University of Zurich[UZH]

MASTER THESIS – Communication Systems Group, Prof. Dr. Burkhard Stiller

MeritMiner4CI: A Novel Approach for Risk Assessment in Cyber Insurance Based on Process Mining

Viktor Matejka
Zurich, Switzerland
Student ID: 17-720-178

Supervisor: Muriel Franco, Eder Scheid
Date of Submission: March 24, 2022

University of Zurich
Department of Informatics (IFI)
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland

ifi

Master Thesis
Communication Systems Group (CSG)
Department of Informatics (IFI)
University of Zurich
Binzmühlestrasse 14, CH-8050 Zürich, Switzerland
URL: http://www.csg.uzh.ch/

Abstract

The market for cyber insurance seems to be at a turning point in the recent years. Premiums are surging at record rates with new claims continuing to be driven by ransomware attacks, as well as by insider threats. The cyber insurance capacity is becoming limited and it is clear that many challenges still need to be tackled in order to avoid market failure. For one, it is clear that traditional risk assessment methods currently applied do not sufficiently address the issues of information asymmetries and the adverse selection and moral hazard that go hand-in hand with them. Therefore, new approaches to audit and assess the level of self-protection of the insureds need to be developed. Taking this into account, this thesis focuses on the assessment of operational cyber risks related to failed internal processes and proposes a novel approach to apply the methods of process mining in cyber insurance. For this purpose, MeritMiner4CI approach was designed and developed. Also, the fundamental challenges and requirements of cyber insurance at the coverage level were clearly mapped to specific process mining methods. The MeritMiner4CI approach was evaluated by conducting a survey with experts and also by considering case study scenarios. The results of this survey provide strong indication that the analyses of the proposed method can be applied by practitioners and have an impact on the ratings of confidence factors that are applied in the industry. Furthermore, quantitative evaluations were conducted to evaluate the performance of the applied methods. Finally, this thesis also highlights how MeritMiner4CI can be integrated with other cybersecurity risk assessment approaches, such as SecRiskAI and MENTOR.

Zusammenfassung

Der Cyber-Versicherungsmarkt scheint in den letzten Jahren einen Wendepunkt erreicht zu haben. Die Prämien steigen auf Rekordhöhe, Ransomware-Angriffe und Insider-Bedrohungen lösen weiterhin neue Schadensfälle aus. Die Cyber-Versicherungskapazität ist begrenzt, und es ist klar, dass noch viele Herausforderungen bewältigt werden müssen, um ein Marktversagen zu vermeiden. Es hat sich gezeigt, dass die derzeit angewandten traditionellen Risikobewertungsmethoden die Problematik der Informationsasymmetrien und der damit einhergehenden adversen Selektion und des moralischen Risikos nicht ausreichend berücksichtigen. Daher müssen neue Ansätze entwickelt werden, um das Niveau des Selbstschutzes der Versicherten zu prüfen und zu bewerten. Unter Berücksichtigung dieser Tatsache konzentriert sich diese Arbeit auf die Bewertung von operationellen Cyber-Risiken im Zusammenhang mit fehlgeschlagenen internen Prozessen und schlägt einen neuartigen Ansatz zur Anwendung der Methoden des Process Mining in der Cyber-Versicherung vor. Zu diesem Zweck wurde der Ansatz MeritMiner4CI entworfen und entwickelt. Der MeritMiner4CI-Ansatz wurde durch eine Expertenbefragung und die Betrachtung von Fallstudien-Szenarien evaluiert, um die Machbarkeit der Implementierung deutlich zu demonstrieren. Die Ergebnisse dieser Umfrage sind ein klarer Hinweis darauf, dass die Analysen der vorgeschlagenen Methode von Experten angewandt werden können und einen Einfluss auf die Bewertungen von Vertrauensfaktoren haben, die in der Industrie verwendet werden. Darüber hinaus wurden quantitative Auswertungen durchgeführt, um die Leistung der angewandten Methoden zu bewerten. Schliesslich wird in dieser Arbeit auch aufgezeigt, wie MeritMiner4CI mit anderen Ansätzen zur Bewertung von Cyber-Sicherheitsrisiken, wie SecRiskAI und MENTOR, integriert werden kann.

Acknowledgments

I would like to thank my supervisor, Muriel Franco, not only for his guidance and support of the thesis, but also during the preceding project on cyber insurance. His continuous feedback and encouragement have been crucial for the thesis. I would also like to thank Prof. Dr. Burkhard Stiller for granting me the possibility to complete my Master thesis at the Communication Systems Group (CSG) at the University of Zurich.

My gratitude also goes to all the underwriters, cyber risk analysts and experts from the industry (incl. those from the CSG) who agreed to take part in the survey and interviews and provided me with crucial feedback and insights.

Finally, I would also like to thank to my family, especially to my brother Milan, who supported and encouraged me as I was working on the thesis.

Contents

Chapter 1

Introduction

Many companies have an understanding of their internal processes that "does not pass a reality check" [1]. In other words, the actual ("as-is") executions of their processes vastly differ from the to-be processes reflected in their internal policies or their often outdated process models. Likewise, their processes might be executed in ways that are not in line with the cyber security best-practices outlined by cyber security frameworks, such as NIST [2], [3], and SEConomy [4]. Finally, processes might often also often be executed in ways that that are in direct compliance violation with established regulatory requirements, such as the General Data Protection Regulation (GDPR) [5], or the Brazilian General Personal Data Protection Law (LGPD) [6].

This is one of the reasons why such companies turn to process management and process improvement. While one can focus on defining a better "de jure" process with traditional process management method, the following question remains: "How can we observe the way how the observed processes are actually executed?". This is the central question of process mining [7]. And, as [1] reports, the prospect of answering this question seems to be a rather compelling proposition as the market for process mining tools grew, according to Gartner [1] from $110 million in 2018 to $320 million in 2019, accounting for a triple digit YoY growth. This market dynamics seems to have recently prompted a number of major technology vendors, such as Salesforce, Microsoft, SAP BPI and many other to make prominent investments in the process mining space [1]. Often, these vendors focus on solutions around operational excellence and improving process efficiency, as well as compliance and conformance checking.

The topic of compliance, or more generally, the overall domain "Governance, Risk and Compliance (GRC)" and "Enterprise Risk Management" (including cyber risk) is, as this thesis further discusses in Chapter 3, a natural candidate for the application of process mining methods and a wide body of research, such as [8], [9] and [10] is available on the matter. It has been shown that process mining can support risk management in different contexts, including risk identification. Once a risk is identified, four fundamental strategies exist how to react to it (see e.g. [11] - we can *avoid it, accept it*, *mitigate it*, or *transfer it*. Precisely the topic of *cyber risk transfer* has been investigated in a previous work conducted in 2021 by Matejka and Huacan Soto [12] within the Communication Systems Group CSG of the University of Zurich UZH. In the aforementioned work, the

authors provided a *Cyber Insurance Framework* and conducted a survey on the cyber insurance market involving some of the largest cyber insurance vendors.

One of the key learnings that we gained in the course of the project was an understanding of how cyber underwriting currently *actually* operates from the perspective of practitioners. As we elaborated on in the preceding project report [12], *underwriter discretion* and *heuristics* often manifest themselves in cyber insurance underwriting practice. From the perspective of why insurers are limited in applying more scalable methods to deal with the complexity - *information asymmetries* seem to be the main challenge and concern of cyber underwriters, based on the results of our survey [12]. Building on that knowledge, it is clear that computational methods that would provide cyber insurance underwriters with objective criteria when making underwriting decisions (such as those on premiums). At the same time, these methods must fit the established industry practices.

For the reasons that were described above, process mining based methods seem to offer considerable promise in that area as they might fit well to the risk assessment process that cyber insurance underwriters apply and base their decisions (on so-called *confidence factors*) on. Process mining could enable cyber insurers to not only consider information provided by the prospective insureds on their defined (*de jure*) processes and risk posture that is often clouded by information asymmetries and moral hazard - it could also enable them to consider and check how those processes run in reality and make assessments based on actual process executions. Possibly leading to better decisions about eligibility for insurance and premiums based on *merit* (*i.e.,* MeritMiner4CI) as well as to *positive*, as opposed to *adverse* self-selection of prospective insureds. For example, it can be hypothesised that insureds who establish that their processes run in a conform and compliant way would be more likely to share this information with cyber insurers, leading to, possibly, more resilient risk pools for the cyber insured. Last but not least, *failed internal policies* and *actions of people* are two of the most prominent operational cyber risks [13] and are often cental for cyber risk assessments by underwriters.

1.1 Motivation and Research Gap

Process mining as a data-analytical method can be traced back, depending on the definition of the term, at least to the 1990s. It is not new [14] and can be considered widely mature both from the research perspective, as well as in practice for a number of use-cases, such as for mining of process-oriented, transactional [15] enterprise systems like SAP, ServiceNow, and Microsoft Dynamics, which aims of getting an understanding of, most typically, Order-to-Cash (O2C) and Procure-to-Pay (P2P) processes. Additionally, process mining has gained wide adoption in the recent years in the auditing, Governance, Risk and Compliance (GRC), and Enterprise Risk Management domains [16].

In this setting, the following questions can be determined: *(i)* if process mining has been shown to have been successfully applied in the domain of *Enterprise Risk Management* in general, could it also be leveraged in the *cyber insurance* context? *(ii)* Could process mining aid with addressing some of the fundamental challenge of information asymmetries? *(iii)* Why is it that the connection has not yet been made in the research on cyber risk

management in the cyber insurance context? *(iv)* What process mining methods could be relevant for cyber insurance underwriting and risk assessment? *(v)* What would be the challenges and limitations of such approach? and finally, *(vi)* as cyber insurance and insurers more generally explore extending their offerings to risk mitigation in addition to risk transfer [12], could process mining at provide for a continuous risk mitigation mechanism and thus increase the overall value proposition and adoption of cyber insurance products? As discussed in the 3 work section, none of these questions seem to have been answered, nor systematically investigated in detail in the literature and this thesis aims to fill this research gap with the MeritMiner4CI approach.

1.2 Description of Work

This section explains in detail what approach will be taken to answer the research need. The goal of this master thesis is to explore the applicability of process mining methods in the cyber insurance and to develop a novel approach to apply these methods in cyber insurance. The approach is then implemented as a prototype solution on the top of the SecRiskAI solution [17] and evaluated, including interviews with a number of domain-experts. The steps that were taken in order to achieve the aforementioned goals as described as follow.

First, the relevance of process mining for cyber insurance was established by conducting a detailed on survey on *(a)* cyber risk and cyber insurance (an more specifically the cyber insurance underwriting and assessment processes) and *(b)* security process management and process mining methods. The key issues of cyber insurance needed to be thoroughly analysed and the potential of specific process mining methods to address these issues understood. Elements of the Cyber Insurance Framework proposed in a previous work [12] were used to support the analysis.

Once the relevance fundamentals were well-understood, literature review on different process mining applications in domains adjacent to cyber insurance (*e.g.,* GRC and general security) was conducted and use-cases at different levels of process abstraction that are valuable for overall cyber risk assessment in cyber insurance were identified. The literature review also confirms the research gap this thesis is addressing, *i.e.,* no available work on process mining in cyber insurance. After the relevant use cases were identified, it was necessary to define a structured approach covering all the steps, actors, analytical methods and metrics needed to successfully assess the risk related to internal processes of a prospective insured with process mining. As a result, the aforementioned MeritMiner4CI approach was designed.

In the next step, a high-level architecture was drafted to determine how the approach can be implemented in a modular way. Considering that reference architecture, the prototype itself was designed and developed in structured way, starting from design and ending with a functional manifestation of the MeritMiner4CI approach in the form of prototype integrated with other contributions to cyber risk assessment developed at the CSG. This prototype demonstrates the feasibility of the underlying approach and implements relevant scenarios. Finally a thorough evaluation was conducted based on a survey, interviews with

experts, designed scenarios, and quantitative evaluation of the implemented algorithms with relevant security process data-sets.

1.3 Thesis Outline

The thesis is organised as follows.

Chapter 2 provides the reader with an introduction to cyber risk and cyber insurance (incl. it's challenges, processes and applied risk assessment methods). Process mining is introduced as well in the context of business process management. The background chapter also establishes the bridge between the two fields.

Chapter 3 is where process mining is presented in different research streams including GRC, security and process performance analysis that are highly relevant for identification of use-cases.

Chapter 4 presents the proposed novel MeritMiner4CI approach and outlines in a structured way how process mining methods can be applied in the cyber insurance underwriting process.

Chapter 5 covers the design and implementation of the prototype (in three layers) aspects of the steps taken can be used to implement the MeritMiner4CI and to prove the feasibility of the approach. Integration with SecRiskAI is also discussed.

Chapter 6 validates the approach from different perspectives. Most importantly, it presents and discusses the results of a survey with experts that were involved in order to validate the concept with designed case study scenarios. It also showcases how the approach can be applied to case studies modelled on real-life scenarios and finally offers a quantitative evaluation of the applied process mining methods from the perspective on established quality criteria.

Chapter 7 discusses the results of the thesis and the implications they might have. Limitations are also clearly outlined.

Chapter 8 is concerned with the summary and conclusion of the thesis and also outlines overall limitations and future work.

Chapter 2

Background

Cyber Insurance is extraordinarily complex, given its interdisciplinary nature involving fields from cyber security, actuarial science, to law and economics [18]. Complex to such an extent that after more than 20 years since the first cyber insurance product was introduced in 1997, some continue to challenge whether a cyber risk is insurable by private companies (even with reinsurance backing) [19]. In an interview with the Financial Times, Mario Greco, the chief executive officer of the Zurich Insurance Group(ZIG), hinted at the possible need to involve state governments: "A connected economy offers lots of opportunities for cyber attacks. A major cyber risk is something only governments can manage" [19]. An alarming take on cyber insurance has also been offered by the risk expert and author Nassim Nicholas Taleb. Replying to a question: "Can you predict cyber terror?", "No", he answered. "That's an Extremistan problem. Individual crime you can profile, but big cyber is too complex." [20]

The following chapter provides a background to understand both the cyber insurance domain and process mining methods. Specific applications of process mining in the more general contexts of cyber risk assessment and auditing are reviewed in the following related work chapter.

The chapter is structured in the following way. First, a definition of cyber risk used thorough the thesis is established and the categories of operational cyber risk highlighted that relate to "actions of people" and "failed internal processes" that, according to [13] contribute the majority of losses. Then, I provide a overview of the currently available estimates of the costs of such risk globally, clearly highlighting the unprecedented magnitude of the problem. From there I will move to provide a brief overview of cyber insurance, contextualising it as one of the four main possible reactions to risk - as a risk transfer mechanism. The focus will be on two key points 1) fundamental challenges and conditions that need to be fulfilled for cyber risk to be insurable and 2) an overview of currently applied cyber-risk assessment methods. The section on cyber insurance is concluded with a simplified formalisation in the form of an economic model, proposed by [21] that clearly delimits the challenge that my thesis is addressing, which is the following: cyber insurance premium can be tied to the amount of self-protection (in order to deal with the related problem of information asymmetries) and that the insurer must audit the self-protection

practices and the level of care that the agent takes to prevent the loss. I also contextualize the thesis at hand among other research conducted at the *CSG* at the *University of Zurich*. Then, centrally, process mining is introduced (in the context of business process management and processes in security) as one of the viable ways to conduct these audits of the "level of self-protection". A method that, as this chapter argues, is especially fit to increase the understanding of operational cyber risk relating to the aforementioned categories of "actions of people" and "failed internal processes", which are clearly the most significant categories of operational cyber risks [13]

As the 2019 research agenda provided in [22] clearly shows, there is a large number of aspects of cyber risk and cyber insurance that are not fully understood and still need to be addressed. These include, to name a few, the measurement and monitoring of risk, opportunities to reduce it, and mechanisms that enable its transfer to third parties. For this reason, please note that this chapter does not have the ambition to provide a holistic overview of cyber insurance, instead only selected aspects relevant for the thesis (especially to risk assessment and risk related to internal processes) are presented and the reader is pointed to other, holistic works on the topic, such as [23],[24], [21], [18] (and many others), as well as to the master project conducted last year at the CSG [12].

2.1 Cyber Risk

2.1.1 Definition and Categorisation of Operational Cyber Risk

Let us start by providing a definition of cyber risk applied in this thesis. The term "cyber" (i.e. "relating to, or involving computers or computer networks" [25]) refers to risk related to electronic events that lead to business disruption, or monetary loss [26]. In other cases "cyber risk" is used as synonym to information security risk [13]. Other researchers, such as Boehme and Kataria, for example, understand cyber risk as risk resulting in failure of information systems [27]

This thesis will follow the definition of "cyber risk" as proposed by [13], who strongly lean back on the definitions applied by regulators of insurance and financial markets (and reflected in frameworks such as Basel II and Solvency II, see [28], [29], [30]), who understand cyber risk as a type of operational risk. In other words, cyber risk refers to "operational risks to information and technology assets that have consequences affecting the confidentiality, availability or integrity of information, or information systems" [31].

Cebula and Young propose the following categories of cyber risk [31] which is a view shared both in [13] and [18] cyber risk into four classes: (1) *actions of people* (which includes, for example, losses of data by employees), (2) *systems and technology failures* (e.g. hardware malfunction), (3) *failed internal processes* (flawed definitions of responsibilities and (4) external events (e.g. natural catastrophes impacting data centers). An overview of this categorization is presented in Figure below. It is key to divide the cyber risk into specific categories in order to investigate the most appropriate risk analysis methods. In this thesis, the categories of failed internal processes and actions of people will be in focus.

1. Actions of People	2. Systems and Technology Failures	3. Failed Internal Processes	4. External Events
1.1 Inadvertent	**2.1 Hardware**	**3.1 Process design or execution**	**4.1 Disasters**
1.1.1 Mistakes	2.1.1 Capacity	3.1.1 Process flow	4.1.1 Weather event
1.1.2 Errors	2.1.2 Performance	3.1.2 Process documentation	4.1.2 Fire
1.1.3 Omissions	2.1.3 Maintenance	3.1.3 Roles and responsibilities	4.1.3 Flood
1.2 Deliberate	2.1.4 Obsolescence	3.1.4 Notifications and alerts	4.1.4 Earthquake
1.2.1 Fraud	**2.2 Software**	3.1.5 Information flow	4.1.5 Unrest
1.2.2 Sabotage	2.2.1 Compatibility	3.1.6 Escalation of issues	4.1.6 Pandemic
1.2.3 Theft	2.2.2 Configuration management	3.1.7 Service level agreements	**4.2 Legal issues**
1.2.4 Vandalism	2.2.3 Change control	3.1.8 Task hand-off	4.2.1 Regulatory compliance
1.3 Inaction	2.2.4 Security settings	**3.2 Process controls**	4.2.2 Legislation
1.3.1 Skills	2.2.5 Coding practices	3.2.1 Status monitoring	4.2.3 Litigation
1.3.2 Knowledge	2.2.6 Testing	3.2.2 Metrics	**4.3 Business issues**
1.3.3 Guidance	**2.3 Systems**	3.2.3 Periodic review	4.3.1 Supplier failure
1.3.4 Availability	2.3.1 Design	3.2.4 Process ownership	4.3.2 Market conditions
	2.3.2 Specifications	**3.3 Supporting processes**	4.3.3 Economic conditions
	2.3.3 Integration	3.3.1 Staffing	**4.4 Service dependencies**
	2.3.4 Complexity	3.3.2 Funding	4.4.1 Utilities
		3.3.3 Training and development	4.4.2 Emergency services
		3.3.4 Procurement	4.4.3 Fuel
			4.4.4 Transportation

Figure 2.1: Risk categories and subcategories according to [31]

2.1.2 Cost of Cyber Risk

Before investigating the cyber insurance market and its challenges, let's establish the magnitude of the problem of cyber risk that we previously defined and categorized. It is often considered as one of the key societal challenges of our time by many organisations, such as (OECD [32], Fed [33], EIOPA [34])

In 2018, acknowledging the wide range of the estimate, which is strongly dependent on input parameters and assumptions, the RAND Corporation [35] put the global cost associated with Cyber Risk somewhere between $275 billion to $6.6 trillion globally and total GDP costs (direct plus systemic) of $799 billion to $22.5 trillion (corresponding to 1.1 to 32.4 percent of GDP).

In 2020, a widely-cited statistic by McAffee [36] estimates the "global losses from cyber crime" to be around $1 trillion, a figure based on interviewing 1,500 IT and line of business decision makers. Compared to research using the same methodology from 2018, this would correspond to more than a 50 percent increase. Accenture [37], focused on the metric of value at risk (from both direct and indirect attacks) and arrives at the figure of US $5.2 trillion over the next five years (in 2019 research). An average G2000 [38] is expected to risk value of $580 million annually, or 2.8% of revenues, accoriding to Accenture.

Coming back to the categorisation of cyber risk in Figure 2.1. It is also interesting to observe to which categories of risk the most losses are associated. Based on an analysis of a sample of data, [22] points out that most of the losses coming from cyber risk relate to actions of people (incl. employees) and then from failed internal processes. It is important to point out this finding as internal processes are the key focus of this thesis.

Spending on cyber security solutions (e.g. for data security, network security, access

Category	N	Mean	Std. dev.	Min.	Quantiles			VaR	TVaR	Max.
					25%	50%	75%	(95%)	(95%)	
Panel A: Cyber versus non-cyber risk										
Cyber Risk	994	40.53	443.88	0.10	0.56	1.87	7.72	89.56	676.88	13,313
Non-Cyber Risk	21,081	99.65	1,160.17	0.10	1.88	6.20	25.37	248.97	1,595.27	89,143
Panel B: Cyber risk subcategories										
Actions of people	903	40.69	463.25	0.10	0.55	1.83	6.87	84.36	679.04	13,313
Systems and technical failure	37	29.07	77.33	0.10	1.10	5.03	11.65	168.95	329.04	370
Failed internal processes	41	47.72	205.92	0.14	0.42	2.04	9.05	158.65	743.40	1,311
External events	13	39.40	115.73	0.28	0.56	1.03	13.77	192.88	422.71	422

Figure 2.2: Losses by risk category based on an analysis of a sample (in million US$) by [22]

management) has been estimated at around 40.8 billion U.S. dollars [39]. A somewhat higher was estimated by [40], who expected the spending on security and risk management to exceed $150 Billion in 2021. Ransomware attacks drive a part of these increases in cyber security spending,, damage costs caused by ransomware attack were most recently estimated [41] and [42] to reach $20 billion by 2021. The figure is 57x higher compared to 2015. Looking at insider threats (including malicious data exfiltration and accidental data), [43] estimates that the increase in losses from such threats between 2018 and 2020 was 47%. Many of these developments are considered to be strongly related to the rise in remote work and other changes to how businesses operate as a result of the COVID-19 pandemic [44].

The report by [41] offers a concerning summary, applying a unique comparison, the rapidly growing global cybercrime costs, accoring to the report, can be seen seen as the largest transfer of economic wealth in history. Costs are expected, in monetary terms, to be larger than the damage inflicted from natural catastrophes and cyber crime will generate more profits than the market for illegal drugs [41]. This clearly outlines the size of the problem at hand. In the next section, I will provide a brief introduction to how cyber insurance aims to contribute to solving that problem.

2.2 Cyber Insurance

The following section aims to provide the reader to provide the reader with an overview of cyber insurance. This was provided, for example, in the Cyber Insurance Framework [12] which covers a number of different aspects of Cyber Insurance in depth. The section on Cyber Insurance in the background chapter of the thesis will focus on selected aspects that highlight including the fundamental challenges of insurability of cyber risk, the typical cyber risk assessment methods and their shortcomings and finally, a formalisation of the cyber insurance model for self-protection in the form of an economic model is to be presented.

2.2.1 Cyber Insurance Market

In a Chicago Fed Letter, [33] Granato and Polacek briefly summarise the history of cyber insurance, which is typically traced back to 1997. In that year, AIG (American International Group) introduced the first internet security liability policy. During the early days of cyber insurance, the policies catered more to IT providers rather than to companies consuming such services [33].

Looking specifically at the size of the cyber insurance market, KPMG [45] predicted the volume of annual premiums to rise from US $2.5bn in 2015 to US $7.5bn by 2020 and eventually reaching US $20 bn in premiums by 2025, which is an estimate shared by MunichRe [46]. The KPMG [45] research also posits that the size of the market for all types of cyber offerings (including both transfer and mitigation) was around $100bn (the topic of the importance of mitigation in cyber insurance products will be discussed in detail in a further section). Finally, for comparison, in a more recent publication by the insurer AmTrust [47] projects the continuation of the growth of the market from $7.8 billion (in 2020) to $20.4 billion by 2025, translating to an expected annualised growth rate of 21.2%. To provide some context for these numbers, the cyber insurance market still constitutes less than 1% of the overall insurance market, but is growing at a rate that is multiple that of the 4-5% growth rate of P&C (property and casulalty) [48].

Marsh highlights in that the cyber insurance market is also growing largely due to premium increases, even as clients try to mitigate these by increasing their retention (deductibles) [49]. Cyber insurance pricing, according to [49] increased an annual rate of 96%, year-over-year (specifically 130% in the US and 92% in the UK), with a whopping 40% increase attributed to the third quarter of 2021. The expectation of both the industry and academia that this would happen has been documented in the preceding master project [12]. To make matters even more complicated, Financial Times reported at the beginning of 2022 [19] that some insurance companies stopped underwriting cyber insurance altogether in 2021, further lowering the available capacity and leading to an increase in prices. As mentioned in the preceding section on cyber risk, the change in the dynamics of the industry has been often explained by the recent increase if remote work, resulting in an explosion of threats from ransomware attacks, phishing etc. [50] and [44]. Evidenced by these massive increases and retreats from the market, it is clear that the assessments of cyber insurance companies systematically underestimated the involved risks and that there is imminent research need to develop new methods and models, should cyber insurance remain sustainable. This has also been a focus topic of a multitude of international organisations, such as the OECD [32], EIOPA [34], or the US Federal Reserve [33].

2.2.2 Overview of Cyber Insurance

A formalisation (see Figure 2.3 of the cyber insurance market can be found in the foundational paper by [23]. Boehme at al. use five components to characterise the market: the networked environment, demand side, supply side (as depicted in 2.3, information structure, and organizational environment. The fundamental dynamics of interdependent security and correlated risk and correlated risk are also integrated into the framework.

This model has served as a basis for a number of works that followed, such as [18] and is the most cited and influential paper in the cyber insurance space identified.

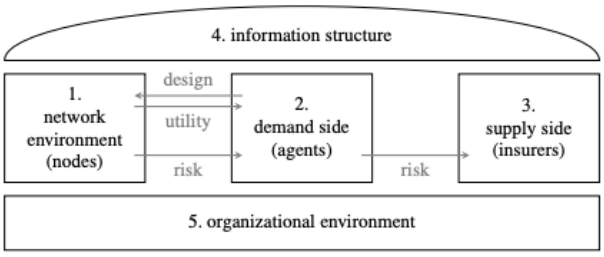

Figure 2.3: Unified Cyber Insurance Framework [23]

Another overview of cyber insurance is offered by the framework constructed in a preceding master project [12] that consists of the three pillars (market model, premium pillar and environment) depicted in 2.4. It is impossible to cover all of the elements in the scope of the thesis, so the following paragraph points out some of the fundamentals.

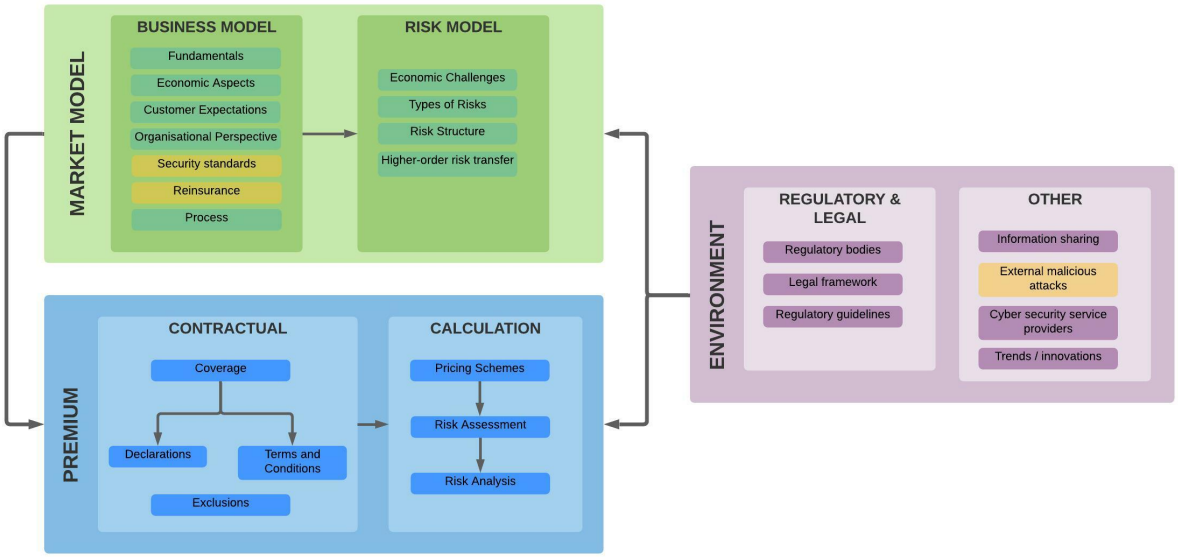

Figure 2.4: Cyber Insurance Framework [12])

The **market model pillar** is divided up to the business model and risk model. It also covers the fundamental challenges (incl. insurability criteria) and economic aspects. The ones relevant for this thesis are presented in the Subsection 2.2.3 on fundamental challenges. Organisational perspective that includes the actors in the market is covered there as well. A summary is presented in the form of an ER-Diagram in Figure 2.5. Last aspect to point out is security standards applied in cyber insurance, which are presented in the context of risk assessment later in this chapter.

Moving on to the **premium pillar**, it can be summarised that cyber-insurance contract typically includes coverage, declarations, exclusions, and terms & conditions. Coverage can be categorised as follows [51]:

1. First party coverage (examples of which include Data Compromise Response, Identity Recovery, Computer Attack, or Cyber Extortion)
2. Third party coverage (examples of which include Data Compromise, Network Security, or Electronic Media)

Typical exclusions from coverage often include: Fines, Injury, War, and Terrorism [51], but might also be related to insufficient security practices, or negligence. However, it is critical to point out that cyber insurance contracts are highly heterogeneous and only limited generalisations can be made. This is further made more complicated by the fact that insurers heavily leverage exclusions [51]. Another element of the premium pillar is the premium calculation, which also encompasses risk assessment and will be presented in that context later in this chapter.

Finally, the **environment** pillar takes into consideration the regulatory and legal aspects as well as other external influences such as interactions with cyber security service providers and innovation. Again, the aspect of regulation is discussed in the context of risk assessment later in this chapter.

2.2.3 Fundamental Challenges of Cyber Insurance

The insurability of cyber risk has been the subject of a number of publications [18]. While there are many challenges, which are outlined in the Table 2.6 below constructed by [13] based on [52], the overall consensus of the research is that cyber risk is generally insurable, but cyber insurers need to work on addressing the issues systematically.

[13] highlight the following main difficulties - 1) randomness of loss occurrence, 2) information asymmetries, and 3) cover limits. Out of these three, this thesis aims to address, most importantly the challenge of information asymmetries by the means of process mining. This is in line with [13], who poses that there is a "need for discovering approaches that can reduce the substantial information asymmetry present with cyber risk". An additional and somewhat overlapping view is offered by [24].There, Gordon provides a framework for integrating cyber insurance to the cyber risk strategy of risk managers and points out the following three key challenges: pricing, adverse selection and moral hazard. Adverse selection refers to the notion that companies with public knowledge of their actual cyber security posture might be more likely to procure cyber insurance if they have a negative private assessment of said posture. Moral hazard might come into play when a cyber insurance coverage may cause the insured to reduce their investment in self-protection (formalisation will be provided in a later section of this chapter). Finally, the pricing issue is predicated on the fact that available actuarial information is still limited and hard to apply [13].

A systematic review of challenges of the cyber insurance market is provided in [53]. This thesis will discuss risk assessment in detail detail and an overview will be provided of the other aspects. The following list will rely on the structure provided by [53] to point out the ones with relevance for the later parts of the thesis. The selected challenges with relevance for this thesis can be grouped to the following high-level categories:

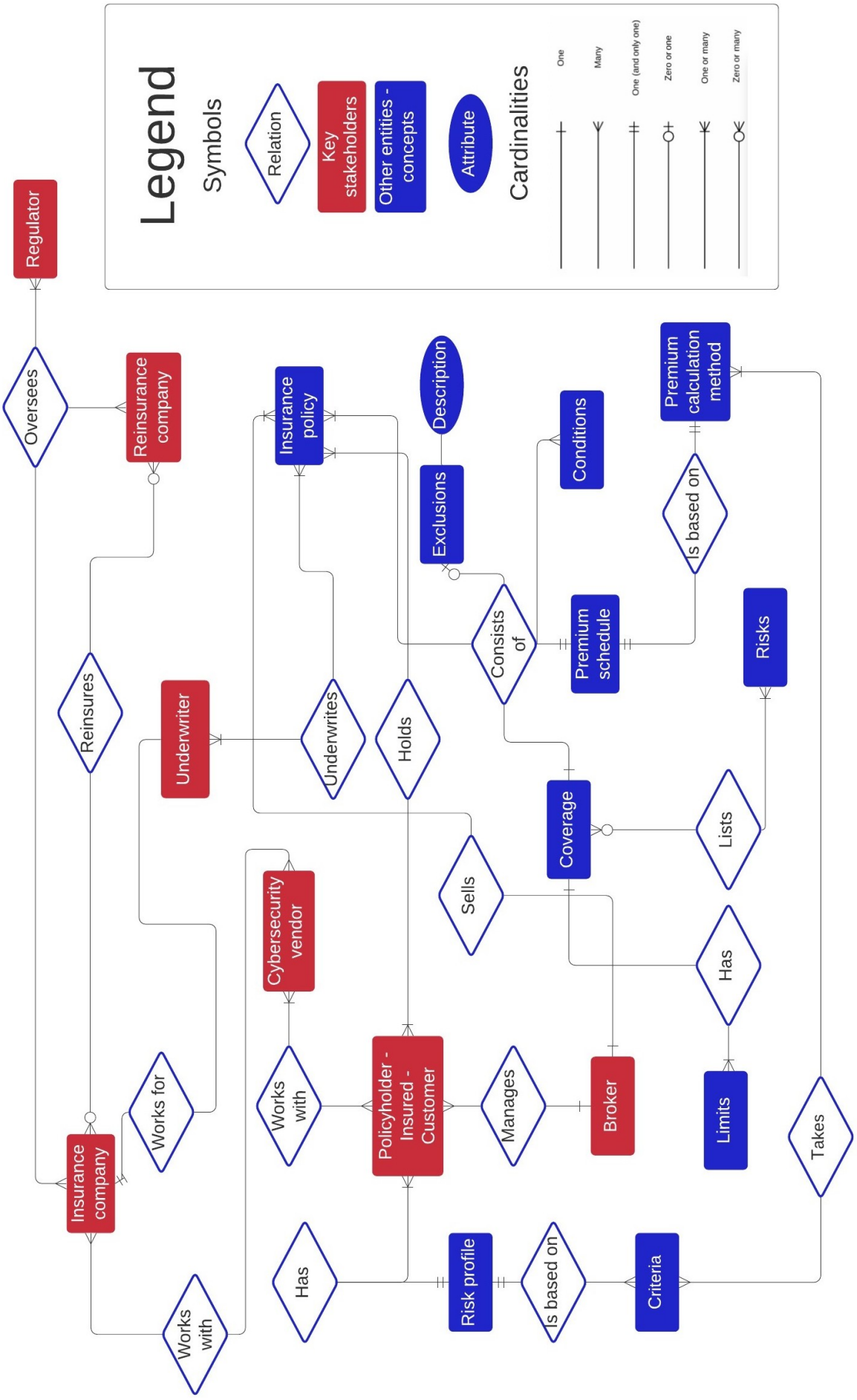

Figure 2.5: Simplified ER-diagram of actors in the CI market [12]

Insurability Criteria			Requirements
Actuarial	(1)	Randomness of loss occurrence	Independence and predictability of loss exposures
	(2)	Maximum possible loss	Manageable
	(3)	Average loss per event	Moderate
	(4)	Loss exposure	Loss exposure must be large
	(5)	Information asymmetry	Moral hazard and adverse selection not excessive
Market	(6)	Insurance premium	Cost recovery and affordable
	(7)	Cover limits	Acceptable
Societal	(8)	Public policy	Consistent with societal value
	(9)	Legal restrictions	Allow the coverage

Figure 2.6: Fundamental insurability criteria for cyber insurance by [52]

1. **Challenges relating to organization eligibility for a cyber insurance coverage**

(a) **Risk assessment method**

A proposal of a novel risk assessment method in the context of cyber insurance is the central focus of this thesis. It aims to address the challenge that in order to decide on the eligibility of an organisation for cyber insurance coverage, one has to select an appropriate method for assessing the risk of the organization [53]. Different trade-offs are connected to each of the available methods. Some methods are simple and inexpensive, such as automated evaluation of questionnaire responses. Other are more thorough but more subjective and hindered by information asymmetries (underwriting meetings). It is also not given that the best risk assessment method would always be most detailed one. Cost aspects need to be considered as well and extensive security audits can not be applied in each case selection of the right risk assessment method therefore needs to be well thought-out for different scenarios and company sizes [53]. An overview of the cyber risk assessment methods available by [18] can be found in Figure 2.7 below.

Phases	Steps	Techniques
Risk Identification	Asset Identification	business documentation meetings/interviews questionnaires/checklists/worksheets knowledge base
	Threat identification	business documentation meetings/interviews questionnaires/checklists/worksheets knowledge base threat trees/FTA/attack trees
	Security/ Vulnerability identification	ETA attack graphs vulnerability scanning penetration testing meetings/interviews questionnaires/checklists/worksheets knowledge base Delphi method
Risk Analysis	Likelihood determination	history/log analysis meetings/interviews questionnaires/checklists/worksheets knowledge base Delphi method
	Impact Determination	meetings/interviews questionnaires/checklists/worksheets knowledge base Delphi method
	Risk Estimation	risk table ALE
Contract Specification	Coverage Specification	selection by agent meetings/interviews game theory
	Premium Estimation	game theory profiling
	Write & Sign Policy	paper work (digital) signature
	Claim Handling	paper work

Figure 2.7: Typical cyber risk assessment methods by [18]

These methods are then often applied evaluate the cyber risk posture of a given organisation against different frameworks that outline sets of practices to follow. [18] (this is either explicit. i.e. prospective insureds are asked to provide confirmation of compliance, or implicit, meaning that criteria in underwriting manuals can be traced back to one or more of the available frameworks) Some of the most-widely applied ones across industries include NIST [2], CIS [3], or ISO 27001 [54]. Because certain cyber insurance policies also cover regulatory fines resulting from cyber incidents, cyber insurers often also focus [51] on checking compliance with different regulatory and compliance guidelines that are industry-specific such as HIPAA in healthcare [55], PCI in the case of payment processors [56]. Other guidelines might be relevant for specific regions only, such as the Sarbanes-Oxley Act (SOX) [57] outlining requirements for financial reporting, California Consumer Privacy Act (CCPA) [58], or GDPR in Europe [5].

From the perspective of the master project [12], it can be pointed out that for small and medium sized enterprises, insurers report tend to use questionnaire data (typically containing no more than 10 questions) and focus on fundamental metrics such as company size, revenue, claims history etc. These are then used as input to models that output the premium (presented in the premium calculation section below). A sample of a cyber insurance application

questionnaire is depicted in Figure 2.8.

#DigitalRisk

U.S. Risk, LLC

Cyber Insurance – Application Form - Short			
Applicant Name:			
Applicant Address:			
Subsidiaries:			
State of Domicile:		Website Address:	
Year Established:		Number of Employees:	
Industry Sector:			
Nature of Business:			

Financial Information	Last Complete Financial Year	Current Year (Estimate)	Next Year (Estimate)
Gross Annual Revenue			
Annual Net Income before Taxes			
Percentage of Gross Annual Revenue - Payment Card			
Percentage of Gross Annual Revenue – Online			

General Information		Yes	No
1	Are all servers, firewalls, etc. located in a purpose-built server room with access restricted to appropriate personnel?		
2	Are backups taken at least weekly and stored in a secure off-site location?		
3	Do you have an email and internet usage policy that has been shared with all employees?		
4	Do you have firewall architecture in place?		
5	Do all systems users have individual, mandatory and non-trivial user IDs and passwords with forced periodic password changes?		
6	Are all PCs and servers protected with up-to-date anti-virus that is updated regularly?		

Data		Number
7	What is the total number of Personal Identifiable Information records stored on your networks?	
8	What is the total number of Social Security Numbers stored on your networks?	
9	What is the total number of Personal Health Information records stored on your networks?	
10	How many payment card transactions do you process annually?	
11	What is the total number of Payment Card records stored on your networks?	

Name	Signature	Position	Date

U.S. Risk – USA Cyber

REV 12.15.17

Figure 2.8: Questionnaire sample([59]

For large enterprises, where the potential risk is multiple orders of magnitudes larger, underwriters typically organise underwriting meetings [60], conduct interviews (focusing on cyber insurance practices as well) and even on-site visits. In any case underwriting discretion is employed based on past experience in a given region [12]. This further contributes to large differences in premium levels offered by different insurers.

A deeper systematic investigation of the actual cyber insurance underwriting process can be found in [51]. Romanovsky et. al. analyse the contents of a number of actual cyber insurance policies from regulatory filings in the NAIC SERFF database [61] (that include the text of the policy, application questionnaires, and rate schedules) with the goal to map what information cyber insurers request from their customers and what methods are used to assess risk and price policies in the context of the underwriting process. One of the key outtakes of the paper is that there certain patterns can be identified, such as the usage of either flat rate pricing, or of base rates with modifications for calculations. It is also pointed out that pricing considering information security posture is in its infancy and not yet widely applied. [51] makes it clear that

risk assessment and premium determination are surrounded by a high level of uncertainty and can not be considered mature.

However, some efforts to innovate the cyber risk assessment process in the cyber insurance context can be identified. For example, Boehme et al. [62] provides a list of alternative data (such as including log data, data from honeypots, proprietary threat intelligence data) that might be leveraged in cyber risk assessment. No details on actual usage or adoption is provided though. Another example of alternative approaches towards cyber risk assessment in insurance is "CyberMatics" [63]. The insurance company AIG works together with with security service and software providers in order to offer enhanced risk assessment for prospective clients. Unfortunately, there are not many details available on the inner workings of the process. According to the available factsheet [63], cyber security providers such as CrowdStrike„ or Darktrace conduct different analyses based on data from the prospective insured and then deliver security recommendations. The prospects do not need to disclose detailed information to the cyber insurance company, rather only data relevant for cyber insurance application is provided by the partner security provider. AIG in turn updates the cyber maturity profile of the company and calculates risk scores. The trend of offering risk mitigation integrated with a cyber insurance offering can be clearly identified in such setup. Finally, a number of vendors focusing on providing cyber risk assessments as a service are available on the market, including BitSight [64], RMS [65] and Advisen [66]. The master project, however, could not identify that such assessments would be widely relied on [12]. No commercial solution focusing on process mining methods with application in cyber insurance was identified.

(b) **Limited historical data** Cyber risks evolve constantly and compared to other types of insurance (such as life or health) only limited data that can be used for loss models is available. Even if data becomes available, the nature of relevant risks might change dramatically in short periods of time. One example for this is the enormous rise in ransomware attacks in recent years. So-called zero-day attacks make it even more difficult to conduct risk assessments on past experiences and data. Furthermore, the incentives to share data, or even to report are weak and in some cases negative. However, there have been notable developments in this area, such as initiatives by the industry sharing at least some of their available data (such as the Cyber Index by Chubb [67]. Different regulations have also been introduced in certain jurisdictions mandating breach reporting (for example [68]).

2. **Challenges relating to insurance contract design**

(a) **Standardization**

Contracts are highly heterogeneous in regards to coverage, limits and exclusions. Both regulators and industry group call for a higher degree of standardisation (see e.g. [32]). However, it can be argued that there are also other factors at play; more standardisation might lead to reduced opportunities for brokers, who play a vital role in the market (see Figure 2.5).

(b) **Terms and conditions** For insured parties, legal aspects of the policy such as terms, conditions and notably exclusions are of utmost importance (see Figure 2.4). While insurers take advantage of these contractual mechanisms to exclude certain losses (such as those caused by nation-state attacks, see e.g. the NotPetya case [69]), they must also design policies in such a way that they are still understandable and commercially attractive to the demand side. The same is true in regards to conditioning coverage based on some information-security related criteria. I.e. while some standards are certainly desirable, the applicable standards must also be general enough to be achievable by the insured.

(c) **Modeling cyber losses** Creating risk models to predict losses for a portfolio is complex and error-prone due to the evolving nature and number of different types of possible claims. Modelling is necessary for setting base rates, limits and exclusions. Please note that actuarial modelling as such is not in focus of this thesis, which focuses on risk assessment at the level of an individual organisation. However, there might be potential to leverage aggregate process data from a portfolio of insureds for the purposes of actuarial modelling. Other researchers, such as [70], [71] and [62] have investigated different modeling approaches to cyber insurance at great depth.

(d) **Coverage** Is related to the difficulty of creating loss models and challenges related to risk assessment. On one hand, insurability criteria [13] have to be met for each coverage component, on the other hand, the overall product still needs to be attractive to be insured.

(e) **Determine premium rates**
Determining premiums has been one of the problems that we mapped in depth in the Cyber Insurance Framework [12], for a summary what a typical premium determination process might look like, please refer to the Table 2.1 below.

Steps in a typical premium calculation
1. Input customer annual revenue
2. Determine overall risk group (0-6)
3. Select applicable coverage
4. Select applicable base rate
5. Select applicable retention
6. Select applicable limits and sub-limits
7. (Optional) Determine business interruption deductible hours
8. Adjust relevant limit modifiers
9. Coverage specific confidence factors
10. Enterprise specific confidence factors
11. Annual premium (business interruption)

Table 2.1: Premium Calculation process, Example of Cyber-Related Business Interruption mapped in [12]

It also has to be noted that the performance of a given insurance company is typically measured by the following ratios, which need to be kept as low as possible for the product to be profitable [72].

1) Loss Ratio (LR) [72]
$LR = \frac{Losses\ an\ insurer\ incurs\ due\ to\ paid\ claims}{Premium\ earned}$
2) Benefits-Expense Ratio (BER) [72]
$BER = \frac{Expenses\ for\ acquiring,\ underwriting,\ and\ servicing\ a\ policy\ over\ the\ net\ charged\ premium}{Premium\ earned}$
3) Combined ratio (CR) [72]
$CR = \frac{Paid\ claims\ +\ Expenses}{Premium\ earned}$
As [73] loss ratios in cyber insurance have been at record highs in 2020, further underlining the need for sound decisions on premiums and enhanced risk assessments.

3. **Challenges once a cyber insurance contract is in place**

   (a) **Insured self-reporting** Self-reporting opens the way to moral hazard and the insured has an incentive to potentially downplay risks in order to get eligibility for coverage and/or lower premiums [18, 53].

   (b) **External security audit** Related to the previous points, regular security audits are required in order to mitigate the risk of moral hazard and adverse selection predicated by self-reporting [18, 53].

   (c) **Lower investments in security** The final challenge will be the focus of the following section containing the formalisation of the problem. Cyber insurance policies need to designed in such a way that they do not negatively impact the incentive for investment in self-protection [21, 53]

2.2.4 Economic Model for Cyber Insurance by Lelarge & Bolot and the Balance Between Self-Protection and Risk Transfer

As previously outlined, there are four fundamental ways to react o to risk as posed by [21]: 1) *avoid the risk*, 2) *retain the risk*, 3) *self-protect and mitigate the risk*, and 4) *transfer the risk*. Cyber Insurance can be clearly categorised as, first of all, a risk transfer mechanism [74] and while this thesis focuses, in line with that, on observing cyber insurance as such, it will, in line with [21], consider the mechanisms by which is can drive the mitigation of risk as well with different incentives Research so far [21], indicates that certain conditions must be in place for the risk transfer mechanism to work in a reliable way that accounts for incentive structures (rules must be in place as well as mechanisms for premium differentiation). Specifically:

1. Premium is **tied to the amount of self-protection**
2. Insurer must **audit the practices** that the insured employs and the level of self-protection must be assessed

Please note the difference between self-insurance (which reduces loss size l, e.g. over-provisioning, DDoS mitigation, PR firms) and self-protection (which reduces loss probability p, e.g. intrusion detection systems, insider threat detection) in the cyber security context [21]. Premium differentiation (and possibly discrimination as will be discussed

later) can mitigate these misalignments of incentives formalised using utility models in [21] . In [75] as cited in [21].

Premium differentiation considering self-protection can be formalised as follows [21] .

$$\varphi[S] = p^{-}l$$

$\varphi[S]$ in this case refers to the (lower) fair premium offered to the **self-protected** agent.

$$\varphi[N] = p^{+}l$$

While $\varphi[N]$ is offered to agents **without self-protection.**

[21] show that premium differentiation might not be enough in many cases and premium discrimination might need to be employed. I.e., contracts might need to be designed in such a way that further premium rebates/loading are offered to different categories of agents seeking protection. Formally, this can be expressed in the following way.

For agents with insufficient self-protection, the offered premium is denoted as $\varphi[N] + \gamma$. For agents with self-protection: $\varphi[S] - \gamma$, , where the addition of $\gamma \geq 0$ denotes a premium loading and subtracting of $\gamma > 0$ refers to premium discount [21] .

While these considerations are rather intuitive, I believe them to be important for the thesis as they formalise the fundamental idea behind advanced risk assessment methods for cyber insurance. Other dynamics that add to the complexity of economic models for CI also been considered in such models (such as interdependent security, or mandatory insurance), but these will not be covered in this chapter. Finally, I want to point out that the literature on the economic models such as the paper by [21] mentioned above as well as [76], call for further development of metrics and techniques for quantifying the level of protection of assessed organisations and for enhancements of techniques aimed at estimating and quantifying potential losses. While my thesis focuses predominantly on the first research need, [4] are concerned with the economic considerations around cyber security and develop guidance to estimate the economic impacts of (lacking) cyber security. It is highlighted there that different categories of cyber security costs should be considered including the Threat Exposure Cost (TEC), Proactive Mitigation Cost (PMC) and Reactive Mitigation Cost (RMC). Failing to consider these costs might lead to systematic under-investment ii self-protection

2.2.5 Other Research on Cyber Risk Assessment and Cyber Insurance at the CSG

Other researchers at the Communication Systems Group (CSG) have also focused on cyber risk assessment (see the *Premium Pillar* in the framework in [12]. For example, a closely related master thesis [17] focused on exploring the applicability of different machine learning methods in the context of cyber risk assessment and proposed the SecRiskAI solution. The developed SecRiskAI solution provides the ability to assess the risk of being targeted by external cyber attacks. The thesis at hand - MeritMiner4CI focuses,

in a complementary way, on the perspective of internal processes that are relevant for risk assessments. While profiles of the companies at risk is taken into consideration in [17] (e.g. based on industry, revenue, number of employees), [17] points out that future work should focus more on investigating more specific risk factors - which is a step in the approach proposed in my thesis.

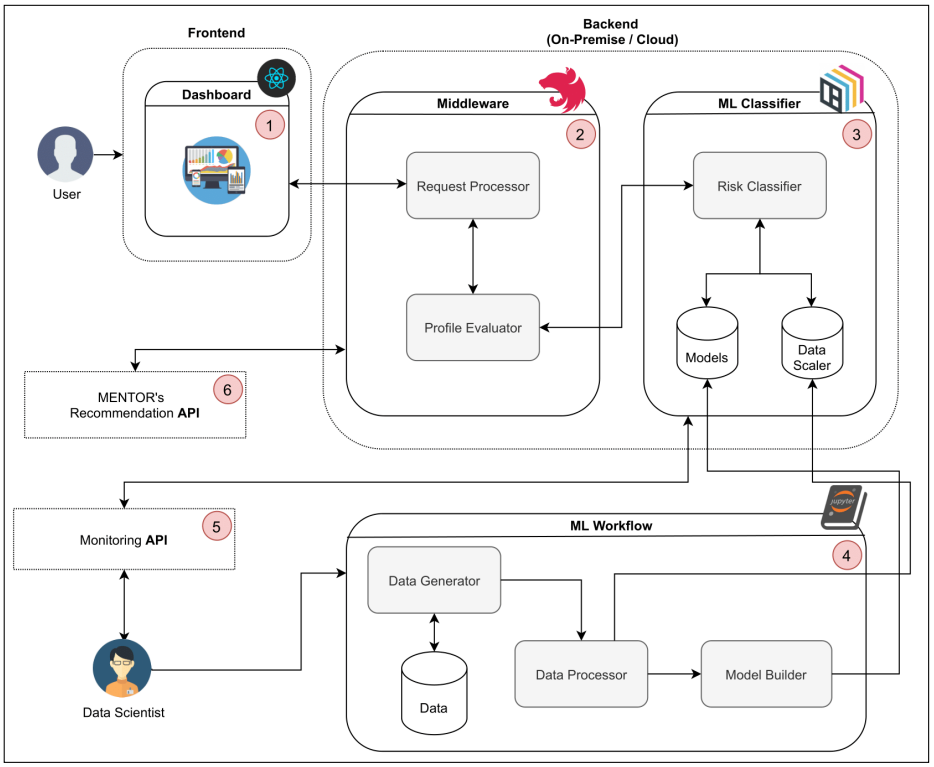

Figure 2.9: Overview of SecRiskAI [17]

Moving on from the research at the CSG focusing on cyber risk assessment, there is also work available by the group on the topic of risk mitigation (see the "Security Standards" and s"External Malicious Attacks" elements in [12]). In [77], the authors focus on decision support for security management by developing MENTOR, a recommender system for security services, which has also been integrated with the aforementioned SecRiskAI in [17]. It is important to point out also the relevance of research on risk mitigation in the cyber insurance context and MENTOR is one example of such approaches. Finally, there has also been specific focus in the CSG group on innovation in the cyber insurance domain. For example, [78] proposed SaCI, which enables to translate relevant information (such as coverage, company security metrics) from cyber insurance contracts to Smart Contracts running on the Blockchain in order to streamline (by the means of increased automation and transparency) the cyber insurance process across the life-cycle of the contract.

2.3 Process Mining

At the highest level, the fundamental idea behind process mining is rather simple - it refers to the analysis of behavior based on event data [79]. The focus of process mining lays on

business processes which Aguilar & Saven define as a "the combination of a set of activities within an enterprise with a structure describing their logical order and dependence whose objective is to produce a desired result". [80].

As depicted in Figure 2.10 below, Aalst [79] positions Process Mining as the "bridge" between data and process sciences. The argument behind that is that process sciences tend to be "naively" model-driven in that the models they are concerned with are not confronted with data about real behavior, whereas data science typically does not consider the process perspective. Process mining enhances the combines these perspectives [81]. Process mining can be therefore seen as complementary to the more general fields such as business intelligence (supported by solutions from vendors like Tableau, Microsoft, SAP or Oracle), data mining and machine learning in that it adds the process perspective (i.e. the perspective of observing the sequences of activities meant to address some organisational goal that typically involves both people and information systems) to these domains that are otherwise data-centric [79]. In that context, the importance of process mining has been further driven by the proliferation [79] of Process-Aware Information Systems (PAIS) [82] in the recent decades. Examples of PAISs include workflow management systems, customer relationship management systems, or ERP systems [79]. Process mining is especially fit to analyse the behavior extracted from such systems and also to challenge the models on which such systems are based in order to drive their further development and configuration with the goal to increase the performance of the processes such systems support [83].

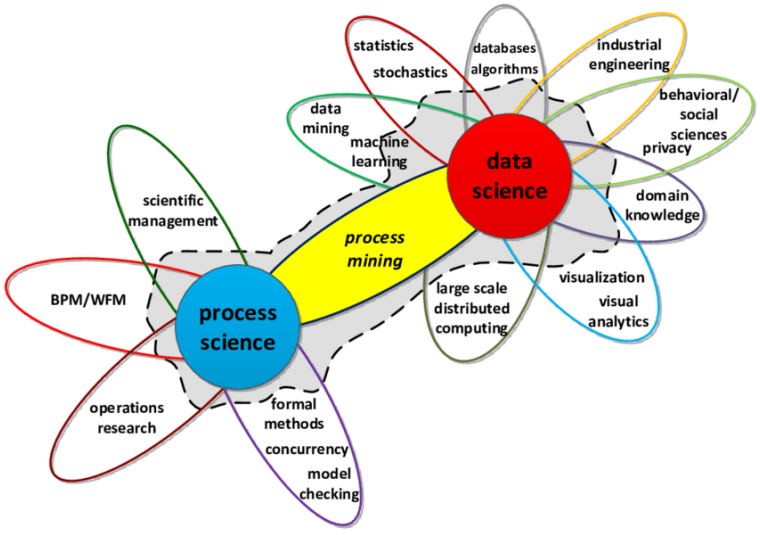

Figure 2.10: Contextualisation of Process Mining [81]

The aforementioned process perspective is arrived at by relating event data about observed behavior to process models, that might be in the form of Petri nets or BPMN models that can be wither discovered [79] using algorithms that are presented in a later section, or constructed manually. In doing that, process mining can analyse processes in a given organisation either from the perspective of process performance, or process conformance. [79] Both of these perspectives will be considered for cyber insurance applications and reflected in the choice of case studies.

2.3.1 Role of Process Models

In order to understand process mining, the notion of business models must be briefly introduced. Process models play an important role, especially in larger organizations and are typically used for one or more of the following reasons: insight, discussion, documentation, verification, performance analysis, animation, or configuration. Such models can be either (1) informal models (such as policies, verbal descriptions, documentation etc.), (2) formal models that typically describe processes in terms of activities, whose ordering describes causal dependencies, (and possibly sub-processes). Formal models take advantage of modelling standards such as BPMN (an example of which is depicted in Figure 2.11), UML, and EPC [79]. Different perspectives can be reflected in formal models - most prominently the control-flow perspective, but also the organisational, data and temporal perspectives.

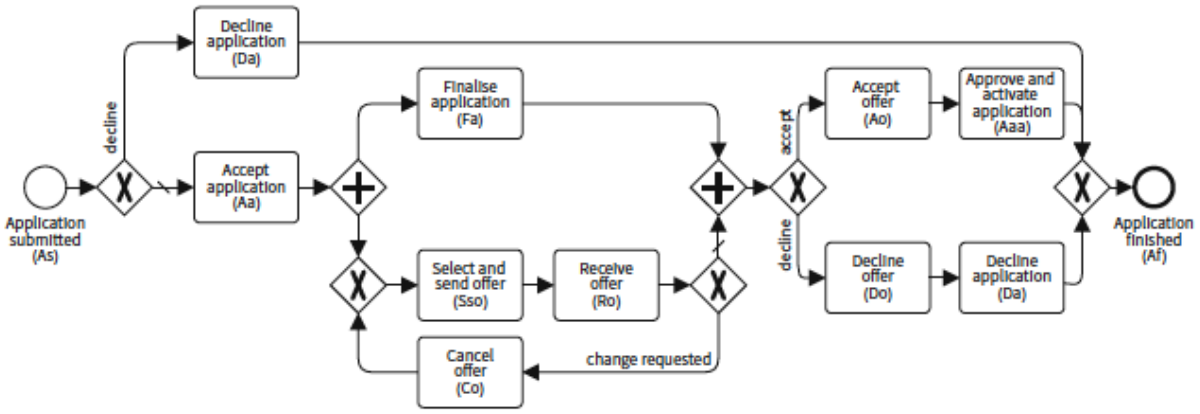

Figure 2.11: Example of a Process Model of a Loan Application from [84]

As [79] points out, the main difficulty in regards to effectively using process models is that they are often disconnected from reality. By defining, process models need to apply some level of abstraction in order to describe complex behavior. Unfortunately, this often results in process models being *idealised versions* [79] of how the actual processes that they are meant to describe are executed in reality. If no attention is given to minimising this gap, the organisation would not be able to trust the process models, or use them for the analysis of the actual behavior in the organisation. And this is exactly one of the main reasons why process mining has enjoyed increased popularity for the following reasons. Analysing business processes is critical for decision makers in organisation, but the value of relying on models only is limited and as the available event data from transnational systems grows exponentially[79], relating these two perspective becomes an attractive proposition. Process Mining refers to the set of techniques that aim to achieve precisely that by discovering actual processes, comparing them to process models and supporting their enhancement [79].

2.3.2 Process Perspective in Security: A Brief Excursion

Before moving on to discuss process mining in more details, in the following section, my aim is to outline how the aforementioned notion of managing business processes and analysing process models is relevant in the context of cyber security. The following paragraph outlines a number of works that argued for applying the business process perspective in that domain.

In [85], the authors advocate for the business process perspective in security engineering. They posit that understanding the nature of the organization is growing in importance compared to focus on technology. Business modelling is important in that regard, because it captures interacting behavior among humans and other agents within an organization (with or without the involvement of technology) and most security threats originate at the level of these interactions. Finally, it is pointed out that the level of business process ([86] in [85]) is an appropriate level at which users can express their security needs and where they feel 'most comfortable'. In order to support this notion, the authors propose the establishment of a business process-driven software development framework integrating security requirements. [87] focus on the security of business processes which they examine from five different perspectives depicted in Figure 2.12 from [88].

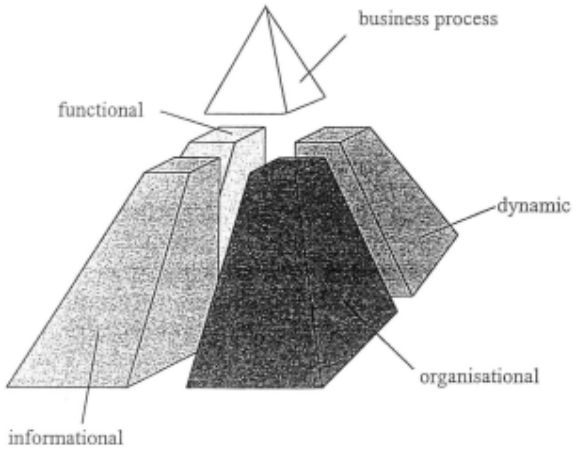

Figure 2.12: Different Perspectives Reflected in Business Process Definitions [88]

Then, [87] proposes the framework shown in Table 2.2 that is structured in three layers and is designed to support the analysis, modeling, and implementation of the security and integrity requirements of business processes.

Layer	Description of content	Representation method	Supporting method
Layer 3	High-level specifications of security requirements of business processes	Graphical	Analyzing methodology and a set of graphical concepts for security semantics Repository of case studies
Layer 2	Detailed specifications of security requirements	Intermediate language	Repository of information on how basic building blocks can be determined from security requirements
layer 3	Security hardware and security software building blocks	Program Program modules Hardware	Repository of hardware and software building blocks (e.g. crypto-library, security APIs, security dongle, etc.)

Table 2.2: Three-Layered Architecture for Process Security Specification According to [87], based on [88]

[89] argue that it is important that both security and business domain experts are able to define their security goals on a common abstract level - at the level of a business process model that can be in turn expressed as service-oriented-architecture. An approach to express security goals at the business process level is also presented. Finally, [90] acknowledge that security per-se does not generate business value, but rather an appropriate investment in security reduces the expected loss of business value. Based on that, it is argued that making a connection between security and underlying connection with business processes provides a basis for the cost-benefit valuation of security. To that end, the paper proposes the "IT-Security Valuation Framework" for the valuation of security measures based on the external value of core business processes. As an example, impact of a security violation can be calculated by simulating a business process modelled using standard approaches and taking into consideration metrics such as a loss of profit resulting from the stop of a business process for a given time, as well as the related employee costs considered in the simulation.

Summarising the aforementioned works, it is clear that the business process perspective (and security at the business process layer of cyber security) is relevant in the security context and one of the interesting question, that the section on process mining in this thesis explores a, given that finding, is what role process mining can play in that domain.

2.3.3 Process Mining Market

The relevance of process mining has been growing in recent years, according to the Harvard Business Review [1] the market for process mining solutions grew from $110 million in 2018 to $320 million in 2019. Two additional examples that illustrate the growing interest in process mining are that (1) the process mining vendor Celonis is currently the highest valued startup company in Germany, according to Pitch Book [91] and (2) many big-tech companies are recognising the importance of process mining by entering the market, typically via acquisitions. For example, the German software company SAP acquired the business process intelligence company Signavio in 2021 [92] and the RPA company UI Path acquired the Process Gold, another process mining vendor [93]

From the perspective of applications, process mining has been proven to work in a number of different domains [94]. From the perspective of general applications, operational excellence initiatives are one of the most typical use-cases. Typically, organisations across industries [95] ranging from healthcare, logistics, financial services to manufacturing (see Figure 2.13) are interested in analysing one or more of their business-critical processes and in uncovering some type of process improvement potential. Such processes can include Order-to-Cash, Procure-to-Pay, or Incident-to-Resolution processes [95].

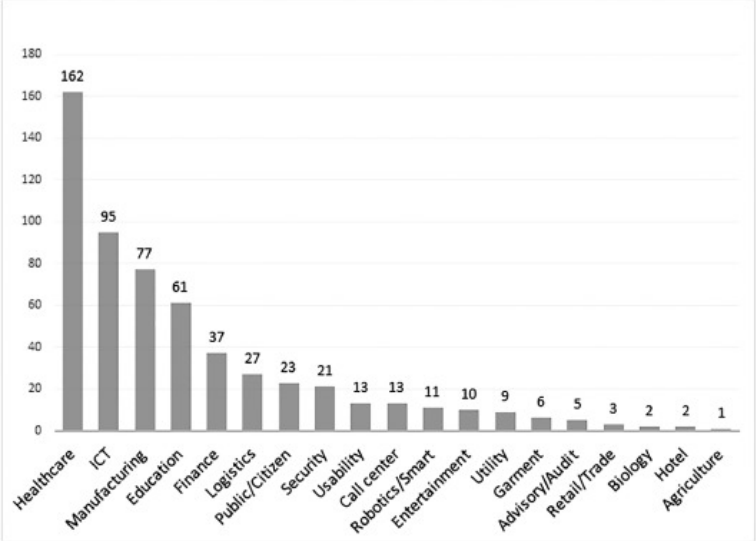

Figure 2.13: Publications in the process mining domain grouped by domains, collected by [95])

As you can see in Figure 2.13, security and auditing domain are also a prominent application of process mining method, some of the relevant publications in that area will be the subject of the Chapter 3.

2.3.4 Fundamental Methods of Process Mining

Now that the notion of business process models (with added focus on security) as well as process mining including it's business value and applications have been introduced, the following section will focus on a brief introduction to process mining methods as reflected in Figure 2.14 [14].

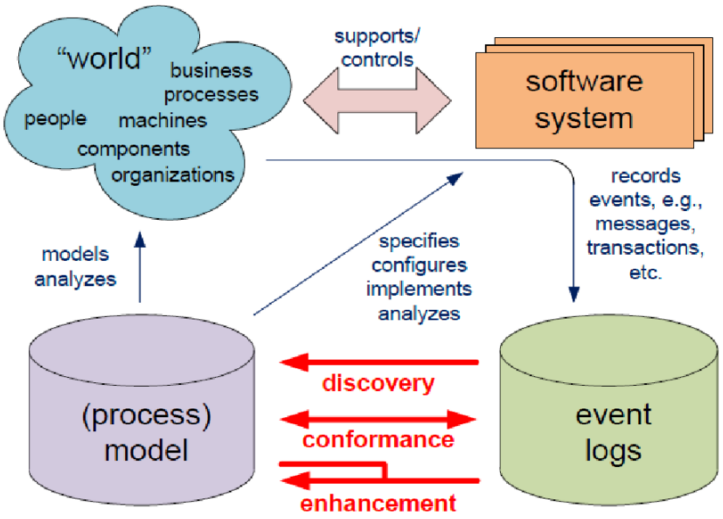

Figure 2.14: Overview of process mining and three main types of process mining [14]

The prerequisite for the application of process mining methods (as illustrated in the Figure 2.15 below) is an event log with (at least) the following data available: Timestamp, Activity and case ID. In addition to that, the event log can also include (or be enhanced with) data on attributes that can be used for filtering purposes (such as employees execution activities, costs related to an activity, organisational units etc.). Attributes can either relate to the case (case-level attributes), or to events associated with a given case (event-level attributes) [14]. It is also crucial to explicitly point out the, somewhat obvious, limitation that behavior that leaves no electronic trace (no event log is available) would not be detectable by process mining. This might be the case, for example, for certain types of physical security breaches, or for certain types of fraud. Indeed, this is true for any other analytical method relying on data from information systems [96]. The generation of an event log depends on the source system. Typically, the source data needs to be transformed in some way to arrive at the event log in the required structure. For some of the more standard enterprise systems, templated transformation scripts are often available, but the task can also be conducted manually by the means of, *e.g.,* custom ETL (extract, transform, load) pipelines that can ensure that the process mining system can conduct analyses on the source data [97]

Event	Application	Offer	Activity	Amount	Signed	Timestamp
...	...	...	...	...	...	...
e_{30}	*O*3521	*A*5636	Select and send offer	€500		Jan 04, 16:32
...	...	...	...	...	...	...
e_{37}	*O*3541	*A*5634	Select and send offer	€1500		Jan 05, 12:32
e_{38}	*O*3521	*A*5636	Receive offer		NO	Jan 05, 12:33
e_{38}	*O*3521	*A*5636	Cancel offer			Jan 05, 12:34
e_{39}	*O*3542	*A*5636	Select and send offer	€500		Jan 05, 13:29
e_{40}	*O*3542	*A*5636	Receive offer		YES	Jan 08, 08:33
e_{41}	*O*3542	*A*5636	Accept offer			Jan 08, 16:34
e_{42}	*O*3541	*A*5634	Receive offer		NO	Jan 10, 10:00
...	...	...	...	...	...	...
e_{54}	*O*3541	*A*5634	Decline offer			Jan 10, 10:04
...	...	...	...	...	...	...

Figure 2.15: Example event log [95]

Process Discovery

Process discovery can be considered as the central process mining method [14]. Where previously the manual creation of process models was necessary, process discovery enables process models in different notations, such as BPMN, process maps, or Petri nets [98] to be discovered from the event log. The goal of process discovery is to discover models that are fit for given purpose, which can be measured by the following criteria [84]: fitness, precision, generalisation and simplicity. The criteria are competing. For example if a model is too general, it might suffer from "underfitting" (allowing for more behavior than in the observed log). On the other hand, too precise models lead to "overfitting" [99].

Many different algorithms have been developed in the last decades that have different focuses and different advantages and disadvantages, An overview can be found in Figure 2.16. Comparative studies on process mining algorithms are available, *e.g.,* in [100], therefore this section does not aim to discuss them in detail.

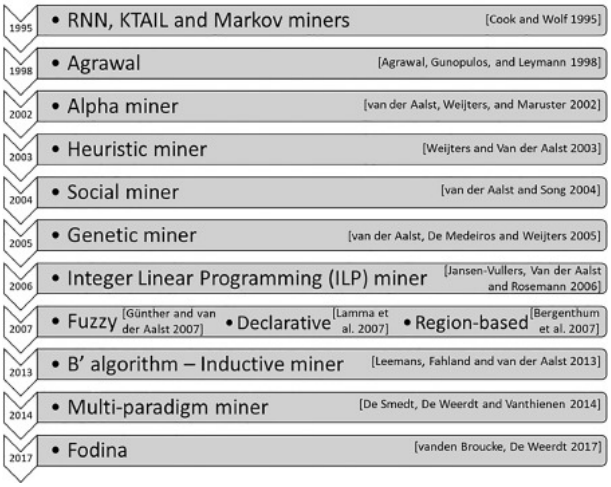

Figure 2.16: Overview and history of process discovery algorithms [95]

For the purposes of this thesis the algorithms alpha miner, inductive miner and heuristic miner are considered, as they have corresponding implementations in the PM4Py [98] library used for the implementation of the approach. The alpha algorithm (formalisation can be found in [14]), can be used to illustrate the idea of process discovery in a simple way, but it does not does not take frequency of traces into account and also provides no soundness guarantee, it can also have issues with loops [14]. In contrast to the Alpha algorithm, the heuristic miner takes into account the frequency of events and sequences and also provides additional parameters which can be used for filtering of infrequent traces [79], which is often useful for large data-sets with noise. Finally, the inductive miner provides fitness and soundness guarantee [98] and is widely used in commercial process mining tools.

Conformance and Compliance Checking

Process mining often serves as the bridge between the processes defined as per policies or regulatory standards and the actual process executions. The approach of relating an event log to either some set of declarative rules, or a process model can be summarised under the term '"conformance checking" mapped by [14]. Figure 2.17 provides an overview of conformance checking methods that can be grouped into three main categories.

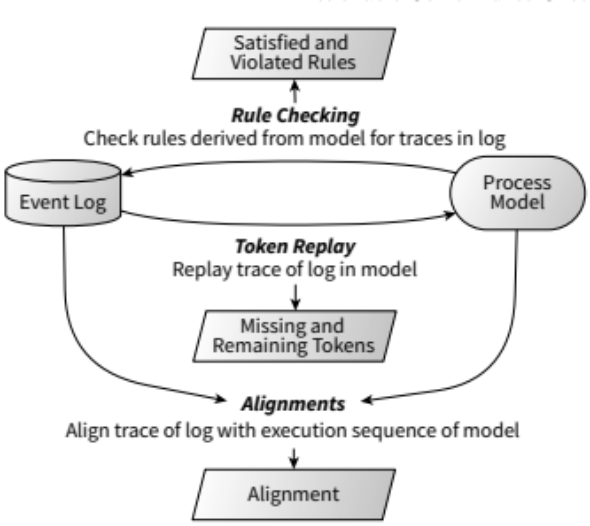

Figure 2.17: Overview of conformance checking approaches [84]

Next, a brief overview and finally a comparison of these three main approaches is provided.

1. **Rule checking** Checking whether the traces of the event log satisfy declarative rules (that are derived from and that capture the behaviour defined by a model) represents the simplest form of conformance checking. These rules might include *cardinality rules*, *precedence and response rules* ,*ordering rules* and *exclusiveness rules* [84]. These rules can, in practice, be formalised based on linear temporal logic (LTL) [84] and checked automatically. However, certain rules can be also simply checked just by the visual analysis of the discovered process. The example in 2.18 illustrates this idea by showing how which traces satisfy a defined cardinality rule.

Activity	*As*	*Da*	*Aa*	*Fa*	*Sso*	*Ro*	*Co*	*Ao*	*Aaa*	*Do*	*Af*
Cardinality:	[1, 1]	[0, 1]	[0, 1]	[0, 1]	[0, *n*]	[0, *n*]	[0, *n*]	[0, 1]	[0, 1]	[0, 1]	[1, 1]
$T_1 = \langle As,Aa,Sso,Ro,Ao,Aaa,Aaa \rangle$											
	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗
$T_2 = \langle As,Sso,Fa,Ro,Co,Ro,Aaa,Af \rangle$											
	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
$T_3 = \langle As,Aa,Sso,Ro,Fa,Ao,Do,Da,Af \rangle$											
	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Figure 2.18: Example of checking traces in the log denoted by T_n against a cardinality rule derived from a process model, violations are denoted by X [84])

2. **Token-based replay**

   Rule checking might provide only a limited conformance feedback and might be hard to implement for large sets of rules. [84]. Replay of the traces of an event log in the process model is the foundational idea behind token-based replay. The goal is to identify whether and to what extent a trace of the event log corresponds to some sequence foreseen the model[14]). We therefore replay each trace in the model (so-called Petri nets are used in token based replay [98]) event by event and check in each step of the replay whether the current state of the model fulfils the conditions for the execution of the the corresponding task. The conformance feedback of token based replay consist of *fitness* measure either at a *local level* (trace), or at a *global level* (log) [84]

3. **Alignments** Alignments [101] are typically considered the most mature of the three conformance checking methods presented [84].

To briefly illustrate the idea of alignments, please consider the following trace in an event log

$$T_1 = \langle As, Aa, Sso, Ro, Ao, Aaa, Aaa \rangle$$

And an execution sequence in a model

$$E_1 = \langle As, Aa, Sso, Ro, Fa, Ao, Aaa, Af \rangle$$

Then, one of the possible alignments is given as follows:

log trace T_1	As	Aa	Sso	Ro	$\gg$	Ao	Aaa	Aaa	$\gg$
execution sequence E_1	As	Aa	Sso	Ro	Fa	Ao	Aaa	$\gg$	Af

Figure 2.19: Example alignment of log trace T_1 against an execution sequence in the process model E_1 [84])

The example alignment in Figure 2.19 comprises of the following:

- Six synchronous moves
- One log move, denoted as (Aaa, $\gg$)
- Two model moves, denoted as ($\gg$, Fa) and ($\gg$, Af).

When checking conformance using the alignment method, cost is assigned to each move. The optimal alignment then aims to minimize the total cost of moves [84] If the alignment of a trace contains synchronous moves only, it can be seen as a valid execution sequence of the model [84]. From the perspective of global conformance measures [84], if the same

is true for all traces in the event log, it can be concluded that the log fits the model and the actual behavior can be well-explained by said model.

For a summary of when each conformance checking method might be applicable, please refer to Table 2.3 that is provided below. It can be concluded that the three conformance checking methods are complementary and each might be suitable for different scenarios.

Method	**Advantages**	**Disadvantages**
Rule checking	Does not require complete model Easy to implement with constraints Suitable for lower number of rules	Not complete conformance check Does not consider Becomes complex and hard to interpret if too many rules are applied
Token-based	Simple Not computationally demanding Allows basic diagnosis Only for Petri nets	Events not in model can't be considered Not suitable for complex processes Early deviations might mask later deviations
Alignments	Severity of deviations configurable High accuracy Model independent Configurable cost function can assign cost to different violations	More computationally demanding

Table 2.3: Comparison of Conformance Checking Methods Based on [84] and [79]

Chapter 3

Related Work

In this chapter, the four main categories of related work investigated will be presented and summarized. As the applications of process mining in cyber insurance have not yet been researched, the review conducted will point to key areas that are overlapping, or adjacent. - namely process mining in cyber security, in *GRC*, general insurance and in the analysis of process performance.

This chapter is structured as follows. After providing a general introduction to anomaly detection and a brief overview of log analysis in general, the first the research stream of applying process mining to cyber security will be presented, followed by an overview of the research on process mining in the areas summarised under the umbrella of the *Governance, Risk and Compliance* (*GRC*) that also encompasses the domains of Risk Management and Audit. Next, in a brief section on security process performance analysis, I will point out that anomaly detection is not the only area of applications of process mining in cyber insurance. Given the criteria in underwriting manuals analysed for this thesis, it is clear that the performance of security and compliance-related processes (such as incident management, help-desk, or even general processes with regulatory or compliance requirements) also needs to be taken into consideration and present selected works in that area. Finally, the so-far limited research on applications of process mining in the insurance domain will be briefly reviewed.

3.1 Process Mining in Cyber Security and Software Reliability analysis

Before focusing on the applications of process-mining methods in cyber security, this section provides a brief overview of general methods of log analysis, which has been a method well-known in the domain of cyber security. [102] conducted a literature review of available research on vulnerability and security log analysis. Their review points out that protection against external threats is typically the focus of cyber security research, while internal threats and intrusion detection get relatively less attention. This is in line with the statistics I pointed out from [13] that evidence the importance of analysing actions

of employees and internal processes. Log analysis is one of the most promising methods to do that, especially because it offers potential for critical automation as systems grow exponentially together with the number of highly qualified experts required to detect and analyse anomalies. [102] then points out the two general categories for intrusion detection identified across current research: *signature-based detection* and a*anomaly-based detection.* Signature-based detection relies on rules as input to determine when to flag which patterns of a log. Such rules can be based on past experience. The simplicity of these methods makes them an attractive choice, but they might fall short when detecting novel (*zero-day*) attacks is required. On the other hand, anomaly-based detection compares logged behavior to other logs in the same stream in order to detect anomalies. The typical shortcoming anomaly-based detection methods of false-positives, while the advantage is that some intrusions that rule-based methods can not detect might be uncovered. Often, the desirable approach is some combination of both methods. Regarding the specific methods employed, the traditional approach has been manual analysis. However, this becomes increasingly more difficult given large amount of data coming from modern systems. [102] therefore points out machine learning, data mining, and text analysis as the main strategies to automate the task. But those methods do not come without challenges. Machine learning, for example, might suffer from performance issues making it difficult to use for real-time applications. Multi-source log analysis, requiring federation and ETL pipelines suffers from a similar issue. High false positive rates, as mentioned above as well as inconsistent formatting of logs further complicate approaches such as clustering traces to detect anomalies.

After having provided a brief overview of log analysis methods and their challenges (many of which are applicable to process mining as well), I will focus specifically on process mining methods in the following sections. As my literature review shows, the research interest in the area has been significantly picking up, especially in the recent years. As I will demonstrate, process mining methods provide analysts with the attractive proposition of being well-suited especially (but not exclusively) for analysing the security posture at the business layer of enterprise architectures, as depicted in Figure 3.1.

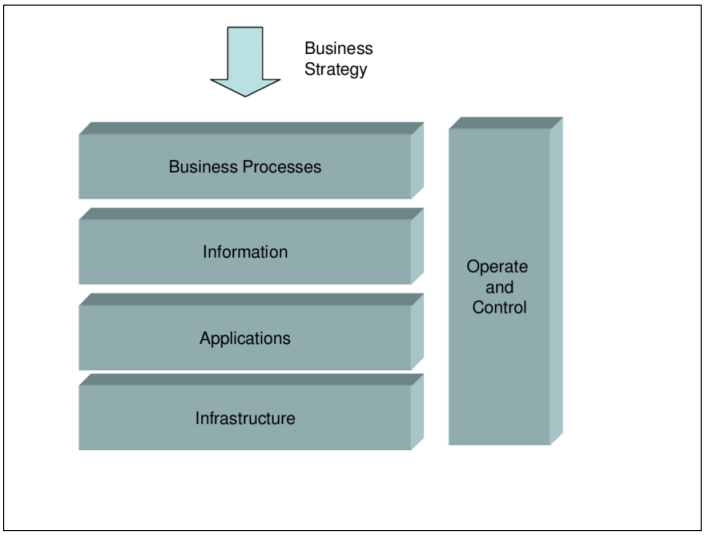

Figure 3.1: Layers of enterprise architectures according to [103]

This thesis provides a systematic review conducted at the intersection of process mining and cyber security as a first step. The first holistic one has been conducted in 2021 by [94]. The authors of the systematic literature review divide the process mining use-cases two the two categories (cyber security and reliability) and cluster the research directions in these area to 9 research streams, as well outline the most relevant works. In the cyber security category, the research streams include *Security of Industrial Control, Security of smartphones, Network traffic security, Web-application security, Attack inspection, Outlier behavior detection and Fraud detection.* For the reliability category, streams of Quality Assurance and Error detection are identified. Interestingly, [94] also introduce a number of further criteria in their review of process mining research in the cyber security domain. Namely, they include target period (past/past & present/present), PM type (referring to one of the three fundamental process mining methods previously introduced in Chapter 2, expert knowledge (whether it is required for a given use-case) and approach to the model analysis (automatic vs. manual). They then evaluate the published research in scope of the review on these dimension. Another categorisation of the research at the intersection of process mining and cyber security is provided by [104], who identifies 7 overall *security categories* to which process mining techniques have been applied. The categorisation according to the systematic literature review by [104] include *Conformance checking, Anomaly Detection, Compliance Control, Fraud Detection, Risk Management and Access Management.* Apart from that, [104] also points to more holistic works that he refers to as Systematic reviews.

The foundation for the applications of process mining in security can arguably be traced back to [105]. Not only this work is widely cited, but it is also often seen as a starting point for many of the works reviewed by this thesis. Specifically, [105] points out that also the literature on security can be split to streams concerned with *(i)* computer security and *(ii)* auditing (corresponding to the structure of this review) and while they are concerned with different level of abstraction, that certain behavioral pattern found in audit logs can point to security violations. Process mining is than established as a fitting method to analyze these trails. In particular, α-algorithm is proposed there be applied to mine the process perspective from and applications of it are discussed. The first one is based on the saydetection of anomalous process executions in the mined *workflow net* based on "playing the token game", while the second focuses on the conformance checking by comparing new audit trails to fragments of a previously discovered model. Based on the analysis, [105] argue, that generally, traces in his analysis, that have trace fitness lower than 80% can be considered anomalous. It is concluded that, the α-algorithm shows clear potential to be applied for scenarios at different layers of security ranging from intrusion detection to electronic fraud. The topic of process mining approaches for detecting security anomalies in the security context has been further investigated by number of papers building on [105], including [106] investigating anomalies in *PAIS* (process-aware information systems), [107] that focuses the application of fuzzy association rules learning and process mining for anomaly detection and [108] in which a general method (depicted in Figure 3.2 below) for anomaly detection using process mining is presented. Especially the last work is of high importance for this thesis as the proposed method is applicable across different security and GRC scenarios and can to different extents be identified in many of the papers reviewed further in this chapter.

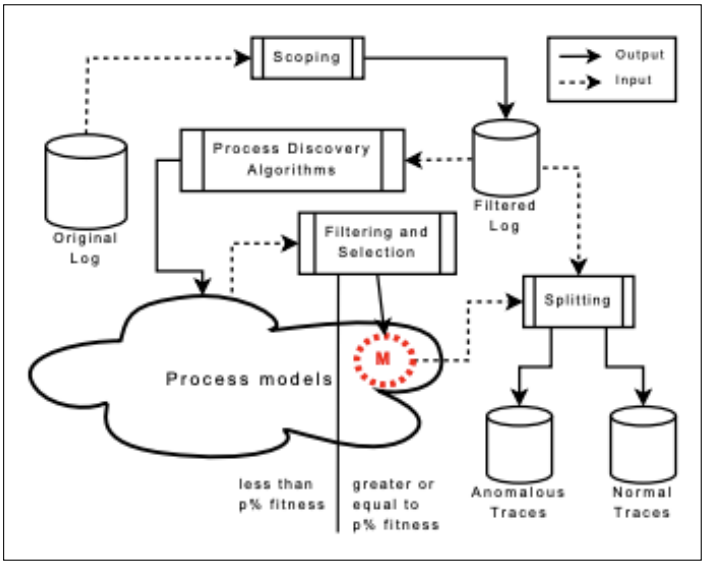

Figure 3.2: Approach to detect anomalies with process mining [108]

A deep-dive on selected applications of process mining in cyber security with specific focus is provided in the rest of this section. Specific scenarios (e.g. ransomware, intrusion detection, and authentication) and the methods that are relevant for the scope of this thesis are pointed out.

In a paper concerned with **ransomware**, which is one of the key topics currently driving pricing pressures in the cyber insurance market, the authors of [109] investigate a method to detect ransomware by combining process mining discovery algorithms (in this specific case the implementation of fuzzy miner in the Disco process mining solution) with classification algorithms such as j48, logistic regression and random forest that were applied to features extracted from the discovery-generated model. While acknowledging that there are many different approaches how ransomware can be detected and distinguished from benign software, the authors conclude that process mining can be suitable to detect ransomware and is useful addition to the methodological toolbox cyber security analysts. Based on analysis of event logs of both benign software behavior and those of 21 different ransomware families collected synthetically on a virtual machine (Windows platform), the authors claim to have achieved a 95% accuracy in detecting ransomware. Finally, they argue that the method is capable of quick detection to the extent that ransomware can be detected before files can be encrypted and the system locked.

In the **intrusion detection** category, multiple works are available proposing approaches to apply process mining to monitor and analyze the processes and events occurring in information systems in order to identify intrusions. For example, [110] advocates for real-time process mining analyses based on discovery and conformance checking methods . While the paper does not provide a detailed proposal for the implementation of such mechanism, the authors suggest that intrusion detection systems based on process mining techniques could be incorporated in both network-based and host-based intrusion detection systems. The main reasons for that is *(i)* the potential for performance improvements compared to more general data mining techniques, possibly allowing for real-time monitoring, *(ii)* The incorporating of delta analyses (with conformance checking) comparing the defined

process and the actual behavior in the systems, and *(iii)* the perspective of how data moves from one point to another and whom it is handled by.

Another contribution related to exploring the potential of PM in intrusion detection, focusing on **smart metering** of critical infrastructures (on the example of smart grids) can be found in [111]. This industry paper is one of key importance for cyber insurance as well, due to potential for catastrophic scenarios due to the risk of a remote turnoff. Interestingly, the paper also points out some of the shortcomings in the methods proposed elsewhere, for example in the works reviewed above concerned with the control-flow perspective [108, 106] and organizational perspective [112]. In summary, the authors point out that anomalies observed previously likely do not correspond to real-life intrusion scenarios and go on to propose methods based on simulations that also take the effects of an intrusion into account based on process descriptions gathered from major energy providers which were then modelled as attack-defence trees (example depicted in Figure 3.3). Suitability of conformance checking is then evaluated for each modelled attack, concluding that process mining can aid in detecting attacks on smart meters. However, for practical applications, combining multiple perspectives (as was pointed out in 2 is required (*e.g.,* control flow, time and organizational perspectives).

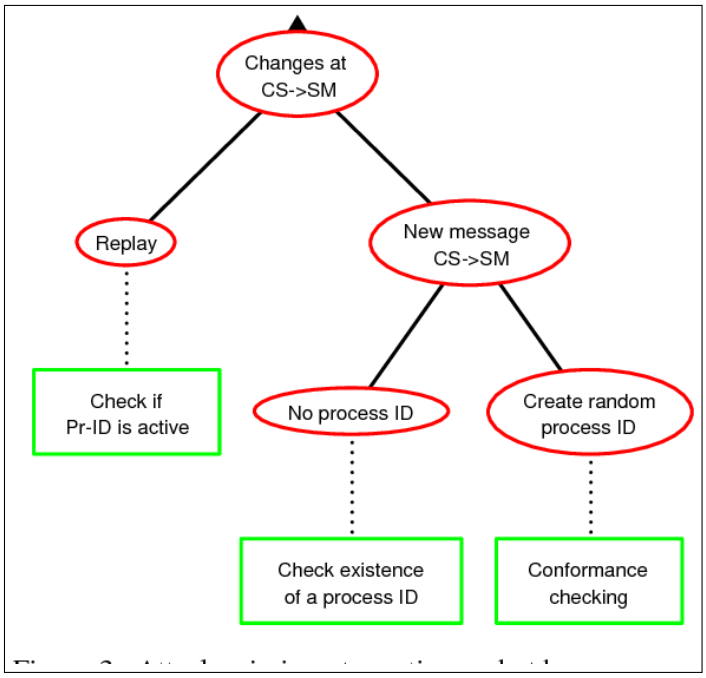

Figure 3.3: Example Attack-Defence Tree Considering Conformance Checking constructed by [111]

Process mining techniques have also been applied to **authentication processes** with the goal of detecting malicious login events and extracting models of behavior of malicious users. [113] first contextualises their proposed approach among the three main concepts of the network security domain - prevention, detection and investigation. The authors argue that most research focuses on the prevention (such as on Software Defined Networking and Network Functions Visualisation) and on detection (e.g. on rule-based or signature based detection, as well as diverse machine-learning techniques).

In this context, according to [113], investigation techniques are not well-investigated due to only limited log-data (with labelled attacks and alerts) available. Based on that, they propose an investigation tool that first, using the α-algorithm, extracts process models from business event logs consisting of a set of sequences of changes. Then they apply conformance to detect issues. The paper concludes that the model constructed by the proposed approach can be used for defending against malicious login events.

Hemmer et al. [114] tackles the major challenge of security management of increasingly complex **IoT (Internet-of-Things)** systems and applications. The nature of IoT devices that tend to be subject to resource constraints makes it difficult to deploy additional mechanisms such as intrusion detection systems to IoT networks. For that reason, authors propose a process mining approach that employs passive collection of data aims to minimize overloads at the network and device level. The key contribution of the paper is that a specific solution architecture is demonstrated and proof-of-concept evaluated, consisting of three main blocks, which include pre-processing, model building block based on process discovery (*i.e.,* inductive miner), and misbehavior detection block. The last one employing methods of conformance checking (with alignments) and relying on conformance metrics to detail whether the model generated in the second block can replay a given trace or log. The authors argue that combining process mining with clustering can be beneficial, based on results of experiments measuring attack detection performance using industrial event logs.

Process mining techniques have also been applied in the domain of web application security. In [115], for example, the authors mined the logs of a production system based on an Apache Tomcat web server and used conformance checking (with trace driven simulation measuring the fitness metric) to detect deviations of behavior observed in the event log generated by fuzzy miner from the use-cases of the system modelled in Unified Modeling Language (*UML*). While the authors also concede that some of the deviations identified corresponded to conform use-cases that the UML model just did not consider, other identified patterns strongly hinted at attacks. For example, Denial of Service attacks attempts were identified, brute-force attacks to get passwords, as well as cross-site scripting patterns.

Also, [116] investigated, at a high-level, the applicability of process mining in Intrusion Detection Systems and highlighted the advantage of relative speed compared to other methods falling under the umbrella of data mining that are also presented in an overview. A brief comparison of process mining algorithms is presented. However, the paper arguably does not go into much detail and no empirical evaluation backing up the reasoning is offered.

3.2 Process Mining in GRC, Risk Management, and Audit

In comparison to the cyber security domain, based on the number of published research works as well as the number of citations, the GRC, Risk Management, and Audit can be considered more mature from the research perspective. In the following section, a systematic literature review concerned with Process mining in GRC, Risk Management,

and Audit will be presented followed by a selection of key papers with relevance for the cyber insurance domain.

The most recent overview of the field is offered by [16], who considers 34 selected papers on the application of process mining in GRC and auditing The types, areas, objectives and frameworks are then mapped and their components classified according to the 6 common phases of process mining projects as outlined in the PM^2 methodology [117], depicted in the Figure 3.4, as components. Afterwards, 32 common sub-components were identified across the selected papers. The components covered include *Planning, Extraction, Data Processing, Mining and Analysis, Evaluation and Process Improvement and Support*. Continuous auditing (with related automation) and algorithms specialised for GRC use-cases are pointed out as research opportunities for the future. It is also pointed out that multiple perspectives are important for GRC analysis and auditing (*e.g.,* control-flow and time perspective). Financial domain auditing has also been identified as the most often investigated domain (15 papers), followed by manufacturing (4 papers) and finally insurance (2 papers).

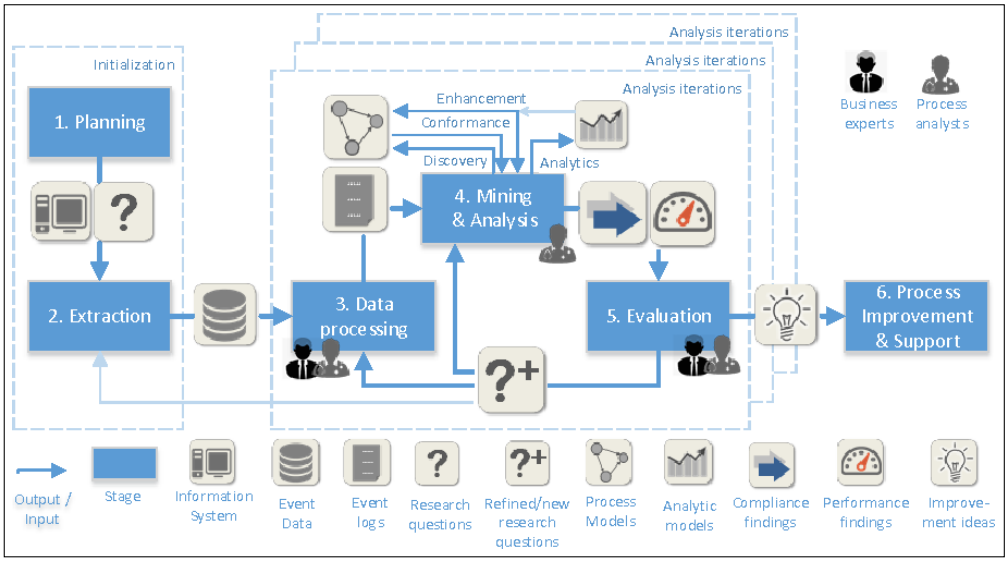

Figure 3.4: PM^2 : A Process Mining Project Methodology [117]

An overview of what value process mining can offer to auditing in general (*i.e.,* not exclusive to security auditing) is provided in [118]. The work maps process mining methods, focusing on discovery and conformance checking, to different auditing challenges. Specifically, the authors argue that using process mining, auditors don't have to only rely on samples of data and can instead observe all events in a given process to audit in a continuous manner [119]. Increasingly more widely available eventlogs serve as a form of business provenance. The paper also proposes an auditing framework based on process mining highlighting how the capability of conformance checking fits well the need of auditors to compare de jure models (as defined by law, policies etc.) with de facto models discovered from actual process executions. LTL rule checking is also mentioned by van der Aalst [118] as a possible mechanism to detect anomalies. Finally, limitations are discussed that include the complexity of extracting data in a reliable manner, as well as the paradoxical challenge that process mining techniques would typically lead to finding more exceptions needing investigations, leading to an increased auditing effort. However, the number of

recent accounting scandals are seen as a justification for improved, rigorous conformance and compliance checking mechanisms.

Zerbino et al. [120] build on top a methodology of the auditing framework presented by van der Aalst in [118] (see Figure 3.5 below). It was developed a process-mining methodology specifically for the audit of information systems (IS). They frame process mining as a so-called *Expert System (ES)* engine that can "extract the actual business rules of the IS and contrast them with rules set out by decision makers". They then contextualise such process mining engine among *Computer-Assisted Audit Tools* (*CAAT*) and Techniques and ESs used for audit and highlight different advantages and disadvantages of each category. However, the main contribution that is highly relevant for the thesis at hand, is their proposal of a process-mining-based IS auditing framework consisting of 5 stages: Justification and planning, Data Extraction, Control-Flow model construction, Model enrichment and conformance checking. The framework is then validated on an auditing use-case with specific findings presented (4 of them likely implying both legal and operational risks). Interestingly, they also provide pragmatic guidance on the choice of discovery algorithms depending on the process structure and complexity.

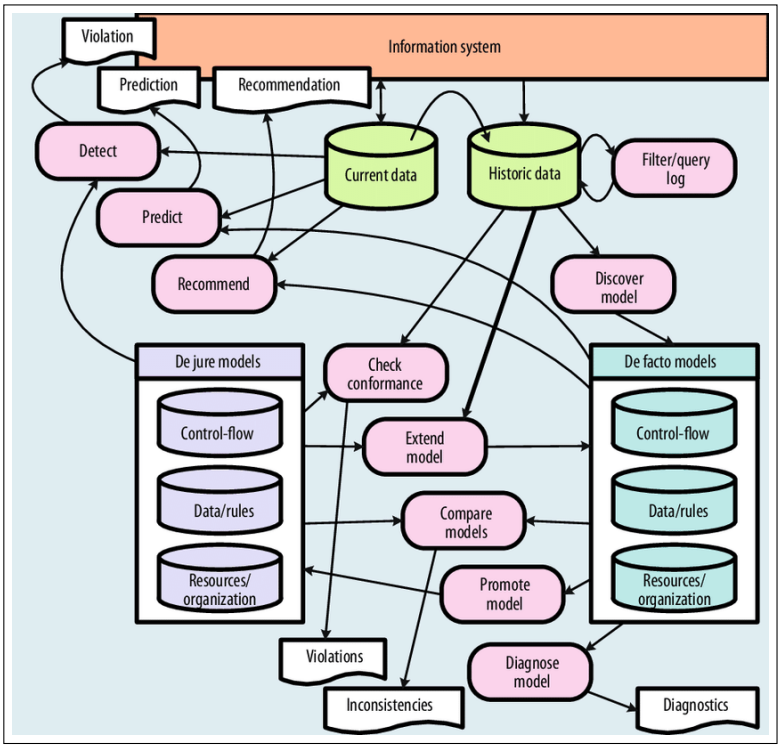

Figure 3.5: Framework for Process-Mining-Based Audit of Information Systems proposed by [118]

Jans et al. [96] provide further reasoning on why process mining is valuable for audit - considering process discovery, conformance checking, performance analysis, social network analysis and decision mining (*i.e.,* focusing on decision points in a discovered process). The reason for this work is reviewed separately is that it strongly focuses on ERP system audit perspective. This is a crucial area for process mining to mention as it has been extensively exploited commercially (see e.g. [92]) as processes having a relation to the

ERP constitute a significant part of enterprise process landscapes. The authors also point to the work conducted in [121], which focuses on mining industry-standard SAP systems. Based on that, it is possible to argue that existing commercial efforts can be leveraged in a synergistic way to enhance the cyber security assessment perspective. Furthermore, [96] introduces the four sources of added value for process mining in auditing: analysing the entire population instead of a sample, inclusion of meta-data independent of the actions of the auditor, effective way to implement the audit risk model and additional ways to conduct audit, such as discovery and social network mining. Finally, a critical point, raised by [96], regarding the principle of deterrence is that certain processes are continuously monitored with process mining and therefore expected to have an effect on organisational behavior.

Additionally, other researchers [122] have also explored how process mining techniques can map directly to COBIT, one of the most widely-accepted enterprise IT governance frameworks. It sees process assessment as one of the components of enterprise capability determination [123]. Specifically, steps in the COBIT assessment process include collecting and validating data [122]. The authors of [122] outline, in a case study, how process mining could aid these COBIT steps. Generally, such alignment with established enterprise risk management, audit and governance frameworks (*e.g.,* ISO 27001, COBIT, ITIL) [120] has an indirect relevance for the applicability of process mining in cyber insurance, as some cyber insurers consider pre-existing certifications and internal audit results in the underwriting process [12].

Focusing on one specific use-case at the intersection of auditing and security, a research group from the Masaryk University tackle the detection of insider threats by applying process mining techniques on audit logs [124, 125] . The insider threat topic is highly relevant for cyber insurers as it is often covered by cyber insurance policies and, according to the *Swiss Cyber Institute*, more than 34% of businesses globally are affected by insider threats [43]. At the same time, the problem of information asymmetries described previously makes it hard to assess the risk of insider threats for a given organisation. [124] argues that process mining can be a suitable alternative to existing insider threat detection approaches that tend to be of complex mathematical nature by providing more accessible interpretation using more user-friendly visualisations. Three specific insider threats use-cases are then designed and analyzed, including *(a)* conformance checking of user application activity, *(b)* visual analysis of data flow, and *(c)* file log analysis for declarative process mining. The data flow scenario will be considered in the cyber insurance context in case study later in the thesis. Heuristic miner from the *PM4Py* [126] package was then applied to discover model from a synthetically generated event log and alignment based conformance checking was applied to calculate fitness of new traces. Threshold of 50% *trace fitness* was chosen to classify traces as potentially malicious. In the use-case for declarative conformance checking, *LTLChecker* from *ProM* [127] was applied to define rules and check the conformance of new traces with them. [124] conclude that process mining techniques are useful insider threat detection from audit logs.

In [128], the academic perspective is combined with practice by a mixed team from the TUM and PwC. The paper focused on measuring deviation from a pre-defined cyber security process based on Identity and Access Management. A reference model is compared with a synthetically generated dataset from which a process is discovered. The discov-

ered process is the compared to the underlying reference model by applying a number of conformance checking methods. Violations of the process-flow are identified and fitness metrics calculated. It is observed that conformance checking can be used in IT Security Auditing and provides value for risk assessments. Again, a similar scenario is validated with experts in the cyber insurance context in the evaluation chapter.

The topic of applying process mining to security audits has been tackled in two related papers by Accorsi and Stocker [112] [129]. They offer a practical approach demonstrated on a case study from the financial sector based on loan application process, outlining the specific steps of conducting security audit exploiting process mining capabilities. The prevalent manual effort, long time to audit, limited automated tools available, and auditing based on samples instead of full audit coverage are mentioned as the problems that process-mining can support solving [112]. First, in [112], the conformance checking methodology is investigated by synthesizing a set of traces that are then tested against pre-defined security requirements. The authors aimed to answer the following questions: Does conformance checking generally allow the testing of traditional security requirements? What kind of properties can be detected by conformance checking? 5 different classes of security requirements based on [87, 130] were then defined and translated to a number of testable constraints. Next, different conformance checking mechanisms were applied to different types of constraints. Accorsi and Stocker, conclude that conformance checking is a powerful tool for the analysis of security requirements, including control flow deviations, separation of duties and obligations. In the related paper [129], discovery methods are also investigated from the perspective of security assessments.

3.3 Process Mining for Performance Analysis

Apart from the aforementioned streams that focused predominantly on conformance checking and that either in the security or more general *GRC* context. Contributions also have to be considered that focus on analysing not only the conformance, but also on the performance of processes relevant for cyber security. One example is [131] that analyses an incident management process, and a problem management process. For the purposes of *ITIL* assessment, [132] rates an incident management process using process mining.

3.4 Process Mining in Insurance

Finally, after covering different application of process mining in different cyber security domains and also in the overall *GRC* domain, it is important to mention the work conducted in [133], which proposes an application of the methods of process mining in general insurance underwriting scenarios. Even though cyber insurance scenarios are not considered or even mentioned, the proposed schema could also be applicable in the cyber insurance cycle.

As a general scenario, [133] assumes the following A service provider company (SP) delivers a sensitive service (or process - P) to a user receiver (UR). This is illustrated on an example of an online car market with sensitive data incl. licence plates that require anonymisation). This sensitive process is then insured by the insurance company (IC). This scenario is shown in Figure 6.4

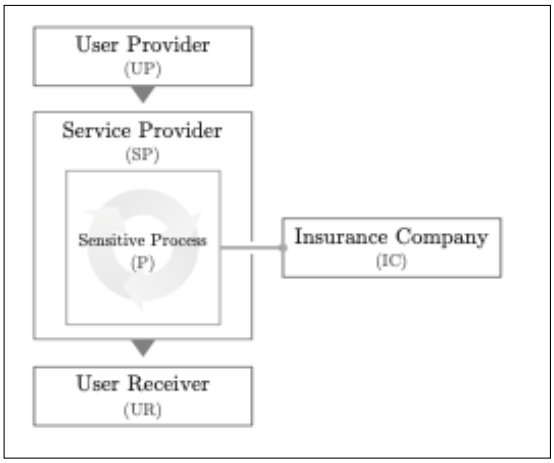

Figure 3.6: Scenario for Insuring Sensitive Processes via Process Mining [133]

The insurance schema is based on the reasoning that the sensitive process (P) has a specific set of steps that need to take place in a predefined order (in the form of a process model). Conformance checking is therefore proposed in [133] as a method to answer the following question. To what degree does the behavior in the modeled process conform to actual process executions reflected in a corresponding event log. The four traditional conformance metrics are then used to serve as quality metrics. As opposed to token-based replay (conformance replay), the alignment-based method is proposed as the more suitable underlying conformance checking mechanism. It is argued that different perspectives could be insured, such as access to sensitive data, sets of specific rules or properties, or the organizational perspective (actors in the process). However, there might be other paths and options as well. This set-up would be reflected in an insurance agreement centered on an agreed-upon formal model, which might aid in reducing ambiguity typical for textual notations and make it easier to automate the formal verification once problems with the process occur. An insurance cycle is then proposed consisting of 4 phases: *Modelling*, *Validation*, *Forensic phase*, and *Re-design* (analogous to process enhancement in [79] . Finally, the approach was validated on a case study, showing that ensuring sensitive business processes is feasible.

3.5 Summary and Research Gap

As the literature review shows, while a number of works are available on process mining applications in cyber security, in the *GRC* domain, as well as in performance analysis of security processes, no systematic exploration and approaches to apply process mining in cyber insurance are available, with the exception of one contribution in [133] focusing on

"sensitive processes" that makes no reference to cyber insurance. Based on the findings from the background chapter, my thesis argues that this is a significant research gap, as the assessment of cyber risk originating from processes is of vast importance and information asymmetries are one of the key challenges of insurability of cyber risk. The following chapter therefore develops the MeritMiner4CI approach to address this research gap.

Chapter 4

MeritMiner4CI Approach

According to the conducted analysis of the literature, process mining methods have never been investigated in the cyber insurance context. I consider this to be surprising because operational cyber risk related to actions of people and failed internal processes constituted are, as measured by aggregate losses from different types of risk, by far the two most severe types of cyber risk [13].

This chapter focuses on the systematic mapping of cyber insurance requirements to process mining methods and then describing the MeritMiner4CI as an approach for the application of process mining in the cyber insurance underwriting process. The chapter concludes with a *BPMN 2.0* process model that demonstrates the steps it proposes.

4.1 Systematic Mapping of Requirements

The following section focuses on clearly highlighting how process mining methods, specifically, map to requirements of the cyber insurance market. The section is structured as follows, first mapping to insurability criteria and fundamental challenges of cyber insurance is provided. Next, the information requirements of cyber insurers are considered, highlighting the inherent process perspective. Then process mining is contextualised in the premium calculation process. Then process mining methods are systematically mapped to different coverage elements of actual cyber insurance policies by the means of confidence factors to specifically highlught in which scenarios such analyses could be highly relevant.

4.1.1 Mapping Process Mining to Insurability Criteria and Fundamental Challenges of Cyber Insurance

: The goal of the MeritMiner4CI approach is to address the fundamental problem of information asymmetries and the issues of moral hazard, adverse selection, and insurance fraud. As the analysis in [134] concludes, under the condition of reduced information asymmetries (partial information asymmetry) and premium-differentiation, the market

can work efficiently, and incentives for self-protection on the side of the insured can be increased. The following section describes, in line with [53], exactly how the proposed approach addresses the aforementioned issue,

Mapping - Risk Assessment Method: Current methods applied are predominantly of qualitative nature: self-assessments are employed, for larger corporate customers, so-called *underwriting meetings* take place, the insurer might interview different stakeholders at a given company and ask a series of questions. Policies and practices are then reviewed as well [12]. As is further confirmed in the evaluation chapter, cyber security experts believe that quantitative analyses of user behavior are highly important and that process mining is one of the candidate methods that can deliver the process perspective.

Mapping - Limited Historical Data: The use of process mining relies on data that already resides in information systems of prospective insured. Therefore it might offer an additional, previously untapped data source for further modelling. In the case of wider adoption, benchmarks of the performance of different processes could be developed and used for the purposes of improvement recommendations.

Mapping - Standardization: From the perspective of standardisation of cyber risk assessments and more clarity on cyber insurance pricing (providing for transparent premium differentiation), rules that can be checked by process mining methods, which can furthermore be mapped to frameworks such as NIST [2], CIS [3], ISO 27001 [54] as well as to regulatory frameworks such as HIPAA [55], PCI [56], *Sarbanes-Oxley Act* (*SOX*) [57], California Consumer Privacy Act (CCPA) [58], or *GDPR* in Europe [5]. For example, you can refer to[31] for a taxonomy of operational cyber risks and mapping to related controls from the *NIST* framework. Such rules can be formalised used to filter out cases violating a rule expressed in some declarative notation. For example, it might be checked whether 4-eyes principle (see separation of duty constraint below) was violated in the process of electronic transfer if the same employee (expressed as event-level attribute) is associated with both the creation of an electronic payment, as well as with its approval. Finally, the insurer can create rule libraries to provide for re-usability of such rules, further addressing the issue of standardisation. Table 4.1 below provides an overview of different types of verifiable constraints that are of relevance and that can be verified with process mining methods.

Process Security Constraint	**Description**
Authorisation	Only authorized individuals can execute tasks
Usage control	Incl. retention and control of use of data
Separation of duty (SoD)	4-eyes principle, aim is typically to reduce fraud
Binding of duties (BoD)	Associating activities only to certain roles.
Conflict of Interest	Preventing non-compliant flow of inf.
Isolation	Preventing interference of process execution

Table 4.1: Possible Process Security Constraints that can be Checked with Process Mining Methods, identified in [112]

Mapping - Insured Self-Reporting: Process mining, as was established in the background chapter, can achieve more transparency about processes in a given company. In the cyber

insurance context, if the insurer, or the auditor is given access to process mining analyses of a given security-relevant process, the self-reporting issue is expected to be further mitigated and a step towards a single-source of truth is taken.

Mapping - External Security Audit: As was demonstrated in the model developed by [21], external auditing is necessary for a functioning cyber insurance market. Traditional audits are not only costly and time-consuming, but they are also only based on samples and rely typically on largely qualitative assessments. In contrast, process mining provides a promising way to achieve wider audit coverage and potential for automation and standardisation of these audits.

4.1.2 Mapping Examples of Process Mining Approaches to Information Requirements

: Now that the proposed approach has been contextualised in the cyber insurance process. Let us move on to map the information required by underwrites to what process mining methods can deliver. The summary in Figure 4.2 below by [60] presents the topics of interest for the cyber underwriter that are requested in a typical underwriting meeting. A similar notion was confirmed during interviews with underwriters in [12].

Step	**Requested information**	**Example verification approach with process mining**
Operational Overview	Sensitive information (PII, PHI), number of records	Analysis of data flow with discovery
Security	Identity management, Incident response, Data Protection, Disaster Recovery, Vendor Compliance, Business Continuity, Patching, Cyber Security Framework	Discovery and Analysis of Incident Response Process, Conformance Check of Identity Management, Process against Cyber Security Framework, Discovery of Patching Process
Privacy	regulatory compliance, third party access, key stakeholders	Check of Process Compliance, with Regulation (e.g. 4-eyes principle)
Governance	OPTIONAL	Conformance Checking Process Security Constraints
Litigation/Claims Activity	N/A	N/A

Table 4.2: Overall Structure of the Underwriting meeting from [60] Mapped to Process Mining Approaches

It is worthy to note that a large (majority) part of the information in the above table can be mapped to some type of an internal process or procedure. Clearly highlighting the relevance of process perspective in cyber risk assessment. This aspect is also validated in the evaluation chapter. Qualitative descriptions of said processes coming from meetings might be misleading. It might be the case that the reality of security related practices, procedures or *processes* do not match with what is actually happening in the organisation from the security perspective. What if the data protection policy exists, but is not followed? What if employees employ workarounds, to e.g. bypass backup, 2FA, or data retention procedures? Maybe regulatory guidelines are not being followed? Processes such as incident management, or help desk might be defined. However, it also might be the case that they run highly inefficient, or even in a non-compliant way. In summary, what is documented and presented to the insurer might not correspond with the reality and might therefore mean that the insured could make a decision based on fundamentally flawed or biased information. This issue maps well to the potential offered process discovery and conformance checking.

Mapping the Proposed Approach to the Premium Calculation Process in Cyber Insurance: Once the information gathering is over, the underwriter, being under time and

resource constraints, needs to decide if a cyber insurance contract will be offered to the prospect and at what cost. The procedure in Table 4.3 was derived in [12]. First, the type of coverage needs to be considered (step 2). Afterwards, as can be seen in the highlighted steps 9 and 10 in Figure 4.3, if information about cyber security posture is gathered in addition to fundamental characteristics of the insured (such as annual revenue and risk group), it influences the ratings of the confidence factors. Therefore, it is proposed that the approach for cyber insurance based on process mining can integrate well with the current cyber insurance underwriting process by means of using the results of process mining analysis to rate these confidence factors and that confidence factors from underwriting manuals are a fitting method to conduct analysis of cyber insurance requirements for cyber risk assessment.

Steps in a typical premium calculation from [12]
1. Input customer annual revenue - in scope of the thesis
2. Determine overall risk group (0-6)
3. Select applicable coverage - covered by the approach
4. Select applicable base rate
5. Select applicable retention
6. Select applicable limits and sub-limits
7. (Optional) Determine business interruption deductible hours
8. Adjust relevant limit modifiers
9. Coverage specific confidence factors - covered by the approach
10. Enterprise specific confidence factors - covered by the approach
11. Annual premium (business interruption)

Table 4.3: Premium Calculation Process Demonstrated on the Example of Business Interruption Coverage from a Cyber Incident [12] with Steps Covered in MeritMiner4CI highlighted in bold and marked

Confidence factors (CF), that are sometimes referred to as risk modifiers, or risk factors, are typically expressed as a float with which the base rate either is multiplied [12]. An example of a base rate is provided in Figure 4.1 below.

Table I - Insuring Agreement Base Rates

Gross Revenue (in 000s)	Contingent Business Interruption and Extra Expenses Base Rates - Hazard Group						
	0	1	2	3	4	5	6
250 and Under	126	140	175	292	438	584	643
500	210	233	292	487	730	974	1,071
1,000	298	331	415	692	1,037	1,383	1,521
3,000	511	568	711	1,185	1,777	2,369	2,606
5,000	692	769	963	1,604	2,406	3,207	3,528
10,000	1,024	1,138	1,425	2,373	3,559	4,745	5,220
20,000	1,489	1,655	2,072	3,451	5,176	6,901	7,592
35,000	2,011	2,236	2,798	4,662	6,991	9,321	10,253
50,000	2,424	2,694	3,370	5,615	8,419	11,225	12,347
100,000	3,289	3,654	4,570	7,615	11,419	15,225	16,747
500,000	5,729	6,374	7,930	13,215	19,779	26,425	29,067
1,000,000	7,929	8,824	11,030	18,365	27,479	36,725	40,417

Gross Revenue (in 000s)	Network Extortion Threat Base Rates - Hazard Group						
	0	1	2	3	4	5	6
250 and Under	63	70	88	146	219	292	321
500	105	117	146	243	365	487	535
1,000	149	166	207	345	518	691	760
3,000	256	284	355	591	888	1,184	1,302
5,000	347	385	481	801	1,202	1,603	1,763
10,000	514	570	712	1,186	1,779	2,373	2,610
20,000	748	830	1,036	1,726	2,589	3,452	3,797
35,000	1,011	1,121	1,399	2,331	3,497	4,663	5,129
50,000	1,217	1,349	1,686	2,807	4,211	5,616	6,178
100,000	1,647	1,829	2,291	3,807	5,716	7,616	8,378
500,000	2,887	3,189	3,971	6,607	9,916	13,216	14,538
1,000,000	4,037	4,439	5,471	9,157	13,766	18,366	20,188

Figure 4.1: Base Rate Example by Chubb (UM1) [135]

The principle of applying confidence factors to the base rate can be expressed using the following simple mechanism: premium loading (increase) is arrived by multiplying base premium by $CF > 1$, $CF < 1$ then leads to premium discount, $CF = 1$ then means that a given confidence factor does not influence the decision on premium in any material way. The selection of confidence factor is typically done based on underwriter discretion. An example of possible confidence factors and their ranges is presented in Figure 4.2 below.

Modifier Description	Range (Low)		Range (High)
Centralized Policies & Procedures	0.75	to	1.25
Financial Condition	0.75	to	1.25
Nature of Operations	0.75	to	1.25
Network Security	0.75	to	1.25
Physical Security	0.75	to	1.25
Quality of Service Provider Contracts	0.75	to	1.25
Quality of Service Providers	0.75	to	1.25
Risk Management Controls	0.75	to	1.25

Figure 4.2: Enterprise specific confidence factors [135])

Furthermore, confidence-factors might typically (e.g. in the UM1 policy used for the case studies) be divided up to two types, depending on whether they, in the case of *enterprise-specific confidence* factors influence the aggregate premium (more general factors) of a given policy, or only premium for a specific coverage in the case of *coverage-specific* confidence factors. However, it needs to be pointed out that there are differences in whether and how confidence factors are used across underwriting manuals and general mapping is not feasible. Therefore, the manual by Chubb (a globally leading cyber insurer) with SERFF reference number *ACEH-131914766* was used for the development and testing of the proposed approach. The mapping would work analogously for other manuals as well.

The following section proposes a mapping of established confidence factors from [135] to examples of process mining.

4.1.3 Mapping of Process Mining Approaches to Confidence Factors:

Mapping Process Mining Analyses to the Rating of Enterprise-Specific Confidence Factors:

Let's first consider enterprise-specific confidence factors. Figure 4.4 summarises an example mapping of these factors to process mining analyses. As you can see, CFs are often not purely technical in nature and very often relate to internal procedures and processes at a general level.

Enterprise Level Confidence Factors	**Example subject of PM analysis**	**Mapping to PM approach**
Centralized Policies & Procedures	Any business-critical process (e.g. incident management)	Benchmark of subsidiaries against global models with conformance checking
Network Security	IAM process	Check compliance with IAM policy or best practice with conformance checking
Risk Management Controls	Any business-critical process	check of event log against risk management policy, or regulation

Table 4.4: Enterprise Specific CFs Mapped to Process Mining Analyses [135]

Mapping Process Mining Analyses to the Rating of Coverage-Specific Confidence Factors

Next, let's consider the confidence factors that are used to adjust premiums for specific coverage types and explore how they could be rated with the support of process mining methods. In Figure 4.5, a **business interruption coverage** scenario is considered. As you can see, the ratings regarding to business interruption, again, in many cases, refer to procedures that can be easily verified with process mining methods.

Business Interruption CF	**Example subject of PM analysis**	**Mapping to PM approach**
Dependency on real time transactions	Order to cash process	Cycle time calculation on the event log
Mirror/Backup Procedures	backup process, data flow	conformance check of backup procedure, data flow
Risk Management for Incident Response Planning	Incident resolution process	Discovery and conformance check of incident management process
Technology Risk Management Process	Patching / update process	conformance check of patching process
Volatility/Recovery in Sales	Order to cash process	analyse number of cases by time

Table 4.5: Business Interruption CFs from [135] Mapped to Process Mining Analyses

As concerns computer fraud coverage that is the subject of the mapping in Figure 4.6, it is again clearly identified that confidence factors can be mapped to analyses of internal processes, such as of password management. Some CFs such as volatility/recovery in sales can also also be investigated as alternatives to determining base rates based on customer revenues. Interesting scenarios could include using the number of cases and transaction amounts in a process to set process-specific limits.

Computer Fraud CF	**Example subject of PM analysis**	**Mapping to PM approach**
Amount of Online Financial Transactions	Event log of payment system	process discovery, number of case, anomaly detection
Network Access Control	IAM process	Check for anomalies in IAM process
Network Intrusion Detection System	User activity event log	Anomaly detection in user activity
Password Management	Password management policy	Conformance check of policy vs. event log of password changes
Volatility/Recovery in Sales	Order to cash process	Analyse number of cases by time

Table 4.6: Computer Fraud CFs from [135] Mapped to Process Mining Analyses

In Figure 4.7, it is proposed how to rate risk for a **cyber incident response coverage**. Sonsider that the cyber incident response again focuses on procedures, as demonstrated in the case studies - see the data flow example.

Cyber Incident Response Fund CF	**Example subject of PM analysis**	**Mapping to PM approach**
Amount of Sensitive Information	Data Flow in Document Management System	Process Discovery and Statistics on Cases in Different Regions
Encryption	Encryption Procedure	Conformance Checking of Activity Logs from Devices Against Rules

Table 4.7: Cyber Incident Response CFs from [135] Mapped to Process Mining Analyses

Figure 4.8 provides a mapping to of approaches to rate risk for **cyber, privacy, and network security liability** coverage. It is clear that the focus of established confidence factors for that coverage revolves around regulations, procedures, measure, practices and processes. All of which are fitting subjects of process mining analyses.

Cyber, Privacy, and Network Security Liability CF	**Example subject of PM analysis**	**Mapping to PM approach**
Compliance with Privacy Regulations	Data flow in an organisation	Conformance check of GDPR rules
User Interactivity	User flow in ecommerce system	Process discovery
Scope of Privacy Regulations	Data retention process	Conformance checking of retention rules
System Management	Patching / update process	Conformance check of patching process
Data Collection Practices	Data retention process	Conformance checking of retention rules

Table 4.8: Cyber, Privacy, and Network Security Liability from [135] Mapped to Process Mining Analyses)

Digital data recovery liability is the focus of Figure 4.9 and a mapping is provided there for a process mining based approach to rate such coverage.

Digital Data Recovery CF	**Example subject of PM analysis**	**Mapping to PM approach**
Backup/Mirror Procedures	Backup process, data flow	Conformance check of backup procedure, data flow
Disaster Recovery Process	Problem management process	Process discovery, conformance with problem management policy
IR Technology(ies)	Incident management process	Discovery and conformance check of incident management process
System Management	Patching / update process	Conformance check of patching process

Table 4.9: Digital Data Recovery Liability confidence factors from [135] Mapped to Process Mining Analyses

Another type of **coverage against funds transfer fraud** is considered in Figure 4.10 below. CFs such as Electronic Transfer Processing Controls and training that are currently typi-

cally evaluated highly subjectively can be investigated with declarative rule checking and process discovery respectively in quantitative manner, leading to increased transparency.

Funds Transfer Fraud CF	**Example subject of PM analysis**	**Mapping to PM approach**
Banking Systems Authentication	Log of login attempts, penetration of 2FA	Discovery of authentication process
Electronic Transfer Processing Controls	Accounts payable process	Conformance check of 4-eyes principle and approval procedure
Online Banking Access Control	Log of login attempts, penetration of 2FA	Discovery of authentication process
Size and Scope of Electronic Transfers	Event log of banking transactions	Discovery and filtering on attributes
Training and Education	eLearning systems	Discovery of user behavior in cyber security eLearning systems

Table 4.10: Funds Transfer Fraud CFs from [135] Mapped to Process Mining Analyses)

Finally, 4.11 investigates **network extortion liability coverage**, which is concerned predominantly with the issue of ransomware attacks. Here, an approach with process mining is expected to require a lower level of abstraction (in the case of IDS). But CFs such as seasonality of sales can be related to widely-adopted process mining methods in the industry.

Network Extortion CF	**Example subject of PM analysis**	**Mapping to PM approach**
Network Extortion Planning	Network-based intrusion detection with conformance checking	Network-based intrusion detection with conformance checking
Seasonality of sales	O2C process	discovery and breakdown by time
Sensitive Information or Services	Data flow	Conformance check of log against best practices

Table 4.11: Network Extortion CFs from [135] Mapped to Process Mining Analyses

4.2 MeritMiner4CI: Proposed Approach

Now that the cyber risk assessment requirements in cyber insurance have been systematically mapped, this section of the thesis proposes a structured approach to address them that consists of two workflows and an iterative security process enhancement component in order generate value by process mining methods in cyber insurance. A periodic analysis in each underwriting circle is proposed, based on which rating of risk modifiers might be changed as the observed behavior evolves. This aims to support the dynamisation of premiums.

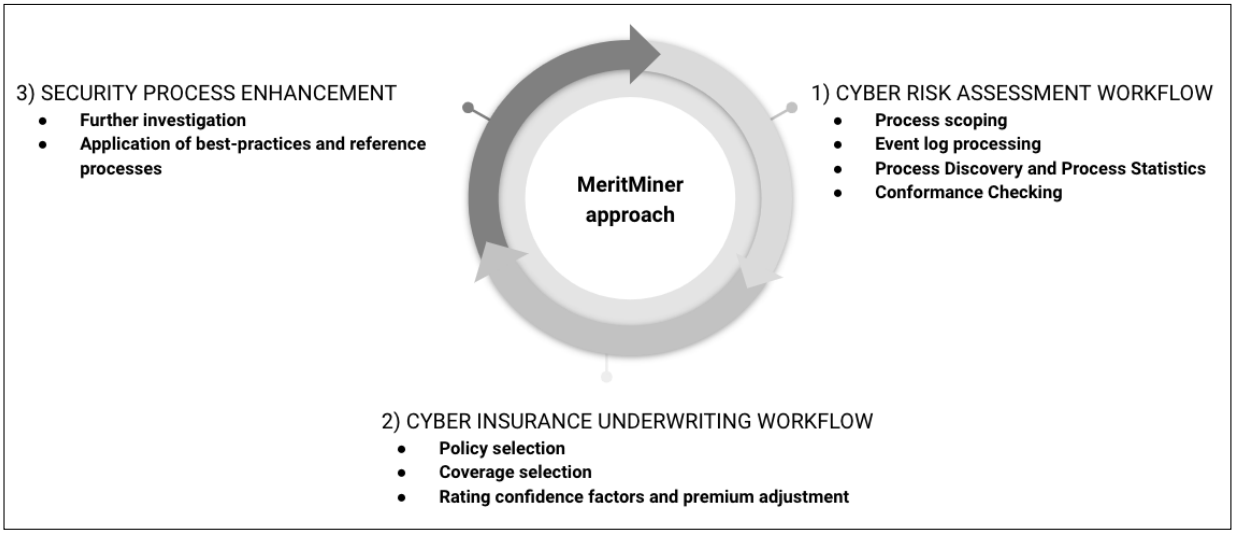

Figure 4.3: High-level overview of the MeritMiner4CI approach

Mapping process mining to the underwriting process: The following mapping to the phases by [18] summarises how the proposed approach maps to each step in the underwriting process. As you can see, all of the three phases are covered.

Cyber Insurance Underwriting Process Step	**Substep**	**Mapping to MeritMiner**
Risk Identification	Threat identification, Security / Vulnerability Identification	Cyber Risk Assessment Workflow
Risk Analysis	Risk estimation	Cyber Risk Assessment Workflow and Cyber Insurance Underwriting Workflow (confidence factor rating)
Establish Contract	Coverage specification Premium Estimation	Cyber Insurance Underwriting Workflow (coverage definition, multiplication of premium with confidence factors)

Table 4.12: Mapping of MeritMiner4CI Approach to[18]

Next, let us investigate what methods specifically are prposed for the applications n cyber insurance. Table 4.13 shows how each of the phases in the cycle map to established process mining methodology by [79] and to each of the established methodological components of process mining that are organised in an iterative fashion.

Process Mining Method	**Mapping to MeritMiner4CI**
Discovery	Cyber Risk Assessment Workflow - Process Discovery
Conformance	Cyber Risk Assessment Workflow - Conformance Checking
Enhancement	Cyber Risk Assessment Workflow - Application of best-practices and reference processes Cyber Insurance Underwriting Workflow - Iterative adjustment of confidence factors

Table 4.13: Alignment of Process Mining Methods by [79] with MeritMiner4CI and it's components

4.2.1 Cyber Risk Assessment Workflow

Cyber Risk Assessment Workflow encompasses the actual steps of the analysis required to generate insights required for ratings in the underwriting workflow. Next, the specific steps are discussed.

Step 1: Process scoping: The scoping of processes is a critical step, arguably the most important one. The foundation for it was laid in the section that mapped process mining

methods to cyber insurance coverage. While most works reviewed in Chapter 3 focus either on conformance checking (e.g. for anomaly detection purposes), or on process performance analysis **separately**, the proposed MeritMiner4CI approach integrates both of these perspectives as depicted as two pillars in Figure 4.4.

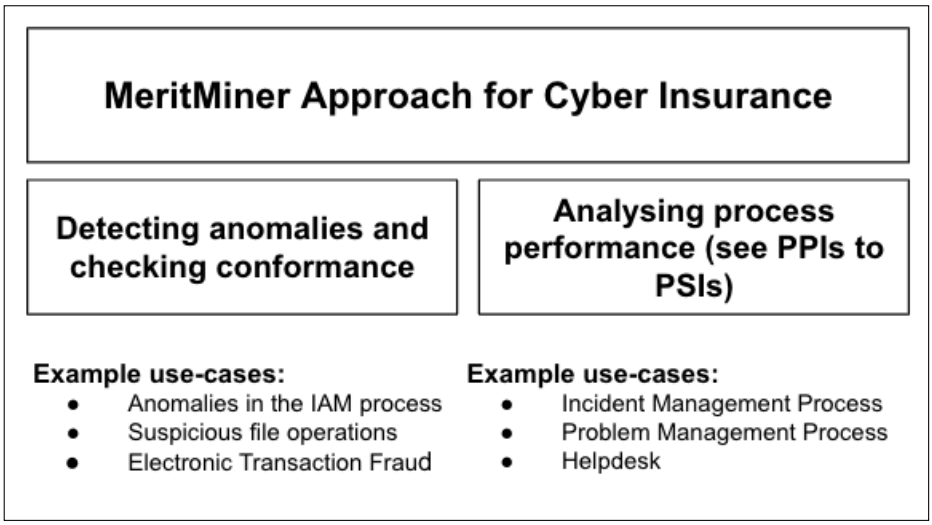

Figure 4.4: Scoping of processes of of the MeritMiner4CI Approach

Step 2: Event Log Processing: Refers to the step of obtaining and processing of event log, which is a necessary step for any further analyses.

Step 3: Process Discovery and Process Statistics: In this step, a process discovery is applied on the event log extracted from the prospective insured's system. In this stage, a preliminary visual analysis of the process flows is conducted and values of metrics (PPIs - process performance indicators) are generated and evaluated. Subject to data limitations given by privacy and compliance regulation, it is proposed that these metrics are collected in a library maintained by the insurer and later used for development of bench-marking mechanisms that can be applied to drive recommendations and process enhancement.

Step 4: Conformance Checking: Conformance checking can be applied either to compare the event log to a predefined model (expressed either as set of rules, or as a process flow diagram). reflecting some type established practice, policy, or regulation. Conformance checking of event log, for example, according to organisational units can also be applied to understand which parts of the organisation exhibit more violations to a globally defined practice. Finally, given correct mapping, reference models defining some standard to be followed can be compared against actual execution of the security-relevant process. Table 4.14 provides details on exactly which methods can be applied for conformance checking in the Cyber Risk Assessment Workflow.

4.2.2 Cyber Insurance Underwriting Workflow

In the second of the proposed workflows, the steps correspond to the mapping discussed in detail in the mapping. In summary, the applicable policy is selected in **Step 1**, enterprise-level risk modifiers are rated in **Step 2**, coverage is selected in **Step 3** and finally in **Step 4**,

MeritMiner4CI conformance checking step	**Conformance checking method**	**Example Mapping to PM approach**
Step 1: Visual analysis	Visual analysis of discovered model	Check of data flow in organisation
Step 2: Rule checking	Rule checking	Checking whether policy regarding sensitive information is followed
Step 3a: Process model/event log comparison	Token-based replay	Replay event log of IAM process against IAM model defined in policy
Step 3b: Process model/event log comparison	Alignments	Diagnose complex incident management process

Table 4.14: Mapping of Conformance Checking in MeritMiner4CI to Conformance Checking Methods by [84]

the confidence factors for each coverage element are rated. These steps are conducted by the underwriter, or responsible analyst, who makes decisions on the rating of confidence factors based on discretion. It expected the rating step might be automated in the future, but given the novel nature of the approach and lack of experience with proposed metrics, the manual rating is believed to integrate better with the current cyber insurance processes discussed above.

4.2.3 Security Process Enhancement

As the Figure 4.3 shows the MeritMiner4CI is iterative and once an application has been either accepted, or rejected, feedback is provided (on why a certain adjustment to the premium was made, or why an application was rejected) to the prospective insured outlining what the improvement potential is. Recommendations on best practices, as well as on, e.g., protection services might be provided. That is in line, e.g. with [77]. If a contract is granted, a continuous monitoring might be instituted, so that improvements might be considered in the next underwriting cycle (which is typically yearly).

4.2.4 Summary of the Logic of the Novel Underwriting Process in BPMN 2.0

Before designing and implementing the prototype itself, it was necessary to design the underlying process that the underwriters would apply when leveraging MeritMiner4CI (from the methodology perspective). The process was designed and implemented in a transparent way using the BPMN 2.0 notation defining the actors, activities, systems and artefacts involved and is demonstrated in Figure 4.5

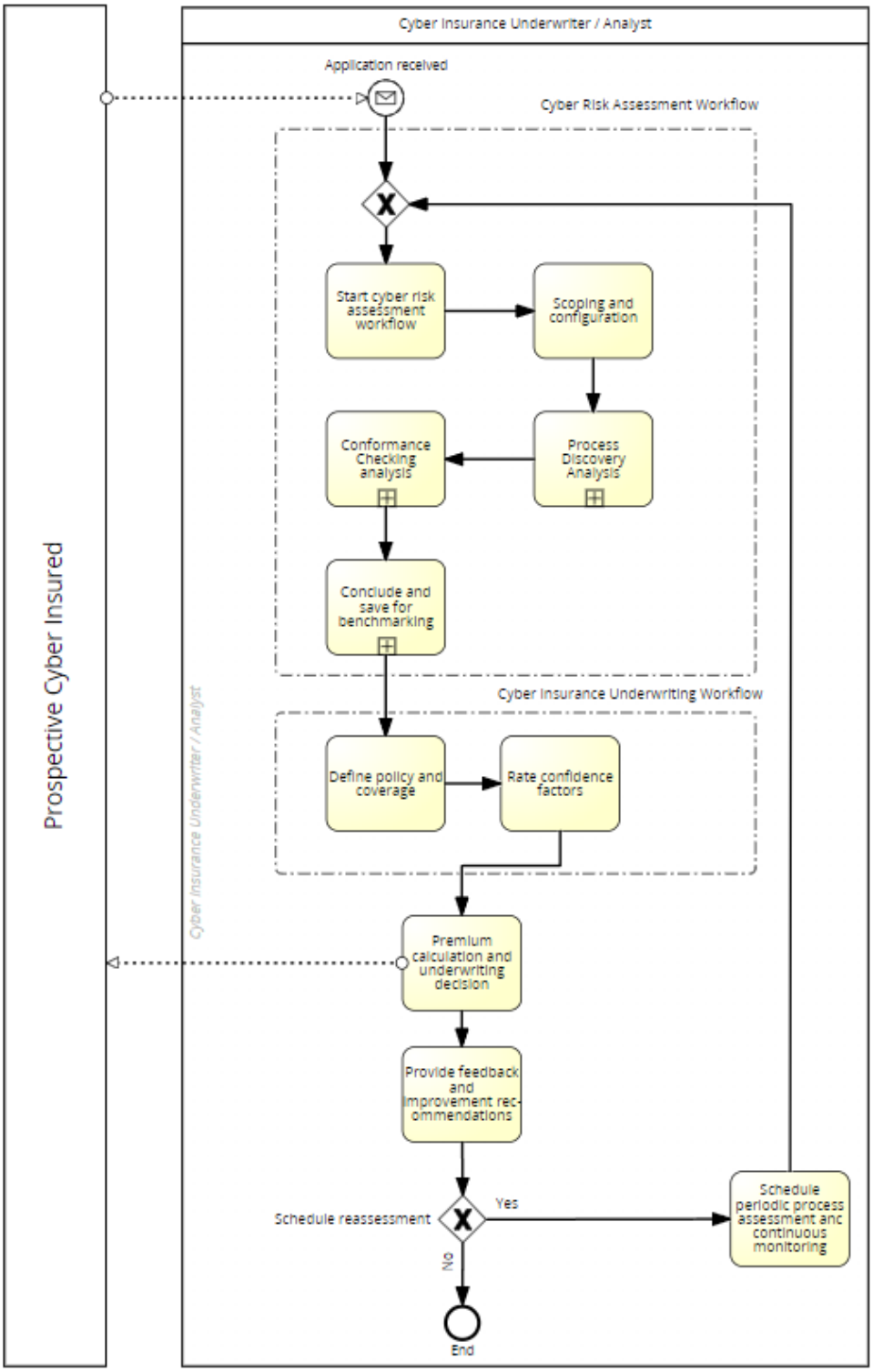

Figure 4.5: BPMN 2.0 Visualisation of the Proposed MeritMiner4CI Approach

Chapter 5

Prototype and Implementation

In the previous chapter, the approach and a methodology for the application of process mining in the cyber insurance context have been proposed. Building on that foundation, the next chapter describes the approach from the preceding chapter that can be implemented in the prototype version of MeritMiner4CI.

It is important to outline the goals of the prototype first. Its main aim is to demonstrate the flow of steps described in the MeritMiner4Ci approach, *i.e.,* demonstrate how process mining could be applied for cyber insurance. The goal is not to provide functionality for deep-dive analyses since there are also many challenges related to creating a data-intensive application. A more suitable approach would have been to exclude the user interface development altogether if that was the main goal of the thesis. The prototype does not consider all configuration, filtering, and production process mining product complexities. Rather, it demonstrates the most fundamental methods of process mining in the form of a cyber risk assessment application with underwriting functionality providing for rating of confidence factors. At the same time, it's structure allows for extension with additional functionality in the future, especially as concerns the depth of analyses as it covers all three layers of the architecture below- the *user layer*, the *business layer* and the *data layer*. The following chapter is structured according to these layers.

5.1 High-level Solution Architecture

Figure 5.1 illustrates the reference architecture that was designed. Each of the components could be implemented in different ways. Especially the process mining functionality is highly complex and therefore, taking advantage of some of the products available on the market is advisable, some of which were described in the background chapter.

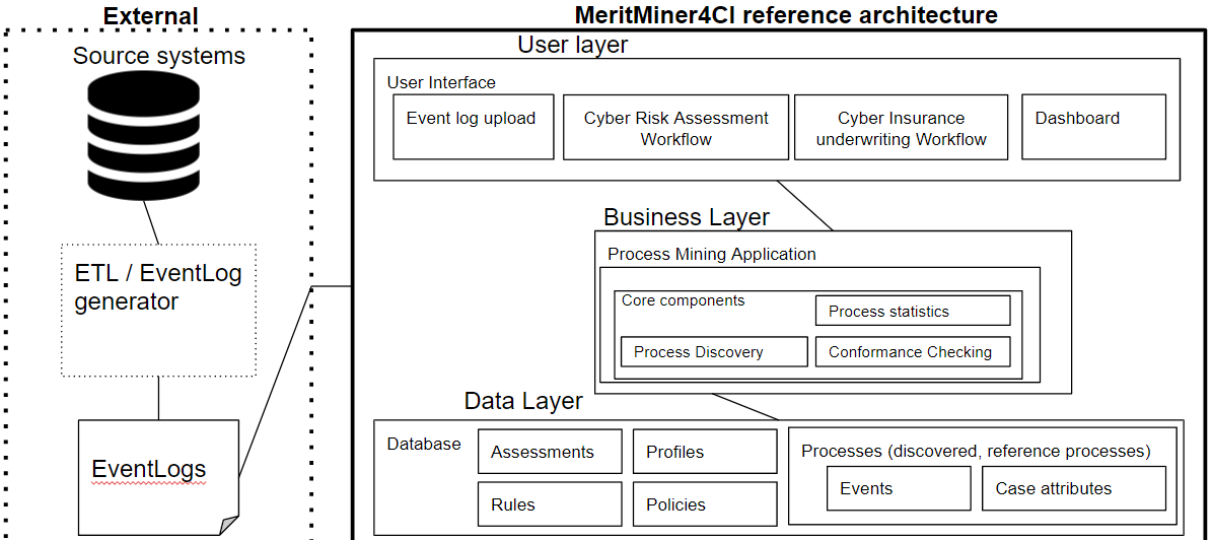

Figure 5.1: Proposed Reference Architecture of the MeritMiner4CI Prototype

5.1.1 User Layer

The user layer is, from the perspective of this thesis, the central component, given its goals. The two main components of the MeritMiner4Ci approach are reflected in the prototype design as follows. (1) The Cyber Assessment Workflow (profile, event logs, discovery and conformance pages) and (2) the Cyber Insurance Underwriting Workflow (underwriting page), the process enhancement aspect lays in the periodic application of the approach in each underwriting cycle and monitoring (represented by a minimalistic dashboard).

The **user layer** can be observed from two perspectives, design and implementation. The **prototype design** was one of the most challenging steps in turning the MeritMiner4CI to a MVP product. This was predicated by the fact that process mining analyses can be prohibitively complex and often include manual steps (including. graph interpretation, manual filtering, conformance analyses against models created manually) that are hard to automate. Furthermore, the overall number of metrics that could be used is high and different metrics might be relevant for different processes. A decision was made to focus on the fundamentals.

Once the overall logic of the workflow was defined, the first iteration of the prototype was developed in the form of a low-fidelity prototype, using the industry-standard prototyping tool Figma [136], which allows to create clickable prototypes that can be tested before actually implementing the underlying logic, enabling an iterative approach to prototype development.

In regards to the implementation of the **user layer** (frontend). The user interface is based on SecRiskAI; an application developed at the CSG by Erion Sula. [17]. This choice was made for the following main reasons. First, this thesis proposes that process mining can be used in conjunction with a number of other methods for cyber risk assessment in the underwriting process and SecRiskAI focuses on external attack predictions, which are

delivered by machine learning models[17]. SecRiskAI. which is available open-source and from it's components, the "frontend" was used as a starting point. The integration with SecRiskAI concerns mainly the user layer (or frontend in SecRiskAI). For this reason, some of the decisions regarding the User Layer were inherited from SecRiskAI. Cyber risk analyst can investigate both the risks from internal processes, as well as access risk assessments on external attacks from SecRiskAI and make use of it's integration with MENTOR API to get protection recommendation. The overall value of the combined application is then more than that of the sum of it's components.

For building the user interface, SecRiskAI and for that matter MeritMiner4CI takes advantage of the React [137] JavaScript library. For example, the frontend of SecRiskAI was bootsrapped with Create React App which, using the following command *npx create-react-app my-app –typescript* [138], can generate the application that uses TypeScript as the default JavaScript syntax that includes the required files and folders, as well as the fundamental configuration required to run the application in a web browser. The main arguments that speak for the usage of react are scalability and flexibility predicated on it's modular nature and on the usage of reusable components [17]. Next, the two logical components of the user layer are discussed, and the designed user flow in the prototype is introduced.

Cyber Risk Analysis Workflow

The Cyber Risk Analysis Workflow from the proposed approach manifests itself in the prototype in the following steps, implemented as separate tabs in the prototype from the perspective of the user layer.

- Profile setup - shared SecRiskAI component
- Event log upload (incl. assignment of event log to the relevant process)
- Discovery (covering also visualisation, statistics and upload of BPMN [139] models)
- Conformance checking

The profile page covers the fundamental information that the insurer would request (no. of employees, revenue, industry). In MeritMiner4CI, this information is therefore taken over to the data layer and persisted.

On the *event logs* page, this profile is then selected representing the prospective insured for which a risk assessment will be created is selected. The underwriter/analyst is also able to upload a pre-processed event log in the XES format, which is a standard in the process mining domain as for event logs [140]. In the background, the event log is persisted in the database and has it available for further analyses. It is also possible to do basic management of event logs directly in the tool (create, delete, preview, select for analysis), as well as map the most important fields for process mining (Case ID, timestamp and event name) to the columns in the event log. shows how an event log can be associated to a business process to be analysed. The limitation of the prototype is the size of event

logs and they need to be provided in the correct structure (in line with XES standard, but csv files can be uploaded as well). Also, in a real-life scenario, some type of ETL pipeline might need to be employed, and event log generation is a significant challenge of it's own. The prototype therefore assumes a pre-existing event log.

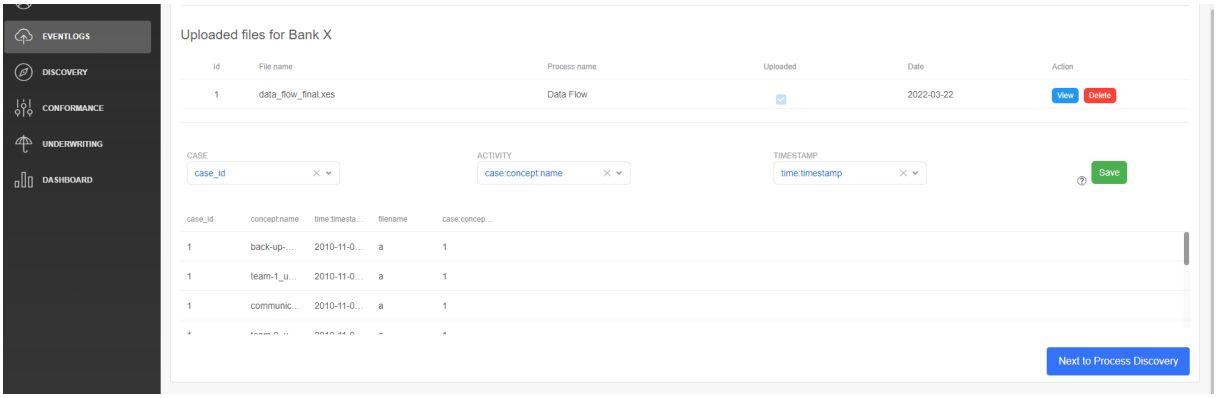

Figure 5.2: Preview of the Event Log Upload Page

Figure 5.3 shows how an event log can be associated to a business process to be analysed.

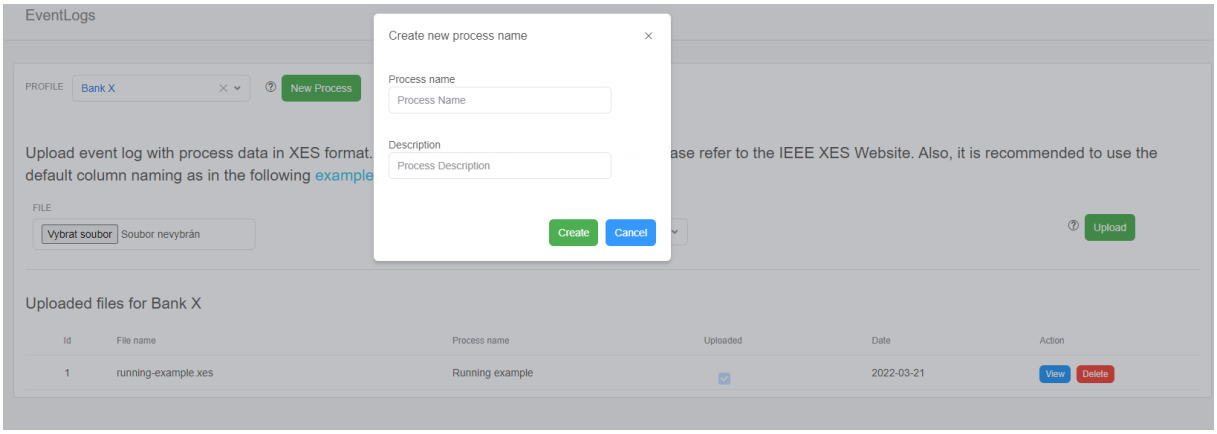

Figure 5.3: Associating an Event Log to a Process

Process Discovery starts with selecting an uploaded event log, users of the tool can run process discovery analysis on the log. Multiple different algorithms (heuristic miner, alpha miner and inductive miner) are available for choice, as well as multiple resulting notations. From the perspective of the user layer, it is necessary to be able to display visualisations (for example process trees as well as well as petri nets, or heuristic nets) outputted by the *pm4py.visualization* package in the business layer. These are critical for visual analysis of the process flows. The example below shows how the user layer requests Graphviz data about of a process tree discovered from a given event log by the inductive miner algorithm.

```
1
2  const getGraphvizData = (id: any) => {
3     const body: any = {
4       eventlogId: eventlogId | 0,
5       algorithm,
6       case: caseParameter,
7       activity: activityParameter,
8       timestamp: timestampParameter,
9       id: id ? id : modelId,
10    };
11
12    ...
13    if (algorithm === "inductive_miner" && processTree) {
14      body.processTree = processTree;
15    }
16    ...
17    setTimeout(
18      () => {
19        fetch(Endpoints.discovery + "gviz/", {
20          method: "POST",
21          body: JSON.stringify(body),
22        })
23          .then((response) => response.json())
24          .then((data) => {
25            console.log("data", data);
26            if (data) {
27              setGraphvizData(data);
28              setError(false);
29            } else {
30              setGraphvizData(undefined);
31              setError(true);
32            }
33          })
34          .catch((err) => {
35            console.error(err);
36            setError(true);
37            setGraphvizData(undefined);
38          });
39      },
40
41      500
42    );
43  };
```

Listing 5.1: Example (shortened) of Requesting GraphViz data from the Discovery Service. The Fetch library is used

The discovery service in the business layer uses the pm4py.visualization package (both discovery and conformance services) is a graph object in *viz* format. Therefore, in order to render GraphViz data in the user layer, the the graphviz-react [141] library was used as can be seen separate TypeScipt component *GraphvizAlgorithm.tsx*.

As concerns displaying process **statistics**, the most fundamental values are displayed, including **rework rates**, **fitness**, **number of cases** , **events** and **median cycle time**. These metrics represent the performance analysis aspect, relevant e.g. for incident management and help-desk process. 5.1.1 shows, how the Fetch library is used to request statistics related to a selected event log from the business layer (Flask API).

```
const getStatisticsData = () => {
   if (!eventlogId) return;

   const url = new URL(Endpoints.discovery + "statistics/");
   const urlSearchParams = new URLSearchParams();
   urlSearchParams.set("eventlogId", eventlogId.toString());
   urlSearchParams.set("case", caseValue);
   urlSearchParams.set("activity", activity);
   urlSearchParams.set("timestamp", timestamp);
   url.search = urlSearchParams.toString();

   fetch(url.href)
     .then((response) => {
       response.json().then((data) => {
         setStatisticsData(data["statistics"]);
         if (data["statistics"][7]) {
           setNodes(JSON.parse(data["statistics"][7]["value"])["nodes"
]);
           setEdges(JSON.parse(data["statistics"][7]["value"])["edges"
]);
         } else {
           setNodes([]);
           setEdges([]);
         }
       });
     })
     .catch((err) => console.error(err));
};
```

Listing 5.2: Example showing how process statistics are requested from the corresponding endpoint in the business layer

Fetching Statistics Data to the User Layer Using the Fetch JavaScript Interface

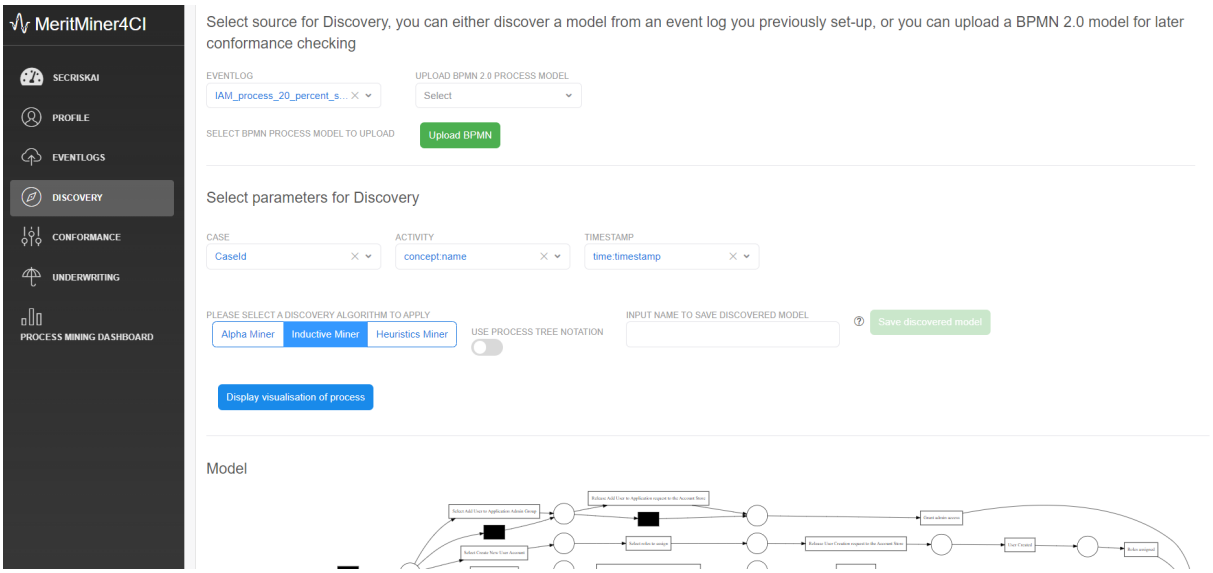

Figure 5.4: Preview of the Discovery page

Conformance Checking Method	PM4Py implementation
Declarative conformance checking / LTL filtering	pm4py.algo.filtering.log.ltl and pm4py.algo.filtering.pandas.ltl
Token-Based Replay	pm4py.algo.conformance.tokenreplay
Alignments	pm4py.algo.conformance.alignments

Table 5.1: Conformance Checking Methods and Corresponding Implementations in PM4Py [98]

Another key feature of process discovery is the upload of reference processes - these can be manually modelled: formal models representing policies, best practices, or reference processes from frameworks discussed in the previous chapter. These processes can also be visualized and used for conformance checking purposes, discussed in detail below. The processes need to be imported in the BPMN 2.0 format on the process discovery page (representing manual process discovery). Also, an additional feature available for event logs containing the attribute "resource" - e.g. an employee. The package pm4py.algo.organizational_mining.sna (implemented in the business layer) is used to generate the handover of work graph is later displayed on the discovery page that the risk analyst can use to understand the social relations in an organisation (e.g. for fraud detection purposes). This represents the organisational perspective discussed previously and can be seen as a form of social network analysis [98].

The final page in the Cyber Risk Analysis workflow is concerned with **conformance checking**. It covers **three steps**. First, a **visualisation** in the form of different types of graphs can be displayed, this represents the simplest conceivable form of conformance checking - visual analysis. Next, event log can be filtered with LTL (Linear Temporal Logic) rules [98]. Using these, the user can run a **declarative conformance check** on the event log can define different rules to check the discovered model against (application of these rules results in filtering out the cases that satisfy the rule that can be applied to generate a subset of the event log). This filtered event log is persisted and can be reused. As an example, one of the implemented rules is the *A eventually B*. The application of this rule is, for example, checking whether a mandatory step, given a preceding step, was executed [79] . For example, a mandatory password change, approval step, or a backup procedure. Rules can also be persisted. Step 3 is then **conformance checking both with token based replay and alignments**, which can effectively compare the event log to a pre-defined model, or a model derived from another log using the *fitness* metric (how much of the behavior in the log can be explained by a given model) [79]. Fitness on the trace level is visualised in two tables, a feature allowing deeper investigation.

Table 5.1 summarises the methods (implemented in the business layer) conformance checking that are available in the user layer.

Cyber Insurance Underwriting Workflow

Once the Cyber Risk Assessment Workflow is finished, the underwriter can move on to the the Cyber Insurance Underwriting Workflow, which is represented by the **underwriting**

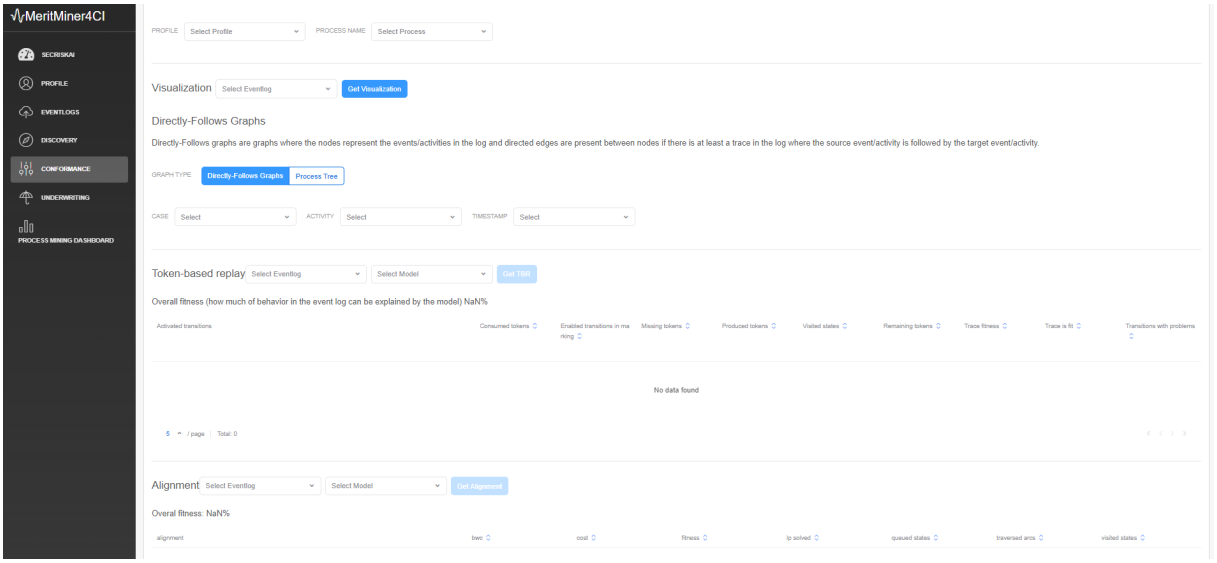

Figure 5.5: Preview of the Conformance Checking Page

page in the user layer. There we can create a minimalistic underwriting assessment based on the preceding process mining analysis as outlined in Chapter 4. To do that, we define customer name and a name of the assessment and also the relevant coverage elements. On both the policy and coverage levels, we can define confidence factors and rate risks. The risk rating is done manually, based on expert assessment of previous analysis. Forr this assessment we do the following. A collected sample of underwriting manual is provided in the repository for reference regarding risk modifiers.

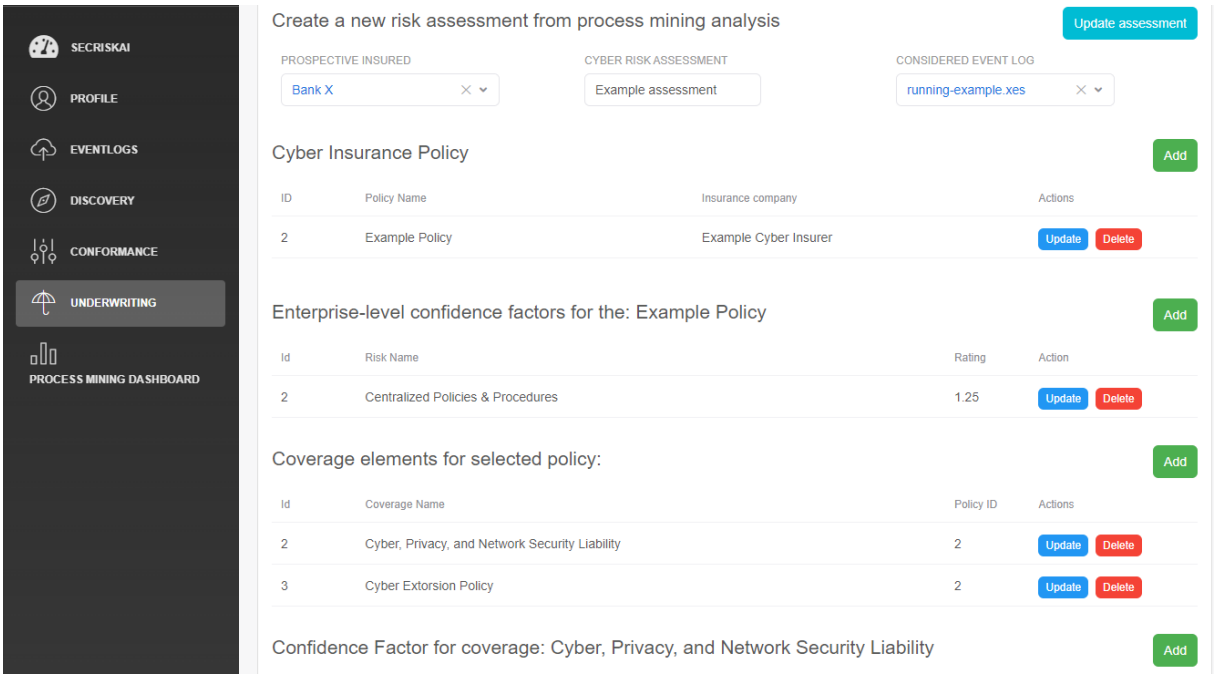

Figure 5.6: Preview of the Underwriting Page

The final step in the prototype is currently a minimalistic dashboard that summarises all the analyses conducted to a certain organisation. There, fundamental metrics are

displayed, and the user can get an overview of the ratings related to each process that was created. This step aims to demonstrate that **continuous monitoring and process enhancement** might be instituted, for example in an annual underwriting cycle. Ideally, improvement recommendations might be delivered based, as well as process benchmarks. Implementation of improvements then should positively influence the premium by lowering risk modifiers. While protection recommendations are covered by [77] on the SecRiskAI dashboard.

5.1.2 Business Layer

The next section briefly summarises the implementation of the business layer. As the fundamental process mining library PM4Py used in the prototype leverages the python programming language.. The decision was made to follow that choice and implement the back-end as a Flask application decoupled from the *nest.js* middle-ware and other building blocks of SecRiskAI, other than the frontend.

The choice of the Flask 2.0 'micro' web development framework [142] was made out of two shortlisted options - either Flask, or Django [143], given the decision to use Python at the back-end. As [100] points out, both framework are mature, well-documented and production read. However, each of the choices has it's advantages and disadvantages as outlined in Table 5.2 that was constructed based on [100]. The decision therefore needs to follow the requirements defined for the prototype. Given the fact that the MeritMiner4CI is a smaller-scale prototype application, prioritising fast prototyping and that an increased flexibility was beneficial.

Framework	**Advantages**	**Disadvantages**
Flask	Best fit for prototyping Quick setup Flexibility	More challenging management and maintenance
Django	Best fit for large-scale projects Batteries-included approach (if features required)	Steeper learning curve Batteries-included approach (possible overhead)

Table 5.2: Comparison of Candidate Frameworks [100]

Flask is a flexible micro-framework [144] and takes advantage of extensions for additional functionality such as database connectivity or building REST APIs. During the implementation of the prototype, multiple extensions were used, including Flask-SQLAlchemy, flask-restx and flaskaccepts. The following section briefly outlines their usage in the prototype. To support the development of the REST API, the Flask-RESTX [145] extension was used as it provides a number of features to support API development. However, the key reasoning behind the choice is that it also provides guidance on structuring more complex APIs by splitting them into reusable namespaces, which is a context that the prototype applies. Additional features include request parsing and, importantly, automated API documentation using Swagger (which is available at http://127.0.0.1:8000/ for the

prototype at hand. Flask-RESTX is a fork of Flask-RESTPlus [146]. `flask_accepts` [147] is a flask extension that was used in the prototype to make it easier to validate inputs and outputs in Flask. This is achieved by providing the following two decorators. 1) the @accepts decorator defines what parameters or schemas is accepted by the endpoint and 2) @responds, which defines how the output should be serialised, as defined in the associated Marshmallow [148] schemas, which the prototype takes advantage of, as reflected in the schema.py files associated with each endpoint.

General Structure of the API

All of the API endpoints and the schemas used can be found in the Swagger documentation. The following section briefly outlines the general structure of the Flask-REST API and provides examples of each component.

Routes and Controllers

In the context of the prototype, @api.route() is a Python decorator that is used by the *Flask* framework that provides us with a way to bind a function ('*controller action*') to a URL. Generally, decorators are used to extend the functionality of a certain functions without making modifications to it [145]. Once a resource is requested (by default, a route only answers to HTTP GET requests), Flask makes an attempt to find a route that matches that resource, and if that resource is found, the associated function is called. This is done in order to provide an abstraction layer on top of the implementation logic. Code snippet 5.1.2 provides an example of an API route that defines the endpoint for policies.

```
1
2  @api.route("policies/")
3  class Policies(Resource):
4      """Policy"""
5
6      @api.expect(policy_filter)
7      @responds(schema=PolicyInfoSchema(many=True), api=api)
8      def get(self):
9          """Get all Policies or filter by assessment id"""
10
11         underwriting_id = policy_filter.parse_args().get("underwritingId
       ")
12         return DashboardService.get_policies(underwriting_id)
13
14     @accepts(schema=PolicySchema, api=api)
15     @responds(schema=PolicyInfoSchema, api=api)
16     def post(self):
17         """Create Policy"""
18
19         return DashboardService.post_policy(request.parsed_obj)
```

Listing 5.3: Example of a controller implemented

One of the key parts of the business layer is the discovery service that contains the three process mining algorithms in scope of the prototype. Table 5.3 lists the algorithms in

Discovery Algorithm	PM4Py implementation used
Alpha Miner	pm4py.algo.discovery.alpha
Inductive Miner	pm4py.algo.discovery.inductive
Heuristic Miner	pm4py.algo.discovery.heuristics

Table 5.3: Used Implementations of Discovery Algorithms from PM4Py [98]

scope, together with their respective implementation. 5.1.2 then presents the implementation of the Discovery Service in the business layer as an example controller action.

The evaluation and discussion on the applicability of different algorithms for different use cases in provided in the quantitative evaluation section in 6.

```
1
2  def process_discovery(
3      log: EventLog, activity: str, algorithm: str = "alpha_miner"
4  ) -> Tuple[PetriNet, Marking, Marking]:
5
6      if "alpha_miner" == algorithm:
7          if activity:
8              parameters = {
9                  alpha_miner.Variants.ALPHA_VERSION_CLASSIC.value.
       Parameters.ACTIVITY_KEY: activity
10             }
11         else:
12             parameters = {}
13         net, initial_marking, final_marking = alpha_miner.apply(
14             log, parameters=parameters
15         )
16     elif "inductive_miner" == algorithm:
17         if activity:
18             parameters = {
19                 inductive_miner.Variants.IMd.value.Parameters.
       ACTIVITY_KEY: activity
20             }
21         else:
22             parameters = {}
23         net, initial_marking, final_marking = inductive_miner.apply(
24             log, parameters=parameters
25         )
26     elif "heuristics_miner" == algorithm:
27         if activity:
28             parameters = {
29                 heuristics_miner.Variants.CLASSIC.value.Parameters.
       ACTIVITY_KEY: activity,
30             }
31         else:
32             parameters = {}
33         net, initial_marking, final_marking = heuristics_miner.apply(
34             log, parameters=parameters
35         )
36
37     return net, initial_marking, final_marking
```

Listing 5.4: Example of a controller action from the DiscoveryService handling the

generation of petri nets

As mentioned above, **controller actions** then correspond to the underlying services in the business layer that are exposed through @api.route(). Figure 5.1.2 below corresponds to the implementation of the alignment functionality. Note that both an uploaded BPMN model converted to a petri net, or a discovered process can be used as conformance artefacts. Fitness and the alignments of traces are evaluated with the functionality from the used process mining package PM4Py.

```
1
2 @staticmethod
3     def get_alignments(params: Dict[str, str]):
4
5         el = EventlogModel.query.get(params["eventlogId"])
6
7         log = xes_importer(el.file)
8
9         if params.get("modelId"):
10             discovery = Discovery.query.get(params["modelId"])
11             if discovery.file_type == "bpmn":
12                 bpmn_graph = bpnm_importer(discovery.file)
13                 net, im, fm = bpmn_converter.apply(bpmn_graph)
14             elif discovery.file_type == "pnml":
15                 net, im, fm = import_petri_from_string(discovery.file)
16
17         aligned_traces = alignments.apply_log(log, net, im, fm)
18
19         fitness = replay_fitness_evaluator.apply(
20             log, net, im, fm, variant=replay_fitness_evaluator.Variants.
    ALIGNMENT_BASED
21         )
22
23         return {
24             "data": json.dumps(aligned_traces),
25             "fitness": fitness.get("percFitTraces"),
26         }
```

Listing 5.5: Example of a Controller Action from the ConformanceService Used for the Alignments Functionality

5.2 Data Layer (persistence)

For the implementation of the data layer the relational Postgres database was chosen [149]. The reason for choosing a relational database, was the fact that event logs are most typically structured as they come from transactional systems. The open-source nature of Postgres, and the size of the community and good documentation were the main contributors for the choice. An alternative, equally feasible solution might have been e.g. a No-SQL database, such as MongoDB. Arguably, such setup might be beneficial in systems with requirements of high scalability with and for sparsely populated, large event logs. But this was not a requirement defined for the prototype. As concerns the

connection between the data layer and the business layer, Flask-SQLAlchemy [150] adds the support for the SQLAlchemy [151] Python SQL toolkit and Object Relational Mapper. The extension was used in order to provide for a simple way to interact with the Postgres database in the *Data Layer* and manipulate database tables using Python classes, objects, and functions. As a database adapter for Python, psycopg2 [152] was applied. Finally, Figure 5.7 provides an overview of the database schema used.

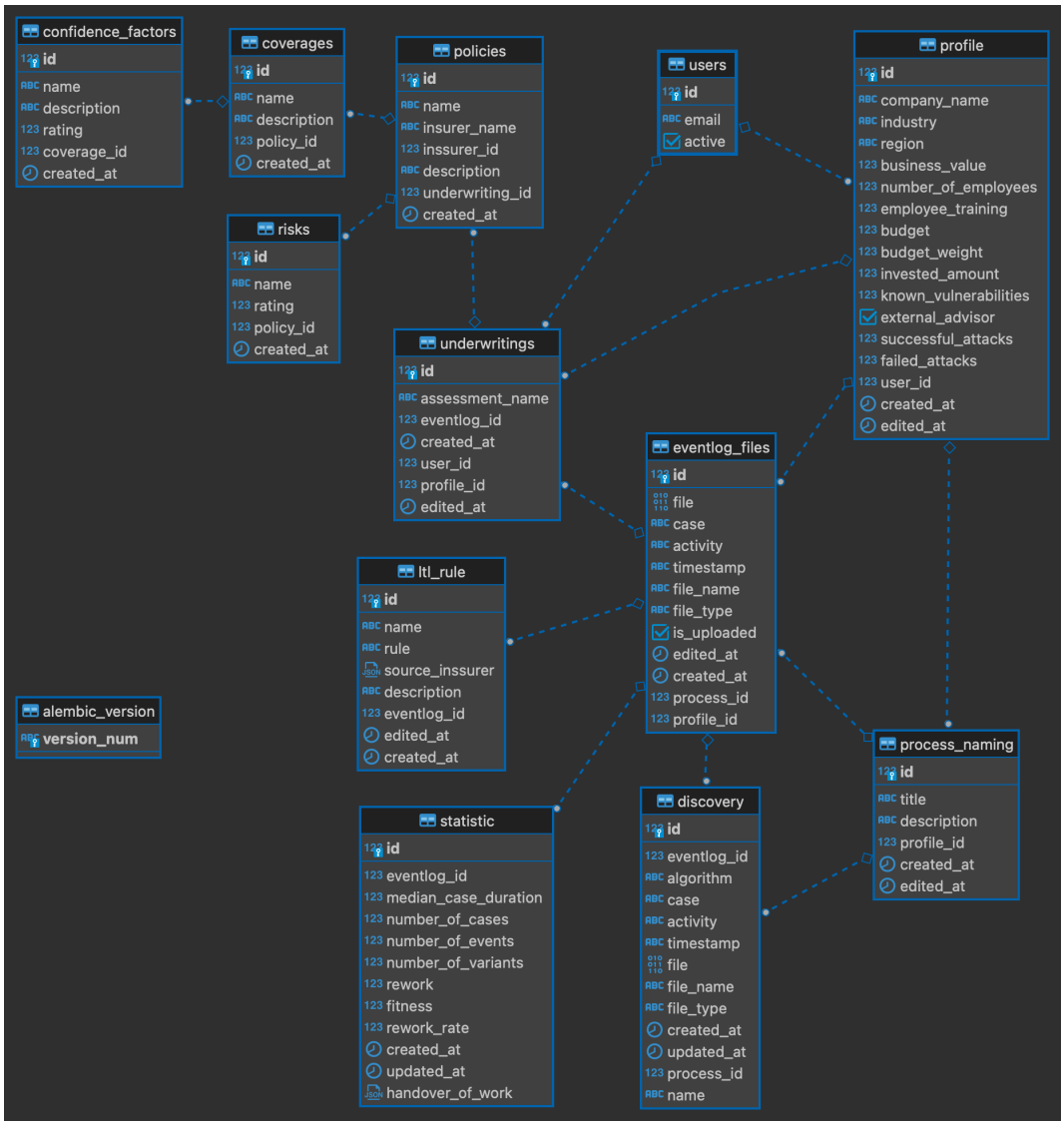

Figure 5.7: Database Schema Used in the Data Layer

Chapter 6

Evaluation

An evaluation of the proposed MeritMiner4CI approach was conducted from three different perspectives. (1) Qualitative evaluation using a survey with 12 cyber security and cyber insurance experts with case study scenarios. With 7 of these participants, interviews were conducted following the questionnaire. (2) Qualitative evaluation of the prototype using scenarios from (1). Finally, (3) quantitative evaluation using security event logs was conducted in order to reflect on which methods might fit which use-cases.

6.1 Questionnaire and Interviews

The following goals were defined for the survey and interviews that correspond to the sections in the questionnaire:

- Evaluate the relevance of data-driven risk assessments and of business process perspective in cyber risk assessment
- Evaluate process discovery, visual analysis of process flows, and of conformance checking via rules from the perspective of cyber risk assessment and evaluate what impact the insight generated by these methods would have on cyber insurance premiums
- Evaluate if conformance checking results influence cyber risk assessments and cyber insurance premiums
- Evaluate how performance analysis of a security-relevant process can influence cyber risk assessment
- Gain additional insights from brief explorative interviews with selected survey participants

In order to allow for the timeline of the survey, the figures used in the surveys were first generated in Jupyter Notebook in an earlier stage of the thesis, rather than implemented in the user interface. However this should have not impact on the validity, as the impact of metrics was in focus and not the evaluation of their implementation.

6.1.1 Selection of Participants

Table 6.1 presented below shows that all of the participants hold roles in the cyber security domain. This is given by the fact that cyber risk expertise of some was the criteria for inclusion of participants in the evaluation. However, while some cyber underwriters took part in the survey as well, it would not be feasible to conduct a survey with underwriters only as the sample would be too low. Already the recruitment of cyber security experts is highly challenging, due to limited number of potential participants. The fact that the group of participants was more diverse does not constitute a large problem because cyber risk assessment for insurance purposes might be conducted by different stakeholders, including external providers, brokers, consultants etc. Participants were recruited from professional networks, via personal network and at the CSG. And came from different regions, including Switzerland, Denmark, the US, the UK, and one participant from India. A complete list of participants, as well as recordings of interviews are available upon request. Survey data is included in the repository.

6.2 Evaluation of Questionnaire Responses

Next, the questions presented to the participants are presented and their answers evaluated. Relevant insights from the interviews are included directly next to the topic related to each survey question and interpretation and short discussion provided.

Introductory Part of the Questionnaire

Question 1: Please select an option that best describes your role and area of expertise

Options offered for question 1:

- **Underwriter Cyber**
- **Underwriter (other P&C)**
- **(Risk) Analyst**
- **Actuarial / Risk Consultant**
- **Insurance Broker / Agent**
- **Reinsurance Underwriter**
- **Product Manager / Specialist / other roles in Product development**
- **Other Cyber Security Specialist / Expert**
- **Other:** Please input the role

Table 6.1 lists how the participants of the survey identified themselves professionally in the first question.

Participant #	**Date**	**Role**
1	2/18/2022	PhD Student on Cyber Security
2	2/18/2022	Researcher Cyber Security Related
3	2/23/2022	Other Cyber Security Specialist / Expert
4	2/25/2022	(Risk) Analyst
5	2/25/2022	Other Cyber Security Specialist / Expert
6	3/1/2022	Actuarial / Risk Consultant
7	3/2/2022	Other Cyber Security Specialist / Expert
8	3/2/2022	Other Cyber Security Specialist / Expert
9	3/3/2022	Other Cyber Security Specialist / Expert
10	3/3/2022	MSc Informatics student (with CI Educational Background)
11	3/3/2022	Insurance Domain Expert Life Health and Property and Casualty and Underwriter
12	3/12/2022	Underwriter (other P&C)

Table 6.1: List of Participants Who Evaluated the MeritMiner4CI Approach

Part 1: Business Process Perspective in Cyber Risk Assessment

The goal of the first section was to evaluate the preliminaries, such as the relevance of the business process perspective in security, as well as the view of the participants on quantitative analyses of behavior in security.

Displayed introduction to part 1: In the following section, you will be asked a series of general questions regarding your views on cyber risk assessment methods and on the analysis of process perspective in the cyber security context.

Question 2: *How would you rate the importance of data mining and general log analysis in the context of cyber risk assessment of individual companies?*

Options offered for question 2:

- 1 **Irrelevant**
- 2
- 3
- 4
- 5 **Critically important**

Answers to question 2:

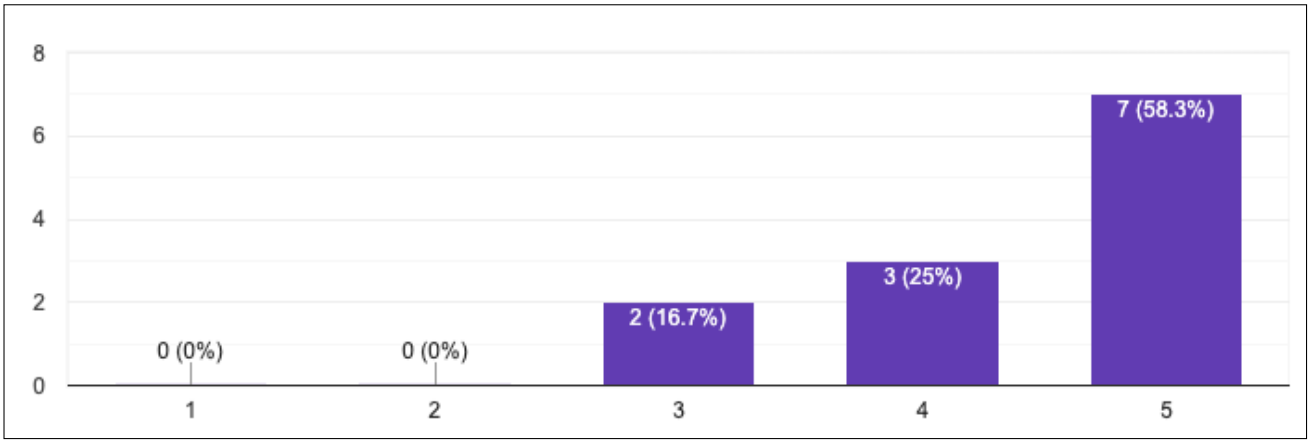

Figure 6.1: Breakdown of answers question 2 of the survey

Summary and Evaluation of the Responses to Question 2:

Regarding the log analysis for cyber risk assessment, most (more than 80% of the participants) considered it critically important, or important. Some of the experts who answered in a neutral way mentioned, for example, that they consider it as one of a whole tool-set of methods. Note that process mining was not evaluated separately, as it was not expected to be widely-known, therefore the term log analysis was chosen. Indeed, it was the case that process mining was not widely known to the participants. With the exception of two more technical participants, who had a general awareness of it. One participant was also aware of task mining.

Question 3: How would you rate the following statement: *Compared to qualitative methods (such as self-assessments, interviews, or analyses of policies), quantitative analyses of actual behavior (from productive systems) are more likely to reduce information asymmetries between the cyber risk analyst (e.g. IT Auditor, Cyber Underwriter, Regulator) and the organization subjected to the analysis.*

Options Offered for Question 3:

- 1 **Irrelevant**
- 2
- 3
- 4
- 5 **Critically important**

Answers to question 3:

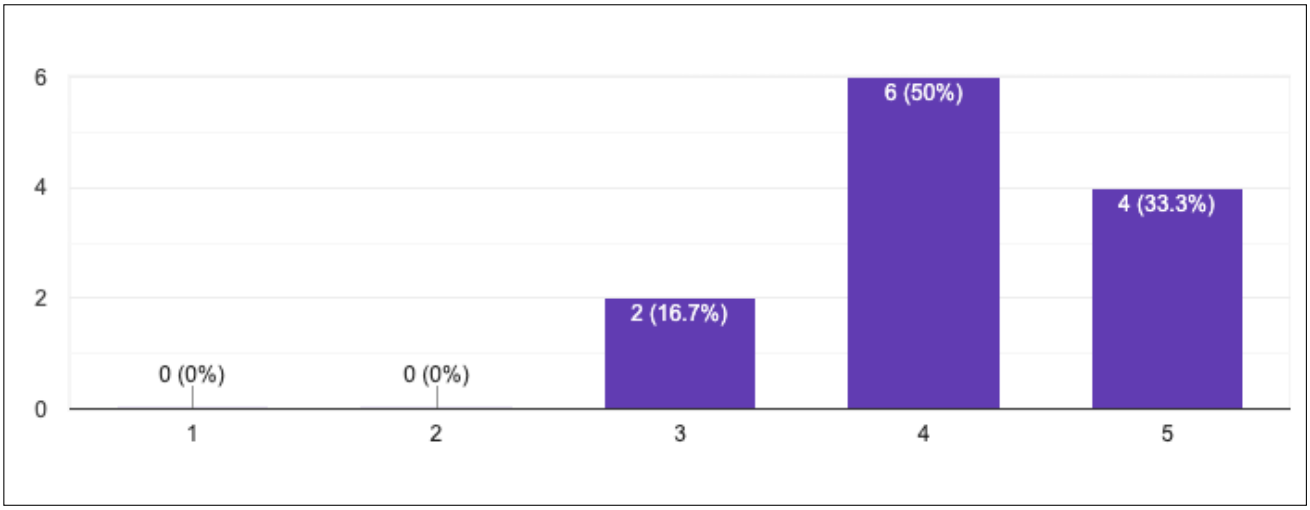

Figure 6.2: Breakdown of answers to question 3 of the survey

Summary and Evaluation of the Responses to Question 3:

Also, the answers to question three lean strongly towards the indication of high-relevance of quantitative analyses of actual behavior. Please note that again, the term process mining was avoided intentionally. Observations from qualitative interviews reveal that some considered the question somewhat confusion at first, as it was, in retrospect, formulated in a complicated way. Interviewees mentioned that they consider both types as necessary and complementary. It can be summarised that interviews indicated a strong current reliance on qualitative assessment. For example, one cyber risk analyst, who noted that he worked predominantly with startups, commented on his perception that analysts can already get a good understanding of organizations' cyber security posture and identify issues with a brief interview Another participant active in the healthcare domain strongly stressed that policies are the focus of her analyses. Data availability was also identified as a topic of concern.

Question 4: How would you rate the following statement: '*Investigating how internal processes of a given company operate is crucial for cyber risk assessment.*'

Options Offered for Question 4:

- 1 **Strongly disagree**
- 2
- 3
- 4
- 5 **Strongly agree**

Answers to question 4:

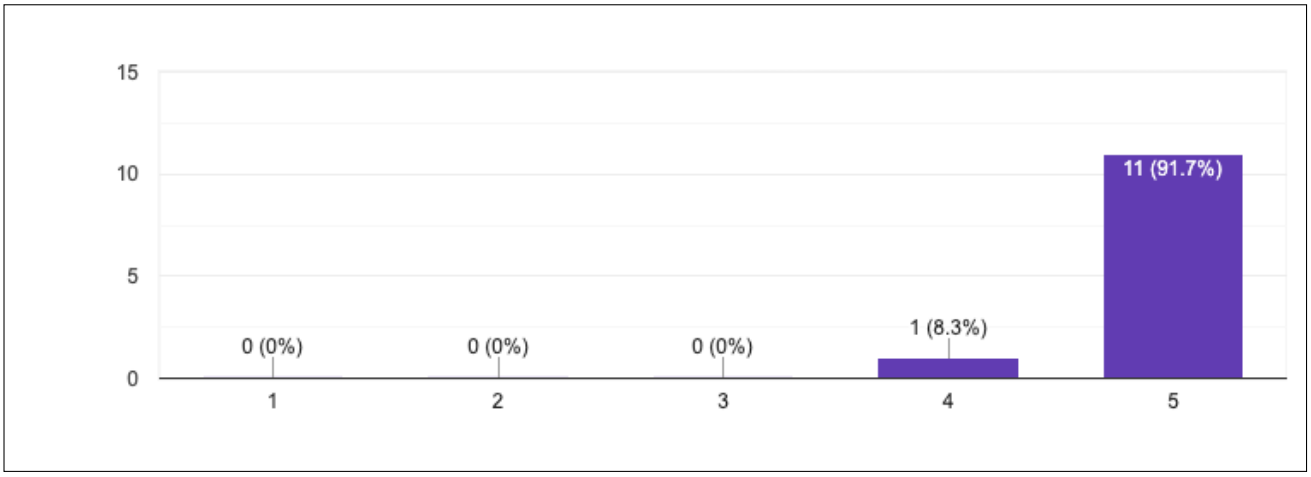

Figure 6.3: Breakdown of Answers to Question 4 of the Survey

Summary and Evaluation of the Responses to Question 4:

The answers to question four depicted in 6.3 showcase unanimous strong agreement on the relevance of business process perspective. This was also clear from qualitative interviews. For example, one cyber risk analyst, mentioned the relevance of incident management and incident response process, help-desk process, as well as disaster recovery.

Part 2: Process Discovery

The next section covers the first scenario that involves process discovery and visual analysis of rules (as a simplified method of conformance checking). Please note that the rules for the data flow are inspired by [124] who proposed an approach to security audits, but both the rules, as well as the data set were adjusted (to be visually interpretable) to fit the scenario better. For example, to make the scenario simple for brief questionnaire, all data in the flow was considered sensitive. The adjusted event log used to generate the visualisation with heuristic miner from the pm4py library is available in the repository. The source for confidence factors displayed was policy retrieved from [135] that correspond to the 3.

Displayed introduction to part 2: In the following section, you will be asked to provide your cyber risk assessment based on visual analysis of process map generated by a process discovery algorithm.

Case Study Scenario 1: Insider-Threat Detection - Suspicious File Operations

Description Presented to Participants:

Let's pose the following set up. As a risk analyst, you are tasked with assessing the cyber risk associated with underwriting a Cyber, Privacy, and Network Security Liability coverage for a private bank. You decide to investigate the flow of sensitive data in the bank.

The IT Compliance responsible of the bank is convinced that their handling of sensitive data runs in a compliant way and provides you with a policy document with the following information on rules that should be followed, that he believes proves compliance:

Rule 0 (start event): Employees begin the day by retrieving the data they need for their responsibilities from a back-up machine. In the unit that you are analyzing, all data retrieved from the machine is considered sensitive.

Rule 1: Users that are part of the same team (e.g. `team-1_user-1` and `team-1_user-4`) can share files with one another directly, but communication between teams needs to go through and be recorded in a communication hub of that team (e.g. `communication-hub_team_1`)

Rule 2: It is strictly forbidden to share sensitive data in a public system (denoted with 'P')

Rule 3: Data is eventually backed up from the communication hubs to the 'back-up machine'. As you want to verify the claim of the IT Compliance responsible, you decide to analyze the event log generated by mining the workstations of the employees. You apply the process discovery method (heuristic miner) on the log and generate the following visualization (process map) of how files move across the organization. Please kindly review it and answer the questions below.

Displayed visualisation: Process Map Discovered with Heuristic Miner

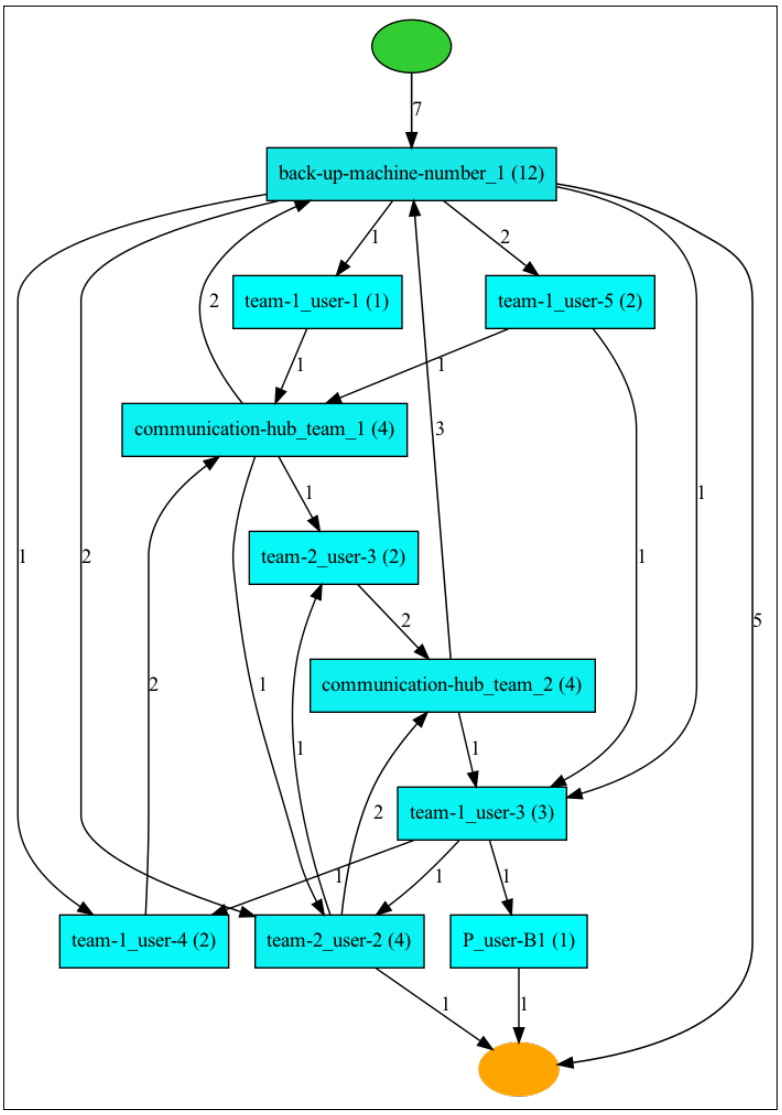

Figure 6.4: Process Map Generated with Heuristic Miner, based on a Dataset Retrieved from [153]

Question 5: For which of the following rules can you identify violations, based on *your interpretation* of the process map:

Options offered for question 5:

- Rule 1: Communication between teams needs to go through and be recorded in a communication hub
- Rule 2: No sharing of data in public systems
- Rule 3: Mandatory eventual back-up of communication hubs
- For none of them

Answers to question 5:

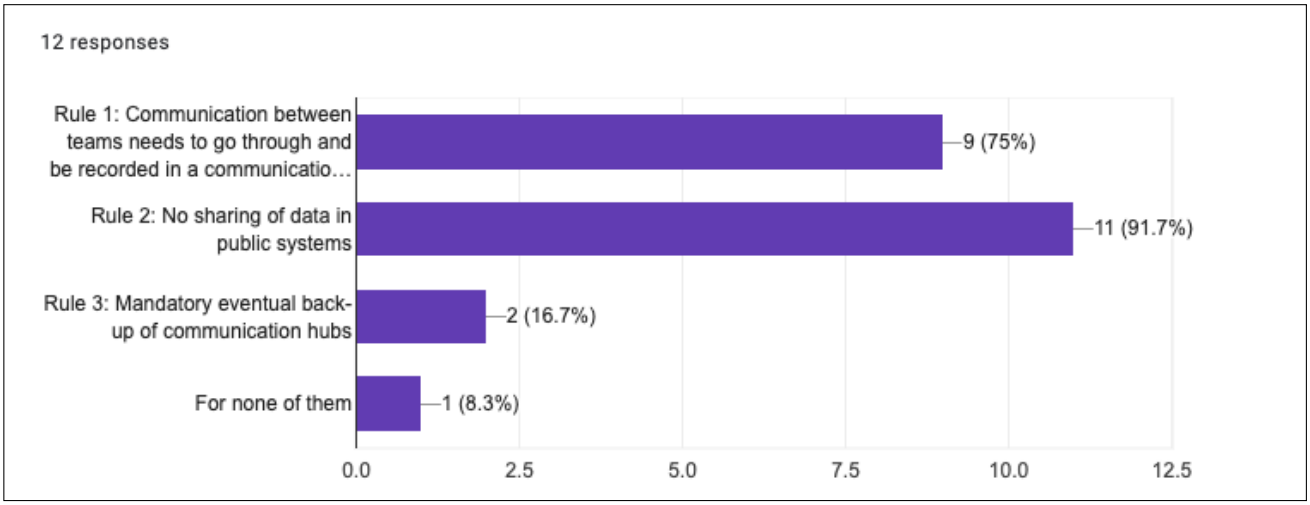

Figure 6.5: Breakdown of Answers to Question 5 of the Survey

Summary and Evaluation of the Responses to Question 5:

The idea of this question was to provide the simplest possible representation of an analysis that could be easily interpretable and would not require too much detail. It was intentional that no information on interpreting the graph was provided at first. From the objective perspective rules 1 and 2 can be considered violated. Rule 3 was not violated as both hubs were backed-up (as you can see on the graph) eventually. As you can see, 11 out of 12 participants, the vast majority, could (correctly) identify violation of rule 2. With rule 1, the number of correct responses was 9, which can be traced back to the relatively higher complexity of checking the rules. On the contrary, the rule that in reality was not violated was marked as such in only 2 cases. In summary, the results strongly indicate that visual abstraction of process flows using process discovery can be interpreted by security experts. The next key question investigates how this finding translates to rating of confidence factors.

Question 6: In case you identified any of the violations outlined above, what (if any) influence does this have on your assessment of the following confidence factors (made available by the Chubb Reinsurance Company)? Note: A positive assessment would potentially lead to a premium discount, whereas a negative assessment to a premium increase.

Options Offered for Question 6:

Coverage-specific modifier	Significantly more favorable rating	Slightly more favorable rating	No influence / no difference in rating	Slightly less favorable rating	Significantly less favorable rating
Handling of sensitive information)					
Backup or Mirror Procedures					
Compliance with privacy regulations					
Risk Management Controls					
Employee Training					
System Management					

Table 6.2: Answer Options to Question 6, Relating to Confidence Factors from Actual Policy, Multi-Line Single-Choice Selection [135]

Answers to Question 6:

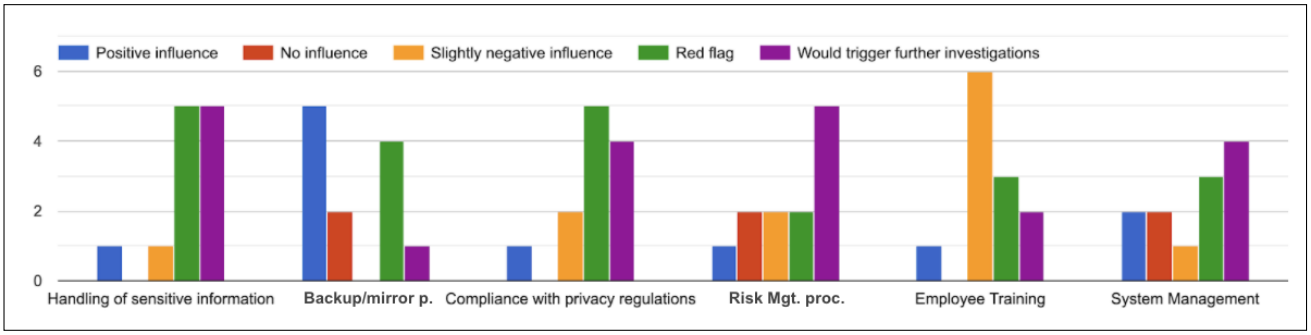

Figure 6.6: Breakdown of Answers to Question 6 of the Survey

Summary and Evaluation of the Responses to Question 6:

In the next question, the participants were asked to express, given the findings from question 5, their rating of a confidence factors from a selected policy from [135]. This was designed to clearly identify a connection between a process mining analysis and a risk assessment impacting a premium. As you can see in Figure 6.6. For 10 out of 12 participants, the analysis would be either a red flag, or trigger further investigations about handling of sensitive information; a similar result can be observed for compliance. On the other hand, the results about confidence factors related to backup policies are

less conclusive, but leaning towards a positive influence given the finding that the backup policy could be verified. An interesting outtake from the interviews was that some experts had difficulties applying a data-driven perspective. The participant who rated all options positively explained that he considered more the existence and validity of the policy, rather than the picture offered by the analysis of the system.

Question 7: Please rate the following statement: *Providing visual abstractions of process flows is a viable way to enable business users to conduct simple cyber risk assessments based on rules.*

Options Offered for Question 7:

- 1 **Strongly disagree**
- 2
- 3
- 4
- 5 **Strongly agree**

Answers to question 7:

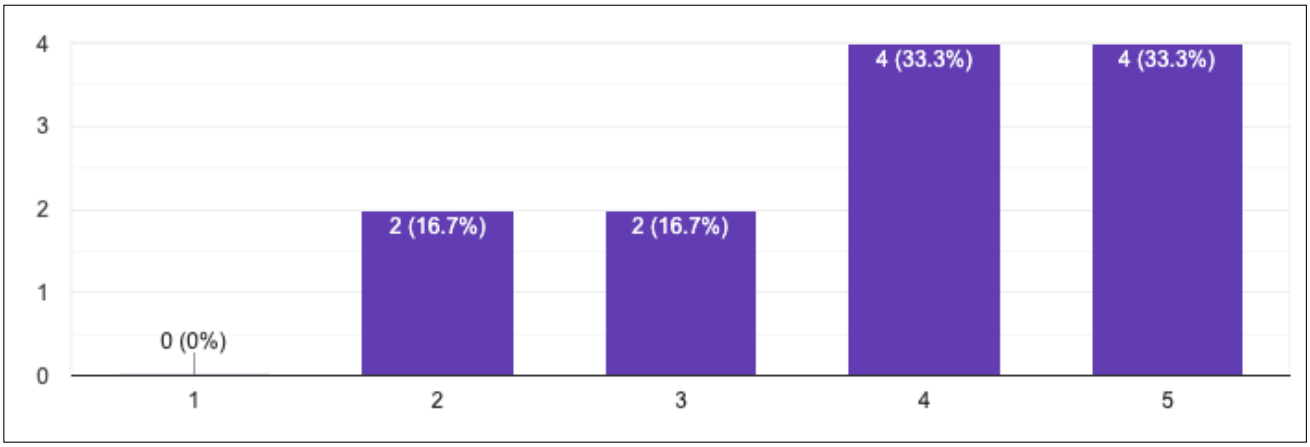

Figure 6.7: Breakdown of Answers to Question 7 of the Survey

Summary and evaluation of the responses to question 7: Finally, participants were asked explicitly to express their opinion about the method as displayed in Figure 6.7. The results can be interpreted as follows: the validity of the method could be confirmed, but with limitations. These include the fact that the scenario was rather simplistic and the complexity of such analysis in a real-life setting would be much higher.

Overall Summary and Evaluation of Scenario 1:

The first scenario indicates that process mining analysis can influence ratings of risk relevant for cyber insurance and the majority of analysts can interpret process mining results without larger difficulty, as evidenced by the fact that violations could be identified in the majority of cases. Indeed, it seems to be the case that analysing processes with process mining methods provided an additional perspective to the analysts.

Part 3: Process Conformance

The next section moves on a somewhat more complex scenario involving conformance checking of an IAM (Identity Access Management) process using a reference process depicted in [112], in order to ensure feasibility of the scenario and relevance of the analysis. The notion of comparing traces in an event log with prescribed formal process. This scenario is also the focus of the summary on case study prototype implementation below.

Introduction to Part 3 Displayed to Participants:

In the following section, you will be asked to assess a simplified scenario in regard to the impact of the findings on your assessment of the cyber risk (modifiers) associated with the process.

Case Study Scenario 2: Identity Access Management, scenario setup displayed to the participants

Let's assume the following scenario. You are provided with the following reference process model by the IT Security team of a major hospital, that reflects their defined Identity Access Management (IAM) practice. Your task is assessing their cyber security posture for the purposes of cyber insurance under time and resource constraints. For the purpose of assessing their access controls, you decide to check whether the prescribed model holds in practice. You therefore decide to extract an event log merging all events from the IAM Tool, Account Store (AD) and Credential store and to check whether that process holds in reality. Please rate the series of statements below, which present you with findings generated by a conformance checking analysis.

Displayed visualisation: Reference model of the IAM process provided by the company

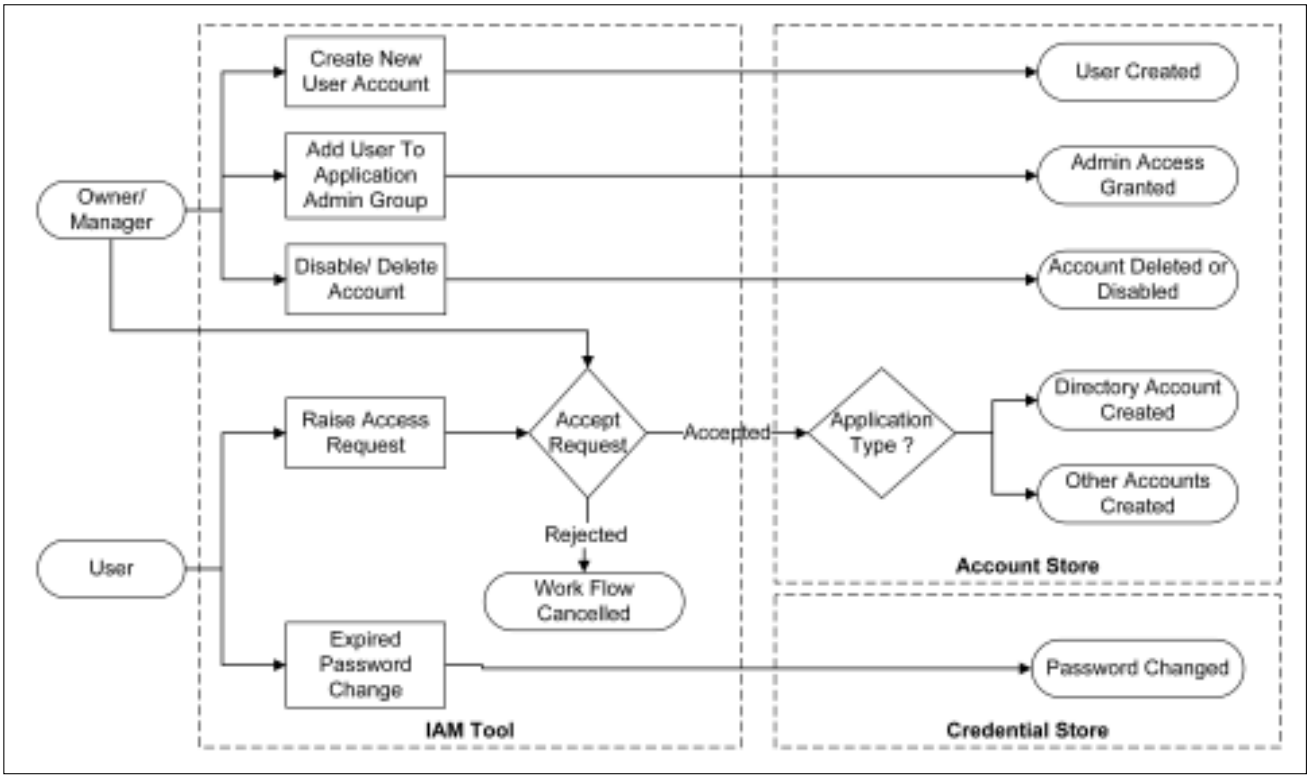

Figure 6.8: Displayed Visualisation for Case Study Scenario 2, reference model based on [112]

Scenario 2, Step 1: Checking Fitness of the Model Against the Event log

First, you decide to use an automated tool convert the prescribed model to a computer-readable format (i.e. petri net) and using automated conformance checking method (token-based-replay) test, whether the process instances in the EventLog conform to the model. You find out that only 40% of the process instances recorded in reality can be explained (replayed in) the model. In other words, the model allows for only 40% of the instances recorded.

Question 8: What impact on your perception of the IAM process does the low fitness of the model have on your perception of the business process from the cyber risk perspective?

Options offered for question 8:

- 1 **No impact / inconclusive / not-interpretable**
- 2
- 3
- 4
- 5 **Points at increased risk**

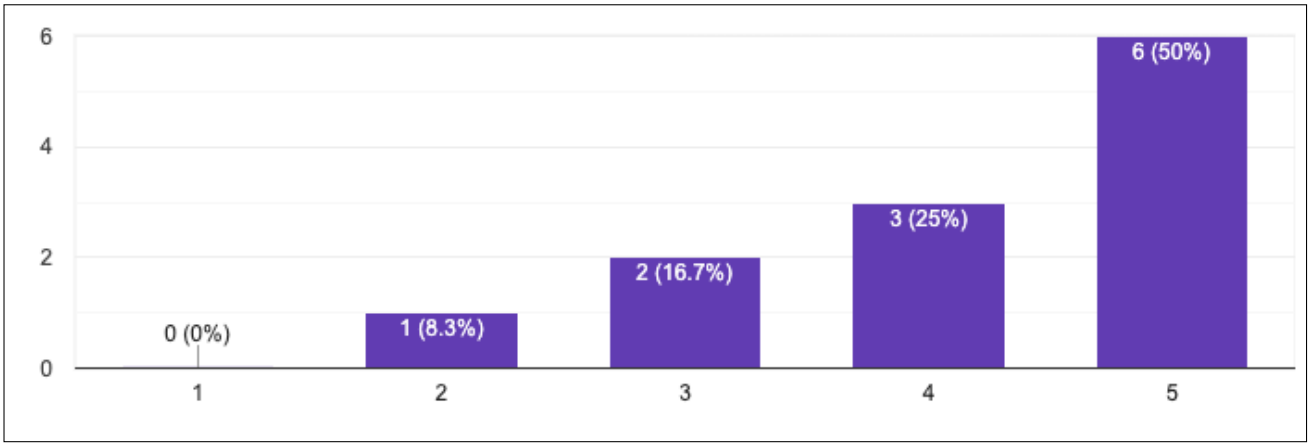

Figure 6.9: Breakdown of Answers to Question 7

Question 9: OPTIONAL: Please shortly comment on your perception of risk based on the metric (participants without responses excluded from the listing)

Response from participant #2: Researcher Cyber Security Related

Certainly a model that is not totally being explained in its full capacity, raises alerts on potential vulnerabilities. Thus, there "might" be risks in any of those processes or the intercommunication between them.

Response from participant #3: Other Cyber Security Specialist / Expert

Low fitness indicated to me poor Capability Maturity Model and lack of automation and process definition, which warrants instances of mistakes and inaccuracies.

Response from participant #5: Other Cyber Security Specialist / Expert

The integration between IAM, credential store and account store is very loosely defined.

Response from participant #6: Actuarial / Risk Consultant

IAM model as depicted in the Figure is perfectly fine but if the model is recording only 40% instances, it means that there is either a problem with model implementation or Conformance method is not appropriate.

Response from participant #7: Other Cyber Security Specialist / Expert

User and Owner roles of User Management System are not perfectly coupled and hence driven by a security policy pointing to considerable risk metric.

Response from participant #8: Other Cyber Security Specialist / Expert

I do not really understand the question. However, only 40% can be explained means that it statistically is impossible to determine the size of the problem. There is something wrong with the control.

Response from participant #9: Other Cyber Security Specialist / Expert

It depends. The 60% gap suggest a process deficiency. However, that 60% would need to be investigated to understand whether there is any additional risk being introduced i.e. users bypassing processes, unauthorised logins/use of applications etc. Equally, the 60% could be normal "noise" which simply isn't aligned with processes.

Response from participant #10: MSc Informatics student (with CI Educational Background)

IAM systems as such constitute an important play an important role on companies to protect an secure sensitive data/information. Violating rules or having a high % of flaws in this system means being highly exposed to cyber risks. It can even escalate depending on the amount of users the organization has and the type of information being shared.

Response from participant #12: Underwriter (other P&C)

The prescribed model allows more than half of the information going unrecorded and we need to analyse the impact of this on hospital administration policies and handling sensitive data of hospital which could impact on hospital authorities and may call for compliance penalty and chances of automation fault also need to be analysed.

Summary and Evaluation of the Responses to Question 8 and 9:

As the responses to questions 8 and 9 clearly show, 9 out of 12 participants evaluated the fitness metric with either 4 or 5, meaning it would negatively influence their risk assessment and increase premium. The responses to the follow-up question clearly validate the relevance of the process mining metric. However, explainability and choice of conformance method remains a question.

Case Study Scenario 2, step 2: You investigate further and, using conformance checking techniques, identify that the low trace fitness can be traced back to a high number of process instances skipping the steps in the IAM system altogether. You identify that 20% of the cases start with manual User Deletion, or manual User Creation events in the Account Store, which is in direct violation of the policy.

Question 10: What Impact does the [ADDITIONAL] Information in step two have on you Assessment of the IAM Process from the Risk Perspective?

Options offered for question 10:

- 1 **No impact / inconclusive / not-interpretable**
- 2
- 3
- 4
- 5 **Points at increased risk**

Answers to Question 10:

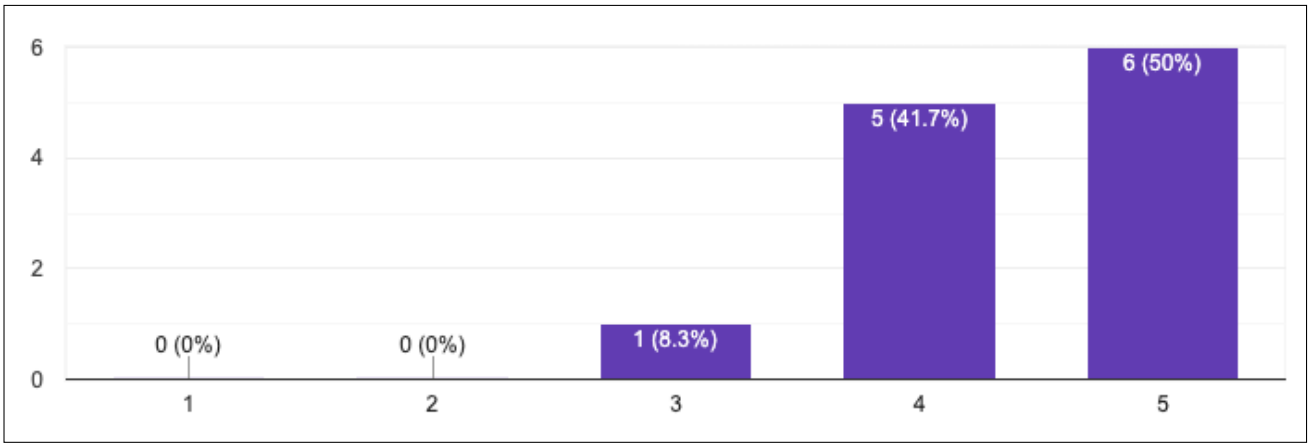

Figure 6.10: Answers to Question 10

Summary and Evaluation of the Responses to Question 10:

As concerns the interpretation of the answers to question 10, it can be considered that it provides further validation for the proposed checking method. Over 90% of participants would rate the risk higher, given additional information.

Question 11: What (if any) impact would the limited information have on your assessment of risk modifiers that might have positive / negative impact on cyber insurance premiums? (modifiers by Chubb Reinsurance)

Options Offered for Question 10:

Coverage-specific modifier	**Significantly more favorable rating**	**Slightly more favorable rating**	**No influence / no difference in rating**	**Slightly less favorable rating**	**Significantly less favorable rating**
Network Security					
Risk Management Controls					
System Management					
Network Access Control					

Table 6.3: Options Offered to Question 11

Answers to Question 11:

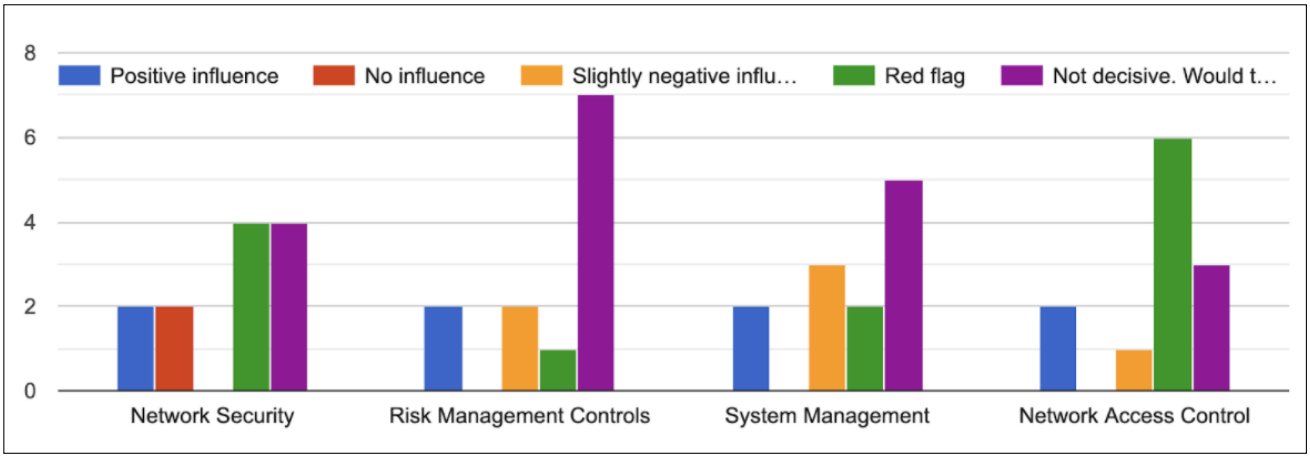

Figure 6.11: Breakdown of Answers to Question 11

Summary and Evaluation of the Responses to Question 11:

When asked about the specific risk modifiers, for risk management control, 7 participants would conduct more in-depth investigation of the matter. From the Network Access Control modifier perspective, half of the participants saw the findings as red flags and for three more, this would lead to further investigations, the results there indicate the validity of the conformance measure. One of the participants who rated the risk more positively then explained, in the follow-up interview, that he considered the existence of the model, rather than the actual executions.

Overall Summary and Evaluation of Scenario 2:

In summary, scenario 2 is in line with my hypotheses. The answers validate the developed concept. The main limitation is the black box nature of the conformance method that needs to be trusted. Discussion on this takes place in the quantitative evaluation.

Section 3: Process Enhancement for Cyber Security

Introduction to part 4: So far, we have focused on applying process mining to detect anomalies and to identify threats. In the final scenario, we will focus on process enhancement and performance analysis of security-relevant processes.

Case Study Scenario 3: IT Incident Response - Major Car-Maker

For the final scenario, we will consider the Incident Management Process (containing events from Acceptance to Resolution) based on a real-life process model and event log. Details are intentionally abstracted away for the purposes of the scenario.

The car manufacturer has a globally defined Incident Management process, which is executed by subsidiaries across the world. In the scenario, we will observe three different countries, which are supported by dedicated local teams. Those teams, however, aim to follow the globally defined process and are otherwise independent of each other.

The set-up of the case study is that we want to rate the risk modifiers, taking the relative posture of other subsidiaries into consideration.In the scenario, we assume that the car manufacturer might arrange a separate cyber insurance agreement for each subsidiary.

The EventLog has been processed, discovery and conformance checking techniques applied, as well as process statistics generated.

Based on the summary of metrics generated below, please fill out the table below. In the open part of the interview, you will be asked on the reasoning behind your choices.

Displayed Visualisation: Summary of the metrics generated with automated process analysis.

Metric	Unites States	France	Germany
Number of incident cases	6126	1799	3625
Number of events (e.g. (re-)assignment, implementation, waiting for assignement)	65659	25253	48520
Median case duration	7 days 12 hours	10 days	21 days
Percentage of rework events	43%	51%	61%
Handovers between teams after first assignent (ping-pong rate)	26%	36%	46%
Number of variants of the process	1871	905	2211
Fitness (percentage of cases that fit into the global corporate pre-defined model)	72%	64%	45%

Figure 6.12: Displayed Visualisation for Case Study Scenario 3

Question 12: Based on the information from the table. Please indicate how you would rate the risk modifiers (as per the Chubb cyber underwriting manual) for Germany relative to other subsidiaries.

Options offered for question 12:

Coverage-specific modifier	Significantly more favorable rating	Slightly more favorable rating	No influence / no difference in rating	Slightly less favorable rating	Significantly less favorable rating
Incident Response Planning					
Risk Management Controls					
Centralised Processes and Procedures					
Employee Training					
Training and Education					
Disaster Recovery					

Table 6.4: Options Offered to Question 12

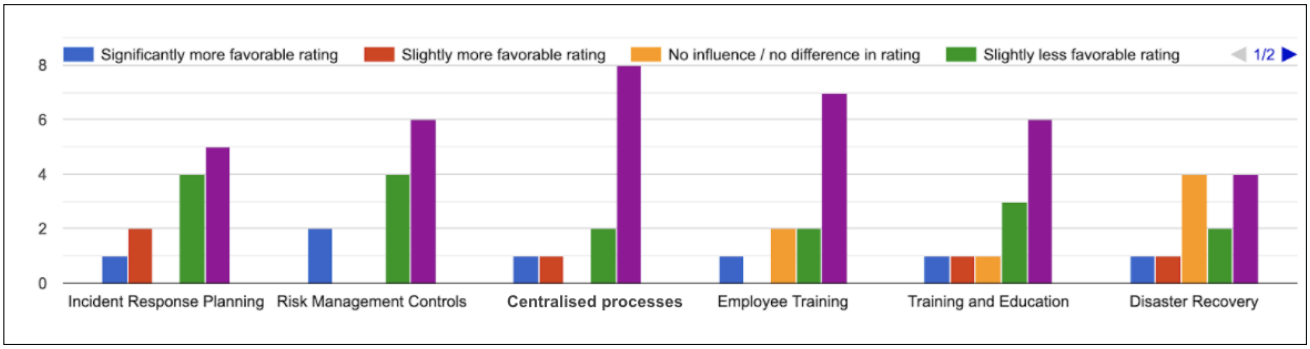

Figure 6.13: Breakdown of Answers to Question 12

Summary and Evaluation of Scenario 3 (Question 12):

The last scenario focused on testing the impact of process performance indicators of a security-relevant process on a cyber risk assessment from the relative perspective. The pink bar in Figure 6.13 corresponds to **Significantly less favorable rating** and was, with one exception, the most selected option for all presented confidence factors. The qualitative interviews then explored which metrics were considered, rework, fitness, case duration and number of variants were all mentioned in the interviews as contributors to the negative ratings. It seems to be the case that fitness of traces in the IAM process was indeed considered relevant, when they were used to reflect on the validity of the policy.

6.3 Case Studies in the Prototype

Case Study Scenario 1: Insider-Threat Detection - Suspicious File Operations

The subject of case study scenario 1 was evaluation of a data flow as outlined above, the following summary briefly show how the scenario can be implemented. A similar

approach was proposed by in security auditing context by [124] and further investigated by [125]. An event log of data flow in an organisation was first retrieved from [153] and then adjusted for easier interpretability. The following procedure outlines how the case study can be executed in the prototype in the following way.

First the event log containing the events in relation to the data flow in the bank can be uploaded and associated with a profile and a business process.

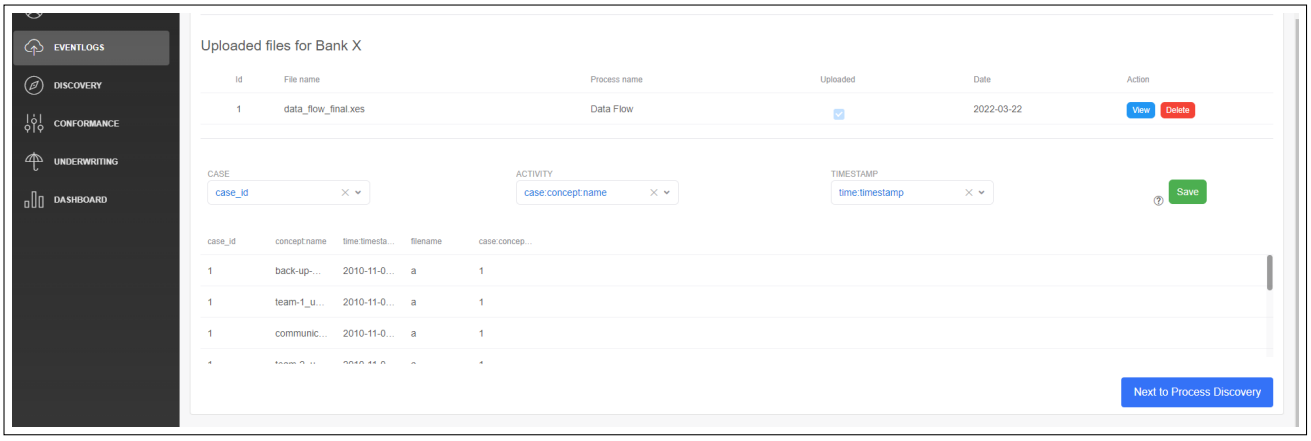

Figure 6.14: Setting up Event Logs for the Case Study Scenarios

Next, the visualisation based on the heuristic miner algorithm can be generated and displayed, which can be used to identify the rule violations.

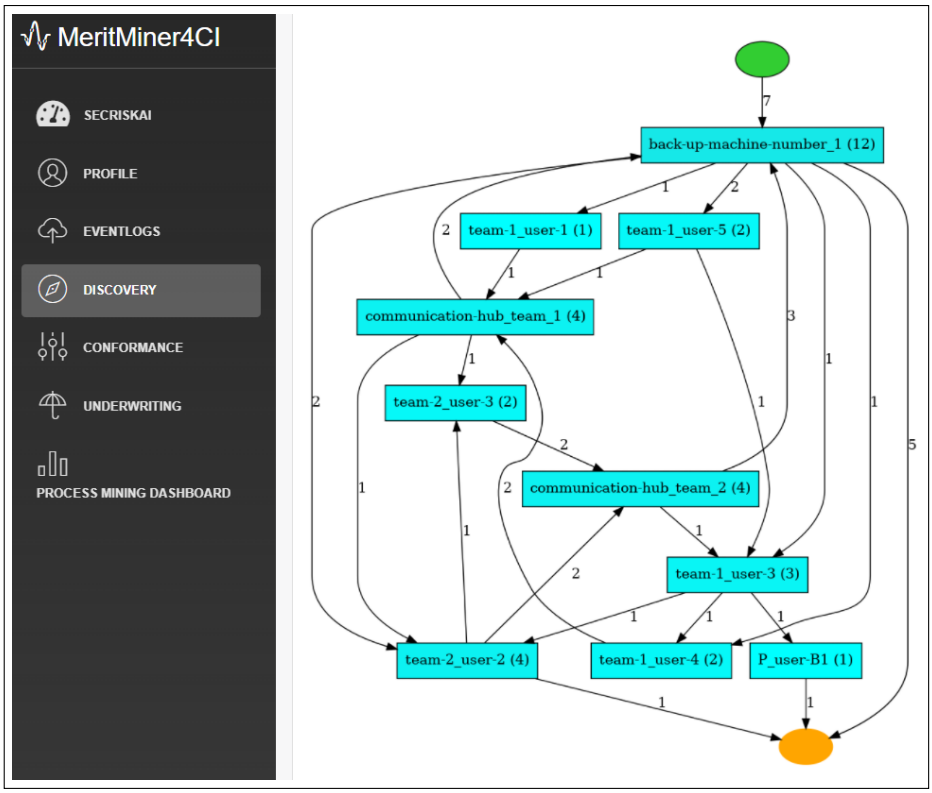

Figure 6.15: Discovered Model for the First Case Study

Finally, process statistics can be generated. Thus covering scenario 1 directly in the prototype. Rating of confidence factors will be presented for the second scenario and work analogously for the data flow and will not be presented separately in this section.

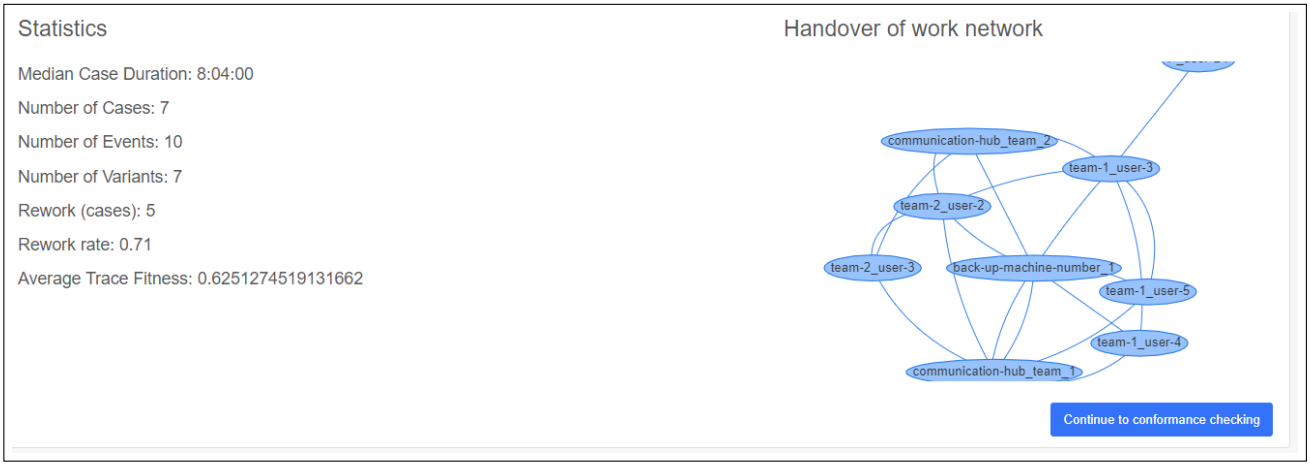

Figure 6.16: Generating Statistics in the Prototype, incl. Handover of Work Visualisation

6.3.1 Case Study Scenario 2: Identity Access Management

For the second scenario to be implemented, the reference process presented in the survey needed to be formalised in the BPMN 2.0 notation. This BMPN 2.0 file was then used to generate aa event log containing anomalous traces in line with [112]. The dataset is provided in the repository, together with configuration and the BPMN 2.0 process. The BPMN file is too large and complex to visualise in the document, therefore it is provided in the original for as an attachment.

Once both the log with anomalous traces and the BPMN process that was modelled is uploaded. A conformance check can be conducted on the conformance page, effectively outputting the corresponding low fitness value given because the log contains events not reflected in model - events reflecting escalation of privileges and creations of accounts that skip the process steps the model expeced in the IAM system (events only occur in the Account Store). A drill-down can be conducted to identify which events do not fit the model using the table.

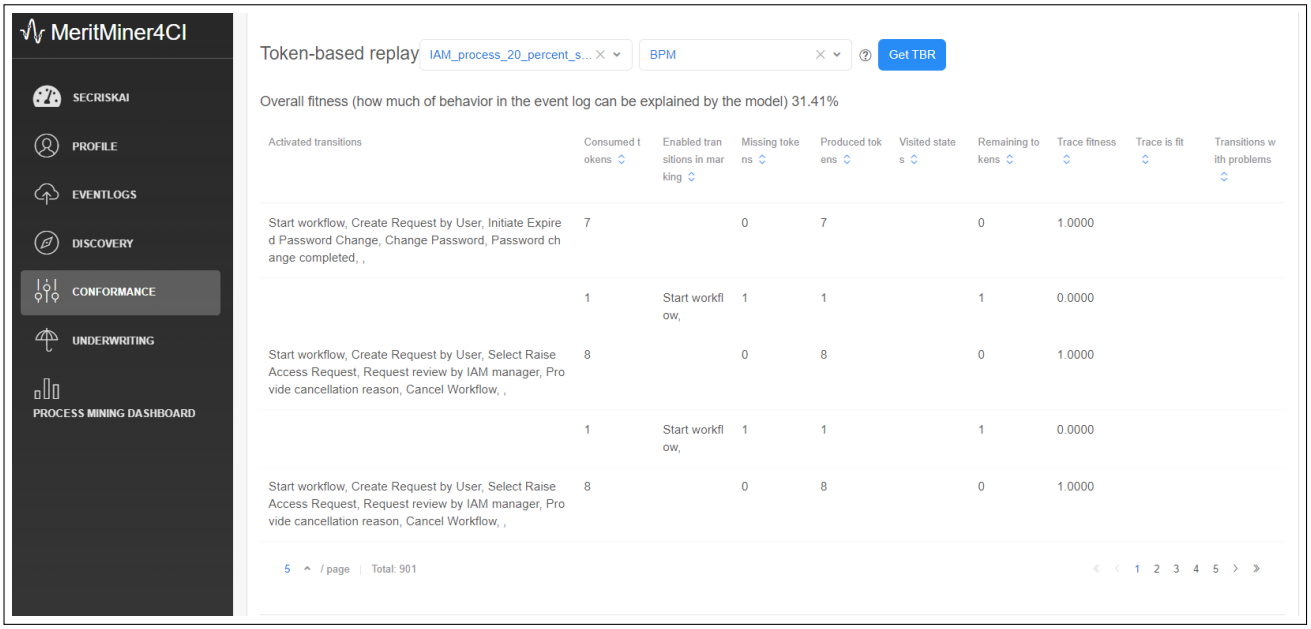

Figure 6.17: Execution of the IAM Case Study on the Conformance Page

Finally, for both case studies, once the Cyber Risk Analysis Workflow is finished, the Cyber Risk Assessment Workflow can start and confidence factors can be rated. Figure 6.18 showcases an example rating in line with the proposed approach. Lastly, the dashboard can be displayed, reflecting an overview of risk modifiers related to processes.

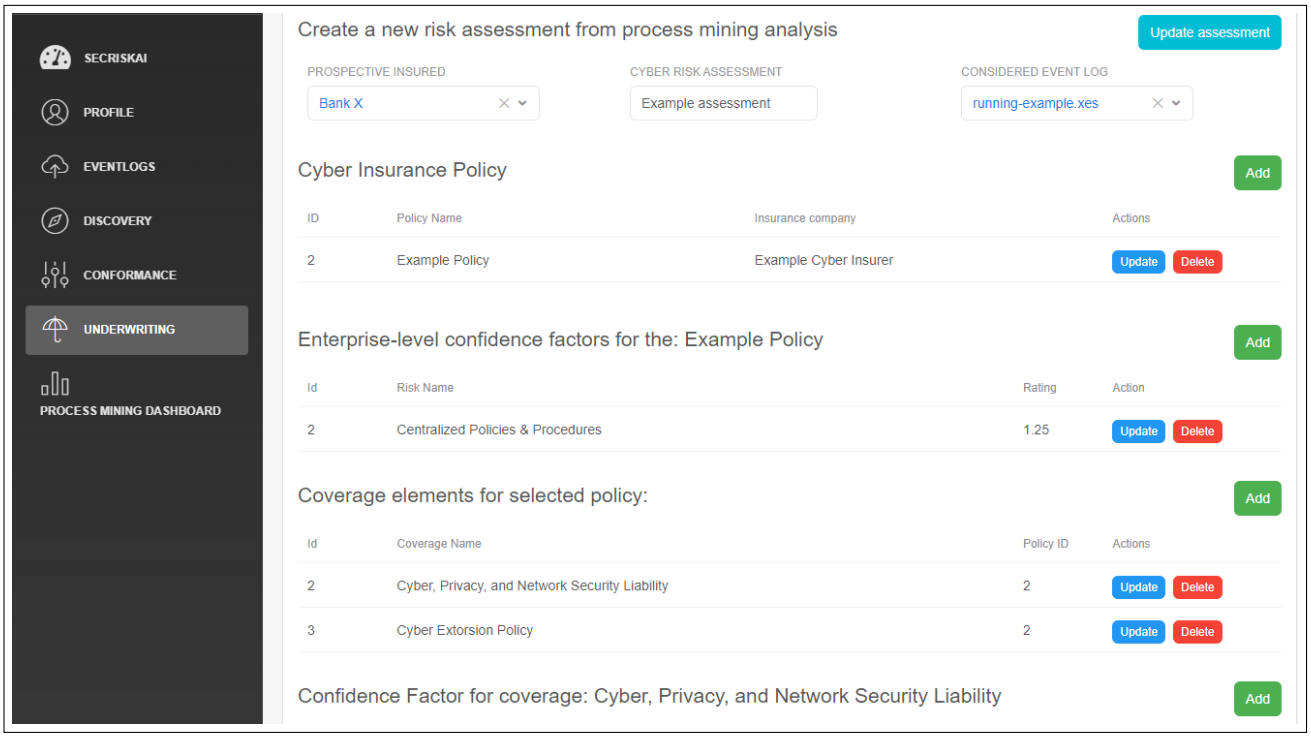

Figure 6.18: Underwriting Dashboard Preview For the Case Studies

6.4 Quantitative Evaluation

For quantitative evaluation, evaluation of different discovery algorithms are evaluated using different model quality dimensions that exist in process mining. Event logs from BPI challenge 2013 (reflecting incident and problem resolution processes) as well as a help-desk event log. Additionally, the event log from BPI challenge 2012 was used as well. On overview of the BPI challenge event logs can be found in [154]. The goal of the quantitative evaluation was to provide an overview, of what quality metric values we can expect from different discovered process modes and to show that they are conflicting criteria; the evaluation takes a relative perspective. The cross-validation method splitting on case ids does not aim to generate insights in absolute terms.

The following metrics were evaluated and can be summarised as follows. A deep investigation of the metrics is offered by e.g. [84].

- **Fitness** - how much the model of the behavior in the log can the model explain.
- **Precision** - how much of the behavior allowed by the model is present in the log
- **Generalization** then measures how many components of the model are present in the event log, some of them can be used infrequently in the actual log that means too general
- **Simplicity** - how complex is the model.

For implementation of the evaluation, the pm4py.algo.evaluation package is applied [98]. A script in the form of a Jupyter notebook was developed to use a K-means cross-validation [155] approach to split the event log into a discovery log (on which algorithm is applied to discover a model, corresponds to train) and a test log, which is then replayed in/aligned with the discovered model. The logic of the evaluation is available in the repository. For each metric the values are then averaged for each combination of algorithm , method and log. Please note that in process mining, the procedure needs to be interpreted carefully and should not be understood in the same way as n the machine learning domain. In general, because event logs can contain noise, cross validation is a good choice, if we want to evaluate process mining algorithms in an objective way. However, one of the problems with cross validation in evaluation of process mining algorithms is the lack of negative examples. Event log on it's own does not provide an indication which behavior is desirable and which not. The following can be concluded: process mining is, in the security domain, not a purely quantitative method. It's best applied together with domain knowledge that can be applied , e.g. for deciding how to identify desirable traces. The following table demonstrate how different process discovery algorithms perform from different perspectives.

	Conformance checking	Alpha	Inductive	Heuristic
0	Token-based replay/Fitness	0.873614	0.999996	0.891460
1	Token-based replay/Precision	0.135788	0.122728	0.812868
2	Token-based replay/Generalization	0.963843	0.964871	0.848563
3	Token-based replay/Simplicity	0.925581	0.599077	0.536971
4	Alignments/Fitness	0.738122	0.999996	NaN
5	Alignments/Precision	0.135788	0.122728	0.812868
6	Alignments/Generalization	0.963843	0.964871	0.839480
7	Alignments/Simplicity	0.925581	0.599077	0.537185

Figure 6.19: Process Mining Evaluation Criteria for the BPI Challenge 2012 Dataset of Loan Applications

	Conformance checking	Alpha	Inductive	Heuristic
0	Token-based replay/Fitness	0.821619	0.999331	0.768829
1	Token-based replay/Precision	0.454918	0.493808	0.894136
2	Token-based replay/Generalization	0.740668	0.832328	0.660124
3	Token-based replay/Simplicity	1.000000	0.652436	0.544279
4	Alignments/Fitness	0.911798	0.998391	NaN
5	Alignments/Precision	0.455714	0.495082	NaN
6	Alignments/Generalization	0.740668	0.832328	0.650133
7	Alignments/Simplicity	1.000000	0.652436	0.545490

Figure 6.20: Process Mining Evaluation Criteria for the Event Log Containing Closed Problems

	Conformance checking	Alpha	Inductive	Heuristic
0	Token-based replay/Fitness	0.873614	0.999996	0.891460
1	Token-based replay/Precision	0.135788	0.122728	0.812868
2	Token-based replay/Generalization	0.963843	0.964871	0.848563
3	Token-based replay/Simplicity	0.925581	0.599077	0.536971
4	Alignments/Fitness	0.738122	0.999996	NaN
5	Alignments/Precision	0.135788	0.122728	0.812868
6	Alignments/Generalization	0.963843	0.964871	0.839480
7	Alignments/Simplicity	0.925581	0.599077	0.537185

Figure 6.21: Process Mining Evaluation Criteria for the BPI Challenge 2013 Event Log Reflecting the Incident Management Process

	Conformance checking	Alpha	Inductive	Heuristic
0	Token-based replay/Fitness	0.705255	0.992719	0.770310
1	Token-based replay/Precision	0.275623	0.385353	0.974501
2	Token-based replay/Generalization	0.596938	0.708574	0.537098
3	Token-based replay/Simplicity	1.000000	0.661939	0.562603
4	Alignments/Fitness	0.752064	0.988174	NaN
5	Alignments/Precision	0.275623	0.385353	NaN
6	Alignments/Generalization	0.596938	0.708574	0.526281
7	Alignments/Simplicity	1.000000	0.661939	0.561709

Figure 6.22: Process Mining Evaluation Criteria for the Helpdesk event log

To summarise the results, my interpretation is the following: which algorithm and conformance to apply when should be decided based on the fundamental business goal and by reflecting on the event log from the business process perspective. In the cyber security domain, we might, for example, want to cover anomaly detection use-cases, or analyse the performance of processes, as is reflected in the chapter outlining the MeritMiner4CI approach.

Let's assume the following. We might decide to apply filtering on an event log and divide cases (analogously to supervised learning) into anomalous and normal traces. Then, we would apply a discovery algorithm on the filtered event log with the goal of discovering a model against which we would check new incoming traces and decide whether they are anomalous, based on trace fitness. In this case, inductive miner (as we can deduce from the tables above) would give us the guarantee that all the previously filtered conform cases are considered in the discovered model, presumably leading to higher accuracy (as all desirable behavior could be explained, if the categorisation was correct). Again, the approach would work best together with a domain expert, or some type of machine learning method providing for the categorisation of which traces are desirable (by the means of some optimisation goal, e.g. cyle time, cost, etc.)) and which are anomalous in order to filter a log for a generation of a best-practice model and possible of a model of violations as well.

But anomaly detection is not the only use case. In other scenarios, we might just want to observe the simplest possible model that explains certain process good enough. For example, we might want to focus only on the most important steps when analysing a helpdesk process with the goal if finding performance optimisation potential. As you can see from my results, the heuristics miner seems to be a choice providing for a good balance between quality criteria, but it does not guarantee soundness, see e.g. [79]. This manifests itself in the NaN values as alignment-based conformance checking requires a sound model. In summary, it is clear that the quality criteria are competing.

Chapter 7

Discussion and limitations

Let us briefly summarise the results of all applied methods to validate the proposed approach. From the perspective of the survey, it was clearly demonstrated that taking the process perspective is relevant. Process mining analyses were to a significant extent able to influence the ratings of established confidence factors once inconsistencies with expected behavior were identified, thus indirectly leading to premium penalties or rebates. Therefore, process mining can support self-protection auditing in line with the proposals by [134] and contribute to addressing the fundamental insurability issues of cyber risk that were reviewed in the background. Supporting evidence was generated that both conformance of processes and performance is relevant to rate cyber security posture, specific scenarios with industry applicability were constructed. These scenarios were then proven to be technically feasible by demonstrating how they can be executed in the designed and developed prototype covering all the steps in the proposed MeritMiner4CI approach. Finally, a brief reflection on process mining algorithms investigated their different quality metrics and evaluated when which of them might be applicable.

As concerns the limitations, it is clear, in line with [53] that there is no universally applicable risk assessment method for all scenarios and all types of organisations. There are vast differences in the relevance of different methods across company sizes. Fundamentally, the underwriter or the cyber risk analyst would decide on the approach and, of course, the incentives must be in place for the organisation to agree to be subjected to this type of security "health check". However, this also has a second aspect to it. If the organisation agrees, it might be a strong signal that *adverse selection*, as issue identified in he background chapter, is not taking place. Currently, in the segment of large enterprises, which were the implicit focus of this thesis (as their scale is more suitable for the application of process-mining methods), '*cyber risk analysis is a people business'*, as one of the underwriters that was interviewed in [12] pointed out. Underwriting meetings take place, questionnaires and expert judgements are applied and this thesis does not see them as replaceable by process mining methods. Instead the approaches are complementary. Related to that, typically process-specific, organisational and domain knowledge is required to interpret process mining results, limiting the potential for high-levels of automation of such analyses. In many cases, process mining would simply serve as a starting point for further investigations (often, experts indicated that process mining results would trigger further investigations). Finally, availability of reliable event logs, as well as of formal

process models is still limited, and data integration and data quality remain an issue, as pointed out e.g. by the survey in [97]. This limitation also manifests itself in the scenarios, which were intentionally constructed with focus on simplicity and relevance for cyber insurance and not from the perspective of the depth of the analysis. The same is true for the prototype, which was designed in a limited scope, focusing on the fundamentals of the proposed approach.

Chapter 8

Summary, Conclusions, and Future Work

The main goal of the thesis was to explore the applicability and role of process mining methods in cyber insurance. Based on the conducted literature review, this thesis seems to be the first work explicitly making the connection with the fundamental challenges of cyber insurance and the potential process mining has to support their mitigation. The thesis clearly answers the question both by evaluating the proposed approach with experts, as well as by proving the technical and methodological feasibility with a proof-of-concept, which, in addition, integrates with other relevant cyber risk assessment projects from the CSG group (*e.g.,* SecRiskAI and MENTOR) that cover the complementary perspectives of assessing the risks related to external cyber attacks that process mining in not well suited to tackle. Also, a mapping of process mining to fundamental challenges of the cyber insurance market and it's current practices, actors and methods. This is done by taking a structured approach and following a clear methodology that proposes methods that cover both process discovery algorithms and conformance checking, to use for specific coverage types and the rating of their relevant confidence factors.

As this thesis has demonstrated, it needs to be pointed out that process mining is suitable to identify whether there are issues and how the actual executions of relevant processes differ to the expectations of the organisations as reflected in policies, models, regulatory guidelines or informal description. However, on it's own, process mining has only limited potential to help explain the discovered conformance and performance issues. This is why the rating of confidence factors is proposed in the MeritMiner4CI approach to be kept manual and based on expert judgement as it fits with the current assessment practices in cyber insurance very well. In the future, different paths could be taken to explore the root-cause analyses of security issues and to make a link between the process perspective and protection recommendations such as those provided by MENTOR for external cyber attacks. This could serve as a basis for new cyber insurance products integrating these methods to support not only risk transfer, but also risk mitigation and process improvement. As concerns the perspective of developing more advanced process-mining analyses, combining process mining with machine learning methods shows promise for a number of different scenarios. These range from root-causes analyses of process inefficiencies, correction recommendations, detecting anomalous behavior by clustering traces to automated event log generation. Combining the perspectives of multiple processes in one process mining analysis seems to be an interesting direction for cyber risk assessment as well.

An interesting research stream that should be explored in the future is the usage of process mining for the quantification of security and for the development of process benchmarks and best practices that the processes of the prospective insureds could be compared against and that would provide a relative perspective and thus more interpretavility. Different existing risk-management libraries could be formalised in process security indicators that could be applied more universally and objectively. The starting point, that one can start seeing in the process mining domain, is the benchmarking of processes at the industry level that are supported by standard software. In the case of cyber security, it is estimated that cyber insurers could play a significant role, as they might provide incentives for process improvement with the premium differentiation mechanism.

Abbreviations

AIG	American International Group
API	Application Programming Interface
BPMN	Business Process Model and Notation
CCPA	California Consumer Privacy Act
CI	Cyber Insurance
CIS	Center for Internet Security
CF	Confidence Factor
CSG	Communication Systems Group
COBIT	Control Objectives for Information and Related Technology
CRM	Customer Relationship Management Software
DDoS	Distributed Denial of Service
EPC	Event-driven process chain
EIOPA	Event-driven process chain
ES	Expert System

ETL	Extract, Transform and Load
ERP	Enterprise Resource Planning
GDPR	General Data Protection Regulation
ISO 27001	(also known as ISO/IEC 27001:2013) is the international standard for information security
ITIL	Information Technology Infrastructure Library
LTL	Linear Temporal Logic
MVP	Minimum Viable Product
NIST	National Institute of Standards and Technology
OECD	Organisation for Economic Co-operation and Development
PAIS	Process-Aware Information System
PCI	Property Casualty Insurers
PM4Py	Process Mining Package for Python
PMC	Proactive Mitigation Cost
PPIs	Process Performance Indicators
RPA	Robotic Process Automation
SOX	Sarbanes-Oxley Act of 2002
RMC	Reactive Mitigation Cost
SP	Service Provider Company

TEC Threat Exposure Cost

List of Figures

List of Tables

Bibliography

[1] Robert Angell and George Lawton. Why process mining is seeing triple-digit growth. *Venture Beat*, 2021.

[2] NIST Special Publication 800-53 Revision 5 Security and Privacy Controls for Information Systems and Organizations Joint Task Force. Technical report, National Institute of Standards and Technology, 2020.

[3] CIS Benchmarks. CIS Benchmarks - Center for Internet Security, 2022. `https://www.cisecurity.org/cis-benchmarks/`, (Last accessed January 2022).

[4] Bruno Rodrigues, Muriel Franco, Geetha Parangi, and Burkhard Stiller. SEConomy: a Framework for the Economic Assessment of Cybersecurity. In *Economics of Grids, Clouds, Systems, and Services, 16th International Conference, GECON 2019, Leeds, UK, September 17â19*, pages 154–166, 2019.

[5] Intersoft Consulting. General Data Protection Regulation (GDPR) â Official Legal Text, 2018. `https://gdpr-info.eu/`.

[6] ANPD. Brazilian General Data Protection Law (LGPD), 2021. `https://iapp.org/resources/article/brazilian-data-protection-law-lgpd-english-translation/`, (Last accessed January 2022).

[7] Wil van der Aalst, Arya Adriansyah, Ana Karla Alves de Medeiros, Franco Arcieri, Thomas Baier, Tobias Blickle, Jagadeesh Chandra Bose, Peter van den Brand, Ronald Brandtjen, Joos Buijs, Andrea Burattin, Josep Carmona, Malu Castellanos, Jan Claes, Jonathan Cook, Nicola Costantini, Francisco Curbera, Ernesto Damiani, Massimiliano de Leoni, Pavlos Delias, Boudewijn F. van Dongen, Marlon Dumas, Schahram Dustdar, Dirk Fahland, Diogo R. Ferreira, Walid Gaaloul, Frank van Geffen, Sukriti Goel, Christian Günther, Antonella Guzzo, Paul Harmon, Arthur ter Hofstede, John Hoogland, Jon Espen Ingvaldsen, Koki Kato, Rudolf Kuhn, Akhil Kumar, Marcello La Rosa, Fabrizio Maggi, Donato Malerba, Ronny S. Mans, Alberto Manuel, Martin McCreesh, Paola Mello, Jan Mendling, Marco Montali, Hamid R. Motahari-Nezhad, Michael zur Muehlen, Jorge Munoz-Gama, Luigi Pontieri, Joel Ribeiro, Anne Rozinat, Hugo Seguel Pérez, Ricardo Seguel Pérez, Marcos Sepúlveda, Jim Sinur, Pnina Soffer, Minseok Song, Alessandro Sperduti, Giovanni Stilo, Casper Stoel, Keith Swenson, Maurizio Talamo, Wei Tan, Chris Turner, Jan Vanthienen, George Varvaressos, Eric Verbeek, Marc Verdonk, Roberto Vigo, Jianmin Wang, Barbara Weber, Matthias Weidlich, Ton Weijters,

Lijie Wen, Michael Westergaard, and Moe Wynn. Process Mining Manifesto. *Lecture Notes in Business Information Processing*, 99 LNBIP(PART 1):169–194, 2011. doi: 10.1007/978-3-642-28108-2{_}19.

[8] Filip Caron, Jan Vanthienen, and Bart Baesens. A Comprehensive Framework for the Application of Process Mining in Risk Management and Compliance Checking. *SSRN Electronic Journal*, 4 2012. doi: 10.2139/SSRN.2244885.

[9] Filip Caron, Jan Vanthienen, and Bart Baesens. Comprehensive rule-based compliance checking and risk management with process mining. *Decision Support Systems*, 54(3):1357–1369, 2 2013. ISSN 0167-9236. doi: 10.1016/J.DSS.2012.12.012.

[10] Filip Caron, Jan Vanthienen, and Bart Baesens. Advances in Rule-Based Process Mining: Applications for Enterprise Risk Management and Auditing. *SSRN Electronic Journal*, 4 2013. doi: 10.2139/SSRN.2246722.

[11] David Hillson. Extending the risk process to manage opportunities. *International Journal of Project Management*, 20(3):235–240, 4 2002. ISSN 0263-7863. doi: 10.1016/S0263-7863(01)00074-6.

[12] Viktor Matejka and Juan Angel Huacan Soto Zurich. A Framework for the Definition and Analysis of Cyber Insurance Requirements. Technical report, University of Zurich, Zurich, 2021. `https://files.ifi.uzh.ch/CSG/staff/franco/extern/theses/MAP-VM-JAHS.pdf`.

[13] Christian Biener, Martin Eling, and Jan Hendrik Wirfs. Insurability of Cyber Risk: An Empirical Analysis. *Geneva Papers on Risk and Insurance: Issues and Practice*, 40(1):131–158, 1 2015. ISSN 14680440. doi: 10.1057/GPP.2014.19.

[14] Wil Van Der Aalst. Process mining. *Communications of the ACM*, 55(8):76–83, 8 2012. doi: 10.1145/2240236.2240257.

[15] W. M.P. Van der Aalst and A. J.M.M. Weijters. Process mining: A research agenda. *Computers in Industry*, 53(3):231–244, 2004. ISSN 01663615. doi: 10.1016/J.COMPIND.2003.10.001.

[16] Johan JC Tambotoh, Harjanto Prabowo, Sani M Isa, and Bonifasius Wahyu Pudjianto. PROCESS MINING IN GOVERNANCE, RISK MANAGEMENT, COMPLIANCE (GRC), AND AUDITING: A SYSTEMATIC LITERATURE REVIEW 1.2. *Journal of Theoretical and Applied Information Technology*, 99(18), 2021. ISSN 1817-3195.

[17] Erion Sula. SecRiskAI: A Machine Learning-based Tool for Cybersecurity Risk Assessment, Master Thesis, 7 2021. `https://files.ifi.uzh.ch/CSG/staff/franco/extern/theses/MA-E-Sula.pdf`.

[18] Angelica Marotta, Fabio Martinelli, Stefano Nanni, Albina Orlando, and Artsiom Yautsiukhin. Cyber-insurance survey. *Computer Science Review*, 24:35–61, 2017.

[19] Oliver Ralph. Companies face soaring prices for cyber insurance, 2022. `https://www.ft.com/content/60ddc050-a846-461a-aa10-5aaabf6b35a5`, (Last accessed January 2022).

[20] Carrier Managment. What Black Swan Author Nassim Taleb Has to Say About Insurance, 2015.

[21] Jean Bolot and Marc Lelarge. Cyber Insurance as an Incentive for Internet Security. In *Seventh Workshop on the Economics of Information Security, Hanover NH(USA), June 25-28*, pages 1–19, 2008.

[22] Martin Eling, Gregory Falco, Danielle Jablanski, and Virginia Miller. A Research Agenda for Cyber Risk and Cyber Insurance. *Conference: Workshop on the Economics of Information Security*, 2019.

[23] Rainer Böhme and Galina Schwartz. Modeling Cyber-Insurance: Towards A Unifying Framework. In *Workshop on the Economics of Information Security (WEIS), Harvard, June 2010*, 2010.

[24] Lawrence A. Gordon, Martin P. Loeb, and Tashfeen Sohail. A framework for using insurance for cyber-risk management. *Communications of the ACM*, 46(3):81–85, 3 2003. ISSN 00010782. doi: 10.1145/636772.636774.

[25] Merriam-Webster Dictionary. Cyber Definition & Meaning - Merriam-Webster, 2022. `https://www.merriam-webster.com/dictionary/cyber`, (Last accessed January 2022).

[26] Arunabha Mukhopadhyay, Samir Sadhukhan, Samir Chatterjee, Debashis Saha, Ambuj Mahanti, and Samir K Sadhukhan. Cyber-risk decision models: To insure IT or not? ? â. *Decision Support Systems*, 56:11–26, 2013. doi: 10.1016/j.dss.2013.04.004.

[27] Rainer Böhme and Gaurav Kataria. On the Limits of Cyber-Insurance. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4083 LNCS:31–40, 2006. ISSN 16113349. doi: 10.1007/11824633{_}4.

[28] European Commission. Risk management and supervision of insurance companies (Solvency 2). Technical report, European Commission, 2021. `https://ec.europa.eu/info/business-economy-euro/banking-and-finance/insurance-and-pensions/risk-management-and-supervision-insurance-companies-solvency-2_en`.

[29] Nadine Gatzert, Andreas Kolb, and Working Paper. Risk Measurement and Management of Operational Risk in Insurance Companies from an Enterprise Perspective. *The Journal of Risk and Insurance*, 2012.

[30] BIS. History of the Basel Committee, 2018. `https://www.bis.org/bcbs/history.htm`, (Last accessed January 2022).

[31] James J Cebula and Lisa R Young. A Taxonomy of Operational Cyber Security Risks CERT ® Program. Technical report, Software Engineering Institute, 2010. `http://www.sei.cmu.edu/library`.

[32] OECD. Encouraging Clarity in Cyber Insurance Coverage - The Role of Public Policy and Regulation, 2020. www.oecd.org/finance/insurance/Encouraging-Clarity-in-Cyber-Insurance-Coverage.pdf.

[33] Andrew Granato and Andy Polacek. The growth and challenges of cyber insurance. *Chicago Fed Letter*, 2019. doi: 10.21033/CFL-2019-426.

[34] Eiopa. Understanding Cyber Insurance-A Structured Dialogue with Insurance Companies. Technical report, The European Insurance and Occupational Pensions Authority, 2018.

[35] Paul Dreyer, Therese Jones, Kelly Klima, Jenny Oberholtzer, Aaron Strong, Jonathan William Welburn, and Zev Winkelman. *Estimating the Global Cost of Cyber Risk: Methodology and Examples*. 2018. www.rand.org/jie/stp.

[36] McAfee and CSIS. New McAfee Report Estimates Global Cybercrime Losses to Exceed $1 Trillion| McAfee Press Release. Technical report, McAfee Corp., 2020. https://www.mcafee.com/de-ch/consumer-corporate/newsroom/press-releases/press-release.html?news_id=6859bd8c-9304-4147-bdab-32b35457e629.

[37] Kelly Bissell, Ryan Lasalle, and Paolo Dal Cin. Cost of Cybercrime Study | 9th Annual | Accenture. Technical report, Accenture, 3 2019. URL https://www.accenture.com/us-en/insights/security/cost-cybercrime-study.

[38] Forbes. The Global 2000: How the World's Biggest Public Companies Endured the Pandemic, 2021. https://www.forbes.com/lists/global2000/#50d7180a5ac0.

[39] Justina Alexandra Sava. Cybersecurity spending worldwide 2021, 2022. https://www.statista.com/statistics/991304/worldwide-cybersecurity-spending/, (Last accessed January 2022).

[40] Gartner Press Release. Gartner Forecasts Worldwide Security and Risk Management Spending to Exceed $150 Billion in 2021, 2021. https://www.gartner.com/en/newsroom/press-releases/2021-05-17-gartner-forecasts-worldwide-security-and-risk-managem, (Last accessed January 2022).

[41] Steve Morgan. Cybercrime To Cost The World $10.5 Trillion Annually By 2025, 2020. https://cybersecurityventures.com/hackerpocalypse-cybercrime-report-2016/, (Last accessed January 2022).

[42] Steve Morgan. Global Ransomware Damage Costs Predicted To Reach $20 Billion (USD) By 2021, 2022. https://cybersecurityventures.com/global-ransomware-damage-costs-predicted-to-reach-20-billion-usd-by-2021/, (Last accessed January 2022).

[43] Swis Cyber Institute. 41 Insider Threat Statistics You Should Care About - Swiss Cyber Institute, 2021. https://swisscyberinstitute.com/blog/41-insider-threat-statistics-you-should-care-about/.

[44] Sarah Stephens. COVID-19: Next Steps for Your Cyber Insurance, 2022. https://www.marsh.com/ie/risks/pandemic/insights/covid-19-next-steps-cyber-insurance.html, (Last accessed January 2022).

[45] KPMG. Cyber Insurance - How Insuretechs Can Unlock The Opportunity. Technical report, KPMG, 2018.

[46] Munich Re. Cyber insurance: Risks and trends 2021, 2021. https://www.munichre.com/topics-online/en/digitalisation/cyber/cyber-insurance-risks-and-trends-2021.html, (Last accessed January 2022).

[47] AmTrust Financial. Cyber Insurance Market Growth, 2021. https://amtrustfinancial.com/blog/agents/growth-of-the-cyber-insurance-market-agents, (Last accessed January 2022).

[48] Stephan Binder, Philipp Klais, and Jörg Mußhoff. Global Insurance Pools statistics and trends: An overview of life, P&C, and health insurance, 2019. https://mck.co/3qxPkf9.

[49] Marsh. Cyber Insurance Market Overview: Fourth Quarter 2021, 2021. https://www.marsh.com/us/services/cyber-risk/insights/cyber-insurance-market-overview-q4-2021.html, (Last accessed January 2022).

[50] Zurich Switzerland. Zurich Cyber Insurance, 2022. https://www.zurich.ch/en/corporate-customers/property-cyber/cyber-insurance, (Last accessed January 2022).

[51] Sasha Romanosky, Lillian Ablon, Andreas Kuehn, and Therese Jones. Content analysis of cyber insurance policies: how do carriers price cyber risk? *Journal of Cybersecurity*, 5(1), 1 2019. ISSN 2057-2085. doi: 10.1093/CYBSEC/TYZ002.

[52] Baruch. Berliner. *Limits of Insurability of Risks*. Prentice-Hall, 1982. ISBN 0135367891.

[53] Baharuddin Aziz, Suhardi, and Kurnia. A systematic literature review of cyber insurance challenges. *2020 International Conference on Information Technology Systems and Innovation, ICITSI 2020 - Proceedings*, pages 357–363, 10 2020. doi: 10.1109/ICITSI50517.2020.9264966.

[54] ISO - ISO/IEC 27001. Information security management, 2022. https://www.iso.org/isoiec-27001-information-security.html, (Last accessed January 2022).

[55] Office for Civil Rights (OCR). Summary of the HIPAA Security Rule | HHS.gov. Technical report, U.S. Department of Health & Human Services website, 2013. https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html.

[56] PCI DSS. Official PCI Security Standards Council Site, 7 2021. https://www.pcisecuritystandards.org/.

[57] Sarbanes-Oxley (SOX). Sarbanes-Oxley (SOX) - KPMG Switzerland, 2022. https://home.kpmg/ch/en/home/services/audit/sarbanes-oxley.html.

[58] State of California - Department of Justice. California Consumer Privacy Act (CCPA) | State of California - Department of Justice - Office of the Attorney General, 2018. https://oag.ca.gov/privacy/ccpa, (Last accessed January 2022).

[59] Steve Beavers. U.S. Risk Cyber Insurance Application Form, 12 2017. https://www.usrisk.com/download/U.S.-Risk-Cyber-Insurance-Application-Form%25E2%2580%2593Short.pdf.

[60] Matt Prevost. How Do Cyber Insurers View the World, 9 2019.

[61] NAIC. SERFF: The System for Electronic Rates & Forms Filing, 5 2021. https://www.serff.com/.

[62] Rainer Boehme, S. Laube, and M. Riek. A Fundamental Approach to Cyber Risk Analysis | Semantic Scholar. *Variance Journal*, 2018.

[63] AIG. CyberMatics® | AIG US, 2020. https://www.aig.com/business/insurance/cyber/cybermatics, (Last accessed January 2022).

[64] Bitsight. Managing Vendor Risks. In *State of North Carolina 2019 Annual Cyber Awareness Symposium*, 2019.

[65] RMS. Risk Management Models, Analytics, Software & Services | RMS, 7 2021. https://www.rms.com/.

[66] Advisen Ltd. Specialty Risk Data Providers - Loss Data and Casualty Data | Advisen Ltd, 7 2021. https://www.advisenltd.com/data/.

[67] Chubb. Chubb Cyber Index, 6 2021. https://www.chubbcyberindex.com/#/splash.

[68] What you need to know about mandatory reporting of breaches of security safeguards. Technical report, Office of the Private Commissioner of Canada, 2021. https://www.priv.gc.ca/en/privacy-topics/business-privacy/safeguards-and-breaches/privacy-breaches/respond-to-a-privacy-breach-at-your-business/gd_pb_201810/, (Last accessed January 2022).

[69] Josephine Wolff. How the NotPetya attack is reshaping cyber insurance, 2021. https://www.brookings.edu/techstream/how-the-notpetya-attack-is-reshaping-cyber-insurance/, (Last accessed January 2022).

[70] Maochao Xu and Lei Hua. Cybersecurity Insurance: Modeling and Pricing. *https://doi.org/10.1080/10920277.2019.1566076*, 23(2):220–249, 4 2019. ISSN 10920277. doi: 10.1080/10920277.2019.1566076.

[71] James Bardopoulos. Cyber-insurance pricing models, 2018.

[72] Adam Barone and Charles Potters. Loss Ratio vs. Combined Ratio: What's the Difference?, 2021. `https://www.investopedia.com/ask/answers/042315/what-difference-between-loss-ratio-and-combined-ratio.asp`, (Last accessed January 2022).

[73] Matt Sheehan. Cyber industry loss ratio at record-high 67% in 2020: Aon - Reinsurance News, 2021. `https://www.reinsurancene.ws/cyber-industry-loss-ratio-at-record-high-67-in-2020-aon/`, (Last accessed January 2022).

[74] Tridib Bandyopadhyay and Vijay Mookerjee. A model to analyze the challenge of using cyber insurance. *Information Systems Frontiers*, 21(2):301–325, 4 2019. ISSN 15729419. doi: 10.1007/S10796-017-9737-3.

[75] Isaac Ehrlich and Gary S. Becker. Market Insurance, Self-Insurance, and Self-Protection. *Journal of Political Economy*, 80(4):623–648, 10 1972. ISSN 0022-3808. doi: 10.1086/259916.

[76] Gregory Falco, Martin Eling, Danielle Jablanski, and Virginia Miller. A Research Agenda for Cyber Risk and Cyber Insurance. *Conference: Workshop on the Economics of Information Security*, 6 2019.

[77] Muriel Figueredo Franco, Bruno Rodrigues, and Burkhard Stiller. MENTOR: The Design and Evaluation of a Protection Services Recommender System. *15th International Conference on Network and Service Management, CNSM 2019*, 10 2019. doi: 10.23919/CNSM46954.2019.9012686.

[78] Muriel Franco, Noah Berni, Eder Scheid, Christian Killer, Bruno Rodrigues, and Burkhard Stiller. SaCI: a Blockchain-based Cyber Insurance Approach for the Deployment and Management of a Contract Coverage. In *GECON 2021 - 18th International Conference on the Economics of Grids, Clouds, Systems and Services*, 2021.

[79] Wil Van der Aalst. Process mining: Data science in action. *Process Mining: Data Science in Action*, pages 1–467, 1 2016. doi: 10.1007/978-3-662-49851-4.

[80] Ruth Sara Aguilar-Savén. Business process modelling: Review and framework. *International Journal of Production Economics*, 90(2):129–149, 7 2004. ISSN 0925-5273. doi: 10.1016/S0925-5273(03)00102-6.

[81] Wil Van Der Aalst and Ernesto Damiani. Processes Meet Big Data: Connecting Data Science with Process Science. *IEEE Transactions on Services Computing*, 8 (6):810–819, 11 2015. ISSN 19391374. doi: 10.1109/TSC.2015.2493732.

[82] IGI Global. What is Process-Aware Information System (PAIS), 2022. `https://www.igi-global.com/dictionary/dynamic-context-aware-process-adaptation/23620`.

[83] Wil M.P. Van Der Aalst. Process-Aware Information Systems: Lessons to Be Learned from Process Mining. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5460 LNCS:1–26, 2009. ISSN 03029743. doi: 10.1007/978-3-642-00899-3{_}1.

[84] Josep Carmona, Boudewijn van Dongen, Andreas Solti, and Matthias Weidlich. Conformance checking: Relating processes and models, relating processes and models. *Conformance Checking: Relating Processes and Models, Relating Processes and Models*, pages 1–277, 11 2018. doi: 10.1007/978-3-319-99414-7.

[85] A. Maña, J. A. Montenegro, C. Rudolph, and J. L. Vivas. A business process-driven approach to security engineering. *Proceedings - International Workshop on Database and Expert Systems Applications, DEXA*, 2003-January:477–481, 2003. ISSN 15294188. doi: 10.1109/DEXA.2003.1232069.

[86] Rayford B. Vaughn, Ronda Henning, and Kevin Fox. An empirical study of industrial security-engineering practices. *Journal of Systems and Software*, 61(3):225–232, 4 2002. ISSN 01641212. doi: 10.1016/S0164-1212(01)00150-9.

[87] Gaby Herrmann and GÃ¼nther Pernul. Viewing Business-Process Security from Different Perspectives. *International Journal of Electronic Commerce*, 3(3):89–103, 2015. ISSN 10864415. doi: 10.1080/10864415.1999.11518343.

[88] Michael Hafner, Ruth Breu, Berthold Agreiter, and Andrea Nowak. SECTET-An Extensible Framework for the Realization of Secure Inter-Organizational Workflows. *Internet Research, Bradford*, 16(5):491–506, 2006.

[89] Christian Wolter, Michael Menzel, Andreas Schaad, Philip Miseldine, and Christoph Meinel. Model-driven business process security requirement specification. *Journal of Systems Architecture*, 55(4):211–223, 4 2009. ISSN 1383-7621. doi: 10.1016/J.SYSARC.2008.10.002.

[90] Thomas Neubauer, Markus Klemen, and Stefan Biffl. Business process-based valuation of IT-security. *7th International Workshop on Economics-Driven Software Engineering Research, EDSER 2005 - International Conference on Software Engineering 2005*, 5 2005. doi: 10.1145/1083091.1083099.

[91] PitchBook. Companies in Germany with the highest venture valuations, 2021. `https://bit.ly/3iPVsvb`.

[92] Scott Carey. SAP finally joins the process mining fray, 2021. `https://www.cio.com/article/188770/sap-finally-joins-the-process-mining-fray.html`.

[93] ProcessGold Acquisition. ProcessGold Acquisition News - E2E Process Aware | UiPath, 2019. `https://www.uipath.com/newsroom/uipath-acquires-process-gold-unparalleled-process-understanding`.

[94] Martin Macak, Lukas Daubner, Mohammadreza Fani Sani, and Barbora Buhnova. Process mining usage in cybersecurity and software reliability analysis: A systematic literature review. *Array*, 13:100120, 3 2022. ISSN 2590-0056. doi: 10.1016/J.ARRAY.2021.100120.

[95] Cleiton dos Santos Garcia, Alex Meincheim, Elio Ribeiro Faria Junior, Marcelo Rosano Dallagassa, Denise Maria Vecino Sato, Deborah Ribeiro Carvalho, Eduardo Alves Portela Santos, and Edson Emilio Scalabrin. Process mining techniques and applications â A systematic mapping study. *Expert Systems with Applications*, 133:260–295, 11 2019. ISSN 09574174. doi: 10.1016/J.ESWA.2019.05.003.

[96] Mieke Jans, Michael Alles, and Miklos Vasarhelyi. The case for process mining in auditing: Sources of value added and areas of application. *International Journal of Accounting Information Systems*, 14(1):1–20, 3 2013. ISSN 1467-0895. doi: 10.1016/J.ACCINF.2012.06.015.

[97] Timotheus Kampik and Mathias Weske. Event Log Generation: An Industry Perspective. 2 2022. doi: 10.48550/arxiv.2202.02539. `https://arxiv.org/abs/2202.02539v1`.

[98] PM4Py. Process Mining for Python, 2022. `https://pm4py.fit.fraunhofer.de/documentation#evaluation`.

[99] A. Rozinat and W. M.P. van der Aalst. Conformance checking of processes based on monitoring real behavior. *Information Systems*, 33(1):64–95, 3 2008. ISSN 0306-4379. doi: 10.1016/J.IS.2007.07.001.

[100] Andre Filipe Domingos Gomes, Ana Cristina Wanzeller Guedes de Lacerda, and Joana Rita da Silva Fialho. Comparative Analysis of Process Mining Algorithms in Python. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, 401 LNICST:27–43, 9 2021. ISSN 1867822X. doi: 10.1007/978-3-030-91421-9{_}3.

[101] A Adriansyah, B F Van Dongen, and W M P Van Der Aalst. Conformance Checking using Cost-Based Fitness Analysis. In *Proceedings - IEEE International Enterprise Distributed Object Computing Workshop, EDOC*, pages 55–64, 2011. `http://www.bpmn.org/`.

[102] Jan Svacina, Jackson Raffety, Connor Woodahl, Brooklynn Stone, Tomas Cerny, Miroslav Bures, Dongwan Shin, Karel Frajtak, and Pavel Tisnovsky. On Vulnerability and Security Log analysis: A Systematic Literature Review on Recent Trends. *Proceedings of the International Conference on Research in Adaptive and Convergent Systems*, 2020. doi: 10.1145/3400286.

[103] Adrian Baldwin. The Layers within an enterprise architecture - Scientific Diagram, 2022. `https://www.researchgate.net/figure/The-Layers-within-an-enterprise-architecture_fig1_242404798`.

[104] Robert Kelemen. Systematic review on process mining and security. *Central and Eastern European eDem and eGov Days*, 325:145–164, 3 2017. ISSN 2663-9394. doi: 10.24989/OCG.V325.13.

[105] W. M.P. Van Der Aalst and A. K.A. De Medeiros. Process Mining and Security: Detecting Anomalous Process Executions and Checking Process Conformance. *Electronic Notes in Theoretical Computer Science*, 121(SPEC. ISS.):3–21, 2 2005. ISSN 1571-0661. doi: 10.1016/J.ENTCS.2004.10.013.

[106] Fabio Bezerra and Jacques Wainer. Algorithms for anomaly detection of traces in logs of process aware information systems. *Information Systems*, 38(1):33–44, 2013. ISSN 03064379. doi: 10.1016/J.IS.2012.04.004.

[107] Riyanarto Sarno, Fernandes Sinaga, and Kelly Rossa Sungkono. Anomaly detection in business processes using process mining and fuzzy association rule learning. *Journal of Big Data 2020 7:1*, 7(1):1–19, 1 2020. ISSN 2196-1115. doi: 10.1186/S40537-019-0277-1.

[108] Fabio Bezerra, Jacques Wainer, and W. M. P. van der Aalst. Anomaly Detection Using Process Mining. *Lecture Notes in Business Information Processing*, 29 LNBIP: 149–161, 2009. doi: 10.1007/978-3-642-01862-6{_}13.

[109] Ala Bahrani and Amir Jalaly Bidgly. Ransomware detection using process mining and classification algorithms. *Proceedings of 16th International ISC Conference on Information Security and Cryptology, ISCISC 2019*, pages 73–77, 8 2019. doi: 10.1109/ISCISC48546.2019.8985149.

[110] Ved Prakash Mishra and Balvinder Shukla. Process Mining in Intrusion Detection-The Need of Current Digital World. *Communications in Computer and Information Science*, 712:238–246, 2017. doi: 10.1007/978-981-10-5780-9{_}22.

[111] Guenther Eibl, Cornelia Ferner, Tobias Hildebrandt, Florian Stertz, Sebastian Burkhart, Stefanie Rinderle-Ma, and Dominik Engel. Exploration of the potential of process mining for intrusion detection in smart metering. *ICISSP 2017 - Proceedings of the 3rd International Conference on Information Systems Security and Privacy*, 2017-January:38–46, 2017. doi: 10.5220/0006103900380046.

[112] Rafael Accorsi and Thomas Stocker. On the exploitation of process mining for security audits: The conformance checking case. *Proceedings of the ACM Symposium on Applied Computing*, pages 1709–1716, 2012. doi: 10.1145/2245276.2232051.

[113] Sofiane Lagraa and Radu State. Process mining-based approach for investigating malicious login events. *Proceedings of IEEE/IFIP Network Operations and Management Symposium 2020: Management in the Age of Softwarization and Artificial Intelligence, NOMS 2020*, 4 2020. doi: 10.1109/NOMS47738.2020.9110301.

[114] Adrien Hemmer, Remi Badonnel, Isabelle Chrisment, and Remi Badonnel. A Process Mining Approach for Supporting IoT Predictive Security. In *OMS 2020 - IEEE/IFIP Network Operations and Management Symposium,Apr 2020, Budapest, Hungary.*, 4 2020.

[115] Simona Bernardi, Raul Piraces Alastuey, and Raquel Trillo-Lado. Using Process Mining and Model-driven Engineering to Enhance Security of Web Information Systems. In *2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2017. `http://mayor2.dia.fi.upm.es/oeg-upm/`.

[116] Ved Prakash Mishra, Joanita Dsouza, and Laura Elizabeth. Analysis and Comparison of Process Mining Algorithms with Application of Process Mining in Intrusion

Detection System. *2018 7th International Conference on Reliability, Infocom Technologies and Optimization: Trends and Future Directions, ICRITO 2018*, pages 613–617, 8 2018. doi: 10.1109/ICRITO.2018.8748748.

[117] Maikel L. Van Eck, Xixi Lu, Sander J.J. Leemans, and Wil M.P. Van Der Aalst. PM2: A Process Mining Project Methodology. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9097:297–313, 2015. ISSN 16113349. doi: 10.1007/978-3-319-19069-3{_}19.

[118] Wil M.P. Van Der Aalst, Kees M. Van Hee, Jan Martijn Van Der Werf, and Marc Verdonk. Auditing 2.0: Using process mining to support tomorrow's auditor. *Computer*, 43(3):90–93, 3 2010. ISSN 00189162. doi: 10.1109/MC.2010.61.

[119] Zabihollah Rezaee, Rick Elam, and Ahmad Sharbatoghlie. Continuous auditing: The audit of the future. *Managerial Auditing Journal*, 16(3):150–158, 4 2001. ISSN 02686902. doi: 10.1108/02686900110385605/FULL/XML.

[120] Pierluigi Zerbino, Davide Aloini, Riccardo Dulmin, and Valeria Mininno. Process-mining-enabled audit of information systems: Methodology and an application. *Expert Systems with Applications*, 110:80–92, 11 2018. ISSN 0957-4174. doi: 10.1016/J.ESWA.2018.05.030.

[121] Nick Gehrke. Basic Principles of Financial Process Mining A Journey through Financial Data in Accounting Information Systems. *AMCIS 2010*, 8 2010.

[122] Atastina Imelda and Kurniati Angelina. Implementing process mining to improve COBIT 5 assessment program or managing operations (Case study: A university blog). *Journal of Theoretical and Applied Information Technology 72(2)*, 72(2): 191–198, 2015.

[123] COBIT. Control Objectives for Information Technologies, ISACA Framework, 2019. `https://www.isaca.org/resources/cobit`, (Last accessed January 2022).

[124] Martin MacAk, Ivan Vanat, Michal Merjavy, Tomas Jevocin, and Barbora Buhnova. Towards Process Mining Utilization in Insider Threat Detection from Audit Logs. *2020 7th International Conference on Social Network Analysis, Management and Security, SNAMS 2020*, 12 2020. doi: 10.1109/SNAMS52053.2020.9336573.

[125] Eduard Šrol. Process mining usage for potential insider threat identification utilizing PM4Py, 2020.

[126] Alessandro Berti, Sebastiaan J. Van Zelst, Wil M.P. Van Der Aalst, and Fraunhofer Gesellschaf. Process Mining for Python (PM4Py): Bridging the Gap Between Process- and Data Science. *CEUR Workshop Proceedings*, 2374:13–16, 5 2019. ISSN 16130073. `https://arxiv.org/abs/1905.06169v1`.

[127] B. F. Van Dongen, A. K.A. De Medeiros, H. M.W. Verbeek, A. J.M.M. Weijters, and W. M.P. Van Der Aalst. The ProM Framework: A New Era in Process Mining Tool Support. *Lecture Notes in Computer Science*, 3536:444–454, 2005. ISSN 03029743. doi: 10.1007/11494744{_}25.

[128] Aynesh Sundararaj, Silvia Knittl, and Jens Grossklags. Challenges in IT Security Processes and Solution Approaches with Process Mining. *Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12386 LNCS:123–138, 2020. ISSN 16113349. doi: 10.1007/978-3-030-59817-4{_}8.

[129] Rafael Accorsi, Thomas Stocker, and GÃ¼nter Müller. On the Exploitation of Process Mining for Security Audits: The Process Discovery Case. In *Proceedings of the ACM Symposium on Applied Computing*, pages 1462–1468, 2013. doi: 10.1145/2480362.2480634.

[130] Vijayalakshmi Atluri and Janice Warner. Security for workflow systems. *Handbook of Database Security: Applications and Trends*, pages 213–230, 2008. doi: 10.1007/978-0-387-48533-1{_}9.

[131] Chang Jae Kang, Young Sik Kang, Yeong Shin Lee, Seonkyu Noh, Hyeong Cheol Kim, Cheol Lim, Juhee Kim, and Regina Hong. Process Mining-based Understanding and Analysis of Volvo IT's Incident and Problem Management Processes. In *The BPI Challenge 2013*, 2013. ISBN 8223000776.

[132] Diogo R Ferreira and Miguel Mira da Silva. Using process mining for ITIL assessment: a case study with incident management, 2008. `https://www.researchgate.net/publication/228816567`.

[133] Jorge Munoz-Gama and Isao Echizen. Insuring sensitive processes through process mining. *Proceedings - IEEE 9th International Conference on Ubiquitous Intelligence and Computing and IEEE 9th International Conference on Autonomic and Trusted Computing, UIC-ATC 2012*, pages 447–454, 2012. doi: 10.1109/UIC-ATC.2012.83.

[134] Marc Lelarge and Jean Bolot. Economic Incentives to Increase Security in the Internet: The Case for Insurance. In *Proceedings - IEEE INFOCOM*, pages 1494–1502, 2009.

[135] Chubb Cyber. Chubb Cyber Enterprise Risk Management Policy Chubb DigiTech ® Enterprise Risk Management Policy. Technical report, NAIC, 2018. `https://filingaccess.serff.com/sfa/sessionExpired.xhtml`.

[136] Wikipedia. Figma (software) - Wikipedia, 2022. `https://en.wikipedia.org/wiki/Figma_(software)`, (Last accessed January 2022).

[137] React. A JavaScript library for building user interfaces, 2022. `https://reactjs.org/`, (Last accessed January 2022).

[138] Facebook Open Source. Create React App, 2022. `https://create-react-app.dev/`, (Last accessed January 2022).

[139] BPMN. BPMN Specification - Business Process Model and Notation, 2019. `https://www.bpmn.org/`, (Last accessed January 2022).

[140] IEEE Task Force on Process Mining. IEEE 1849-2016 XES Standard, 2022. `https://xes-standard.org/`.

[141] TS. graphviz-react - npm, 2021. https://www.npmjs.com/package/graphviz-react, (Last accessed January 2022).

[142] Pallets. Extensions â Flask Documentation (2.0.x), 2010. https://flask.palletsprojects.com/en/2.0.x/extensions/, (Last accessed January 2022).

[143] Django. The web framework for perfectionists with deadlines, 2022. https://www.djangoproject.com/.

[144] Flask. Quickstart â Flask Documentation (2.0.x), 2010. https://flask.palletsprojects.com/en/2.0.x/quickstart/.

[145] Flask-RESTX. Flask-RESTX 0.5.2.dev documentation, 2020. https://flask-restx.readthedocs.io/en/latest/, (Last accessed January 2022).

[146] Flask-RESTPlus. Flask-RESTPlusâs documentation! â Flask-RESTPlus 0.13.0 documentation, 2014. https://flask-restplus.readthedocs.io/en/stable/, (Last accessed January 2022).

[147] apryor6. Easy, opinionated Flask input/output handling mixing Marshmallow with flask-restx, 2022. https://github.com/apryor6/flask_accepts, (Last accessed January 2022).

[148] Steven Loria. marshmallow: simplified object serialization â marshmallow 3.15.0 documentation, 2021. https://marshmallow.readthedocs.io/en/stable/, (Last accessed January 2022).

[149] PostgreSQL. PostgreSQL: The world's most advanced open source database, 2022. https://www.postgresql.org/.

[150] Pallets. Flask-SQLAlchemy â Flask-SQLAlchemy Documentation (2.x), 2010. https://flask-sqlalchemy.palletsprojects.com/en/2.x/, (Last accessed January 2022).

[151] SQLAlchemy. SQLAlchemy - The Database Toolkit for Python, 2022. https://www.sqlalchemy.org/.

[152] psycopg2. PyPI, 2022. https://pypi.org/project/psycopg2/.

[153] MacakM. lasaris/Security-audit-logs-for-Process-Mining, 2020. https://github.com/lasaris/Security-audit-logs-for-Process-Mining, (Last accessed January 2022).

[154] Iezalde F. Lopes and Diogo R. Ferreira. A Survey of Process Mining Competitions: The BPI Challenges 2011â2018. *Lecture Notes in Business Information Processing*, 362 LNBIP:263–274, 2019. ISSN 18651356. doi: 10.1007/978-3-030-37453-2{_}22.

[155] Annne Rozinat, Alves A k Modeitos, CW Guenther, and AJMM Weijters. (PDF) Towards an evaluation framework for process mining algorithms. *Reactivity of Solids*, 2007.

Appendix A

Installation Guidelines

The repository contains a **README.md** that also provides further details on how to set-up the prototype. The prerequisite for the installation of the prototype to have access to the GitHub and also the latest version of Docker (`https://www.docker.com`) installed.

The steps can be summarised as follows:

Step 1: Access Github and clone the repository at `https://github.com/viktor-matejka/meritminer4cyberinsurance.git`.

The repository also contains documentation on the installation.

Step 2: Once you have cloned the repository, navigate to the **./meritminer4cyberinsurance** folder and execute the following steps. Create .env file using the **.env.example** in the folder as template, by removing the .example extension.

Then, in your terminal, execute the following commands:

```
docker-compose up -d
```

Listing A.1: Run docker-compose

```
docker-compose run api sh -c "flask db upgrade"
```

Listing A.2: Create database

```
docker-compose run api sh -c "python manage.py seed_db"
```

Listing A.3: Create first user for the database

```
docker-compose run api sh -c "python manage.py seed_profiles"
```

Listing A.4: Prefill Database with Example Profiles

Now you should be able to access the user layer (frontend) here running at (`http://127.0.0.1:3001`). Accessing (`http://127.0.0.1:8000`) from the browser then displays the Swagger documentation of the API. Datasets for testing are available in the **examples-for-prototype-testing** folder. Apart from .XES files reflecting event logs, a .BPMN process is available to test the conformance checking functionality of the IAM scenario.

Appendix B

Contents of the Repository

The repository contains the following content:

1. Application source code
2. In the additional content folder, you can find

- Configurations used for synthetic log generation
- Datasets for prototype testing
- Datasets for quantitative testing with evaluation script
- Implementation of the MeritMinerCI Approach BPMN
- Script used for quality metrics in quantitative evaluation
- Survey questionnaire and data
- Modelled reference process for the IAM conformance chacking use case
- Underwriting Manual for reference regarding confidence factor rating