checking*, i.e., monitoring and inspecting whether the real execution of the process conforms to the corresponding designed (or discovered) reference process model, and (3) *process enhancement*, which improves and enriches a process model based on the related event data [10].

2

*M. Macak et al.*  *Array 13 (2022) 100120*

|  |  |  |
| --- | --- | --- |
| **Table 1**  Summary of review protocol. |  |  |
| Initial security papers | IEEE Xplore  ScienceDirect  Springer Link  ACM DL  Web of Science  Scopus  **Total** | 33  272  350  97  29  79  **860** |
| Initial reliability papers | IEEE Xplore  ScienceDirect  Springer Link  ACM DL  Web of Science  Scopus  **Total** | 28  835  524  274  62  110  **1833** |
| Unique papers |  | 1967 |
| Filtered papers | 1st Stage — title & abstract 2nd Stage — Full text | 121  30 |
| Snowballed papers | Google search  References | 2  3 |
| Relevant papers | Security  Reliability  **Total** | 25  10  **35** |

be found in the review by Leitner and Rinderle-Ma [46], which fo-cuses on security in Process-Aware Information Systems (PAIS), where process mining is however only one of many methods considered. Moreover, as the review covered results from 1993 to 2012, many recent approaches are missing. In [47], Kelemen provided an overview of the usage of process mining for security in the public sector domain, published between 2000 and 2016. Based on this review, the author provides a set of topics that are dominant in the identified research papers, together with challenges and future research directions. The paper, however, only considers the public sector domain (explicitly in its inclusion criteria). Hence, our work, which is an enhancement to our previous study [26], which was limited in scope to the cybersecurity area, continues this research path with a broader study, including more recent publications and a wider spectrum of domains.

Moving away from surveys focusing on process mining, there are many surveys on specific aspects of cybersecurity or software relia-bility. However, they tend to focus on particular domains or tech-niques [48–51]. Also, an overview of existing surveys in cybersecu-rity [52] does not include focus on process mining.

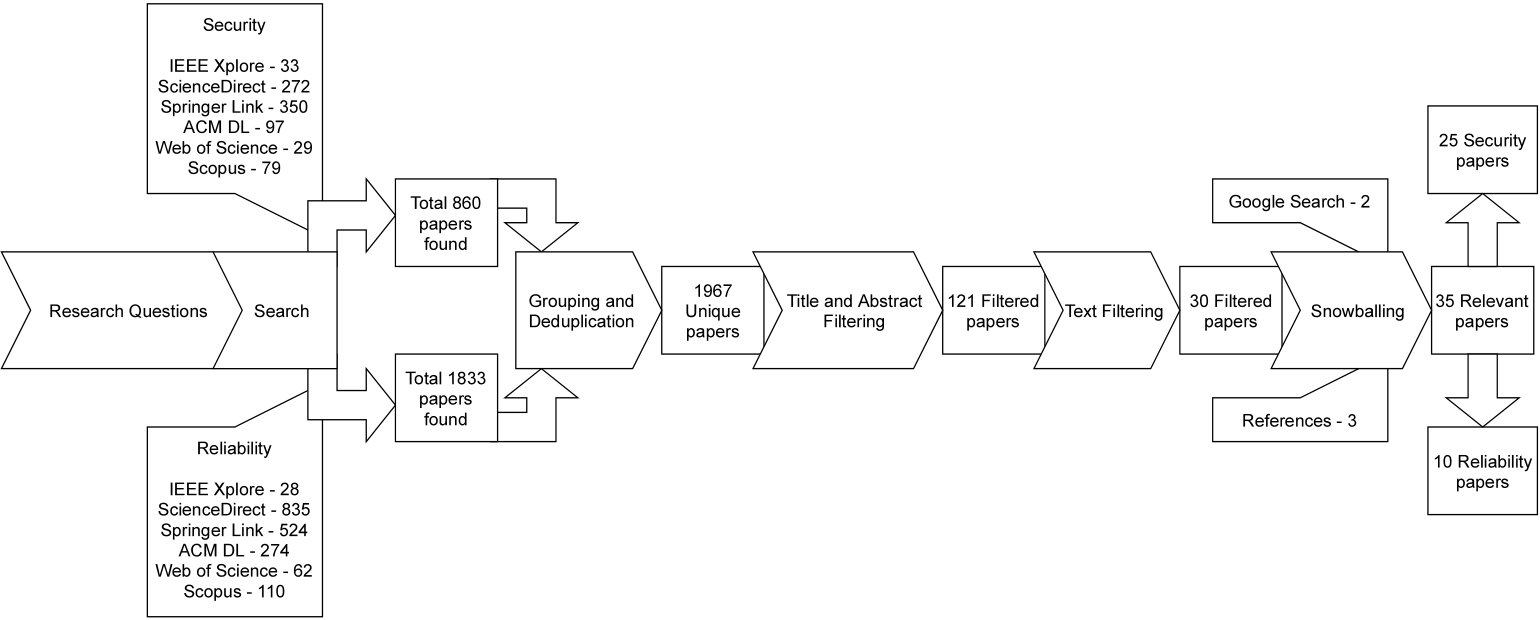
With the missing comprehensive systematic literature review of process mining usage for cybersecurity and software reliability, it is hard to understand where and how the process mining could help the researchers and practitioners. The focus on process mining applications for these tasks in different research directions is an important aspect to consider. Moreover, none of the existing reviews covers the application of process mining in software reliability, which makes the systematic literature review presented in this paper a valuable complement of the current state of the art.

**4. Methodology**

The goal of this study is to review recent research in cybersecurity and software reliability that employs process mining. To this end, we formulate a strategy to guide this study, based on the Kitchenham and Charters guidelines for systematic literature reviews [53]. This includes a digital library search for current literature, snowballing, and manual searching. Within the review, we first consider the cybersecurity and software reliability separately and merge them in later stages. Our review methodology is visualized in Fig. 1, together with the results of each conducted search and processing stage. See Table 1 for a summary of the review protocol containing the number of publications.

3

*M. Macak et al.*  *Array 13 (2022) 100120*



**Fig. 1.** The strategy of the review with the number of publications at each stage.

review. Therefore, the next stage was to filter out false-positive results from the collection.

There are two steps for filtering the collection. In the first step, the only parameters considered were paper title, abstract, keywords, and conclusion of the paper. The papers were labeled by their topic, which is cybersecurity or software reliability. The following second step of reading the full paper to confirm or correct the labeling. Any relabeling was confirmed with an independent opinion by another team member. To maintain the fairness of the filtering, several inclusion and exclusion criteria were formulated. Conforming to one or more of the exclusion criteria results in discarding the paper. The paper is kept if and only if it conforms to at least one inclusion criterion. Otherwise, if it conforms to neither, it is discarded. Additionally, if there are multiple papers from the same authors that are directly related (i.e., follow-up of the same idea, but not a new topic), only the most recent or most extensive is kept. The *inclusion criteria* are that the work:

• deals with the detection, analysis, modeling, recovery, or avoid- ance of cyberattack, malware, fraud, fault, or error;  
• focuses on cybersecurity-related processes;  
• aims at the reliability of software systems;  
• is concerned with anomalous behavior regarding cybersecurity or software reliability.

The *exclusion criteria* are that the work:

• does not utilize process mining;  
• does not consider software systems as the primary target of the approach;  
• is focused only on reliable performance/effectiveness;  
• is not a full paper (i.e., is only a keynote abstract, chapter frag- ment, or encyclopedia article).

Each filtering step was done by a different researcher. In case of any doubts, the paper was marked and discussed by two more researchers to reach the consensus. To further avoid bias in filtering, a sample of 10% papers filtered out in the first phase was revived by another researcher. **Snowballing** The last stage is the snowballing, which includes papers referred by papers kept after the second step of filtering. Each snowballed paper was considered in terms of inclusion and exclu-sion criteria, and within the 2014-to-2020 limit, and English-written condition.

The final collection of papers was further supplemented by man-ual, non-systematic search using search engines like Google Scholar.7 Additional papers were also added after consultation with domain experts. Nevertheless, even those papers conformed to the inclusion and exclusion criteria.

7 <https://scholar.google.com/>.

4

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *M. Macak et al.* | | |  |  |  | *Array 13 (2022) 100120* |
| **Table 2**  Cybersecurity and software reliability papers included in the review. | | |
| Paper | Category | Research direction | Target period | PM Typea | Expert knowledge | Model analysis |
| [54]  [55]  [56]  [57]  [58]  [22]  [59]  [60]  [61]  [62]  [63]  [24]  [64]  [65]  [66]  [67]  [9]  [68]  [69]  [70]  [23]  [71]  [72]  [73]  [74]  [75]  [76]  [77]  [25]  [78]  [79]  [80]  [81]  [82]  [83] | Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Cybersecurity Reliability  Reliability  Reliability  Reliability  Reliability  Reliability  Reliability  Reliability  Reliability  Reliability | Security of ICS  Security of ICS  Security of smartphones  Security of smartphones  Security of smartphones  Network traffic security  Network traffic security  Network traffic security  Web-application security  Web-application security  Web-application security  Web-application security  Attack inspection  Attack inspection  Attack inspection  Outlier user behavior detection Outlier user behavior detection Outlier user behavior detection Outlier user behavior detection Outlier user behavior detection Outlier user behavior detection Fraud detection  Fraud detection  Fraud detection  Fraud detection  Quality assurance  Quality assurance  Quality assurance  Error detection  Error detection  Error detection  Error detection  Error detection  Error detection  Error detection | Past  Past  Past  Past  Past  Past  Past  Present  Past  Past  Past  Past  Past  Past  Past  Past  Past  Present  Present  Past  Past  Past & present Past & present Past  Past  Past  Past  Past  Present  Past & present Past  Past & present Present  Past  Past | PD  PD & CC  PD  PD  PD & CC  PD  PD  PD & CC  PD  PD & CC  PD & PE  PD & CC & PE PD  PD  PD  PD & CC  PD  PD & CC & PE CC  CC  CC & PE  CC  CC  PD & CC  PD  CC  PE  PD  CC  PD & CC  PD  PD & CC  CC  PD  PD | ✓ | Automatically Automatically Automatically Automatically Automatically Manually  Manually  Both  Automatically Automatically Manually  Manually  Manually  Automatically Manually  Both  Manually  Automatically Automatically Automatically Automatically Automatically Automatically Automatically Manually  Automatically Automatically Manually  Automatically Automatically Manually  Automatically Automatically Automatically Manually |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✗ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✗ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |
| ✓ |

aPD — process discovery, CC — conformance checking, PE — process enhancement.

*5.1. Security of industrial control systems*

Industrial control systems (ICS) are present in critical domains like manufacturing, transportation, and energy sector, where they are responsible for production, monitoring, and control. Naturally, security in these domains is vital as successful attacks could cause significant money loss, physical damage, or injury. The domains are often paired with the term critical infrastructures, emphasizing extensive need for dependability.

Specifically, Bernardi et al. [54] studied the detection of anomalous behavior in energy usage from smart meter readings. They classify the readings based on the expert knowledge into several levels. Then, they use process mining for discovering the behavior of customers over time by looking at how the levels of energy usage were changing. They performed process discovery for several periods with the same length. The output is a sequence of time-evolving graphs over a bigger period. Therefore, in this sequence, they can compare the graphs with each other and find anomalous periods with potential security implications. Two similarity measures were used, i.e., the Hamming distance and cosine similarity [84].

In addition to process discovery, Myers et al. [85] applied the conformance checking method in their work. They firstly investigate the suitability of process-mining discovery algorithms for the detection of cyberattacks in industrial control systems. To this end, they compare five algorithms by the ability to create a usable model, the accuracy and the simplicity of it. Then, with conformance checking, they evaluate the most suitable process discovery algorithm by comparing the number of detected anomalous cases, trace fitness, and time. As a result, they suggest that the Inductive Miner with perfect fitness is the most suitable algorithm in this regard. Based on this paper, in their later work [55], they introduced a method for identifying anomalies in ICS and SCADA

5

*M. Macak et al.*  *Array 13 (2022) 100120*

|  |  |
| --- | --- |
| **Table 3**  Use cases of papers included in the review. | |
| Paper | Use case |
| [54]  [55]  [56]  [57]  [58]  [22]  [59]  [60]  [61]  [62]  [63]  [24]  [64]  [65]  [66]  [67]  [9]  [68]  [69]  [70]  [23]  [71]  [72]  [73]  [74]  [75]  [76]  [77]  [25]  [78]  [79]  [80]  [81]  [82]  [83] | Integrity attack detection  Detection of cyberattacks  Malware detection  Malware detection  Attack detection  Detection of spam attacks  Alert visualization  Classification of attack traces  Enforcing of security policies  Detection of abnormal behavior on social networks Detection of attacks and new behavior  Deviation detection in IS audits  Analysis of attack process  Ransomware detection  Attack vector identification  Insider threat detection  Detection of abnormal behavior  Detection of tampered data  Automatic resolving of issues  Security-critical deviations detection  Detection of non-conforming user behavior  Classification of traces  Detection of security breaches and frauds  Process-based fraud detection  Fraud detection  Diagnosing service implementation  Isolating and replicating failure as a test case  Detection of defects in acceptance testing  Runtime verification of behavioral properties  Detecting failures during a rolling upgrade  Identification of bugs in evolving software  Detection of application failures  Monitoring of correct configuration  Monitoring of workflow resilience  Auditing of smart contract executions |

families. It was also proven that it could be used in phylogeny tracking because it could identify variants of the same malware family. They also performed the evaluation on the transformed dataset, where they did current anti-malware obfuscations. Their approach greatly reduced the number of false negatives in this case.

Phylogenic analysis and malware family detection were also per-formed by Cimino et al. [57]. In this work, process discovery was used to obtain temporal logic formulae which were used in formal model verification. Each path of the discovered process model is transformed into a temporal logic formula. This approach was also evaluated and confirmed to offer an effective solution for this problem in smartphone applications.

The process data from activity logs of Android devices were also used by process mining analysis by Hluchy and Habala [58], who applied conformance checking in addition to process discovery. From the phone logs, they chose OS-generated information about specific performed actions, browser history, and network connection log. Then, they performed an attack in which the user activated a malicious URL, which resulted in downloading personal user data via a known vul-nerability, and its discovered model. The used conformance checking technique considers this process model for the offline detection of that attack from examined smartphone logs. This approach raised many false positives, which they concluded is caused because of the simplicity of the attack and the low quality of Android logs.

All the approaches within this direction deal with the automatic analysis of past events. Notably, there is an interesting usage of the declarative approach in two approaches [56,57], which can be at-tributed to diverse options in smartphone usage. This is further sup-ported by the third work [58], where the focused model covers mali-cious usage instead of more traditional non-malicious. It is shown that process mining neatly exploits the available data but the number of approaches is surprisingly low given the wide range of available data discussed in the papers.

6

*M. Macak et al.*  *Array 13 (2022) 100120*

like authorization constraints. The model is then transferred into a reachability graph that represents all possible valid executions of the web application workflow. Using this method, a run-time monitor is synthesized. This monitor is used by a proxy between the user and web application. Based on its output, the proxy either forwards the user requests to the application or drop them.

The security of web applications, specifically web information sys-tems, was also considered in work by Bernardi et al. [63], who pro-posed a method for improving the security of these systems. This approach utilizes process mining and model-driven engineering. First, they specify the system behavior with Unified Modeling Language, from which they automatically generate a formal model. At the same time, they preprocess the obtained logs of the system to get the event logs, which can be used by process mining techniques. To identify devia-tions, they use ProM [87] visualizations. Those deviations were filtered, and based on the output of the fuzzy mining discovery technique [33], they could be classified as an attack or the new behavior. The classifier, in this case, is the HTTP request code. This method was applied to a web information system for managing the publications.

Similarly to the previous work, Zerbino et al. [24] proposed a process mining methodology in which they manually detect deviations. In this case, it was used for audits information systems, and they used the Disco tool [88]. First, they discovered a process model from historical data. Using Disco, this model can be automatically enriched with other perspectives, like time or organizational perspective. With this tool’s visualizations, they manually detected several deviations in a case study. They mentioned process mining advantages over other audit approaches, like better depth of analysis, broader scope, and easier automation.

Web-application security direction contains several approaches, all utilizing process discovery but often combined with other techniques. Remarkably, there is greater utilization of process enhancement [24, 63], mainly regarding the time perspective, but only for manual anal-ysis. This direction shows a wide variety of process mining usage for cybersecurity, prompting to utilize the ideas for different domains.

*5.5. Attack inspection*

The direction of attack inspection consists of inspecting how the specific attack is performed, supporting a better understanding and preventing future attacks. In this direction, the found papers only utilize the process discovery technique for the analysis of past events.

Viticchie et al. [64] used process mining for the investigation of the process of the attacks on a small application with different levels of protection, which was performed as an experiment. Every participant filled the report in which their attack strategy was described. These texts were annotated, and process discovery was used on the traces of annotations. The discovered model was used to find out the attack process, the differences between successful and unsuccessful attack processes, and whether this process was influenced by the level of protection used in the application software.

Attacks were also inspected by Macak et al. [66], but a the different point of view. This work is focused on the unintentional insider attack vector identification. Process discovery is used on the event logs pro-duced by the simulation games platform which simulate the working environment for players.

The other approach in this domain is aiming for the inspection and detection of ransomware. Bahrani and Bidgly [65] created event logs from harmless applications and ransomware families. Those logs contained three types of registry events. Then, for each software, they discovered a process model, and from each process model, they ex-tracted the frequencies of each transition. These data then can be used by a classification algorithm to identify ransomware in the application software. Thanks to this, no expert knowledge is required to use this approach.

7

*M. Macak et al.*  *Array 13 (2022) 100120*

human intervention in Service Desk because of automatic resolving of user issues.

There are several papers that are detecting the outliers in the processes. However, in this literature review, we exclude those that do not have the security as their primary goal in their proposed approach. Alizadeh et al. [23] proposed an auditing approach that combines data and process perspectives to detect non-conforming user behavior with conformance checking. This approach can identify the previously undiscovered deviations.

In this direction, conformance checking is prevalent, emphasiz-ing its effectiveness in outlier behavior detection. Additionally, all approaches except one [23] use expert knowledge for the analysis. Remarkably, there is a wide variety of use cases and techniques, despite this research direction being the largest one found. We speculate that the cause is that user behavior can be very unpredictable, forming a complex process, but at the same time, it can be very harmful in cybersecurity. This variability indicates a potential for future research as the methods are not stale yet.

*5.7. Fraud detection*

Fraud detection is a specific research direction aiming to detect false pretenses of entities. Typical examples are unusual processes, violating a rule or policy. The found approaches utilizing process mining in the cybersecurity context tend to employ conformance checking with expert knowledge.

Fazzinga et al. [92] proposed a method for online and offline classification of event log traces as potential security breaches. They create a security breach model, which is used later in conformance checking. In their work, they are trying to solve the problem with the mapping between high-level and low-level operations. It is important in this scenario because they claim that security breach models are typically described as high-level activities, but log traces are typically performed as low-level operations. They used a probabilistic approach in the created model and the following conformance checking. In their following paper [93], they proposed a classification framework that combines their previous work [92], a model-driven method with an example-driven classification. In a model-driven approach, a security breach model is created, and incoming traces are classified based on conformance checking. In the example-driven approach, a set of previously labeled traces is used for later classification. This approach was experimentally validated in their next paper [71].

Security breaches were also the aim of the work of Böhmer and Rinderle-Ma [72], who proposed an anomaly detection strategy in process execution events to prevent not only security breaches but also frauds. They include the control flow, time, and resource per-spective into one anomaly detection approach. They try to detect point, contextual, and collective anomalies. They also try dealing with unexpected events that might not indicate an anomaly. They propose to construct a likelihood graph, which represents the likelihood of the expected execution of events. Firstly, they create a basic likelihood graph with activities and their probabilities. Then they extend it with other perspectives. For detecting the anomaly, they compare the likeli-hood of the currently executing event with the likelihood of recorded comparable events based on the likelihood graph.

Baader and Krcmar [74] were also using process mining for fraud detection. Specifically, they present a method for reducing the number of false positives in this detection. They combine the red-flag approach and process mining for identification and visualization of possible undesired process instances. Potential frauds are identified and then visualized with the fuzzy miner. Their approach was compared to two other approaches, and they got a lower number of false positives. However, they detected over half fewer frauds than one chosen variant. They discuss that process mining gives several other advantages to a classical red flag approach, such as easier dashboard analysis and visualization.

8

*M. Macak et al.*  *Array 13 (2022) 100120*

Service orientation is a topic of interest by van der Werf and Verbeek [81], who present the application of process mining for con-tinuous auditing of service behavior. The aim is to monitor violations of security requirements in support of configuration management. To this end, the authors present a tool that analyzes event logs from executions across the service landscape using semantic process mining, which combines process mining with semantic web techniques. A key concept is enabling relations between elements in the event log and other elements using annotation rules. The tool utilizes process mining to check conformance to defined constants. Concretely, reliability and security-related constraints are discussed.

Software design and runtime behavior, albeit in the case of adap-tive middleware, is considered by Rosa et al. [25], who presents a solution for its implementation and execution. Besides covering the basic requirements of adaptive middleware systems, the authors in-clude techniques for verification of their implementation. The process mining is used for runtime verification of behavioral properties of an implemented adaptive middleware system. The behavior is verified by conformance checking against its specification in LTL. In the case of undesired behavior, the proposed solution automatically creates and executes an adaptation that mitigates the issue. Thus, making the system more robust.

Detection of application failures is the topic of highly detailed work by Pecchia et al. [80]. The proposed approach consists of multiple steps that utilize different methods of process mining. The main ingredient of the approach is the application logs. Those are firstly used to construct a reference model capturing normal behavior using process discovery. Subsequently, the logs are used, possibly in real-time, for conformance checking against the reference model to discover failures in the application. The work is highly detailed in both implementation and evaluation. Furthermore, it considers the presence of noise in application logs.

Applications of process mining in the field of software maintenance are explored by Gupta [79]. The author presents four applications of process mining assisting in improving various processes. Concretely, management of software projects, ticket resolution, and software bug detection. The latter is highly relevant for increasing the overall reli-ability of the software at hand. In this case, two models are obtained by performing process-discovery algorithms based on execution logs —the current stable and the new software versions. The models are then compared for differences. However, the paper does not go into much detail about the actual technique used and serves more as an outline for future work.

The work by Xu et al. [78] presents an error detection method aimed at sporadic operations like software deployment. The specific use case presented in the paper is a rolling upgrade in a cloud environment. In order to achieve accurate detection, the method first creates a process model of the operation from regular, un-anomalous logs using process discovery. Afterward, the conformance checking is used as part of the analysis to determine if the process is running correctly. Here, process mining is just one step in a more sophisticated analysis.

Resilience monitoring of executable business processes, with the focus on time, is discussed by Zahoransky et al. [82]. The authors present a mathematical framework to define the notion of resilience in workflows. The main idea is to automatically extract the probability distribution of time characteristics of a given workflow. Process mining is used to obtain the time information from an event log, which is then used to calculate the probability distributions. The approach can be utilized to predict and monitor the resilience indicators of the executed workflow, possibly enacting timely countermeasures if the levels drop. Auditing, specifically aimed at blockchain smart contracts, is a focus of the work by Corradini et al. [83]. The authors devised a method that extracts the transaction logs, cluster them by the sender to create an event log in XES format, discovers a model, and analyzes it. Process mining, concretely the process discovery technique, is used in the third phase to generate three candidate models using different

9

*M. Macak et al.*  *Array 13 (2022) 100120*

**7. Results**

This section formulates the insights from the literature review into the answers to our research questions.

*7.1. [RQ1] What are the research directions in which the process mining is used to ensure cybersecurity and software reliability?*

Within our review, we have identified 35 approaches employing process mining in cybersecurity or software reliability, which we have structured according to the direction of the research problem they are addressing. These include the security of industrial control systems, security of smartphone devices, web-application security, network traf-fic security, attack inspection, outlier user behavior, fraud detection, error detection, and quality assurance. Although this division is by no means perfect, it gives a useful understanding at what research problems the process mining is directed in terms of cybersecurity and software reliability.

In the case of cybersecurity, the most popular research approaches (with the highest number of publications) were targeting general di-rection towards either detection of outlier user behavior or frauds. The next most popular direction was the cybersecurity of network elements, namely focusing on the applications related to networking in connection with websites, information systems, and other technologies that primarily communicate through networks. The research in the remaining directions is rather sparse.

The most popular research direction in the case of software relia-bility is the error detection. In this case, process mining is typically utilized to monitor runtime behavior in search of errors or abnormal behavior. Curiously, in the quality assurance research direction, the utilization is very similar, however focusing on aiding the development of the systems, rather than its operation and maintenance. Indeed, some outlier approaches can be found, like generating test cases. Overall, in software reliability, process mining is primarily used in a very narrow sense to detect faults, errors, and failures. The usage for analysis of reliability attributes is, barring one paper, unexplored.

Surprisingly, we did not discover a holistic approach combining both fields, as stressed by [2]. However, the work by van der Werf and Verbeek [81] contains features of both fields. For the rest of the papers, there are some significant similarities where researchers took analogous approaches across the fields. The most significant overlap lies in detecting various anomalous or undesired behavior, which shares a similar impact while having different root causes. Indeed, utilizing process mining to analyze behavior is a major strength of process mining. Additionally, process mining is utilized for inspection and diagnosis across the fields, focusing on attacks and errors, respectively. This hints towards similar techniques for general faults regardless of cause.

On the other hand, we noticed some differences regarding the use of process mining between the two fields. First and foremost, cyber-security seems to be more popular with more than twice the number of papers compared to software reliability. Furthermore, we found out that the application domains (e.g., smart grids, mobile devices, finan-cial systems) in the cybersecurity field are more diverse. In contrast, the distinguishing factor within software reliability is technology or platform.

Domains that are generally dominant in process mining fields in the general process mining realm were surprisingly not found among the results. Those domains include healthcare, manufacturing, education, finance, and logistics. Thus, we assume that in these domains, even though they are generally popular, the process mining techniques have not yet been properly explored for the purpose of cybersecurity or software reliability. Alternatively, they could be employing general domain non-specific techniques and thus not published mentioning a particular domain.

10

*M. Macak et al.*  *Array 13 (2022) 100120*

|  |  |
| --- | --- |
| **Table 4**  Techniques used in papers included in the review. | |
| Paper | Techniques |
| [54]  [55]  [56]  [57]  [58]  [22]  [59]  [60]  [61]  [62]  [63]  [24]  [64]  [65]  [66]  [67]  [9]  [68]  [69]  [70]  [23]  [71]  [72]  [73]  [74]  [75]  [76]  [77]  [25]  [78]  [79]  [80]  [81]  [82]  [83] | Fuzzy miner (PD)  Inductive miner (PD), A\*-based (CC)  Declare miner (PD)  Fuzzy miner (PD)  Inductive miner (PD), A\*-based (CC)  Fuzzy miner (PD)  Fuzzy miner (PD)  Online heuristic miner (PD), alignment-based (CC)  Alpha miner (PD)  Genetic process mining (PD), not specified (CC)  Fuzzy miner (PD)  Fuzzy miner (PD), manual (CC)  Fuzzy miner (PD)  Fuzzy miner (PD)  Fuzzy miner (PD)  Heuristic miner (PD), fuzzy miner (PD), LTL checker (CC) Fuzzy miner (PD)  *𝜏* algorithm (PD), custom algorithm (PD), not specified (CC) Not specified  Alignment-based (CC)  Custom  Custom  Custom  Not specified  Fuzzy miner (PD)  Alignment-based (CC)  Custom (PE)  Fuzzy miner (PD)  LTL checker (CC)  Fuzzy miner (PD), token-based (CC)  Not specified  Fuzzy miner (PD), token-based (CC)  Fuzzy miner (PD)  Not specified  Heuristic miner (PD), inductive miner (PD), split miner (PD) |

popular usage of this library in the future, as it supports multiple process mining algorithms and is not so limited as analytical tools. Regarding the conformance checking, we did not find any prevalent technique. Often they were not specified or were customly created for the given task. The most used techniques were alignment-based [37].

*7.3. [RQ3] What are the current gaps and possible research directions in the usage of process mining for cybersecurity and software reliability?*

Research gaps and possible research directions can be observed in multiple aspects. First, there is a gap in the usage of process mining in cybersecurity and software reliability in several domains that otherwise do utilize process mining (for other applications), like healthcare, man-ufacturing, education, and logistics. Although we found some research papers focused on these areas, they were dealing with safety-critical processes involving humans and not computer systems [98,99], or just presented upcoming research direction [97]. Secondly, we detected a substantial untapped potential in the application of process mining within a larger context of cybersecurity and software reliability. Areas such as availability, robustness, or resilience are rather unexplored with respect to process mining, although they are implied in literature or applied in a non-IT field. In summary, and based on the found papers, we see a great potential of process mining for cybersecurity and software reliability. Daunting future directions could utilize the good practice from non-IT areas, where faults, anomalies, safety, or robustness of a process is evaluated.

Furthermore, we want to emphasize a set of criteria required to approach cybersecurity and software reliability problems in a process-aware manner. We believe this will help other researchers discover new use cases for process mining. These criteria are:

• Events can be ordered into a sequence, and membership of the event in the sequence can be determined.

• The beginning and end of the process need to be clearly definable.

11

*M. Macak et al.*  *Array 13 (2022) 100120*

papers filtered out in the first phase was double-checked by another researcher. In this review, we formulated a strategy to guide the re-view, which is based on the existing guideline for systematic literature reviews [53]. In the reliability search, we included all popular terms which are connected to the reliability and software dependability.

*8.3. External validity threats*

The search is limited to the up-to-date papers over the last six years, from 2014 to 2020. The primary motivation for the restriction to this period is to focus on the most recent research and applications. Secon-darily, we expected the more mature approaches among these recent publications. Lastly, the lower bound, the year 2014, also corresponds to a local peak in the number of publications within the Dimensions8 dataset, showing a slight change of trend.

To evaluate the number of potentially missed publications from the filtered period, we performed an additional search with the same parameters apart from the period. Based on the same duplicity and rel-evancy ratio assumption, we approximated the filtering of 14 relevant articles published before 2014.

reliability. Primarily, we would encourage the research community to deeper investigate the domain of critical information infrastructures, which might benefit significantly from more advanced cybersecurity and software reliability techniques. The real-time analysis of systems has a strong potential to utilize the advantages of process mining techniques. Furthermore, it would be indeed beneficial in the advanced detection and prevention of cyberattacks and incidents, enhancing it with a process-oriented approach.

**Declaration of competing interest**

The authors declare that they have no known competing finan-cial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Acknowledgment**

This research was supported by ERDF ‘‘CyberSecurity, CyberCrime and Critical Information Infrastructures Center of Excellence’’ (No. CZ.02.1.01/0.0/0.0/16\_019/0000822).

*8.4. Conclusion validity threats*  **References**

To ensure the correctness of the extracted data, we formulated the strategy of this literature review and established categories of interest which had to be described for each relevant research paper. The popularity of identified domains was also cross-checked with more general reviews on process mining. Moreover, the interpretation of the data was extensively discussed. Despite our systematic approach, some trends, patterns, and research gaps could have been missed in the review. On the other hand, we believe that the value of the provided summary in this work is not primarily in the research gaps, but in overview of the existing body of knowledge in applying process mining in the area of cybersecurity and software reliability. In this way, we believe we can facilitate the understanding of what attempts exist so that it is easier for the reader to see where they can build on what is existing and where they need to start building their approach from scratch.

**9. Conclusion**

In this paper, we conducted a systematic literature review to provide an overview of research papers that use process mining for cybersecu-rity and software reliability. In accordance with our initial assumption, process mining techniques have been used for this purpose. Papers summarized in this work show that original research advancements of mentioned directions are possible thanks to process mining. However, the coverage is still rather limited, with numerous research gaps, espe-cially in software reliability. As such, there is much potential in both cybersecurity and software reliability applications of process mining. We identified nine major research directions, discussed how they fit in the overall landscape and presented how they utilize the process mining for these purposes. While we found that the process mining research directions for cybersecurity and software reliability are taking quite different focus, there are similarities in approaches and usage of process mining techniques. For example, a detection of anomalies follows a similar approach regardless of their cause (e.g., errors, frauds). Most importantly, they contribute to the dependability of the systems, which must consider both areas. We demonstrated the feasibility of using process mining with this goal.

Based on this systematic literature review, we pointed out a set of possible process mining research directions that can be taken to tackle the state-of-the-art challenges in cybersecurity and software

8 <https://app.dimensions.ai/>.

12

*M. Macak et al.*  *Array 13 (2022) 100120*

[18] dos Santos Garcia C, Meincheim A, Junior ERF, Dallagassa MR, Sato DMV, Carvalho DR, et al. Process mining techniques and applications - A systematic mapping study. Expert Syst Appl 2019;133:260–95. [http://dx.doi.org/10.1016/ j.eswa.2019.05.003](http://dx.doi.org/10.1016/j.eswa.2019.05.003).

[19] [Reinkemeyer L. Pr](http://dx.doi.org/10.1016/j.eswa.2019.05.003)[ocess mining in action: principles, use cases and outlook.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb19)  [Springer Nature; 2020.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb19)

[20] [Keith B, Vega V. Process mining applications in software engineering. In: International conference on software process improvement. Springer; 2016, p. 47–56.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb20)

[21] [Elkoumy G, Fahrenkrog-Petersen SA, Sani MF, Koschmider A, Mannhardt F, Von Voigt SNn, Rafiei M, Waldthausen LV. Privacy and confidentiality in process mining: Threats and research challenges. ACM Trans Manage Inf Syst 2021;13(1).](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb21)

[22] Bustos-Jiménez J, Saint-Pierre C, Graves A. Applying process mining techniques to dns traces analysis. In: 2014 33rd international conference of the chilean computer science society (sccc). 2014, p. 12–6. [http://dx.doi.org/10.1109/ SCCC.2014.9](http://dx.doi.org/10.1109/SCCC.2014.9).

[23] Alizadeh M, Lu X, Fahland D, Zannone N, van der Aalst WM. Linking data and process perspectives for conformance analysis. Comput Secur 2018;73:172–93. <http://dx.doi.org/10.1016/j.cose.2017.10.010>.

[24] [Zerbino P, Aloini D, Dulmin R, Mininno](http://dx.doi.org/10.1016/j.cose.2017.10.010) V. Process-mining-enabled audit of information systems: Methodology and an application. Expert Syst Appl 2018;110:80–92. <http://dx.doi.org/10.1016/j.eswa.2018.05.030>.

[25] Rosa NS, Campo[s GM, Cavalcanti DJ. Lightweight formalisati](http://dx.doi.org/10.1016/j.eswa.2018.05.030)on of adaptive middleware. J Syst Archit 2019;97:54–64. [http://dx.doi.org/10.1016/j.sysarc. 2018.12.002](http://dx.doi.org/10.1016/j.sysarc.2018.12.002).

[26] [Macak M, Daubner L, Fani Sani M, Buhnova B. Cybersecurity analysis via process mining: A systematic literature review. In: Advanced data mining and applications. Springer International Publishing; 2021.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb26)

[27] [Cook JE, Wolf AL. Automating process discovery thr](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb26)ough event-data analysis.

In: Proceedings of the 17th international conference on software engineering.

Icse ’95, New York, NY, USA: Association for Computing Machinery; 1995, p. 73–82. <http://dx.doi.org/10.1145/225014.225021>.

[28] [Datta A](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb28)[. Automating the discovery of as-is business](http://dx.doi.org/10.1145/225014.225021) [process models: Probabilistic](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb28)  [and algorithmic approaches. Inf Syst Res 1998;9(3):275–301.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb28)

[29] [Agrawal R, Gunopulos D, Leymann F. Mining process mode](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb28)[ls from workflow logs. In: Schek H-J, Alonso G, Saltor F, Ramos I, editors. Advances in database technology — edbt’98. Berlin, Heidelberg: Springer Berlin Heidelberg; 1998, p. 467–83.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb29)

[30] [van der Aalst W, Weijters T, Maruster L. Workflow mining: Discovering process](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb30)  [models from event logs. IEEE Trans Knowl Data Eng 2004;16(9):1128–42.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb30)

[31] [van Dongen BF, De Medeiros AA, Wen L. Process mining: Overview and out](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb30)[look of petri net discovery algorithms. In: Transactions on petri nets and other models of concurrency ii. Springer; 2009, p. 225–42.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb31)

[32] [Weijters A, van der Aalst WM, De Medeiros AA. Proce](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb31)[ss mining with the heuris-tics miner-algorithm. Tech. rep. wp 166, Technische Universiteit Eindhoven; 2006, p. 1–34.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb32)

[33] [Günther CW, van der Aalst WMP. Fuzzy mining – adaptive process simplifica-tion based on multi-perspective metrics. In: Alonso G, Dadam P, Rosemann M, editors. Business process management. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007, p. 328–43.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb33)

[34] [Lamma E, Mello P, Montali M](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb33)[, Riguzzi F, Storari S. Inducing declarative logic-based models from labeled traces. In: Alonso G, Dadam P, Rosemann M, editors.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb34)

[Business process management. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007, p. 344–59.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb34)

[35] [Carmona J, van](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb34) [Dongen B, Solti A, Weidlich M. Conformance checking.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb35)  [Springer; 2018.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb35)

[36] [Rozinat A, van](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb35) [der Aalst WM. Conformance checking of processes based on](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb36)  [monitoring real behavior. Inf Syst 2008;33(1):64–95.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb36)

[37] [van der Aalst W, Adriansyah A, van Dongen B. Rep](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb36)[laying history on process models for conformance checking and performance analysis. Wiley Interdiscip Rev: Data Min Knowl Discov 2012;2(2):182–92.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb37)

[38] [Fahland D, van der Aalst WM. Model repair—aligning process models to reality.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb38)  [Inf Syst 2015;47:220–43.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb38)

[39] [Burattin A, Sperduti A,](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb38) [Veluscek M. Business models enhancement through](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb39)  [discovery of roles. In: CIDM. 2013, p. 103–10.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb39)

[40] [Jaisook P, Premchaiswadi W. Time performance analysis of medical treatment processes by using disco. In: 2015 13th international conference on ict and knowledge engineering (ict & knowledge engineering 2015). IEEE; 2015, p. 110–5.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb40)

[41] Kurniati AP, Johnson O, Hogg D, Hall G. Process mining in oncology: A literature review. In: 2016 6th international conference on information commu-nication and management (icicm). 2016, p. 291–7. [http://dx.doi.org/10.1109/ INFOCOMAN.2016.7784260](http://dx.doi.org/10.1109/INFOCOMAN.2016.7784260).

[42] [Williams R, Rojas E, Peek N, Johnson OA. Process mining in primary care: A](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb42)  [literature review. Stud Health Technol Inform 2018;247:376–80.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb42)

[43] [Bogarín A, Cerezo R, Romero C. A survey on educational process mining. Wiley](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb43)  [Interdiscip Rev: Data Min Knowl Discov 2018;8(1):e1230.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb43)

13

*M. Macak et al.*  *Array 13 (2022) 100120*

[67] Macak M, Vanat I, Merjavy M, Jevocin T, Buhnova B. Towards process mining utilization in insider threat detection from audit logs. In: 2020 seventh international conference on social networks analysis, management and security (snams). 2020, p. 1–6. <http://dx.doi.org/10.1109/SNAMS52053.2020.9336573>. [68] Li C, Ge J, Li Z, Huang [L, Yang H, Luo B. Monitoring interactions across mult](http://dx.doi.org/10.1109/SNAMS52053.2020.9336573)i business processes with token carried data. IEEE Trans Serv Comput 2018;1. <http://dx.doi.org/10.1109/TSC.2016.2645690>.

[69] [Talamo M, Povilionis A, Arcieri F, Schunck](http://dx.doi.org/10.1109/TSC.2016.2645690) CH. Providing online operational support for distributed, security sensitive electronic business processes. In: 2015 international carnahan conference on security technology (iccst). 2015, p. 49–54. <http://dx.doi.org/10.1109/CCST.2015.7389656>.

[70] [Salnitri M, Alizadeh M, Giovanella D, Zannone N, Giorgini P. From security-by-design to the identification of security-critical deviations in process executions.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb70)

[In: International conference on advanced information systems engineering. Springer; 2018, p. 218–34.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb70)

[71] [Fazzinga B, Folino F, Furf](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb70)aro F, Pontieri L. An ensemble-based approach to the security-oriented classification of low-level log traces. Expert Syst Appl 2020;153:113386. <http://dx.doi.org/10.1016/j.eswa.2020.113386>.

[72] [Böhmer K, Rinderle-Ma S. Multi-perspective anomaly detection in business process execution events. In: Otm confederated international conferences’’ on the move to meaningful internet systems’’. Springer; 2016, p. 80–98.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb72)

[73] [Huda S, Ahmad T, Sarno R, Santoso HA. Identification of proc](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb72)ess-based fraud patterns in credit application. In: 2014 2nd international conference on information and communication technology (icoict). 2014, p. 84–9. [http: //dx.doi.org/10.1109/ICoICT.2014.6914045](http://dx.doi.org/10.1109/ICoICT.2014.6914045).

[74] [Baader G, Krcmar H. Reducing false positive](http://dx.doi.org/10.1109/ICoICT.2014.6914045)s in fraud detection: Combining the red flag approach with process mining. Int J Account Inf Syst 2018;31:1–16. <http://dx.doi.org/10.1016/j.accinf.2018.03.004>.

[75] [Stahl C, van der Aalst WM. Behavioral service substitution. In: Web services](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb75)  [foundations. Springer; 2014, p. 215–44.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb75)

[76] [Lübke D. Extracting and conserving p](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb75)roduction data as test cases in exe-cutable business process architectures. Procedia Comput Sci 2017;121:1006–13.

<http://dx.doi.org/10.1016/j.procs.2017.11.130>, cENTERIS 2017 - International [Conference on EN- TERprise Information Sy](http://dx.doi.org/10.1016/j.procs.2017.11.130)stems / ProjMAN 2017 - Inter-national Confer- ence on Project MANagement / HCist 2017 - International Conference on Health and Social Care Information Systems and Technologies, CEN- TERIS/ProjMAN/HCist 2017.

[77] Rubin VA, Mitsyuk AA, Lomazova IA, van der Aalst WMP. Process mining can be applied to software too!. In: Proceedings of the 8th acm/ieee international symposium on empirical software engineering and measurement. Esem ’14, New York, NY, USA: Association for Computing Machinery; 2014, [http://dx.doi.org/ 10.1145/2652524.2652583](http://dx.doi.org/10.1145/2652524.2652583).

[78] [Xu X, Zhu L, Weber I,](http://dx.doi.org/10.1145/2652524.2652583) Bass L, Sun D. Pod-diagnosis: Error diagnosis of sporadic operations on cloud applications. In: 2014 44th annual ieee/ifip international conference on dependable systems and networks. 2014, p. 252–63. <http://dx.doi.org/10.1109/DSN.2014.94>.

[79] [Gupta M. Improving software maintenan](http://dx.doi.org/10.1109/DSN.2014.94)ce using process mining and predictive analytics. In: 2017 ieee international conference on software maintenance and evolution (icsme). 2017, p. 681–6. <http://dx.doi.org/10.1109/ICSME.2017.39>. [80] Pecchia A, Weber I, Cinque M, M[a Y. Discovering process models for th](http://dx.doi.org/10.1109/ICSME.2017.39)e analysis of application failures under uncertainty of event logs. Knowl-Based Syst 2020;189:105054. <http://dx.doi.org/10.1016/j.knosys.2019.105054>.

[81] [van der Werf JME, Verbeek H. Online compliance monitoring of service land-scapes. In: International conference on business process management. Springer; 2014, p. 89–95.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb81)

[82] [Zahoransky RM, Koslowski T, Accorsi R. Toward resilience assessment in business process architectures. In: International conference on computer safety, reliability, and security. Springer; 2014, p. 360–70.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb82)

[83] [Corradini F, Marcantoni F, Morichetta A, Polini A, Re B, Sampaolo M.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb83)

[Enabling auditing of smart contracts through process mining. In: From software engineering to formal methods and tools, and back. Springer; 2019, p. 467–80.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb83) [84] [Choi S-S, Cha S-H, Tappert CC. A survey of binary similarity and distance](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb83) [measures. J Syst Cybern Inform 2010;8(1):43–8.](http://refhub.elsevier.com/S2590-0056(21)00057-6/sb84)

14