A SYSTEMATIC REVIEW ON PROCESS MINING AND SECURITY

Robert Kelemen[1]

Abstract
Security is an important issue that every organisation should address. One approach to secure systems could be the use of process mining techniques. Process mining is an emerging discipline which can extract knowledge from event logs that are available in information systems. In the security context, process mining is used to analyse security trails to detect anomalies in process execution. While some of the data mining projects are already implemented in banking, insurance and telecom sector the interesting question is: What is happening in public sector? This paper investigates the research on process mining techniques in security domain and tries to discover the examples of its implementation especially in public sector.

The systematic review has been conducted in order to give an overview of state-of-the-art process mining techniques used in security context, to classify the main areas of development, algorithms, tools and to identify possible future research course.

1. Introduction

The basic idea for this research was to explore the relevant databases and find out what are the common ways of implementing the process mining techniques in the field of security, with special emphasis on usage in public domain. Nowadays, the security field is very interesting and challenging. Especially regarding the interconnected systems. One thing is sure, organizations use more and more different systems, generate more and more data, and inevitably, more and more data about data (event logs) and, despite all the security systems implemented, they are still vulnerable. The security is becoming the key challenge. Usually, analysing the event logs is a very boring job, it takes a lot of time and it is usually performed with the focus on one system only. That could cause security breaches in systems. One approach to raise security of information systems could be to use process-mining techniques. Therefore, it is necessary to define the process mining.

1.1 Process mining

The organizations use different systems and generate event log data. The challenge is to exploit event data in a meaningful way, to provide insights, identify bottlenecks, anticipate problems, record policy violations, recommend countermeasures, and streamline processes [1]. This is the W. M. P. van der Alst's explanation of the "philosophy" of process mining.

A group of 75 people from more than 50 organizations, in the IEE Task Force on Process Mining created the Process Mining Manifesto. This is the set of guiding principles and challenges whose purpose is to serve as a guide for software developers, scientist, consultants, business managers and end-users [2]. Process mining is a relatively new research discipline that comes between computational intelligence and data mining on the one hand, and process modelling and analysis on

[1] Varazdin County, Franjevacki trg 7, 42000 Varazdin, Croatia, robert.kelemen@vzz.hr

the other hand [2]. The idea of process mining is to discover, monitor and improve real processes by extracting knowledge from event logs readily available in today's (information) systems. Process mining includes (automated) process discovery, conformance checking, social network/organizational mining, automated construction of simulation models, model extension, model repair, case prediction, and history-based recommendations [2]. Process mining provides an important bridge between data mining and business process modelling and analysis [2]. A Starting point for process mining is an event log. All process mining techniques assume that it is possible to sequentially record events so that each event refers to an activity and is related to a particular case [3]. Event logs may store additional information such as the resource executing or initiating an activity, the timestamp of an event, or data elements recorded with an event [3].

Some authors [4] have already undertaken the systematic review on process mining. Their task has been to provide summary of the current trends in process mining practice and highlight the need for the future research. The research has been focused on process mining, the problems within process mining and issues to be solved regarding automatization, methods and standardization. The systematic review on security in Process-Aware Information systems (PAIS) was performed by Leitner and Rinderle-Ma [5] in 2014. The objective was to investigate research on security in PAIS and to establish common understanding of terminology in this context. It investigates which security controls are currently applied in PAIS and utilized security controls [5]. In conclusion, the authors underline that they want to detect unauthorized access or misuse of permissions in RBAC models using process mining techniques

1.2 Information security

Information security defined by International Organization for Standardization [6] is the preservation of confidentiality, integrity and availability of information. Internal and external parties can use International Standard [7] to assess the organization's ability to meet the organization's own information security requirements.

It is essential that an organization identifies its security requirements. There are three main sources of security requirement [8]:

1. the assessment of risks to the organization, taking into account the organization's overall business strategy and objectives. Threats to assets are identified, vulnerability to and likelihood of occurrence is evaluated and potential impact is estimated through risk assessment;

2. the legal, statutory, regulatory and contractual requirements that an organization, its trading partners, contractors and service providers have to satisfy, and their socio-cultural environment;

3. the set of principles, objectives and business requirements for information handling, processing, storing, communicating and archiving that an organization has developed to support its operations.

When considering an enterprise information system, security plays a role at different levels, i.e., from the level of UNIX processes to the level of interorganizational business processes [9]. Security policies may refer to things ranging from cryptography and role-based access control to auditing and the four eyes principle. Security violations may be conducted by hackers but also by white-

collar criminals. Literature on security can be split into computer security and auditing [9]. Although computer security and auditing are at very different levels, the absence or presence of certain behavioural patterns may indicate security violations. Therefore, audit trails can be useful. Fortunately, many enterprise information systems store relevant events in some structured form [9].

An Information Security Management System (ISMS) is a systematic approach for establishing, implementing, operating, monitoring, reviewing, maintaining and improving an organization's information security to achieve business objectives [6]. It is based upon a risk assessment and the organization's risk acceptance levels designed to effectively treat and manage risks [6]. Information Security Management System (ISMS) involves the following essential components [10]: 1) Information security policies; 2) Organization of information security; 3) Human resource security; 4) Asset management, 5) Access control; 6) Cryptography; 7) Physical and environmental security; 8) Operations security; 9) Communications security; 10) System acquisition, development and maintenance; 11) Supplier relationships; 12) Information security incident management; 13) Information security aspects of business continuity management; 14) Compliance.

Therefore, in this research the security will imply the usage of any of above mentioned terms defined in [6].

1.3 Contribution

This paper provides a systematic literature review [11] on process mining techniques adopted in the security domain with special emphasis on public sector implementations. To achieve this goal the following research questions are formulated:

A. What are the prevailing topics in research papers on the process mining usage in security?

B. What are the main challenges of process mining usage in security domain?

C. What are the main possibilities for future work or identified research areas?

The rest of the paper is structured as follows: the research methodology is described in Section 2, the main explanation of research results is presented in Section 3 and finally, concluding remarks in Section 4.

2. Research methodology

To gain insight on process mining implemented in the security domain a literature review has been conducted according to the general systematic review steps proposed by Kitchenham [11]. The most common reasons for undertaking a systematic review are [11]: 1) to summarise the existing evidence concerning treatment or technology e.g. to summarise the empirical evidence of the benefits and limitations of a specific agile method; 2) to identify any gaps in current research in order to suggest areas for further investigation; 3) To provide a framework/background in order to appropriately position new research activities.

2.1 Literature search

The first step in literature research was to define the keywords which will be used in database search. The first keyword was "process mining" and the second "security" as the general term which

includes all the essential components. The content analysis method has been used for the paper analysis and establishing categories and then counting instances that fall into each category [12]. Consequently, the database query is as follows: "process mining" AND "security".

The second step was to define the relevant databases to perform search: 1) Web of Science; 2) Computer Science Bibliography – DBLP; 3) Science Direct; 4) IEEE Computer Society; 5) ACM; 6) Springer; 7) Google Scholar.

For all the databases, the search query was limited to: proceedings papers and articles, timespan from 2000 -2016, and categories: Computer science information systems, Computer Science theory methods.

The inclusion and exclusion criteria for article selection are presented in Table 1. Based on these criteria all the articles have been selected. The whole research has been conducted in three phases. The first phase objective has been to rise queries in databases with defined keywords and titles and keywords analysis. The articles matching the criteria have been selected. The second phase has taken into consideration the abstracts and duplicates. If the paper has matched criteria, the third phase has been performed – the full text analysis. The papers that have not matched the criteria have also been excluded from the research.

	PHASE 1	**PHASE 2**	**PHASE 3**
INCLUSION CRITERIA	Title indicates that the paper is about the process mining and security Keywords indicate that the paper is about the process mining and security	Abstract of articles indicates that the paper is about the process mining and security	Topics on the process mining usage in security examples of implementation in the public sector The main challenges of process mining usage in security domain The main possibilities of future work or identified research areas
EXCLUSION CRITERIA	Title indicates that the paper is about another topic, it can include process mining but without security Keywords indicate that the paper is about another topic Book Chapter PhD or Master Thesis	Abstract of articles indicates that the paper is not related to the topic No abstract available Duplicates excluded	There are no process mining techniques related to the security There are no security issues covered

Table 1: Paper inclusion and exclusion criteria

An initial search was performed during November 2016 and the initial number of 393 papers were discovered. The results of the first phase – rising queries with keywords in databases are presented in Table 2.

	DATA SOURCE	**NO. PAPERS**
1	Web Of Science	14
2	DBLP	4
3	Science Direct	151
4	IEEE Computer Society	22
5	ACM	7
6	Springer	72
7	Google Scholar	126
	TOTAL	**393**

Table 2: First phase – rising queries with keywords in databases

The second selection was based on the analysis of the abstracts and removing duplicates. After the second phase 60 articles remained. The third selection was performed by full text analysis and final number of papers ready for analysis was 40. The results are presented in table 3.

	DATA SOURCE	NO. PAPERS-1ST PHASE RESULTS	NO. PAPERS-2ND PHASE RESULTS	NO. PAPERS - 3RD PHASE RESULTS
1	Web Of Science	14	4	1
2	DBLP	4	0	0
3	Science Direct	151	9	5
4	IEEE Computer Society	22	8	7
5	ACM	7	2	0
6	Springer	72	2	2
7	Google Scholar	126	35	25
	TOTAL	**393**	**60**	**40**

Table 3: The number of papers according to phases

3. Research Results

The analysis of 40 identified articles has shown that in time span 2000 - 2016 the number of articles relating to process mining and security has been rising since 2012. – Table 4.

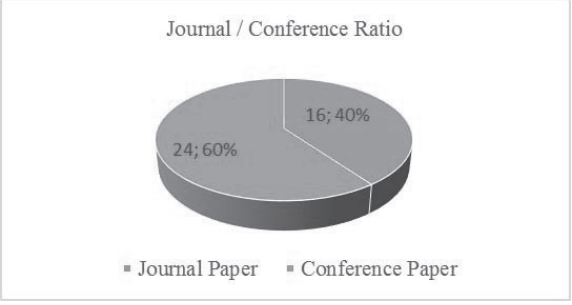

Figure 1: Journal/conference ratio

Figure 1 presents the ratio of the papers published for conferences (60%) to the papers published in journals (40%).

YEAR	NO. PAPERS
2004	1
2005	1
2008	4
2009	2
2010	1
2011	3
2012	6
2013	6
2014	5
2015	4
2016	7
TOTAL	**40**

Table 4: The number of papers per year

Table 5 shows the Journals and the number of papers published in them according to the criteria of this research. It is very interesting that 24% of all the papers have been published in a really acknowledged journal - Information Systems journal, whose Impact factor is 1.832.

JOURNAL	NO. PAPERS	IF	5 YR IF	SNIP	SJR
Information Systems	4	1.832	2.105	3.111	1.202
International Journal of Business Process Integration and Management	2	-	-	-	-
Computers & Security	1	1.640	1.783	2.563	1.020
Computing	1	0.872	1.144	0.956	0.440
Decision Support Systems	1	2.604	3.271	2,271	2.262
EURASIP Journal on Information Security	1	-	-	-	0.295
Expert systems with applications	1	2.981	2.879	2.561	1.839
Information and Software Technology	1	1.569	2.016	3.163	0.920
International Journal of Communication Networks and Information Security (IJCNIS)	1	-	-	-	-
International Journal of Computer, Electrical, Automation, Control and Information Engineering	1	-	-	-	-
Journal IDA	1	0.631	-	-	-
Journal of Information and Data Management	1	-	-	-	-
TOTAL	**16**				

Table 5: The number of papers per journal

The most popular conferences to publish research papers in the scope of this particular research are: ACM symposium on Applied computing, Business Process Management Workshops, IEEE Conference on Communications and Network Security (CNS), International Carnahan Conference on Security Technology (ICCST) with two published papers. The list of conferences with number of published papers is presented in Table 6.

1) In the *Conformance checking* category, the three main development areas can be identified: different approaches to conformance checking, proposing new frameworks and algorithms and visualization. An approach to check if the data recorded in the event logs of a process aware information systems (PAIS) conforms to the corresponding process-related Role-based access control (RBAC) model is presented [20]. The process-related RBAC models are automatically transformed to corresponding Linear Temporal Logic (LTL) rules which are used to check the event logs for violations of the policies that are defined via the RBAC model [20]. The results of this conformance check can serve as basis for security and domain experts to detect violations [20]. In future work, authors plan to integrate presented work into related approaches for analysing the control flow [20].

VENUES	NO. PAPERS
ACM symposium on Applied computing	2
Business Process Management Workshops	2
IEEE Conference on Communications and Network Security (CNS)	2
International Carnahan Conference on Security Technology (ICCST)	2
Advances on P2P, Parallel, Grid, Cloud and Internet Computing	1
Asia-Pacific Software Engineering Conference (APSEC)	1
CEUR Workshop Proceedings	1
Intelligence and Security Informatics	1
International Conference of the Chilean Computer Science Society (SCCC)	1
International Conference on Computer, Control, Informatics and its Applications (IC3INA)	1
International Conference on Enterprise Information Systems	1
International Conference on Information and Communication Technology (ICoICT)	1
International ISC Conference on Information Security and Cryptology (ISCISC)	1
International Symposium on Intelligent Systems and Informatics (SISY)	1
International Workshop on Business Process Modelling, Development and Support/14th Conference on Exploring Modelling Methods for Systems Analysis and Design	1
International Workshop on Database and Expert Systems Application	1
International Workshop on Security Issues with Petri Nets and other Computational Models (WISP 2004)	1
Management Intelligent Systems	1
On the Move to Meaningful Internet Systems: OTM 2011	1
Simposio Brasileiro de Sistemas de Informaçao	1
TOTAL	**24**

Table 6: The number of papers per conference

A novel technique to identify potential causes of failures in business processes based on event logs has been proposed [29]. Four types of causes can be identified and missing or unnecessary activities and behavioural patterns that differ from each other in the control flow or the time perspective can be found [29]. An approach to categorize deviations, which enables auditors to quickly gain an overview of different types of existing deviations along with their frequencies is proposed [33]. Categorizing deviating process instances can also give an insight for assessing the risk at case level [33]. An application of process mining for the financial audit has been presented in [34]. The conformance analysis can be used as an audit technique in the execution of the financial audit [34]. Using a generalized and simplified process model of an organisation's procurement process, and the event log of SAP R/3, a number of deviating process instances and fitting classes of transactions have been detected [34]. The paper Enhancing Mobile Device Security with Process Mining [43] presents a research project which uses mining of processes found in mobile device activity logs and analysis of those processes. The authors report on the lack of quality data in the area of simulated attacks. The target platform - Android - lacks capabilities required for gathering of more detailed data about processes [43]. The advantage of such process mining is its ability to perform heuristic analysis and to detect multi-channel attacks, which are difficult to catch using traditional methods. Process mining can be used to detect not only actual attacks, but also any other behaviour which could cause harm [43]. If the analysis of activity logs can be performed off-site but in near real time, it can also cover a large number of known models of harmful activity. Performed in the mobile device itself, the performance barrier of these devices is soon met [43]. An approach and a

supporting tool for the evaluation of the overall process risk and the prediction of process outcomes based on the analysis of information recorded in event logs has been presented [48]. It can help managers evaluate the overall risk exposure of their business processes, track the evolution of overall process risk, identify changes and predict process outcomes based on the current value of overall process risk [48]. Development of new techniques for both measuring and visualizing non-conformance and the support for further modelling languages is needed [48].

A framework which structures the field of process deviation analysis and a general outline to detect high-level process deviations has been formulated [23]. The change of patterns in PAIS provides an interesting starting point for future research [23]. A Proces Aware Host-based Intrusion Detection (PAHID) model has been introduced [25]. The model uses both anomaly detection and misuse detection techniques to provide more efficiency. Organizational perspective is considered to detect more attacks. The model is automated and flexible and can deal with large logs. In future works case perspective should also be considered to provide more accuracy and to perform an evaluation on a set of real log [25]. A methodology and infrastructure for securing and validating inter-organizational business processes with high-throughput at run time is presented [27]. The method uses process mining techniques and can be used to validate and enforce security rules as well as contractual and legal requirements and is specifically adapted to distributed processes [27]. It can be used to rollback multi-step, multi-party transactions after an anomaly is detected [27]. It has been implemented for the distributed issuing infrastructure of electronic identity cards in Italy [27]. The traditional deviation detection approaches have problems in situations where event logs contain a variety of process behaviour [35]. A novel algorithm named cyclic SC which is faster than cluster-based approaches and more accurate than model-based approaches has been proposed [35]. The framework is configurable and it is used to create a concrete approach for detecting deviations from control-flow perspective [35]. A novel conformance algorithm that balances the deviations with respect to all perspectives (control-flow Data dependencies, resource assignments and time constraints) based on a customizable cost function has been proposed [36]. Some of the future work possibilities are investigating the nature and effects of different cost functions, looking at the specific alignments to dig into specific deviations at the case level, improving the visualization so that it is easier to explore a large set of alignments. In process security checking, the conformance of a case depends on the behaviour observed in other cases that are being executed [36].

A novel approach for visualizing database intrusion detection using process mining techniques modelling low-level event logs has been proposed [19]. The visualization will be able to help security officers who might not know deeply the complex system, identify the true positive detection and eliminate the false positive results [19]. One of our future research courses might be to investigate heuristics to embed them in this algorithm for the intrusion detection purpose [19].

2) The papers in the *Anomaly detection* category can be logically divided in following areas: algorithms, models and implementations. It is argued that both aspects (discovery and delta analysis) are relevant for computer security and auditing [9]. In the context of security, the concept of process mining and the α-algorithm is examined to discover a process that describes all possible behaviours [9]. The α-algorithm discovers a net that models all acceptable behaviour whenever the complete log given as input has only acceptable audit trails and the discovered net is a sound WF-net [9]. Once the net is discovered, the conformance of every new audit trail can be verified by playing the "token game". The anomalous audit trails do not correspond to possible firing sequences in the "token game" for the discovered net [9]. The "token game" detects the point in which the audit trail diverges from the normal behaviour and also allows the real time verification of trails [9]. The challenges Detecting Anomalous Process Execution and Checking Process Conformance are

highlighted. The authors predict that organizations will increasingly need to store and monitor audit trails in order to detect intrusion on low level and fraud detection on high-level security. The report on ways to automate and control business processes, and also to track misuse of their systems has been presented [15]. The control provided by normative systems may compromise the necessary flexibility to companies. The approach to identify anomalous traces, which may represent a misuse, has also been presented. The ProM framework has been described and a real application of approach has been carried out with a real log from Dutch municipality [15]. Since the presented anomaly detection approach is limited to the control-flow perspective, the data and organizational perspectives should also be considered to provide more accuracy [15]. The automated solution might be implemented using genetic algorithms as well. The research gap has been reported in anomaly detection area in the context of PAIS [16]. The two different approaches for detecting anomalous traces in a log using an algorithm based on Sampling and an algorithm based on Threshold has been presented. The research can be resumed on the assessment of other process mining algorithms and the development of other "noise" metric [16]. More detailed research regarding anomaly detection methods and assessment of algorithms has been presented [18]. Four algorithms for detecting anomalies in logs of PAIS are discussed [21]. One of the algorithms only marks as potential anomalies traces that are frequent in the log, the other three algorithms: threshold, iterative and sampling are based on mining a process model from the log, or a subset of it [21]. The research was limited because only the control-flow perspective has been adopted thus its approach is incomplete [21]. An anomaly detection algorithm for logs of PAIS based on four different metrics: fitness, structural appropriateness, behavioural appropriateness, and size has been proposed [24]. Such an algorithm is important in application scenarios where a flexible and secure business process is essential [24]. The anomaly detection algorithm was based on the process mining α-algorithm. It is important to note that the accuracy of algorithms is strictly related to the following components: (i) the process mining algorithm; (ii) the metric used to evaluate the compliance variance between two logs (with and without anomalous traces); and (iii) the threshold value used to define the compliance variance limit for logs without anomalous traces [24]. The future research should consider the assessment of other process mining algorithms (e. g. α – algorithm extensions), other metrics, and a deeper study of threshold values [24]. A Dynamic Threshold Algorithm for anomaly detection of traces in PAS logs is presented in order to provide a solution to balance the trade-off between flexibility and security [28]. This paper emphasises anomalies as frauds. This algorithm has a statistically significant better accuracy for both dataset of logs against the algorithms proposed in [21]. The proposal for the future work is to combine the results of Sampling and Dynamic Threshold Algorithms into a single decision [28].

In [26] a genetic-based anomaly detection model for logs of PAIS has been presented. This model is appropriate for all application domains to provide a trade-off between flexibility and security [26]. The proposed anomaly detection approach is concerned with the control-flow perspective [26]. Therefore, in the future, data and organizational perspectives will be considered to provide more accuracy [26]. An effective Business Process Mining Based Insider Threat Detection system has been introduced [30]. The system uses genetic mining method to discover the control-flow model of the business process. On that basis, the system further mines the tree-structured operator's behaviour profile. Additionally, it also gets the normal data about performance on specific events through statistical method and expert knowledge [30]. Possible abnormal behaviours that a malicious operator may perform when conducting an insider attack are analysed, as well as the influences of these behaviours on business activities [30].

A research to study the DNS traces using process mining instead of traditional statistical approach has been presented [39]. One distinctive feature is the representation of type of each message as a

node; this is different from the usual way of representing the data where each host is a node and the messages are represented as edges. This data representation could also be applied to other types of data workflows, such as HTTP and SSH negotiation [39]. This could give a new insights into how much implementations differ from the standards, as well as into discovering or detecting different types of attacks [39]. It is planned to study how this graph representation of DNS traces can be analysed by process mining and other technologies to discover new patterns and behaviours that are hard to identify otherwise: Different types of attacks may lead to different types of graphs [39]. Obtaining this information in real time or near real time could provide more information to take specific countermeasures to avoid the attack, whether by automatic or manual systems [39]. Similar to [39], an approach in using Passive Testing (used in protocol and software conformance checking) and Process Mining (used in enterprise workflow analysis) techniques for analysing DNS operation traces has been presented [44]. This approach was applied over a Day in Internet Life DNS traces for showing how easily a mail bonnet attack can be discovered [44]. As future work, it is planned to compare the results obtained using PT/PM algorithms with other techniques such as passive/active testing and automating tools for checking the conformance of the protocols [44]. A paper on different usage of process mining techniques in order to present an application of process cube to software defect resolution process to analyse and compare process data from a multi-dimensional perspective is proposed [42]. Each process cube cell is defined by metrics from multiple process mining perspectives like control flow, time, conformance and organizational perspective [42]. The process cube with 9 dimensions: issue report timestamp, priority, state, closed status, OS, component, bug type, reporter and owner is defined [42]. OLAP cube operations: slice, dice, roll-up and drill-down, and create materialized sublog for each cell are applied [42].

3) The *Compliance control* category can be structured in following basic areas: the framework development, and automation with security. A process mining as a basis for various security audits of business process and corresponding business process management systems has been reported [13]. The process discovery has been evolved beyond original α-algorithm, and there are several process discovery methods used depending on the focus of the analysis [13]. The three drawbacks are reported [13]: 1) a lack of tools capable of analysing the structures produced by process discovery algorithms, so that it is largely manual; 2) support for several desirable structures relevant for security analysis can still not be drawn from event logs; 3) precision issues regarding the structures are still missing. This indicates some of the future research course regarding "security analytics" which in general is a powerful basis for risk analysis [13]. A compliance monitoring framework that tackles three major challenges: Identification and Monitoring of Individual Activations of a Compliance Rule, Proactive Prevention of Violations and Root Cause Identification in Case of Violations has been proposed [22]. The framework enables the identification of all activations of a compliance rule and enables to "initiate" a Compliance Rule Graph (CRG) each time a new activation is observed and thus to individually monitor the activations [22]. In the future, efficiency and further addressing the interplay of CRGs should be improved. A framework for context-based analysis of transaction data to validate and secure inter-organizational business processes have been proposed [40]. The analysis is based on process mining techniques and uses observations taken at all relevant communication layers which are combined with semantic analysis [40]. The presented context based analysis allows the simple implementation of complex security and compliance policies [40]. A framework for Compliance Monitoring Functionalities (CMF) that enables the systematic comparison of existing and new approaches for monitoring compliance rules over business processes during runtime has been defined [46]. The framework consists of ten Compliance Monitoring Functionalities (CMFs) and includes requirements for the constraint modelling notation, requirements with respect to the execution and user requirements [46]. The work can be further extended in several directions, e.g., to cross-

organizational or configurable processes [46]. An incremental approach to check the conformance of a process model and an event log has been proposed [47]. The fitness between the log and the model is measured and the appropriateness of the model can be analysed with respect to the log [47]. Appropriateness can be evaluated from both a structural and a behavioural perspective [47]. To operationalize the ideas a Conformance Checker has been implemented within the ProM framework, and it has been evaluated using artificial and real-life event logs [47]. Future work will aim at the development of new techniques for both measuring and visualizing non-conformance, and at the support of further modelling languages.

A novel system for the provision of efficient operational support for distributed and security sensitive business processes in which automated process validation and extensive troubleshooting functionality is closely integrated with IT Service Desk (SD) operations has been presented [41]. That system's capabilities enable the efficient and automated detection and resolution of anomalies in distributed business processes [41]. The system improves Service Desk (SD) performance and increases process security through real time compliance checking [41]. The system further reduces security risks associated with SD operations and SD staff-user interactions including social engineering attacks as it allows SD staff to cross-reference user provided information with system reports [41].

4) In the *Fraud detection* category the identified papers describe the new detection methods of Process-based Fraud (PBF) and implementations. Process mining implementation in the context of transaction fraud detection has been presented in [17]. Although tools are available, they are still quite under-developed and there is need to enhance tools like ProM to better automate the audit process and to visualize results for management [17]. The ontology-based process modelling to model and capture the business process anomalies and the method of multi-level class association rule learning (ML-CARL) to detect fraud in business process has been proposed [31]. A New Method for Occupational Fraud Detection in Process Aware Information Systems is proposed [32]. In this approach, a process model is mined and its structures specified, a numerical vector for each process instance using the structures is built and then outliers in the vectors using statistical information specified [32]. These outliers are suspected of frauds. The method is focused solely on the sequence of activities. To improve the method, some other process mining algorithms, real logs time of activity execution and unauthorized performers should be performed [32]. Ahmad and Sarno have indicated that none of the several earlier proposed detection methods of Process-based Fraud (PBF) presents identification of PBF attributes and pattern clearly [38]. In order to detect PBF a PBF table pattern is required with a set of PBF attributes which are proposed in research [38]. Some future work should focus on designing a more effective pattern of PBF with domain expert [38]. A comprehensive rule-based compliance checking approach as a possible solution to eliminate the limited fit has been proposed [45]. The approach enables analysts to uncover compliance failures as well as to identify and assess potential risks [45]. The major opportunities in the research area are: effectiveness, persuasive evidence and audit independence, assumption (data quality) [45]. Improvements can be found in the ability to take additional data into account, the reduction of possible distortions (including over specification) and no need for generalization [45]. The identified challenges are: distortions in interpretation and pattern design and continuous monitoring/auditing [45]. A logical future step is to further test this approach and to tackle the identified challenges focusing on the development of a continuous monitoring/auditing approach based on process mining techniques [45].

5) In the *Risk management* category, the following areas have been identified. The first area is to propose a comprehensive process mining applicability framework that provides a clear guidance as

well as a common language for business process mining in the context of enterprise risk management and a broader governance, risk and compliance (GRC) setting [49]. The applicability framework consists of three interrelated dimensions: (i) Process Mining Techniques Dimension (Process Discovery & Visualization, Conformance Checking & Delta Analysis and Rule-Based Property Verification); (ii) Control Functions Dimension; (iii) Control Function Activities Dimension: Discusses the potential application areas of process mining for GRC activities: Risk Identification & Assessment, Control Activities and Information, Documentation & Communication [49]. The second area is more theoretical: to provide an overview of scientific research efforts regarding the integration of security and risk considerations into business process management [14] and to provide a review of existing literature in financial fraud detection and compare their findings [50]. Some challenges [14] were identified: 1) Consideration of different impact perspectives; 2) Occurrence probabilities; 3) Extension of security/dependability attributes (availability, confidentiality, integrity, accountability, safety, etc.); 4) Efficient resource allocation taking security aspects into account; 5) Improvements on the current business process notations to facilitate risk/security evaluation; 6) Providing metrics on the security robustness of business processes.

6) In the *Access Management* category the authors reports that hardly any supportive means for the automated detection and refinement as well as management of identity and access management (IAM) policies are available [37]. A dynamic policy management process (DPMP) which structures the activities required for policy management in identity and access management environments into four phases [37] has been proposed. It facilitates a mining engine which generates policy recommendations based on contextual data of employees and further presents gathered results to human IAM engineers [37]. For future work, it is planned to extend the DPMP in order to improve the representation and management of policy recommendations and provide an analysis of policy interdependencies [37].

7) The first *systematic literature review* on security in PAIS [5] have taken into the consideration different aspects of security in PAIS. In the future work, they aim at working towards closing the gap between security research in Information Systems and PAIS, to concentrate on some of the open issues outlined in the paper such as the development of detection and reaction controls and to investigate the evaluation of inter-instance constraints with mining techniques [5].

4. Conclusion

In this research 393 papers have been proven to fulfil the criteria claiming that process mining techniques must be implemented in the field of security with special emphasis on implementation in public domain. The systematic review method has been chosen and in three iterations based on clearly defined criteria 40 papers have been selected. At first sight, it was concerning that the total number of 40 papers is not impressive in such a field, but with ongoing research the systematic review paper [5] has been discovered where authors have underlined that some future work should further explore the process mining as the next big topic. That finding has confirmed the right course of this research which is combining process mining with security and has also revealed that there could be some room for further investigations.

At the beginning of the research three research questions have been formulated:

A. What are the prevailing topics in research papers on the process mining usage in security?

B. What are the main challenges of process mining usage in security domain?

C. What are the main possibilities for future work or identified research areas?

The answer to the first research question has been given using the content analysis. In selected papers 6 security categories have been identified: conformance checking; anomaly detection; compliance control; fraud detection; risk management; access management. The most popular security categories are Conformance checking which has been analysed in the 30% of all the published papers and the Anomaly detection with 30% of all published papers. The complete list of the prevailing topics in research papers on the process mining usage in security is presented in Table 7.

No.	SECURITY CATEGORY	NO. PAPERS
1	Conformance checking [19], [20], [23], [25], [27], [29], [33], [34], [35], [36], [43], [48]	12
2	Anomaly detection [9], [15], [16], [18], [21], [24], [26], [28], [30], [39], [42], [44]	12
3	Compliance control [13], [22], [40], [41], [46], [47]	6
4	Fraud detection [17], [31], [32], [38], [45]	5
5	Risk management [14], [49], [50]	3
6	Access management [37]	1
7	Systematic review [5]	1
	TOTAL	**40**

Table 7: The number of papers per Security Category

The second research question has dealt with the main challenges of process mining usage in security domain. The answer to this question has been provided in the text analysis of selected papers where it was possible to extract this information.

The challenges Detecting Anomalous Process Execution and Checking Process Conformance have been highlighted. Some research has been incomplete because only one perspective has been adopted in the research.

The lack of tools capable of analysing the structures produced by process discovery algorithms has been pointed out. Support for several desirable structures relevant for security analysis (e.g. such as process structures with data and role hierarchies) could still not be drawn from event logs. Precision issues regarding the structures are, for the most approaches, still missing. Extension of security/dependability attributes (availability, confidentiality, integrity, accountability, safety, etc.) is needed. Efficient resource allocation taking security aspects into account is necessary. Improvements on the current business process notations to facilitate risk/security evaluation should be undertaken. Providing metrics on the security robustness of business processes is necessary.

In the paper Intelligent financial fraud detection: A comprehensive review [45] of some of the key issues associated with financial fraud detection and suggested areas for future research are presented as follows:

- Typical classification problems: CI and data mining-based financial fraud detection is subject to the same issues as other classification problems, such as feature selection, parameter tuning, and analysis of the problem domain.

- Fraud types and detection methods: Financial fraud is a diverse field and there has been a large imbalance in both fraud types and detection methods studied: some have been studied extensively while others, such as hybrid methods, have only been looked at superficially.

- Privacy considerations: Financial fraud is a sensitive topic and stakeholders are reluctant to share information on the subject. This has led to experimental issues such as under sampling.

- Computational performance: As a high-cost problem it is desirable for financial fraud to be detected immediately. Very little research has been conducted on the computational performance of fraud detection methods for use in real-time situations.

- Evolving problem: Fraudsters are continually modifying their techniques to remain undetected. As such detection methods are required to be able to constantly adapt to new fraud techniques. Disproportionate misclassification costs: Fraud detection is primarily a classification problem with a vast difference in misclassification costs. Research on the performance of detection methods with respect to this factor is an area which needs further attention.

- Generic framework: Given that there are many varieties of fraud, a generic framework which can be applied to multiple fraud categories would be valuable.

The third research question has investigated which are the main possibilities of future work or identified research areas?

The authors predict that organizations will increasingly need to store and monitor audit trails in order to detect intrusion on low level and fraud detection on high-level security. Since the presented anomaly detection approach is limited to the control-flow perspective, the data and organizational perspectives should also be considered to provide more accuracy. One of the future work proposals is to combine the results of Sampling and Dynamic Threshold Algorithms into a single decision for anomaly detection in PAS logs [28].

Bustos-Jimnez et al. suggests further research in process mining and other technologies in order to analyse graph representation of DNS traces to discover new patterns and behaviours that are hard to identify otherwise: Different types of attacks may lead to different types of graphs [39]. Obtaining this information in real time or near real time could provide more information to take specific countermeasures to avoid the attack, whether by automatic or manual systems [39].

Process mining has been used as a basis for various security audits of business process and corresponding business process management systems. However, there is a lack of tools capable of analysing the structures produced by process discovery algorithms, so that it is largely manual [13]. Secondly, support for several desirable structures relevant for security analysis [13]. Thirdly, precision issues regarding the structures are, for the most approaches, still missing [13].

Although tools for fraud detection are available, they are still quite unreliable and such tools like ProM should be enhanced in order to better automate the audit process and to visualize results for management. Some authors suggest that the future research should consider the assessment of other process mining algorithms (e. g.: extensions of α -algorithm), other metrics, and a deeper study of threshold values. Some future work should be directed towards designing a more effective pattern of PBF with domain expert.

Possible future research directions are: investigating the nature and effects of different cost functions, looking at the specific alignments to dig into specific deviations at the case level, improving the visualization so that it is easier to explore a large set of alignments.

The development of new techniques for both measuring and visualizing non-conformance with the support of further modelling languages are needed.

Regarding implementation of process mining in public sector, only two papers have explicitly stated that real application has been carried out. The first identified paper was Anomaly Detection Using Process Mining [15]. The ProM framework has been described and a real application of approach has been carried out with a real log from Dutch municipality [15], the second paper was Balanced multi-perspective checking of process conformance [36] where the process mining algorithm is used in management of road traffic fines.

5. References

[1] VAN DER ALST, W. M. P.: *Process Mining - Discovery, Conformance and Enhancement of Business Processes*. Berlin: Springer-Verlag Berlin Heidelberg, 2011.

[2] VAN DER AALST, W. *et al.,*: Process Mining Manifesto, in *Business Process Management Workshops*, vol. 99, F. Daniel, K. Barkaoui, and S. Dustdar, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 169–194.

[3] VAN DER AALST, W. M. P. and DUSTDAR, S.: Process Mining Put into Context, *IEEE Internet Comput.*, vol. 16, no. 1, pp. 82–86, Jan. 2012.

[4] TIWARI, A. TURNER, C. J. and MAJEED, B.: A review of business process mining: state-of-the-art and future trends, *Bus. Process Manag. J.*, vol. 14, no. 1, pp. 5–22, Feb. 2008.

[5] LEITNER, M. and RINDERLE-MA, S.: A systematic review on security in Process-Aware Information Systems – Constitution, challenges, and future directions, *Inf. Softw. Technol.*, vol. 56, no. 3, pp. 273–293, Mar. 2014.

[6] Information technology — Security techniques — Information security management systems — Overview and vocabulary, ISO/IEC 27000:2014. International Organization for Standardization, 2014.

[7] Information technology – Security techniques – Information security management systems – Requirements (ISO/IEC 27001:2013). International Organization for Standardization, 2013.

[8] Information technology – Security techniques – Code of practice for information security controls (ISO/IEC 27002:2013+Cor 1:2014). International Organization for Standardization, 2014.

[9] VAN DER AALST, W. M. P. and DE MEDEIROS, A. K. A.: Process Mining and Security: Detecting Anomalous Process Executions and Checking Process Conformance, *Electron. Notes Theor. Comput. Sci.*, vol. 121, pp. 3–21, Feb. 2005.

[10] BSI-Standard 100-1, Information Security Management Systems (ISMS). Bundesamt für Sicherheit in der Informationstechnik (BSI), Godesberger Allee 185 -189, 53175 Bonn, 2008.

[11] KITCHENHAM, B.: Procedures for performing systematic reviews, *Keele UK Keele Univ.*, vol. 33, p. 2004, 2004.

[12] SILVERMAN, D., *Interpreting Qualitative Data*, 5th ed. Los Angeles: SAGE.

[13] ACCORSI, R., STOCKER, T. and MÜLLER, G.: On the Exploitation of Process Mining for Security Audits: The Process Discovery Case, in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, New York, NY, USA, 2013, pp. 1462–1468.

[14] JAKOUBI, S., TJOA, S., GOLUCH, G. and QUIRCHMAYR, G.: A Survey of Scientific Approaches Considering the Integration of Security and Risk Aspects into Business Process Management, 2009, pp. 127–132.

[15] BEZERRA, F., WAINER, J. and VAN DER AALST, W. M. P.: Anomaly Detection Using Process Mining, in *Enterprise, Business-Process and Information Systems Modeling*, vol. 29, T. Halpin, J. Krogstie, S. Nurcan, E. Proper, R. Schmidt, P. Soffer, and R. Ukor, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 149–161.

[16] BEZERRA, F. and WAINER, J.: Anomaly detection algorithms in logs of process aware systems, 2008, p. 951.

[17] JANS, M., VAN DER WERF, J. M., LYBAERT, N. and VANHOOF, K.: A business process mining application for internal transaction fraud mitigation, *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13351–13359, Sep. 2011.

[18] BEZERRA, F. and WAINER J.: Anomaly Detection Algorithms in Business Process Logs, in *Proceedings of the Tenth International Conference on Enterprise Information Systems,* Barcelona, 2008, vol. AIDSS.

[19] HUYNH, V. H. and LE, A. N. T.: Process Mining and Security: Visualization in Database Intrusion Detection, in *Intelligence and Security Informatics*, vol. 7299, M. Chau, G. A. Wang, W. T. Yue, and H. Chen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 81–95.

[20] BAUMGRASS, A., BAIER, T., MENDLING J. and STREMBECK, M.: Conformance Checking of RBAC Policies in Process-Aware Information Systems, in *Business Process*

Management Workshops, vol. 100, F. Daniel, K. Barkaoui, and S. Dustdar, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 435–446.

[21] BEZERRA, F. and WAINER, J.: Algorithms for anomaly detection of traces in logs of process aware information systems, *Inf. Syst.*, vol. 38, no. 1, pp. 33–44, Mar. 2013.

[22] LY, L. T., RINDERLE-MA, S., KNUPLESCH, D. and DADAM, P.: Monitoring Business Process Compliance Using Compliance Rule Graphs, in *On the Move to Meaningful Internet Systems: OTM 2011*, vol. 7044, R. Meersman, T. Dillon, P. Herrero, A. Kumar, M. Reichert, L. Qing, B.-C. Ooi, E. Damiani, D. C. Schmidt, J. White, M. Hauswirth, P. Hitzler, and M. Mohania, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 82–99.

[23] DEPAIRE B., SWINNEN J., JANS M. and VANHOOF, K.: A Process Deviation Analysis Framework, in *Business Process Management Workshops*, vol. 132, M. La Rosa and P. Soffer, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 701–706.

[24] BEZERRA, F. and WAINER, J.: Fraud detection in process aware systems, 2008, p. 254.

[25] JALALI, H. and BARAANI, A.: Process Aware Host-based Intrusion Detection Model, *Int. J. Commun. Netw. Inf. Secur. IJCNIS*, vol. 4, no. 2, Aug. 2012.

[26] JALALI, H. and BARAANI, A.: Genetic-based anomaly detection in logs of process aware systems, *Int. J. Comput. Electr. Autom. Control Inf. Eng.*, vol. 4, no. 4, pp. 692–697, 2010.

[27] TALAMO, M., ARCIERI F., SCHUNCK, C. H. and D'IDDIO, A. C.: Conformance checking of electronic business processes to secure distributed transactions, 2013, pp. 1–6.

[28] BEZERRA, F. and WAINER, J.: A Dynamic Threshold Algorithm for Anomaly Detection in Logs of Process Aware Systems, *J. Inf. Data Manag.*, vol. 3, no. 3, Sep. 2012, p. 316.

[29] CALDERÓN-RUIZ, G. and SEPÚLVEDA, M.: Automatic discovery of failures in business processes using Process Mining techniques, in *Proceedings of the IX Simposio Brasileño de Sistemas de Información*, Joao Pessoa, Brasil, vol. 1, 2013, pp. 439–450.

[30] ZHU T., GUO Y., MA J. and JU A.: Business Process Mining based Insider Threat Detection System, in *Advances on P2P, Parallel, Grid, Cloud and Internet Computing*, vol. 1, F. Xhafa, L. Barolli, and F. Amato, Eds. Cham: Springer International Publishing, 2017, pp. 467–478.

[31] SARNO, R. and SINAGA, F. P.: Business process anomaly detection using ontology-based process modelling and Multi-Level Class Association Rule Learning, 2015, pp. 12–17.

[32] MARDANI, S. and SHAHRIARI, H. R.: A new method for occupational fraud detection in process aware information systems, 2013, pp. 1–5.

[33] HOSSEINPOUR, M. and JANS, M.: Categorizing Identified Deviations for Auditing, in *Proceedings of the 6th International Symposium on Data-driven Process Discovery and Analysis (SIMPDA 2016)*, Graz, Austria, 2016, pp. 125–129.

[34] HAKVOORT, R. and SLUITER, A.: Process Mining: Conformance analysis from a financial audit perspective, *Int. J. Bus. Process Integr. Manag.*, vol. X, 2008.

[35] LI, G.: A Framework for Detecting Deviations in Complex Event Logs, *IDA*, 2016.

[36] MANNHARDT, F., DE LEONI, M., REIJERS, H. A. and VAN DER AALST, W. M. P.: Balanced multi-perspective checking of process conformance, *Computing*, vol. 98, no. 4, pp. 407–437, Apr. 2016.

[37] HUMMER, M., KUNZ, M., NETTER, M., FUCHS, L. and PERNUL, G.: Adaptive identity and access management—contextual data based policies, *EURASIP J. Inf. Secur.*, vol. 2016, no. 1, Dec. 2016.

[38] HUDA, S., AHMAD, T., SARNO, R. and SANTOSO, H. A.: Identification of process-based fraud patterns in credit application, 2014, pp. 84–89.

[39] BUSTOS-JIMENEZ, J., SAINT-PIERRE, C. and GRAVES, A.: Applying Process Mining Techniques to DNS Traces Analysis, 2014, pp. 12–16.

[40] TALAMO, M., ARCIERI, F., SCHUNCK, C. H. and D'IDDIO, A. C.: Providing context-based security for inter-organizational electronic business processes, 2013, pp. 393–394.

[41] TALAMO, M., POVILIONIS, A., ARCIERI, F. and SCHUNCK, C. H.: Providing online operational support for distributed, security sensitive electronic business processes, 2015, pp. 49–54.

[42] GUPTA, M. and SUREKA, A.: Process Cube for Software Defect Resolution, 2014, pp. 239–246.

[43] HLUCHY, L. and HABALA, O.: Enhancing mobile device security with process mining, 2016, pp. 181–184.

[44] SAINT-PIERRE C., CIFUENTES F. and BUSTOS-JIMENEZ J.: Detecting anomalies in DNS protocol traces via Passive Testing and Process Mining, 2014, pp. 520–521.

[45] WEST J. and BHATTACHARYA M.: Intelligent financial fraud detection: A comprehensive review, *Comput. Secur.*, vol. 57, Mar. 2016, pp. 47–66.

[46] CARON F., VANTHIENEN J. and BAESENS B.: Comprehensive rule-based compliance checking and risk management with process mining, *Decis. Support Syst.*, vol. 54, no. 3, Feb. 2013, pp. 1357–1369.

[47] LY, L. T., MAGGI, F. M., MONTALI, M., RINDERLE-MA, S. and VAN DER AALST, W. M. P.: Compliance monitoring in business processes: Functionalities, application, and tool-support, *Inf. Syst.*, vol. 54, Dec. 2015, pp. 209–234.

[48] ROZINAT, A. and VAN DER AALST, W. M. P.: Conformance checking of processes based on monitoring real behavior, *Inf. Syst.*, vol. 33, no. 1, Mar. 2008, pp. 64–95.

[49] PIKA, A., VAN DER AALST, W. M. P., WYNN, M. T., FIDGE, C. J. and TER HOFSTEDE, A. H. M.: Evaluating and predicting overall process risk using event logs, *Inf. Sci.*, vol. 352–353, Jul. 2016, pp. 98–120.

[50] CARON, F., VANTHIENEN, J. and BAESENS, B.: Rule-Based Business Process Mining: Applications for Management, in *Management Intelligent Systems*, vol. 171, J. Casillas, F. J. Martínez-López, and J. M. Corchado Rodríguez, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 273–282.