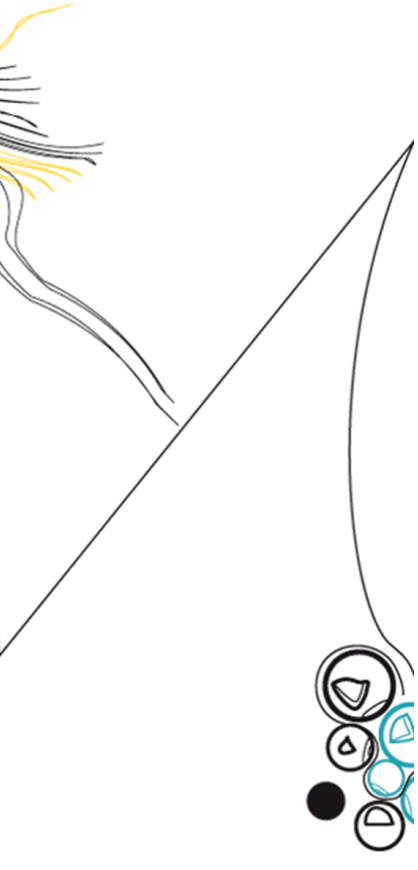
UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering, Mathematics & Computer Science

Analytic-Driven Decision Support in Cybersecurity : Towards Effective IP Risk Management Decision Making Process

Rutuja Dolas

M.Sc. Final Thesis
in collaboration with



Supervisors:
Han Gijsbers (Cyber Risk Manager, NXP)

dr. Abhishta A. (Abhishta)

dr. Bukhsh F.A. (Faiza)

Faculty of Electrical Engineering,
Mathematics and Computer Science
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Acknowledgements

The past two years have been an incredible journey at the University of Twente. It's been a time of great ups and downs, new challenges, and discoveries. However, I've grown a lot during this time, both personally and professionally and I'll always be grateful for that. Throughout this journey, I have had the opportunity to work with incredible people and have benefited greatly from their advice and experience, and I would like to take this opportunity to express my heartfelt gratitude to those who assisted me and contributed in various ways and aspects.

First and foremost, I'd like to thank my supervisor Han Gijsbers and my colleagues Peter Valent, Olivier, Arun, and Eef from the IRM cybersecurity team in NXP Semiconductors N.V. for believing in me and mentoring me during my thesis internship period. Special thank you to Sam Flower and Bettina Nyeste for helping me out in this whole process with my doubts and setting up technical things for me. Thank you all for your help and for being such amazing mentors and friends. I really enjoyed working with you all.

Next, I'd like to thank my university supervisors for the thesis, Prof. dr. Abhishta A. and Prof. dr. Bukhsh F.A., for their unending support, encouragement, and valuable insights throughout this entire process. They both gave me sound advice and always pointed me in the right direction.

Most importantly, I would like to thank my parents for believing in me and encouraging me to pursue a master's degree abroad and my brother for pushing me beyond my limits. And a special thanks to my friends for their constant encouragement and my boyfriend for his unending love and support throughout this journey.

Rutuja Dolas
Enschede, January 11, 2023

Document Change Control

Document Version	Revision Date	Description of Change
Version 1.0	14-09-2022	Addition of Chapters 1 and 2 from SLR. Added sections - company background, and problem statement in Chapter 1
Version 2.0	03-10-2022	Addition of research methodology, contributions, and key deliverables in Chapter 1. Started with Chapter 3 - Addition of analytic maturity model, high-level kri goals, security analytics architecture
Version 3.0	17-10-2022	Completion of Chapter 4
Version 4.0	31-10-2022	Changes to Chapters 1,2,3,4,5 as per feedback
Version 5.0	14-11-2022	Changes to Chapters 1,2,3,4,5 as per feedback and Addition of Chapter 6
Version 6.0	28-11-2022	Changes to Chapters 4,5,6 as per feedback and addition of Chapter 7,8
Version 7.0	12-12-2022	Finalizing all Chapters for pre-green light
Version 8.0	15-12-2022	Restructuring last 3 Chapters as per feedback and submission of final thesis document for green light

Abbreviations

DLP - Data Loss Prevention
DSRM - Design Science Research Methodology
FBA - Forcepoint User and Entity behavior analytics
IP - Intellectual Property
IRM - Insider Risk Management
SLR - Systematic Literature Review
UBA - User and Behaviour Analytics
UEBA - User Entity Behaviour Analytics

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

This chapter includes motivation, information about the collaborative company, and background information to understand the scope and context of the problem statement. Finally, an overview of the research objective and proposed research questions followed by methodology and thesis outline.

1.1 Motivation

The digital revolution of the business ecosystem has created new opportunities for organizational growth and transformation across industries. As cybersecurity threats emerge and proliferate, the digital revolution has developed new risks to business operations. The majority of government and business decision-makers are related to external cyber-attacks such as unauthorized network access, distributed denial of service attacks, viruses, trojan horses, worms, etc. Around 10% of their IT budget is spent on cyber security to secure their assets to defend the networks from outside assault [1]. However, both external and insider threats are significant, with insider attacks causing more damage than outsider attacks. This means that any insider who has authorized access to an organization's intellectual property (IP) puts the organization's IP security at risk.

With the recent Covid-19 pandemic, the global work landscape has been altered forever. More than 70% of the employees worked remotely. According to Gartner, [2], after the pandemic, nearly half of all businesses intended to permanently integrate remote work into their corporate cultures. While this may be good news for employees, it is a nightmare for cybersecurity and information technology professionals. Insiders are the most vulnerable component of the security landscape, as they acquire intellectual property, personal customer information, employee information, or trade secrets. Therefore, telecommuting and remote work pose significant security challenges for organizations [2].

According to IBM and Morning Consult study, [3] on Work from Home (WFH), 53% of employees working remotely utilize unreliable tools and personal devices for work. This occurs when employees work remotely with little to no supervision. In addition to that, some employees may use unsecured public WiFi networks increasing the opportunity for hackers, which eventually results in a threat to an organization.

Insider attacks are the most expensive type of information security breach according to IBM's data breach 2022 report [4], with an average cost per insider incident of USD 4.91 million. As a result, insider threats are particularly damaging and also difficult to manage because they are committed by individuals who already have access to sensitive information. Insiders have a competitive advantage over external attackers because they know where the data is and how to obtain it. To protect the enterprise perimeter, security and risk management leaders must understand and address the threat of insider risk. As a result, strong tactics, techniques, and procedures should be used to address these issues. This new wave of change initiated even more improvements in cyber security procedures concerning insiders including appropriate risk analysis, mitigation methods, business processes, new policies, and employee education.

1.2 Company Information

This thesis work has been carried out in collaboration with NXP Semiconductors N.V., within their IT cyber security department. NXP is a semiconductor designer and manufacturer based company in Eindhoven, Netherlands. They develop technology solutions for the automotive, industrial, internet of things, mobile, and communication infrastructure markets. Within NXP, the IT cyber security team is responsible for Information technology security, IP protection & compliance, and Identity & access management. Their main focus is on the protection and cyber resilience of three main business risks viz., revenue generation, competitiveness, brand, and loyalty.

As a result, this research was performed for the Insider Risk Management (IRM) cyber security team within NXP, which aims to enable data analytic capabilities within cyber security to make better decisions while managing insider risks.

1.2.1 Background

NXP has initiated a program to protect its Intellectual Property (IP). Their goal is to strengthen their capabilities to prevent, detect, and respond to any misuse or misconduct of NXP Intellectual Property by insiders that can have a negative impact on NXP's market position, regulatory compliance, reputation, operations, supply chain,

and/or financial risk. NXP wants to protect its IP from being exfiltrated by insiders. To prevent this from happening, NXP has started an Insider Risk Management (IRM) program, that monitors employees who work with IP. To enable the monitoring capability within NXP on IP protection and Insider risk an Employee Behavior Monitoring solution is used, known as Forcepoint User and Entity Behavior Analytics (FBA).

FBA is a type of User and Entity Behavior Analytics (UEBA) real-time cybersecurity solution that uses analytics to achieve an understanding of how users (humans) and entities (servers and networked devices) in an organization generally behave in order to detect and respond to anomalous activity. The platform combines structured and unstructured data to provide a comprehensive view of complex human activity, habits, and long-term trends that constitute a human risk. As a result, FBA aids in identifying potential sources of data exfiltration, critical IP loss, and preventing bad actors from accessing critical assets and systems by understanding attributes such as typical access patterns.

1.3 Problem Statement

Provided the background information and context of the scope, after a systematic literature review, the following arguments form the problem of this research. Even though the monitoring solution (UEBA) has a lot to offer, it also has its challenges.

Firstly, from the literature one of the drawbacks of UEBA, is that it cannot identify long-term sophisticated low and slow attacks because they have no day-to-day impact and appear to be non-existent. It prioritizes only the most reliable and high-level risk security alerts. Using machine learning techniques involves many drawbacks as well such as the limitation of dealing with users and developers who are privileged and also knowledgeable insiders. These users have special cases because their job functions frequently require unusual behaviors, which makes it difficult to create a baseline for the algorithms.

In practice, the monitoring software lacks the ability to create insight into general behaviors within the organization. For example, an overview of IP data movement within or outside the organizations, type of devices or software used to transfer IP data, one-time or recurring user policy violation, etc. Such trend analytic insights are not possible within NXP's Forcepoint user and entity behavior analytics (FBA) solution since it prioritizes the most reliable and high-level risk security alerts [5]. As a result, such analysis is frequently done manually within the NXP cyber security team, with limited success. However, low-risk events, on the other hand, and detailed insights into them, can play an important role in better decision-making. These insights can help businesses make better decisions over time, such as determining whether training, cybersecurity policies, or other tools are required to mitigate

the risk of intellectual property loss. Creating insights on such broad organizational behaviors remains difficult with the current monitoring solution. In addition, due to NXP legal policies regarding FBA, some functions of the monitoring solution, such as natural language processing, text identification technology, etc are restricted. As a result, their ability to detect malicious behavior at an early stage is limited.

To overcome these difficulties enterprises need a new cybersecurity strategy. Therefore, this empirical research focuses on the risk associated with insiders and seeks to analyze and validate the results shown in the literature with the current practices, extending the reach of the research and improving current capabilities that assist organizations in mitigating the risk level of insider threat by combining the existing monitoring solution with an analytically driven approach to help businesses make better decisions regarding insider risk IP management.

1.4 Research Objective

The high-level goal of this study is to design and enable data analytic capabilities as decision support in cyber security to manage insider IP risks. Hereby, the main research question can be defined as:

How can organizations make better decisions by adopting an analytically driven strategy in cyber security to manage and minimize the risks from the insiders?

The main research objective for this thesis is: *To deliver a solution using an analytics-driven decision support approach that can be used in combination with current user behavioral monitoring systems in cybersecurity to assist organizations in making better decisions to manage and minimize the risk from insiders.*

1.4.1 Knowledge questions

The following knowledge questions for our research have been developed to fulfill the aforementioned research objective while also allowing for further discussion based on the findings.

Research Objective 1 (RO1): To develop an effective solution and achieve the desired results, it is essential to conduct research on how insiders are defined, their causes, and the types of threats insiders are capable of posing. Therefore, the first research question is as follows:

RQ1: What are the security threats posed by insiders?

SQ1.1: How is an insider defined?

SQ1.2: What are the causes of the insider threat and risk problem?

SQ1.3: Which type of threats are insiders capable of making?

Research Objective 2 (RO2): In recent years, particularly following the pandemic due to WFH culture, may have heightened the risk that careless or malicious insiders can pose to crucial assets and data. In addition to that, the enormous economic uncertainty, and the loss of important assets such as research and innovation, trade secrets, and critical materials can inhibit an organization's recovery. Also, organizations can face difficulties to combat and manage insider threats due to insufficient knowledge among the employees or improper tools. Therefore, it is important to learn about the recent cyber risks posed by insiders to an organization's Intellectual Property (IP), its consequences, and the challenges organizations face when dealing with insider risk management.

RQ2: Which recent cyber risks to intellectual property have been posed by insiders in an organization?

SQ2.1: What is the impact on the organization when it comes to insider attacks on their intellectual property?

SQ2.2: What are the challenges faced by organizations to combat & manage insider risks?

Research Objective 3 (RO3): To prevent, detect, and respond to potential insider threats, organizations have become more proactive. Various mitigation solutions are initiated to stop the forward momentum of asset exploitation, which can range from data exfiltration or system compromise to sabotage or even workplace violence. As a result, we need to know more about the existing decision-making solutions that are used in organizations as well as their limitations.

RQ3: What are the existing decision-making analytical solutions used in an organization to manage insider risks?

SQ3.1: What are the limitations of the current decision-making analytical solutions to manage insider risks?

1.4.2 Design questions

Research Objective 4 (RO4): To enable and design an analytically driven decision support solution that can be used in combination with current user behavioral monitoring systems in cybersecurity to assist organizations in making better decisions

while managing and minimizing the risk from insiders. Following that, the proposed solution needs to be evaluated along with the results. Therefore we propose the following design question:

RQ4: How to design and implement an analytic-driven solution to effectively support cybersecurity decisions in order to manage insider risks ?

Lastly, to evaluate the solution the sub-questions are stated as follows:

SQ4.1: How successfully does the proposed solution provide value to the organization's business in managing insider risks?

SQ4.2: What kinds of cybersecurity decisions and strategies can be made after analyzing the results from the proposed solution to manage insiders who pose a risk to the organization?

RQ1, RQ2, and RQ3 aim to understand the background of how insiders are defined, their causes, consequences, and challenges faced by organizations in managing insider risks and existing decision-making solutions while RQ4 is intended to reflect on how to design and enable an analytic-driven solution in cybersecurity followed by the evaluation of the results.

1.5 Scientific and Practical Contribution

The research is relevant from the perspective of practitioners and researchers. These contributions aim to improve the current knowledge gap in literature and practice.

1.5.1 Scientific contribution

1. Several IS research schools and individual scholars are focused on insider threats and how to mitigate the threats. However, to the best of our knowledge, there has not been much progress in insider risk management, which refers to strategies for managing risks after the data has been exposed by the insiders. This research will be an adding value to the research and organizations to make better decisions while managing insider risks effectively.

2. Furthermore, this research proposes an analytics maturity model and a sustainable security data analytics architecture, which considers the perspectives of different stakeholders and provides guidelines for data analytics and cybersecurity practitioners.

3. It contributes to knowledge by proposing a solution that includes using elements from the existing monitoring solutions and combining them in a way that evaded so far the attention of scholars. This solution was evaluated by means of experts and business people's opinions regarding its usefulness and usability.

4. At last, we further figure out a number of open issues and propose future research directions to motivate security analytics IP risk management research.

1.5.2 Practical contribution

Insider threats and risks have gained prominence in recent years as a result of their high value in times of pandemic. Reducing business risk caused by someone with authority over systems and data can be a challenging issue for any firm. Therefore, the analytical-driven design and approach proposed in this research will help the organizations to:

1. Examine the added value of analytical-driven capabilities in cyber security.

2. Analytics maturity model will help the organization to understand the current state of analytics and what steps need to be taken to take analytics to a new level.

3. Sustainable data analytics architecture and KRI's will enable cybersecurity insider risk management to leverage current and future efforts in managing insider risks.

4. The implemented models and business reports will help in assisting decision-makers to make better decisions, strategies, and policies in an organization to prevent such risks from occurring.

5. Proposed risk matrix will help in prioritizing the risk to make better decisions.

6. Better visibility on IP movement from the organization and overview trend analysis related to insider risk activities.

1.6 Research Methodology

This research adheres to the Design Science Research Methodology (DSRM) defined by Peffers et al. (2007). Figure 1.1 shows five steps: identification of the problem and motivation, defining objectives of the solution, design and development, demonstration, evaluation, and communication.

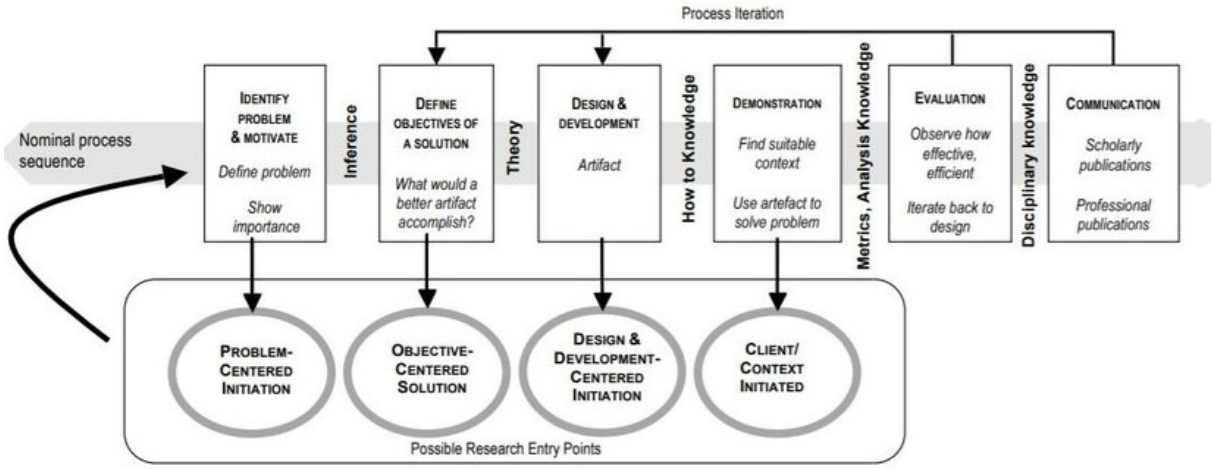


Figure 1.1: Design science research methodology process (DSRM) model

1. **Problem identification and motivation**:
   The first step of the DSRM process is to identify the problem and justify it with motivation. The thesis aims to provide a clear overview of the problem identification and motivation, which can be found in Chapter 1, as well as a factual investigation, which occurs in Chapter 2.

2. **Define the objectives for a solution**: The research objectives for the problem identified in the first phase are determined. These objectives can be qualitative or quantifiable, illustrating how the proposed solution outperforms current ones or describing how the suggested technique can aid in resolving problems that have never been addressed previously, as seen in Section 1.4. At this stage, a systematic literature review is also performed, with the goal of aggregating all current content for the research question and objectives and assisting in developing evidence-based guidelines for practitioners.

   As per the guidelines discussed in [6], the following steps are carried out for a systematic literature review:

   (a) Formulating research questions and objectives: The first step is to formulate well-defined research questions and objectives.

   (b) Developing the search:

      i. Selection of library & formulating the search string(s).

      ii. Gathering literature from the libraries using the search string(s).

   (c) Screening and selection of articles:

i. Inclusion and exclusion criteria are outlined and applied to the gathered literature.

(d) Extracting, analyzing, and synthesizing data: The finalized literature is gathered after screening for further analyzing and synthesizing the answers to the proposed research questions.

Chapter 2, which thoroughly reviews the available literature, provides detailed responses to all of the knowledge research questions RQ1, RQ2, and RQ3.

3. **Design and development**: This phase will include designing a security data analytics architecture and analytics maturity model, understanding the data extraction process and preparation, and finally, developing business reports. These activities are shown in Chapters 4 & 5.

4. **Demonstration and evaluation**: The adoption of the analytics-driven strategy in cybersecurity and business reports will create a practical environment in which research may be applied. The research findings will be displayed and evaluated by the IRM cybersecurity team to determine the added value it offers in making business decisions regarding insider risk management which can be seen in Chapters 6 & 7.

5. **Communication**: The final stage of the research is to share the research process and results. The report contains challenges, artifacts, novelty, and other pertinent information that might assist the organization, researchers, and audiences in understanding the research problem and answers in a nutshell which is presented in Chapter 8.

1.6.1 Scope

This section describes the scope of this research. The extensive research is carried out in collaboration with NXP's Insider risk management (IRM) cybersecurity team, which focuses only on risks posed by insiders to the organization's intellectual property. Due to the market's competitive nature and the initial concern of how to secure intellectual property and information, this research within NXP focuses primarily on managing and minimizing insider risks that could result in information disclosure. Therefore this research only focuses on the risks and threats that are posed by insiders. Also, numerous approaches and methodologies for better decision-making processes in cyber security are presented in the literature. Since we intend to work on real-world use cases involving real user behavioral data, the research focuses on the current solutions and their limitations related to user behavior analytics and risk management. Finally, the models developed as part of the research focus solely on

data exfiltration issues faced in the organization by insiders, but this research study can be extended as future work to analyze models related to malicious, compromised users, or other insider risk problems.

1.7 Expected Deliverable

The following are the expected key deliverables for this research:

1. To design and propose an analytic maturity model, followed by assessing NXP's current organizational analytics maturity in the cybersecurity domain.
2. To design a sustainable data pipeline security analytic architecture.
3. To understand the data extraction process and understand the data preparation steps followed by defining high-level key risk indicators (KRI) and developing the business reports.
4. Evaluation of the proposed solution followed by conclusion and discussions about the results and future research directions pertinent to insider risk management.

1.8 Thesis Outline

The remaining chapters of the thesis are organized as follows: we review the literature to understand the background of the defined problem, explain the research methodology, design the artifact, prepare the data, develop the solution, and finally, evaluate the artifact in terms of the objectives of our study, followed by conclusion and discussion. Figure 1.2 depicts the structure of the thesis document.

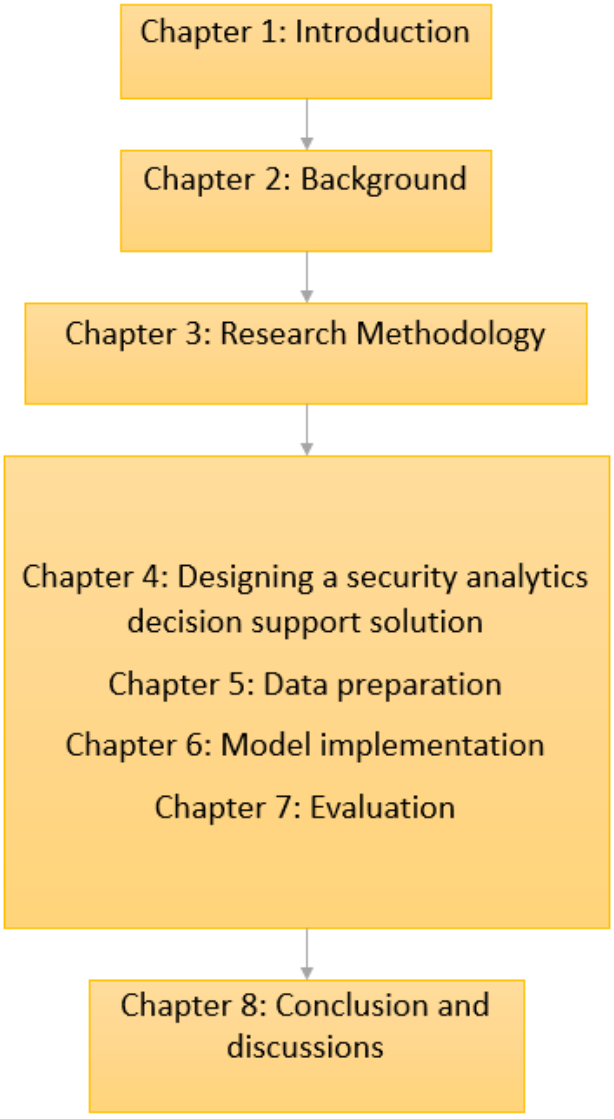


Figure 1.2: Thesis outline

Chapter 2

BACKGROUND

This section of the paper provides a systematic literature review. We followed the guidelines of [6]. In Section 2.1 of this chapter, we conduct a systematic literature review in order to answer RQ1, RQ2 and RQ3 outlined in Chapter 1. RQ1 seeks to clarify how insiders are defined, their causes, and the types of threats they are capable of posing. RQ2 examines recent insider cyber attacks on intellectual property, as well as the consequences and challenges that organizations face in combating and managing insider risks. Finally, RQ3 examines the existing decision-making solutions that organizations use to manage and minimize insider problems followed by a discussion. Section 2.3 summarizes the main takeaways of the systematic literature review of the proposed research questions. Further, we will discuss the case study of NXP and understand the current capabilities of the organization with respect to insider risk management. Finally, we will describe the maturity framework, which will be used to design a maturity model in the next chapters.

2.1 Literature Review Methodology

2.1.1 Research Questions

The knowledge question and the sub-questions we posed as seen in Section 1.4 were as follows:

RQ1: What are the security threats posed by insiders?

SQ1.1: How is an insider defined?

SQ1.2: What are the causes of the insider threat and risk problem?

SQ1.3: Which type of threats are insiders capable of making?

RQ2: Which recent cyber threats to intellectual property have been posed by insiders in an organization?

SQ2.1: What is the impact on the organization when it comes to insider attacks on their intellectual property?

SQ2.2: What are the challenges faced by organizations to combat & manage insider risks?

RQ3: What are the existing decision-making analytical solutions used in an organization to manage insider threats?

SQ3.1: What are the limitations of the current decision-making analytical solutions to manage insider risks?

2.1.2 Search Strategy

We selected IEEE Explore [1] and Scopus [2] as the digital libraries for this review. A digital library is used to collect relevant scholarly articles and to answer research questions. These libraries contain publications from significant journals and conferences, providing you with access to a wide range of articles on the subject. In order to develop a search string, a series of keyword variations were evaluated using synonyms from the literature [6]. Two search queries were obtained after several iterations. The search in IEEE Explore and Scopus was done on June 25, 2022, and we used the following search string(SS) in the article title, abstract, or keywords:

Search String 1: "security" AND "analytics" AND ("insider" AND ("risk" OR "threat" OR "attack"))

Search String 2: "security" AND "insider" AND ("user behavior" OR "decision making")

These search strings were chosen because they gave significantly better results for the research questions posed in Section 1.4. As per the scope discussed in subsection 1.4.1, our focus is on insiders, insider threat, and usage of user behavior data in security analytics decision-making solutions to manage insider risks and both the search strings gave better results for the proposed research questions. The potential keyword combinations that are pertinent to our research questions were useful in gathering study results that can be used to address our research questions. Instead of plain search strings, boolean AND and OR conjunctions with

[1] https://ieeexplore.ieee.org
[2] https://www.scopus.com

keywords are used. Furthermore, we used the potential keywords to make sure that no relevant material was overlooked. To filter relevant studies that are directly related to the research questions, inclusion and exclusion criteria for the resulting search query were developed and applied to both databases.

2.1.3 Inclusion and exclusion criteria

The studies were first chosen based on their titles. Establishing inclusion and exclusion criteria reduces the possibility of subject damage, increases the likelihood of obtaining reliable and reproducible data, and prevents the exploitation of weaker individuals. We used the following inclusion criteria (IC's) and exclusion criteria (EC's).

The following are the inclusion criteria:

- IC1 The paper has an immediate connection to the subject of our review. This indicates that we include papers that specifically suggest proper techniques or a development of an already established approach in the field of cyber security analytics. We also include articles that evaluate the efficacy of the current approaches using comparison studies, case studies, and experiments.
- IC2 limit by publication year, starting in 2012 and ending in 2022
- IC3 The paper refers to the research questions.
- IC4 The article appears in a conference and peer-reviewed journal.
- IC5 The paper is written in English.
- IC6 The paper is available to download.

The following are the exclusion criteria:

- EC1 The same-titled or identically written articles are not included.
- EC2 Off-topic articles in relation to the research questions.
- EC3 Articles that are too brief or incomplete.

Since the number of articles obtained from IEEE and Scopus was large, additional restrictions were imposed to define the scope of the studies: limit by IEEE subject and Scopus' subject area, i.e. Computer Science, Engineering, or Business.

2.1.4 Search results

IEEE returned 42 results for the first string and 169 results for the second string. The first string from Scopus returned 69 results, and the second returned 199 results. In order to find any studies that we might have overlooked during our search, we looked through the reference lists of the studies we had gathered, which also added some new results. A total of 399 articles were identified and 180 unrelated and duplicate articles were removed, leaving a total of 219 articles for screening. Of these 219 articles, 31 articles were accessed for eligibility. Reasons for rejecting an article included the article being too specific on a specific topic or articles which were off-topic from the research questions also taking into consideration the inclusion & exclusion criteria. 21 additional literature studies were chosen as grey literature since they gave a more detailed source of information than a scientific study which included blogs, websites, theses and dissertations, Forrester research studies, etc. Figure 2.1, depicts a mind map overview of the search strategy related to the topics of the research questions raised during the literature review and Appendix A.1 contains a complete list of all papers reviewed.

2.1.5 Data extraction from the selected articles

We found 52 relevant studies and used the following data extraction strategy. For RQ1, we included papers with proper definitions of insiders and discussions of the different types of insiders as well as papers classifying different insider threats to an organization in our data extraction strategy. For RQ2, the impacts and challenges that organizations face to combat and manage insider risks were examined. For RQ3, various decision-making analytical methods, their limitations, and the evidence produced by the method's application were analyzed and classified. Following that, our qualitative and theoretical data synthesis strategy included the following: We began by looking for the most important findings, or lessons learned in each paper. These are usually the answers to the research questions presented in the research context or are summed up in the paper's conclusion section. We organized the 52 papers into clusters after determining their findings. This also aided us in categorizing our research. Concept-centric matrix in Appendix A.3 shows which concepts from the gathered literature are discussing which aspects of the proposed research questions.

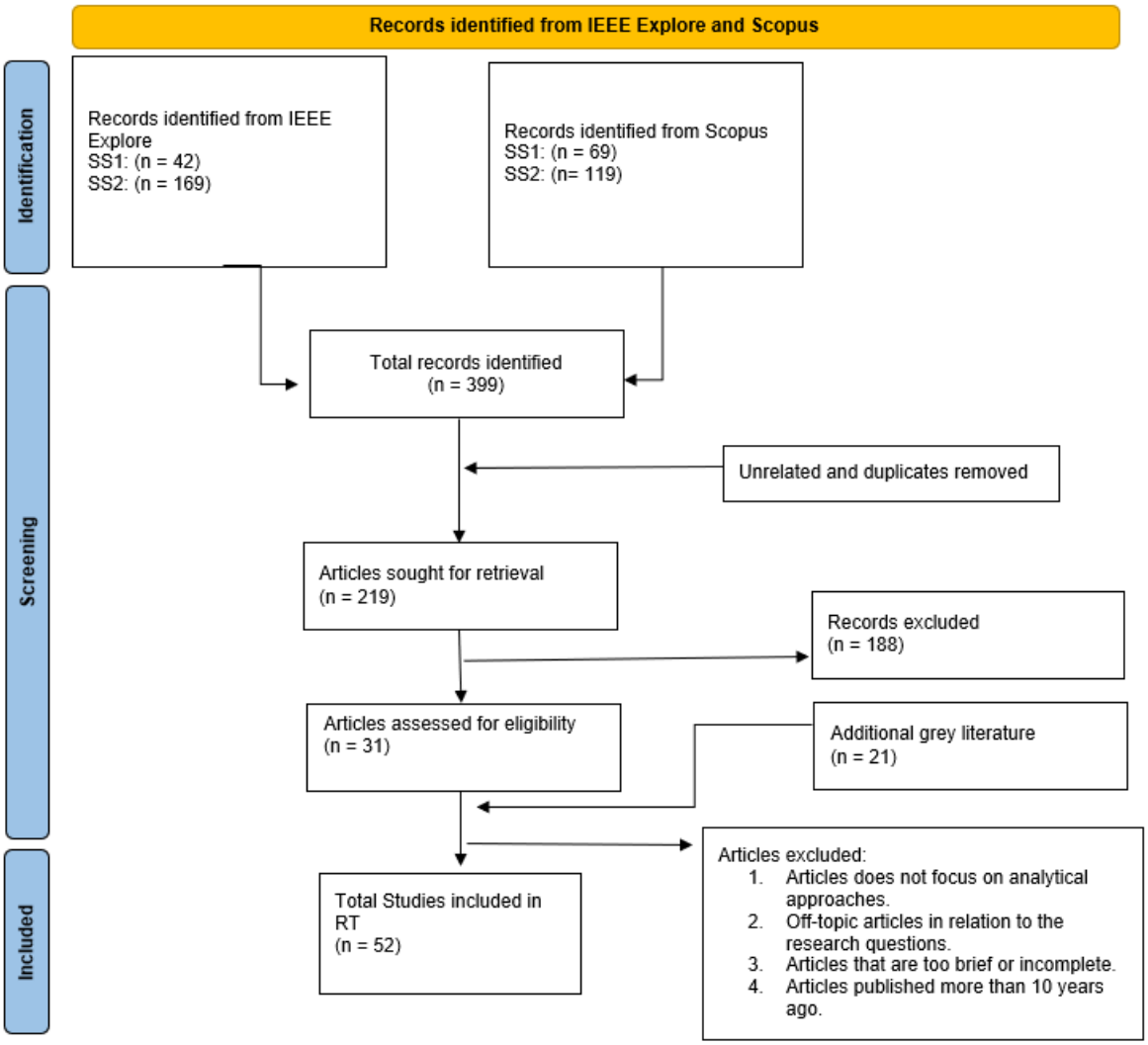


Figure 2.1: Results of literature selection

2.2 Literature Review Results

This section summarizes our results for the RQs defined in Section 1.4. Section 2.2.1 explores research on insiders, their causes and the types of threats insiders pose to the organization and answers RQ1. Section 2.2.2 answers RQ2 by summing up the issues that organizations face when it comes to insider risks to their intellectual property. Finally, Section 2.2.3, answers RQ3 and recognizes and evaluates the current solutions/methodologies as well as their limitations.

2.2.1 Insider Problem

Insider threat and risk have gained a lot of attention recently in the literature. The insider threat problem is one whose severity is continually expanding and which causes serious harm to both organizations and businesses. Therefore, the first research question **RQ1: "What are the security threats posed by insiders?"**.

Insider definition

Cybersecurity and infrastructure security agency (CISA) [7], defines insider as: “Any person who has or has had authorized access to or knowledge of an organization’s resources, including personnel, facilities, information, equipment, networks, and systems, is considered an insider.” Definitions from the literature are summarized in Table 2.1.

Reference	Insider definition
[8]	"An employee, contractor, or business partner who purposefully overstepped or misused their access to an organization’s network, system, or data in a way that jeopardized the confidentiality, integrity, or availability of the information or information systems of the organization."
[9]	"An insider is a trusted individual who has authorized access to a company’s employees, facilities, and information assets."
[10]	"Insiders can be current or former employees, as well as stakeholders, contractors, managers, or anyone else with access to sensitive systems and knowledge of how the organization runs."
[11]	"Insider generally refers to a static definition of a person using words such as access, knowledge, trust, or security policy."

Table 2.1: Definitions from the literature for ‘Insider’

The definitions discussed provides an overview of several essential traits that set insiders apart from outsiders. The following is a description of the essential insider traits:

1. Trustworthy:
   Insiders are trustworthy individuals who are typically employees, but they could also be contract workers and consultants, helpers, or personnel from third-party business partners with which the organization has a formal or informal business arrangement.

2. Authorized and privileged access to IP:
   Insiders have access to confidential information such as personal customer data, intellectual property of the organization, trade secrets, etc. Insiders have genuine reasons to perform sensitive tasks that require privileges in addition to having authorized access. Insiders are in a position where misuse, whether intentional or unintentional, is easily possible as a result of this combination.

3. Skilled and knowledgeable:
   Insiders may not only hide behind the network's built-in defensive security mechanisms, but they may also have intimate knowledge of the network [12]. Insiders are skilled given that they are already aware of inside security policies and procedures, and they can thus violate them. This increases the likelihood of going unnoticed.

4. Better at perceiving the risk:
   Insiders perceive risk more precisely than outsiders. Insiders are more likely to successfully bypass organizational controls, cover their tracks, and go undetected.

5. Security boundary:
   Insiders work within the organization's security boundary [9]. The organization's security measures protect against outsider threats by building a 'perimeter' around its assets that includes defenses against potential outsider attacks. This, nevertheless, does not address the insider threat issue, which is the risk that insiders will use their rights to compromise and misuse the organization's assets.

An insider may be motivated by a number of primary and secondary factors, such as monetary gain, retaliation, rage, thrill, pressure, betrayal, discontentment, jealousy, organizational politics, and acknowledgment [13]. In theory, there are various types of insiders. [14], classifies insiders as inadvertent insiders, malicious insiders, disgruntled workers, and contractors. Inadvertent insiders are sloppy personnel who cause unintentional security violations. Malicious insiders frequently utilize legitimate accounts and have permission to access information systems or data. It is difficult to tell whether their intentions and actions are good or bad. Regular employees (49%) and privileged IT users (59%) are the most common insider threat

actors [15], followed by contractors (52%). Insider attacks are primarily enabled by a rise in the number of users with unneeded access rights, an increase in the number of devices accessing confidential information, a rapid increase in technological complexities, a lack of user awareness and training, and an increase in sensitive data.

Insider threat and risk definition

Insider threats and risks endanger the safety and security of an organization's infrastructure. According to research, insider threats pose a higher amount of risk than outsider attacks. Table 2.2, summarizes the definitions of insider threat and risk found in the literature.

Reference	Term	Insider definition
[16]	Insider risk	"Insider risk occurs when any data exposure endangers an organization's and its employees, customers, or partners' well-being. Unlike insider threat, which focuses on specific users, insider risk focuses on data first and foremost."
[16]	Insider threat	"A variety of technical, behavioral, and organizational issues influence insider threats, and these must be addressed through policies, procedures, and technologies."
[17]	Insider threat	"A person who has, or previously had, authorized access to information, facilities, networks, people, or resources; and who intentionally or unintentionally commits: acts that violate law or policy and cause harm either directly or indirectly by destroying, losing, or degrading information, assets, or capabilities of the government or of a company; or destructive acts, such as inflicting bodily harm on coworkers."
[18]	Insider threat	"When it comes to the attacker abusing his privileges, the insider threat necessitates three factors: An insider attacker must have a reason to attack ('a motive'), a target to attack ('an opportunity'), and the ability to launch an attack ('a capability')."
[19]	Insider threat	"Insider threat refers to risky behaviors that trusted individuals may engage in that jeopardize the organization or its employees or lead in an unauthorized action that benefits the individual. The insider danger arises whenever human conduct deviates from accepted guidelines, whether out of malice or disdain for security rules."
[20]	Insider threat	"Based on their intent, insider threats can be classified as malicious or unintentional. A malicious insider tries to gain access to and potentially harm an organization on purpose. On the other hand, an unintentional threat refers to instances where harm is caused by an insider who has no malicious intent."

Table 2.2: Definitions from the literature for 'Insider risk and threat'

Classification of insider threats to the organization

The objectives of insider threats might vary greatly, depending on their nature. Not all insiders pose an insider threat. It is comforting to know that the majority of employees, contractors, and consultants can be trusted and thus share the very same interest in protecting the organization's valuable information. However, misuse of information and systems stored within them cannot be ruled out. Although most of the existing research on the insider threat concern refers to these exceptions as malicious, it should be observed that not all cases of misuse are motivated by malicious intents.

According to [20], a malicious insider's common types include espionage, sabotage, intellectual property theft, fraud, the destruction of competitive advantage, the corruption of critical data, and the disclosure of sensitive information. Espionage or spying is when an insider takes knowledge from a corporation for a different organization, such as a government agency or rival business, and sabotage is when an insider uses his or her IT knowledge and experience to launch an attack on a person or organization.

According to [21], the insider threat is divided into seven sub-categories viz., insider IT sabotage, insider IT fraud, insider theft of intellectual property, insider social engineering, unintentional insider threat incident, an insider in cloud computing and insider national security. Insider social engineering (SE) occurs when a malicious insider psychologically manipulates another innocent worker without their knowledge in order to disclose sensitive information or perform an action that will harm the organization's IT, network infrastructure, applications, or services. An incident involving an inadvertent insider threat occurs when an authorized user unintentionally damages the organization's IT and networking infrastructures, applications, or services without intending to launch a hostile assault [22].

Insider threats to National Security (NS) occur when an insider uses their authorized access to pose a threat to or harm the NS of a nation. This threat may include harm to the country as a result of espionage/spying, sabotage, disclosure of national security information, or the loss or deterioration of departmental assets or capabilities. Their primary targets are NS classified information.

2.2.2 Rise in Intellectual Property (IP) risk posed by insiders within the organization

The second research question **RQ2: "Which recent cyber risks to intellectual property have been posed by insiders in an organization?"**.

According to the 'Insider Report 2021' [23], 98% of organizations are vulnerable to insider attacks. 49% of organizations are unable to identify insider threats or detect them only when the data has left the organization. Insider threats frequently lead to financial loss and harm to an organization's reputation. These losses frequently reach hundreds of millions of dollars and get worse every year. Insider danger is thus a significant issue that needs to be addressed [14].

Insider intellectual property (IP) theft occurs when an insider uses the IT infrastructure to conduct espionage or steal information created and owned by the organization in which he works. Insider thieves of intellectual property in general are current employees or employees on leave who work in the office and have authorized access to IP. They typically work as scientists, programmers, engineers, or salespeople during the day and do not require tools to launch an attack. Source codes, business plans, strategic plans, product information, and customer information are the primary targets of insiders [16].

There are numerous instances of insider danger in recent years which are summarized in Table 2.3.

Reference	Organization	Year	Insider Attack
[24]	Ubiquiti Networks	2021	Former cloud lead stole the company's private information by abusing his GitHub server and Amazon Web Services (AWS) credentials. One of the employee published some of the stolen files while working from home and demanded the business pay a $2 million ransom. According to the Justice dept, the stock price dropped 20% and the company's share price was reduced by $4 billion as a result of all the negative coverage.
[25]	Pfizer	2021	One of the employee used her company laptop to upload 12,000 secret files to a private Google Drive account. Monitoring tools were set up immediately before the incident picked up. Trade secrets pertaining to the COVID-19 vaccine as well as its research were also included in the sensitive records, along with details on drug development. Prior to the incident, the employee received a job offer from another organization, according to Reuters' reporting on Pfizer.
[26]	Twitter	2020	Twitter attack demonstrated the potential severity of the insider threat by coercing insiders to change user accounts to their financial advantage. The sums stolen were relatively small (around $120,000 in Bitcoin), but they temporarily reduced Twitter's market value by nearly $1 billion, causing embarrassment and inconvenience to the company and the individuals whose accounts were compromised.
[27]	Vodafone	2013	The organization suffered a data breach caused by an insider who had intimate knowledge of their IT infrastructure and system, and he was able to take a copy of more than two million customers' records, including customer names, addresses, dates of birth, and bank account details.

Table 2.3: Insider attacks in recent years

According to the research '2022 cost of insider threats' global report conducted by Ponemon institute [28], insider threats have grown as a result of the rapid digital transformation over the last two years. Organizations are realizing that the conventional approach to data security is no longer adequate, from the use of personal devices to the increased use of the cloud and work from anywhere. According to

the findings, 56% of incidents experienced by organizations represented in this research were caused by negligence, with an average annual cost of $ 6.6 million to remediate the incident.

Consequences of insider attacks on organizations and the challenges that organizations face in managing insider risk

Insider attacks can have a significant impact on the organization. In addition to the asset's lost value from being destroyed or revealed, they may also suffer immediate losses. According to the annual report 2018 by data security company Clearswift [29], 88% of respondents (organizations) reported having encountered IT or data protection incidents in the previous 12 months, with insiders being responsible for 73% of these breaches. Thus, addressing insider threats and risks is the top priority for achieving full networked infrastructure protection. The impacts can generally fall into categories like revenue, business process, reputation, organization, culture, and liabilities.

According to [10], insider attacks can cause serious and expensive damage to an organization. Among the impacts are:

1. Important data theft and financial loss.
2. Negative media coverage.
3. Reputational harm and competitive disadvantage.
4. Legal impact which can include penalties and legal defense expenses related to claims made by people and organizations impacted by data breaches.

According to a new survey-based report by Forrester [30], insider threats were to blame for the majority (59%) of EMEA data security incidents in the past year, while 70% of firms in the region lack a strategy for such a risk category. Because of the quick adoption of remote work, many employees are currently not covered by the standard security measures used by firms, making it more difficult to identify and stop internal threats.

Dealing with an insider attack can also be costly. Smaller organizations may face budgetary difficulties. The recent technological advancements and changes can also pose challenges. According to report [31], when an organization moved to the cloud, it became more difficult to identify insider threats. Another barrier to insider threat management is a lack of sufficient staff and security tool knowledge.

Internal users can legitimately access crucial systems, which makes them invisible to intrusion detection systems and firewalls, making insider threats difficult to detect. The security of an organization's data is significantly at risk due to the lack

of awareness of insider threats. Insiders can pose a significant threat to organizations, so it is critical that organizations incorporate problems insiders can possess into their complete data protection strategy.

2.2.3 Analytical Techniques for Decision-Making in Cybersecurity

The third research question: **"What are the existing decision-making analytical solutions used in an organization to manage insider risks?"**.

There have been numerous cases of cybercrime discovered in the areas of information misuse, security attacks, and so on. Authentication, Access Control, Anti-Virus, Firewalls, Intrusion Detection Systems, and Security Information and Event Management (SIEM) are some of the security tools available to help organizations control and mitigate information misuse and threats to their systems. However, while they are effective at detecting outside threats, they frequently fail to detect insider threats.

Security analytics is a proactive strategy in cybersecurity that employs data collection, aggregation, and data analysis to perform critical security functions such as cyber threat detection, analysis, and mitigation. The data sources can be user log data, audit logs, user activity data, system logins, etc. A wide range of research related to detecting insider threats has been conducted in the literature. One of the research's key findings is that the four most commonly used decision-making methods in cybersecurity & analytics are user behavior analytics, data mining, artificial intelligence, and deep learning. Table 2.4 shows the number of articles retrieved for each method.

Sr. No.	Security Data Analytical Methods	Article ID
1	User Behavioral Security Analytics (Machine Learning - Supervised, Unsupervised & Semi-Supervised Learning Algorithms)	[32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [10] , [44], [45], [46]
2	Artificial intelligence (AI) Analytics	[47], [48]
3	Deep Learning	[49]
4	Data Mining	[50], [8]

Table 2.4: Decision making methods from literature

The most widely used tool in cybersecurity to detect insider threats and risks are User and Behaviour Analytics (UBA) & User and Entity Behaviour Analytics (UEBA). Analytic methods such as machine learning, statistical models, rules, and threat signatures are used with UBA and UEBA solutions to isolate anomalies. UEBA

systems can identify insider threats, malware, and sophisticated assaults. They offer investigative insights so analysts can immediately verify and neutralize threats prior to causing more harm, as well as the knowledge to detect odd activity in real-time.

2.3 Discussion

This Section summarizes the main takeaways of the systematic literature review of the proposed research questions discussed in Section 2.1.1.

RQ1: "What are the security threats posed by insiders?"

- The major difference between insiders and outsiders is that insiders are trustworthy. Insiders can include employees, but also contractors, consultants, workers, and outside business partners as a result of cross-company cooperation (e.g., outsourcing activities). Authentic access to an organization's information is granted to trusted insiders. Insiders are also more likely to break security rules and procedures because they are aware of them.
- Insider threat can be summed up as the possibility that dependable workers, temporary assistants, contractors, or consultants could gain access to sensitive information, exploit security flaws, and then break the organization's security policy.
- Insider risk occurs when any data exposure endangers an organization's and its employees, customers, or partner's well-being. Insider risks focus on data first and foremost.
- Classifications of insider threats show that they can be categorized as sabotage, fraud, IP theft, espionage, social engineering, unintentional insider threat incident, or insider national security. These goals can range from the release of information (e.g. profit or theft) to alteration (e.g. fraud) and interference or destruction of data (e.g. sabotage). These findings are directly related to information security properties such as confidentiality, integrity, and availability. Furthermore, the classifications show that threats can be motivated by personal motives (e.g., economic gain, vengeance) or by a third-party provider (i.e. information leakage).

RQ2: "Which recent cyber risks to intellectual property have been posed by insiders in an organization?"

- Insider attacks could have a significant impact on a company. Organizations could indeed suffer instant losses of inherent worth as well as lost revenue

in addition to the lost worth of the asset that was eliminated, disclosed, or destroyed. Because internal users have valid access to critical systems, they are undetectable to traditional security solutions such as firewalls and intrusion detection systems. Also insider threats can be difficult to deal with and manage due to budgetary constraints, lack of employees, and inadequate tools. When attempting to defend against external threats, organizations must keep in mind that insider threats also can cause harm.

RQ3: "What are the existing decision-making analytical solutions used in an organization to manage insider risks?"

- As a form of large-scale data analytics, multiple proposals have been made to use machine learning and anomaly detection methodologies to make automated decisions about which insiders are behaving strangely or maliciously. The different methods and the data sources used from the reviewed literature are briefly defined in Appendix A.2. By using machine learning and behavioral analytics to analyze users, devices, and things, UEBA systems can identify insider threats, malware, and sophisticated assaults. They offer investigative insights so analysts can immediately verify and neutralize threats prior to causing more harm, as well as the knowledge to detect odd activity in real-time.

2.3.1 Research Gap

Gap in literature

Different proposals as seen in Appendix A.2, have been made to use machine learning and anomaly detection techniques to make automated decisions, to know which insiders are acting suspiciously or maliciously in the context of large-scale data analytics. Furthermore, while many studies on the topic of cyber security analytics have been conducted using tools like UBA, UEBA, and other analytic monitoring solutions, little has been explored about the limitations of the user behavioral analytic approach and how it can be made more feasible for better decision-making processes. Insider risk, unlike insider threats, is not a well-defined concept. Therefore, the majority of proposals presented focused on insider threats, with little to no discussion on how to minimize risks once they had been committed by insiders. Also from the literature, one of the drawbacks of UEBA is that it cannot identify long-term sophisticated low and slow attacks because they have no day-to-day impact and appear to be non-existent also, UEBA prioritizes only the most reliable and high-level risk security alerts which makes UEBA less efficient in managing and minimizing insider risks.

Gap in practice

Focusing on the FBA solutions with respect to NXP use case, it is challenging to determine their genuine worth when some of them simply take a small subset of actions into account or do not offer validation in a real-life setting. It is widely acknowledged that this presents numerous challenges, such as detailed meaningful insights on general behaviors being limited within FBA solutions. For example, to track the IP data movement within or outside the organizations, the type of devices or software used mostly to transfer IP data, etc. Such insights are not possible within FBA since it prioritizes on most reliable and high-level risk security alerts. Also, issues like a baseline and low and slow attacks are also experienced in practice [5]. However, low-level risk events and detailed insights about them can play an important role in a better decision-making process. These insights can help businesses make better decisions over time and determine whether training, cybersecurity policies, or other tools are needed in mitigating the potential loss of intellectual property.

To conclude, this research will help to address the limitations specific to data exfiltration problems within User and entity behavior analytics solutions and will help to improve the feasibility of the decision-making process within the organization of the IRM cybersecurity team in NXP, for further directions and approaches that can be used to aid decision-makers. Therefore further research is focused on how organizations can make better decisions by adopting an analytical-driven approach which can be used in combination with current user behavioral monitoring systems in cybersecurity to assist organizations in making better decisions to manage and minimize the risk from insiders. With this combined method of decision-making along with UEBA about possible risks, the analytical solution will help to capture the human reasoning for the decision process more accurately, provide real-time statistics and insightful data, reduce false positives flagged by the system and allow business and the organization to gain up-to-date knowledge to make better decisions.

2.4 NXP Case Study

This Section provides background information on the case study company, NXP Semiconductors N.V., including its current insider risk management solution and capabilities. The context for this master's research project is provided by NXP. NXP has initiated a program to protect its intellectual property (IP). Their goal is to strengthen their capabilities to prevent, detect, and respond to any misuse or misconduct of NXP Intellectual Property (IP) by insiders that have a negative impact on NXP's market position, regulatory compliance, reputation, operations, supply chain, and/or

financial risk. Their expertise and market position in the production of automotive semiconductor chips makes them a prime target for foreign intelligence services. Therefore, NXP wants to protect its Intellectual Property from being exfiltrated by insiders.

To prevent this from happening, NXP has started Insider Behaviour management and Data loss prevention (DLP) as a part of its Insider Risk Management (IRM) program. Insider Behaviour Management mitigates the risk of individuals, who use their access, either maliciously or unintentionally, to act in a way that could negatively affect an organization. DLP mitigates the risk of data loss, misuse, or access by unauthorized users. The DLP Solution is currently being planned and is expected to be implemented by the fourth quarter of 2022. Figure 2.2, shows the overview solution plan of the Insider risk management (IRM) program within NXP.

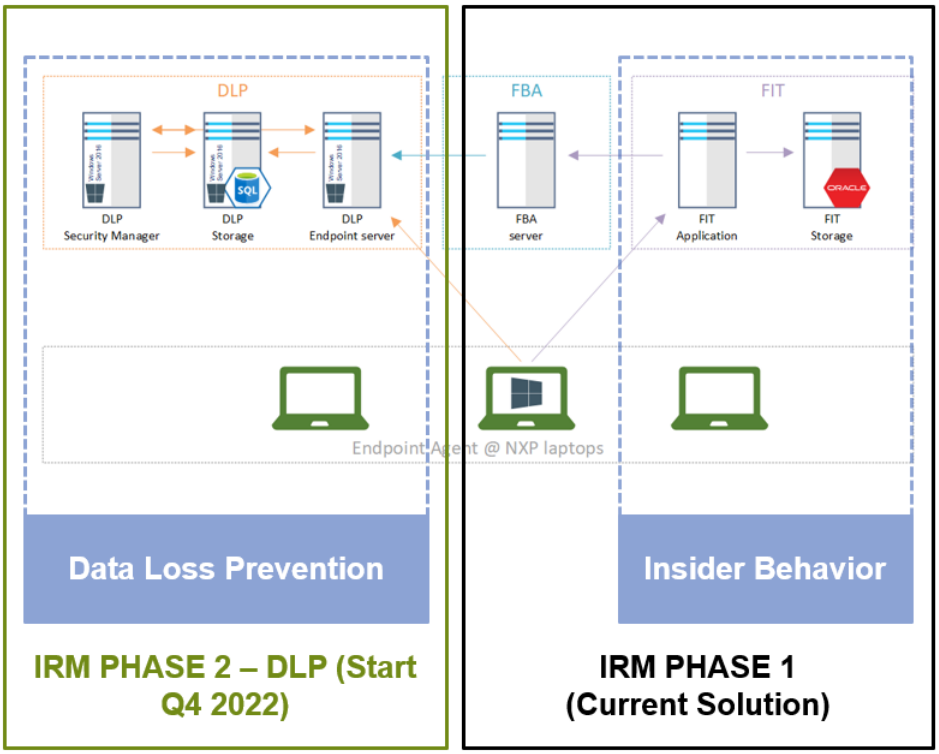


Figure 2.2: NXP - IRM Solution Plan

To enable the monitoring capability on IP protection and insider risk behavior an Employee Behavior Monitoring solution software is used within NXP, known as Forcepoint user and entity behavior analytics (FBA). FBA enables security teams to monitor high-risk behavior within the enterprise on a proactive basis.

Intellectual Property (IP) Theft

Intellectual Property (IP) theft occurs when someone steals a company's creative expressions, designs, inventions, or trade secrets, which are collectively known as intellectual property. In brief, intellectual property is an organization's intangible property. Insiders or malicious insiders who believe they have a right to information are frequently the perpetrators of intellectual property theft. Figure 2.3, presents three common scenarios of insider-threat activity. The following are the possible scenarios for IP theft:

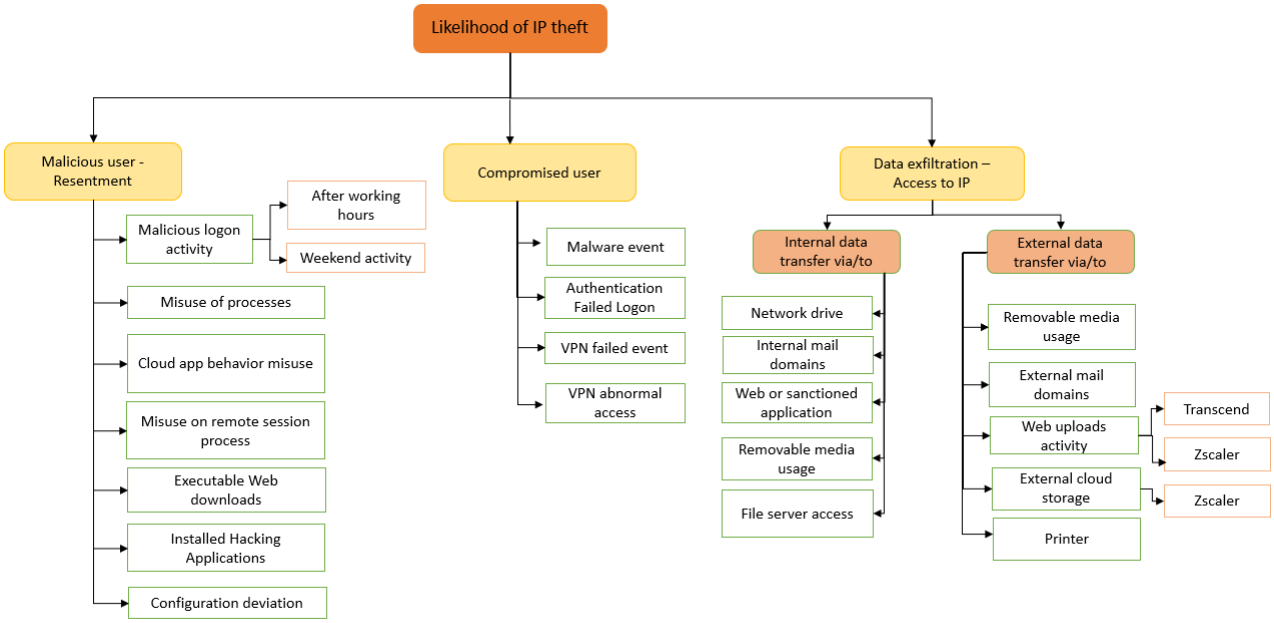


Figure 2.3: IP theft-tree generation

1. Compromised User - A user of which the user account is being accessed (used) by someone else.
   Examples: Password attack, taking over the computer if left unattended, malware on the computer, etc.

2. Malicious User - A user that performs activities that harms the company.
   Examples: Installing and executing malicious software, tampering with behavior monitoring software, malicious login activities, etc.

3. Data Exfiltration - Data exfiltration occurs when malware and/or a malicious user perform an unauthorized data transfer from a computer. Data Exfiltration Events are events created by actions that may lead to (un)intentional exfiltration of (disruptive) IP data.
   Example: Copy (disruptive) IP to personal USB storage, cloud, transcend, etc.

We present a simplified subset of nodes for the sake of this discussion; in reality, the tree would contain many more connected nodes. At the lowest level of the tree are the leaf nodes that are directly measurable. However, these models are measured within FBA. **As part of our research, we will focus on Insider Behaviour Management and scenarios involving data exfiltration concerning external IP movement.** To be considered since this research is in its early stages of enabling analytics in cybersecurity, the organization wants to currently focus more on data exfiltration external events, with the possibility of implementing other scenarios and models in the future. The models are designed to support reasoning within and between the various tiers. The data exfiltration external IP data movement models which will be used as a part of this study are as follows:

Sr. No.	Model	Description
1	IP File copied to removable media	Movement of IP files to removable media storage (HD, USB, Floppy, CD/DVD)
2	External IP data movement to external mail domains	Movement of IP files via attachments to external mail domains
3	External IP Data Movement to web upload	Movement of IP files over external web application domains
4	IP Data Movement to External Cloud Storage	Movement of IP files to external cloud storage
5	IP Data movement to printer	Movement of IP files to printer
6	External IP Data Movement to Zscaler	Movement of Data on web application domains
7	External IP Data Movement to Intercom Transcend	Movement of IP files through Intercom transcend (a website used by people in NXP to safely transfer data(IP) via the web)

Table 2.5: Data exfiltration - External IP data movement models

2.5 Analytics Maturity Framework

The first step in enabling analytics in any organizational domain is to determine the project's current level of maturity. A maturity model identifies areas where you can improve in order to achieve a higher level of maturity in the business. In this Section, we will define the analytics maturity framework from the literature to help us to design an analytics maturity model for NXP and then identify NXP's current maturity state in cybersecurity analytics.

Analytics maturity is defined as an organization's ability to integrate, manage, and leverage all relevant internal and external data sources at key decision points [51]. It entails creating an ecosystem that allows for insight and action. In other

words, analytics maturity entails technologies, data management, analytics, governance, and organizational components as well as technology. Creating and instilling an analytics culture in an organization can take years. Analytics maturity describes the extent to which an organization uses tools, people, processes, and strategy to manage and analyze data in order to inform business decisions. This transformation is guided by maturity models. Therefore, understanding an organization's ability to use data analytics to increase innovation and competitive advantage necessitates an assessment of its position on the analytics continuum. The use of data analytics in an enterprise does not evolve in a linear way. The order and intensity with which organizational changes are implemented may vary depending on the specificity of the organization and the business context. **In this research study, we will use the most commonly used model framework of the 'Gartner analytics ascendency model' [52] as shown in Figure 2.4.**

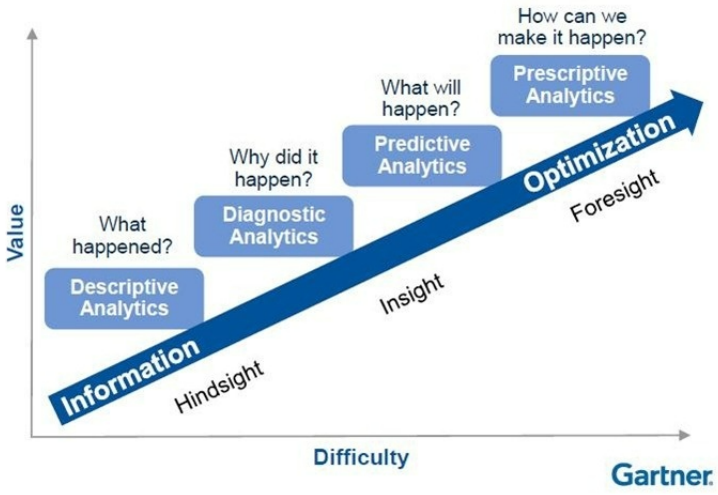


Figure 2.4: Gartner analytics ascendency model

The analytics maturity model depicts an organization's analytics maturity stages, beginning with descriptive analytics and finale with prescriptive analytics. Each step along this path moves the organization closer to solutions that enable thoroughly based decisions to be made more quickly (on-demand enterprise). The approach is divided into four categories and in analytics practice, these categories co-exist and complement each other :

1. Descriptive analytics:
   During this stage, knowledge can be gained primarily from reports and descriptions. This is normally done manually. Most businesses already have several dashboards that display information about specific developments and events. Organizations can make decisions and take actions based on a combination of these insights.

2. Diagnostic analytics:
   During this phase, information is gathered to identify the cause of specific events. In this phase, organizations can have a course of action, however, would have no idea how likely this strategy will succeed. Data is systematized, analyzed, and interpreted, and technology is being used at this stage to identify and explain patterns and dependencies in available data.

3. Predictive analytics:
   The primary goal of predictive analytics is to provide forecasts along with, the cause of a specific problem as well as the actions to be taken to prevent it in the future. Technologically advanced tools assess opportunities and risks, allowing for the prediction of future outcomes. Massive amounts of current and historical data from a variety of sources can be processed to generate models, simulation models, and forecasts, identify trends, as well as provide insights for more precise and efficient business practices.

4. Prescriptive analytics:
   The main goal is to automatically recommend the best course of action and suggest optimization options based on massive amounts of historical data, real-time data feeds, and details about the outcomes of previous decisions. Delegate decision-making authority to a system. And the system can do so based on the information gathered in the previous phases.

2.6 Summary and Conclusion

In this chapter, we covered the systematic protocol used in Section 2.1 for the literature review of this thesis. The literature review conducted in Section 2.2 provided adequate knowledge to understand better how insiders and the risks and threats they posed are discussed. As shown, it has gained a lot of popularity, especially after the pandemic, and has a great impact on organizations, therefore proper steps should be taken by businesses to reorganize their security protection and IT strategies to assist reduce these occurrences. It also aided in understanding the several techniques used to tackle the security concerns posed by insiders. Several security data analytical methods have been applied in practice, with different levels of success or results each. In Section 2.3 we shared our main findings regarding potential challenges and scientific gaps that were identified. In Section 2.4, we provided context for NXP's current and future plans and solutions for managing insider risks. Finally, in Section 2.5 we discussed the analytics maturity framework, which would be helpful in the further sections to design the maturity model for the organization and to understand their current maturity level.

Chapter 3

RESEARCH METHODOLOGY

This chapter explains the methodology used to achieve the research objectives and answer the research questions defined in Section 1.4. The first section describes the overall methodological approach that was used in this research. Following that, we will summarize and map our research outline with the research methodology.

3.1 Research Design

To fulfill the goal, we choose Design Science Research Methodology (DSRM) defined by Peffers et al. (2007) [53] for our thesis because it aligns with the overall objectives, i.e., we intend to address and solve a specific problem by creating an artifact that can be seen in Chapter 4. As shown in Figure 3.1, it consists of six steps: (i) identification of the problem and motivation, (ii) defining objectives of the solution, (iii) design and development, (iv) demonstration, (v) evaluation and (vi) communication; and four possible entry points: problem-centered initiation, objective-centered solution, design, and development-centered initiation, and client/context initiation.

The **problem-centered initiation** entry point is the appropriate starting point for our research because the concept emerges from an observation of a problem and has a clear gap in the literature has been identified, as seen in Section 2.3.1 that needs to be addressed, which being the lack of current methods and tools to assess an organization's decision-making process for minimizing insider risks.

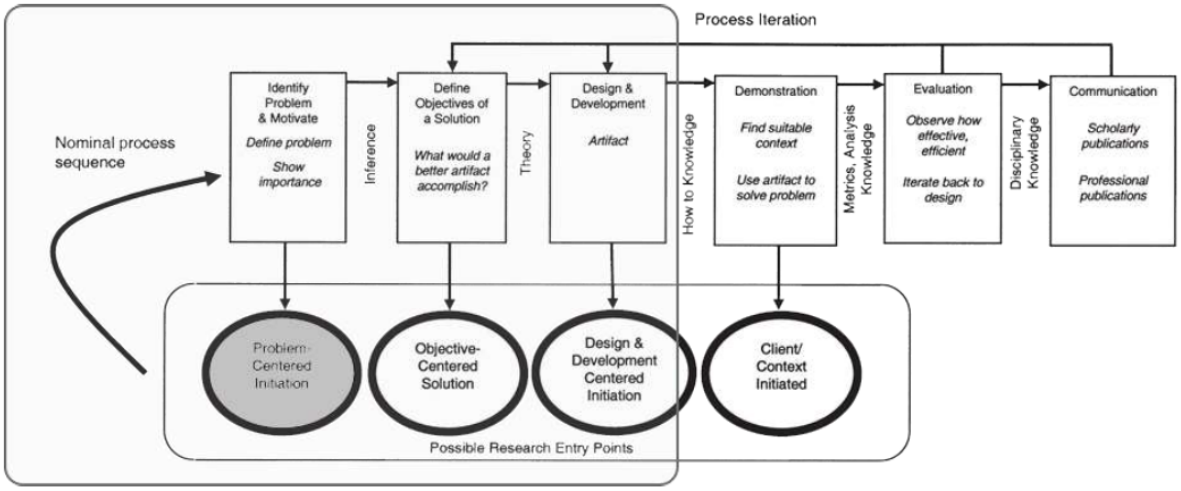


Figure 3.1: Design science research methodology process (DSRM) model

3.1.1 Problem Identification and Motivation

In Sections 1.1 and 1.3 of Chapter 1, we already stated the problem and motivation for this research study that insiders pose a high risk to organizations and can have serious consequences if not handled properly. To overcome these difficulties enterprises need a new cybersecurity strategy.

3.1.2 Define Objectives of a Solution

The objective of the solution is discussed in Section 1.4 of Chapter 1. As stated, the main objective is - *To deliver a solution using an analytics-driven decision support approach that can be used in combination with current user behavioral monitoring systems in cybersecurity to assist organizations in making better decisions to manage and minimize the risk from insiders.*

Therefore, this empirical research focuses on the risk associated with insiders and seeks to analyze and validate the results shown in the literature with the current practices, extending the reach of the research and improving current capabilities that assist organizations in mitigating the risk level of insider threat. We performed an extensive systematic literature study to get a more complete view of the consequences and solutions used to minimize insider risks. We identified the importance and necessity of bridging the gap, which is discussed in Chapter 2. The following subsections will further elaborate on the implementation and application of the remaining stages of the DSRM concerning the research study design.

3.1.3 Design and Development

The third step of the DSRM process involves creating an artifact, i.e., the solution to the problem. The solution can be in the form of models, concepts, or other methods. Peffers et al. [53] define the artifact more broadly as "any designed object in which a research contribution is embedded in the design". The artifact created in this thesis is to design and develop a data analytic decision support solution within cyber security and is covered in Chapters 4 and 5. The following are the design and development steps that will be taken as part of this thesis :

1. Design and propose an analytics maturity model. Adapted from Gartner's analytics framework which is discussed in Section 2.5, this model will help to measure and characterize how well NXP uses its resources to extract value from data, as well as serve as a guide throughout the entire process of integrating analytics decision support system into cybersecurity.

2. Design and propose a real-time processing data pipeline security analytic architecture. This architecture will assist the organization in bringing data from various sources together to provide a complete and accurate dataset for business intelligence (BI), data analysis, and other applications and business processes.

3. Identify key risk indicators and understand the data preparation process. KRI will benefit the organization by providing advance notice of potential risks that could harm it; insight into possible flaws in an organization's monitoring and control tools; and ongoing risk monitoring in between risk assessments. The following steps are followed for data preparation:

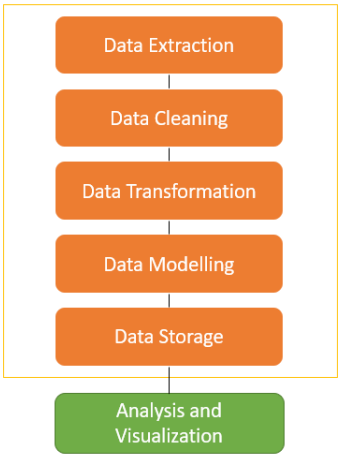


Figure 3.2: Stepwise Data Analysis Process

- Data extraction: Retrieving data from data sources for further data processing or data storage.
- Data cleaning: Involves repairing or removing inaccurate, corrupted, incorrectly structured, redundant, or incomplete data from datasets.
- Data transformation: Finally, data is converted and structured into a usable format that can be analyzed to aid decision-making processes and propel an organization's growth.

4. Development:

   For the development phase as part of our research, we will use **Power BI**, which is a component of Microsoft Power platform with a primary focus on business intelligence.

   - Data modeling: Datasets are loaded in Power BI for further structuring. For that purpose, the data modeling method is used for arranging data in a database that improves the structuring and organization of the database's contents.
   - Data storage: The data needs to be stored in a centralized location for that purpose there is a need for a data warehouse. A data warehouse is a system that stores information from both internal and external databases.
   - Analysis and visualization: Finally, Power BI is used to load the data and generate reports. With the help of business reports, businesses and the cybersecurity team will be able to make better, more informed, data-driven decisions. Because they are dynamic, interactive, and display near real-time data, they will assist in gaining a more precise, in-the-moment understanding of what is going on in the organization in terms of insider risks and navigating rapid, sometimes difficult changes.

Weekly meetings were held with the NXP-IRM cybersecurity team to discuss the progress of the design and development of the models, and changes were implemented in response to feedback.

3.1.4 Demonstration and Evaluation

The following step is divided into two steps. First, step (iv) shows how the artifact can be used to solve one or more of the problems identified (step (i)). This can be achieved by using the artifact, e.g., in experiments, case studies, or proof of concept [53]. Second, step (v) examines the artifact's performance in achieving the defined objectives.

Prototype: For our research study, since we are using real-time data, this model will benefit the organization by providing real-time data insights into managing insider risks. As a result, the decision is made to rely on expert opinions to evaluate and validate the model. The model will be explained to the experts for them to understand how it would work in practice, after which the evaluation and validation process will begin. Chapter 6 describes the implementation of a prototype using a proof-by-prototyping approach.

Evaluation: This thesis includes evaluation methods in the form of questionnaires and interviews. For this purpose, a group of experts is gathered, and the prototype is introduced to them. Following a tour of the prototype, the same experts are encouraged to participate in the questionnaire. Next, to validate the usefulness and applicability of the proposed models, the content of the research study will also be evaluated, e.g., is the proposed analytics maturity model and data pipeline efficient, etc. The following section describes, in brief, the characteristics of the data collection method for the evaluation of the research study :

- Questionnaires and Interviews: A questionnaire is a type of data collection method that consists of a series of questions or other types of items designed to collect useful information that can be analyzed and can be self-administered or administered by a researcher. It is helpful to the reader when authors describe the contents of the survey questionnaire so that the reader can interpret and evaluate the potential for errors of validity and reliability. It can be in the form of paper, personal interviews, or sent via email. The interview is a data-gathering strategy that focuses on the interviewer and interviewee's verbal engagement with the goal of developing knowledge in a certain area or topic. For our thesis evaluation, we favored semi-structured interviews because it allows them to ask both the pre-prepared questions as well as go further into areas that are important to interviewees.

In this thesis, we used researcher-administered questionnaires and semi-structured interviews with NXP-IRM cybersecurity experts. The findings of the questionnaires have served to outline the model and thesis evaluation and conclusion. In addition, we conducted one round of semi-structured interviews with the IRM team to gather critical information about the model's evaluation. An expert in the field will be interviewed to assess the model's usefulness, efficacy, and understandability. Chapter 7 presents details on the evaluation. Following that, Chapter 8 discusses the conclusion and discussion of the identified underlying problem.

3.1.5 Communication

Step (vi) entails communicating the solution, for example, in research publications or to the management of the organization. On one hand, the thesis itself communicates the solution when mapped to this thesis. It will also aid the organization in future developments. Furthermore, the work presented has been shared in the university research repositories.

3.2 Research Methodology Summary

This section summarizes how DSRM was implemented in this study. The investigation began with a systematic review of the literature. The literature review can be considered as step (i) problem identification and motivation of the DSRM process. A few gaps were identified in this process. Taking into consideration the research limitations, the author identified the gaps that will be investigated as a part of the research. After the research goal has been formed, an artifact is created in Chapters 4 and 5. Finally, in Chapter 6 and Chapter 7, the artifact demonstration and evaluation respectively are discussed. Table 3.1 show the mapping of the DSRM approach along with the research methods used in this research study.

Sr. No.	Chapter No.	Research Method	DSRM Step	Research Questions
1	1 - Introduction	-	(i) Problem identification and Motivation (ii) Define the objectives for a solution	RQ1, RQ2, RQ3
2	2 - Background and Context	Systematic Literature Review [6], NXP Case Study	(i) Problem identification and Motivation (ii) Define the objectives for a solution	RQ1, RQ2, RQ3
3	3 - Research Methodology	DSRM - Peffers et al. (2007) [53]	(i) Problem identification and Motivation (ii) Define the objectives for a solution	RQ1, RQ2, RQ3
4	4 - Designing a security analytics decision support solution 5 - Data preparation	Gartner's Framework [52], NXP Case Study	(iii) Design and development	RQ4
5	6 - Model implementation 7 - Evaluation	Expert opinion method (TAR)	(iv) Demonstration and (v) Evaluation	RQ4
6	8 - Conclusion and discussion	-	(vi) Communication	All

Table 3.1: Chapter mapping to DSRM

Chapter 4

DESIGNING A SECURITY ANALYTICS DECISION SUPPORT SOLUTION

In this chapter, firstly we will propose an analytics maturity model in Section 4.1, which will help the organization to understand the current state of analytics and what steps need to be taken to take analytics in cybersecurity to a new level along with the data pipeline architecture which will help in the data extraction process. For our proposed model, we use Gartner's analytics maturity framework, as discussed in Section 2.5. In Section 4.2, we will explain the data pipeline architecture which will assist organizations and businesses in breaking down information silos, understanding data movement, and obtaining value from their data in the form of insights and analytics. Finally, in Section 4.3 we will define the high-level key risk indicators for achieving the business goals, followed by a conclusion. This analysis is carried out in order to extend input for the final design of the artifact based on the company case NXP Semiconductors N.V. We would be answering RQ4 of our thesis beginning with this and the upcoming chapters.

4.1 Proposed Analytics Maturity Model

In this section, we will propose an analytics maturity model which is adapted from Gartner's framework as discussed in Section 2.5. The framework approach is divided into four categories which are descriptive, diagnostic, predictive, and prescriptive analytics, and in analytics practice, these categories co-exist and complement each other.

We will use a similar approach to create the maturity model and explain each step in brief with respect to organizational and technological perspectives. The defined

organizational and technological perspectives will aid in viewing the organization's high-level organizational and technical views. Also in our model, we constructed five categories to assess the NXP-IRM cybersecurity team's analytics maturity level.

The proposed analytics model has a series of steps approach to improve an organization's capabilities with specific reference to detect and mitigate insider risks with respect to integrating analytics in cybersecurity to develop a strategy that will grow with evolving risk priorities and evolve to the desired level of maturity. This model will help to measure and characterize how well NXP uses its resources to extract value from data and also act as a guide during the whole process of integrating analytics in cybersecurity. It will also help to understand the current state of analytics and what steps need to be taken to take analytics to a new level. It evolves from simple to more complicated types of analysis, with the working assumption that the more complex types of analytics bring more value. The stages of analytics maturity needed in cybersecurity are as follows:

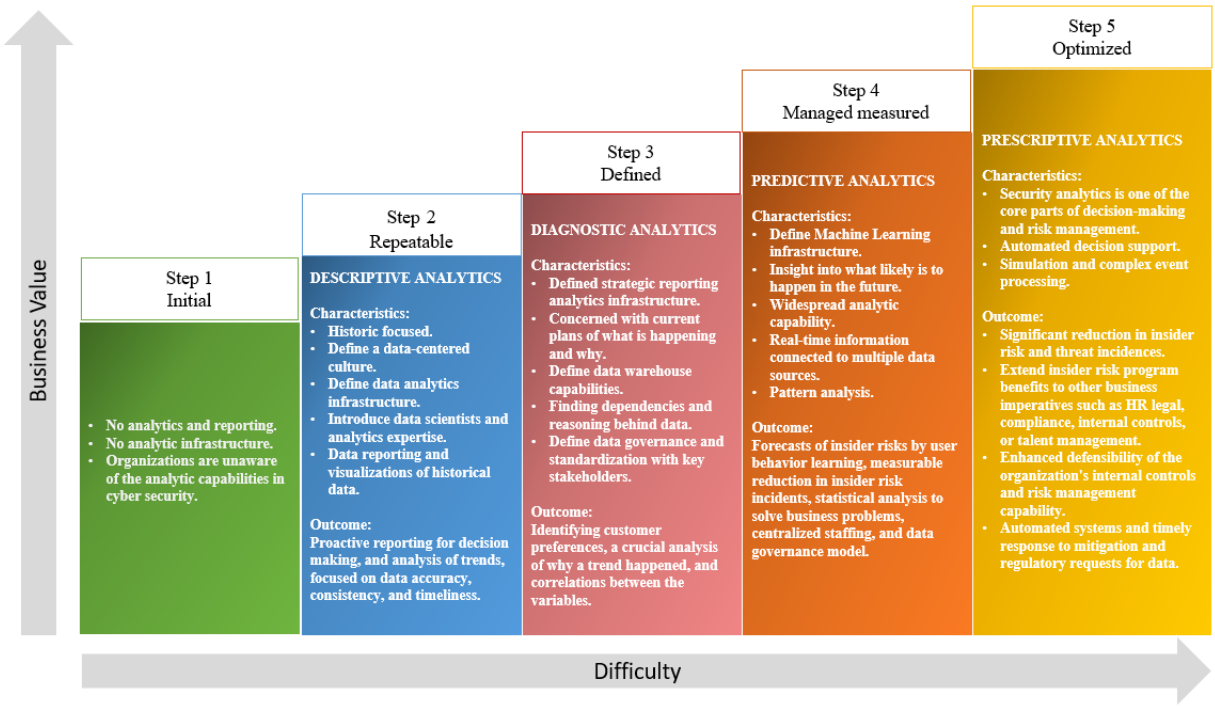


Figure 4.1: Proposed Analytics Maturity Model

4.1.1 Initial

The first stage of a maturity model is the 'Initial' which is a starting point of a new process. At this stage, organizations do not have an analytical strategy or infrastructure. Also, some organizations might still be unaware of the value of analytics in cybersecurity. The main challenge here is a lack of vision and understanding of the

significance of analytics in cybersecurity. Changing management's mindset and explaining the value of analytics would be a great place to start on the path to analytics maturity.

4.1.2 Repeatable

At this stage, businesses can utilize some software to collect historical data and present it in a more understandable format and make decision-makers analyze the value by themselves. The main goal is to dig into the details of the events that have already happened. This will help organizations to understand what events led to the success or failure of a particular action and how to achieve more successful outcomes in the future. Further, we will describe the level from the following measures with respect to different perspectives.

Organizational measure :

- Define a road map for integration steps of analytics in cybersecurity.
- Define goals and expected outcomes expected from this initiative.
- Building a data-centric culture within cybersecurity.

Technological measure :

- Define data pipelines - Data management can be defined depending on the size and technological awareness of the organizations
- Define a centralized data infrastructure.
- Data aggregation - At this stage, the data is siloed. Important data from past events can be identified and aggregated for further visualizations and reporting.
- Explore some kind of data analytics such as data reporting and visualizations of historical data to gain insights into what has happened.

4.1.3 Defined

Businesses start to realize the value of analytics and involve technologies to interpret data more accurately. It is concerned with the current plans and diagnostic analytics which will allow the organizations to evaluate the effectiveness of the analytics initiatives by comparing data-driven insights with business outcomes. In this

stage, more complex technology can be used to detect the dependencies and patterns. Statistical operations can be done to figure out the core of the problem. This stage is concerned with respect to what is happening and why. Further, we will describe the level from the following measures with respect to different perspectives.

Organizational measure :

- Data-driven culture needs to be maintained.
- Define key stakeholders and customers and a budget for the analytics integration.
- Define data governance framework.
- Hire expert people in the fields of data engineering, data science, and data/business analytics.

Technological measure :

- Integration of advanced analytics software.
- Diagnostic reporting with help of BI tools and ETL systems to handle data flow and usage of more complex data mining techniques
- Enable a centralized data warehouse or data lake for data storage.

4.1.4 Managed Measured

This stage is the starting point of advanced analytics. It involves providing forecasts for the outcomes of the planned initiatives. The overall data infrastructure is more sophisticated. This stage is concerned with respect to what is going to happen. Further, we will describe the level from the following measures with respect to different perspectives.

Organizational measure :

- Integration of analytics expertise in cybersecurity teams.
- Define funding and ROI for automated systems.
- Enhance monitoring systems.

Technological measure :

- Introduce MLOps and DataOps.
- Define an machine learning infrastructure.

4.1.5 Optimized

At its most advanced level, analytics goes beyond predictive modeling to automatically recommend the best course of action and optimization options based on massive amounts of historical data, real-time data feeds, and information about the outcomes of past decisions. Further, we will describe the level from the following measures with respect to different perspectives.

Organizational measure :

- Analytics becomes fully automated, provides decision support, and is seen as a competitive weapon.
- Since analytics is a booming technology, there will always be new ways that can be introduced.
- Introduce policies on the basis of new decisions.

Technological measure :

- Optimization techniques and decision management technology.
- Real-time data and analysis for better decision-making process.

At this stage, the main challenges that an organization can face are not related to further development, but rather to maintaining and optimizing their analytics infrastructure. There can be issues that can include data privacy, lack of knowledge and specialists, data security, etc. So, at this point, organizations should focus on developing their expertise in data science and engineering, protecting customer private data, and ensuring the security of their intellectual property.

NXP - Current Analytics Maturity Level

The proposed analytics maturity model, as discussed in Section 4.1, will assist NXP in understanding the current state of analytics and what steps can be taken to take cybersecurity to the next level. With the help of this research, the NXP-IRM cybersecurity team is stepping into the second stage of maturity and is aware of the value analytics can bring to cybersecurity. **Therefore, with respect to the proposed maturity model, this research is the implementation of moving one level higher i.e. level 2 in the analytics maturity.**

4.2 Data Pipeline Architecture Overview

A data pipeline is a series of processes that collect, modify, and deliver data from beginning to end. The organization will benefit from copying or moving data from one source to another so that it can be stored, used for data analytics, or merged with other data. It will also help businesses in breaking down information silos and obtain value from their data in the form of insights and analytics. The real-time processing data pipeline proposed in this research is composed of two major components:

1. Forcepoint user and entity behavior analytics (FBA) Software Component
    - Data Processing and Ingestion System
    - Application System
2. Data Analytics Component
    - Data Sources and Data Storage
    - Data Visualization component

Figure 4.2 shows the overall architecture of our proposed data analytics pipeline. [54] explains a detailed description of the rest of the components in FBA.

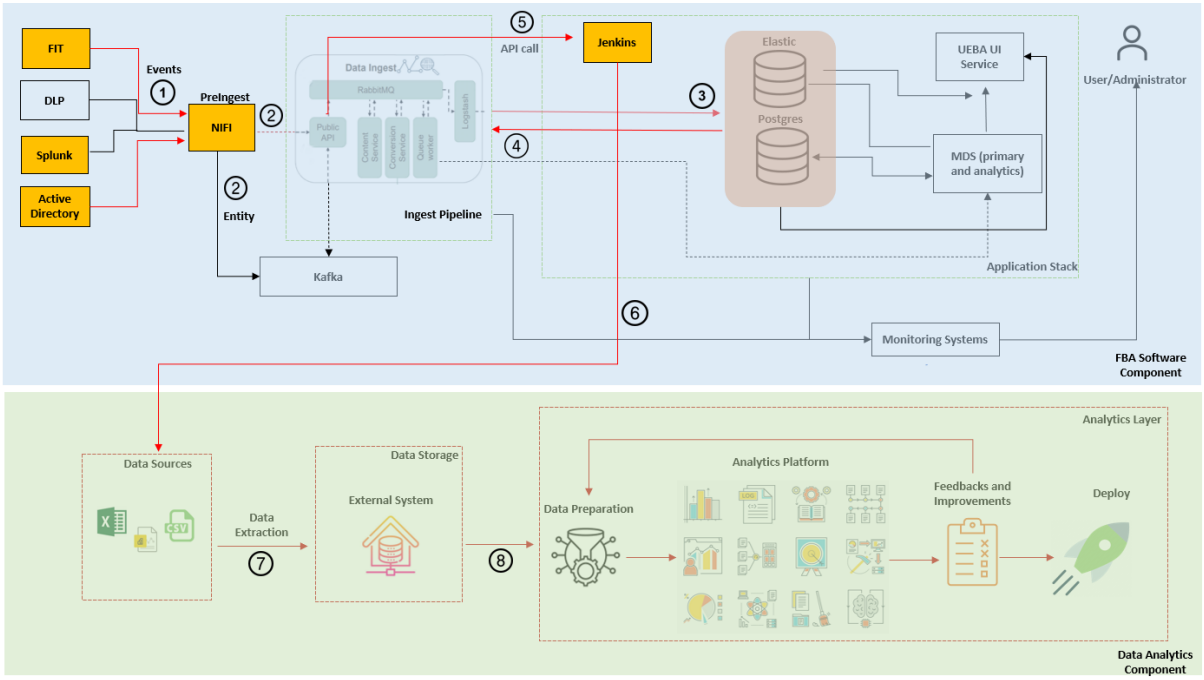


Figure 4.2: Proposed data analytics pipeline architecture

In the following sections, we will go over each of these components in detail.

4.2.1 FBA Software Component

User and Entity Behavior Analytics (UEBA) is a branch of cybersecurity that focuses on analyzing activity within your network environment, specifically user activity, device usage patterns, and security events, to assist businesses in detecting potential insider risks and compromised accounts.

For monitoring insider risks, NXP employs **Forcepoint user and entity behavior analytics (FBA)** [55], a third-party software, as a UEBA Security Solution. FBA is a powerful behavior analysis platform that enables security teams to monitor high-risk behavior on a proactive basis. The industry-leading platform from Forcepoint UEBA combines structured and unstructured data to offer comprehensive visibility into the complex human behavior, patterns, and long-term trends that make up human risk. Further, we will go over the process of extracting data from the monitoring system for our research purposes in detail.

Data Processing and Ingestion System

1. Data Processing System

   (a) Forcepoint Insider Threat (FIT): FIT is a software suite that collects system **'Events'** from endpoints and sends them to FBA for analysis. FIT is installed on the systems of the entity that has access to the organization's IP data. Policies within FIT are deployed to identify disgruntled behavior and data exfiltration of company data across multiple channels. Deep Analytics is run to determine the riskiness of a user's behavior. FIT collects events on any risky events carried out by the entity and sent to the Data ingestion system.

   (b) Active directory: The Active directory has all the raw data of the **'Entity'** that has FIT installed in their system along with all their organization-related information such as department, team, manager name, country code, etc.

   (c) Splunk: The zscaler logs are collected in splunk. Splunk is a software for finding, monitoring, and analyzing machine-generated data through a web-based interface.

2. Data Ingestion System

   (a) Data ingest process starts with pre-ingest systems. These systems pull or receive raw data about entities (users, systems, and resources) and Events (the interactions between users, systems, and resources) from external sources.

(b) Apache NiFi is used to transform raw client data into the FBA format, and reliable, secure data transfer between client sources and the Public API. NiFi is also used for data enrichment, preparation of data, conversions between formats, extracting and parsing, and routing decisions. The pre-ingest systems (NiFi) transform the data so it aligns with the FBA data model, then submit it to the public API.

(c) In FBA, Apache Kafka is used as a data ingestion system for a real-time processing system. Apache Kafka is a high-throughput distributed publish-subscribe messaging system. In Apache Kafka, multiple producers publish the message on a topic and multiple consumers who subscribe to that topic can also consume the messages. Kafka pushes the endpoint data to the Public API for Ingest.

(d) RabbitMQ is a message queuing system with an architecture based on the submission and retrieval of messages in first-in-first-out order. It moves the data through multiple services and processors, which clean up, normalize, and analyze the data.

(e) Logstash prepares the data for storage and transmits it to the Elasticsearch cluster. The Event and Entity data and associated analytics results are stored in the Elasticsearch and Postgres databases respectively.

(f) The Entity resolution systems(Rose and Postgres) contain a map of known Entities and Entity identifiers. For our data extraction, we are using the Postgres server for the extraction of the **'Entity'** dataset.

(g) Monitoring systems contain stack metrics and log files.

Application System

1. Database Systems

   (a) Elastic: Elastic Search is a document database with strong text search capabilities. This is where the database of record for **'Event'** dataset is stored. Therefore the 'Event' dataset is extracted via API call from Jenkins to Elastic search.

   (b) Postgres: Postgres is where the database of record for **'Entity'** dataset and UI user management is stored. Therefore the 'Entity' dataset is extracted directly from Postgres via SSH.

2. Jenkins Server
   Jenkins helps in coordinating automated software tests and builds, or to define

'jobs'. In NXP, Jenkins functions as a recurring job scheduler, essentially acting as a centralized crontab.

API Call : Application Programming Interfaces, are software-to-software communication interfaces. They enable different applications to communicate with one another and exchange information or functionality. An API call is a process by which a client application sends a request to an API, which then retrieves the requested data from an external server or program and returns it to the client.

As a result, we will use Jenkins to make an API call to the database storage in FBA to extract **'Event'** datasets. As shown in Table 5.1, jobs are triggered by providing a query for each model as an input. While running the job Jenkins makes an API call to the database system. After completion of the job, the datasets are extracted from the FBA backend by using SSH File transfer protocol and are then securely stored in a centralized location for further analysis.

4.2.2 Data Analytics Component

The Data Analytics component is the part where the data will be extracted from the FBA software component for further storage and analysis to gain valuable insights.

Data sources

The main data source for our research study is FBA. Datasets extracted from the monitoring system using Jenkins are of the file type .xlsx and.csv and the data extraction method is explained further in Section 5.1.

Data storage

The extracted data needs to be stored in a centralized location for which we will need a data warehouse. It is a type of data management system that enables and supports business intelligence (BI) activities, particularly analytics. Data warehouses exist solely to perform queries and analysis on large amounts of historical data.

For our research, we used **Sharepoint** as a data warehouse with restricted access, with a direct connection to Power BI allowing the organization to use and gain insight from the data. Sharepoint has an effective collaboration and communication space, good security because only authorized members can view the data, and is simple to use.

Data Visualization component

The visualization part of the platform displays the real-time dashboard based on real-time processed data, which helps in both decision-making and visualization purposes. It is more important than ever in our increasingly data-driven world to have accessible ways to view and understand data. Many businesses use data visualization in the form of dashboards to analyze and share information with the goal of making data more accessible and understandable.

The data visualization software we would be using is Microsoft **Power BI**. Its primary focus is business intelligence, and it is a collection of software services, apps, and connectors that work together to transform disparate data sources into coherent, visually immersive, and interactive insights. The software allows us to connect to Sharepoint which then will be used to connect to the datasets stored in Sharepoint with Power BI for further visualization.

4.3 Key Risk Indicators

In this section, we will define the high-level key risk indicators for achieving the organization's business goals with respect to each model discussed in Section 2.5.

Key Risk Indicators (KRIs) is a metrics that organizations use to provide an early warning of rising risk exposures in different areas of the enterprise. It is important that business objectives are clearly communicated throughout the organization's business goals so that when employees understand and are accountable for their particular KRIs, the corporation's larger goals remain top of mind. Benefits of KRIs include advance notice of the potential risks that could damage the organization or insight into possible weaknesses in an organization's monitoring and control tools. Following is the list of all the key risk indicators for all the models.

Sr. No.	KRI	Description	Rationale for measuring this KRI
1	IP data movement within the organization viz., Business Unit, Business Line, Profit Center	The number of events triggered per model and per Business Unit, Business Line, and Profit Center	This metric assesses the risk of how IP data moves within the organization's hierarchy
2	IP data movement per country	The number of events triggered per country	This metric assesses the risk of how IP data moves geographically
3	Number of events triggered	Number of malicious actions triggered by entity	This metric assesses the risk of the events triggered per model
4	Number of users involved	The number of users violating the policies per model	This metric assesses the number of specific users engaging in such malicious behavior
5	File types of the IP Files transferred	Filetype of IP data transferred to removable media, Zscaler, transcend and external web uploads per model	This metric evaluates the IP data file types used to transfer IP data
6	External mail domains to which IP Files are transferred	External mail domains to which IP data is transferred per model	This metric assesses the external mail domains to whom IP data was transferred outside the organization
7	Trend analysis on users - One time or Recurring activity/event per user per model	Event per user per model over the year	This metric evaluates to understand if the user activity with respect to not following the policies and making malicious attempts is a one-time or recurring activity
8	External cloud domains to which IP Files are transferred	External cloud domains e.g Google Drive, iCloud, etc to which IP data is transferred per model	This metric assesses the external cloud domains used by the entity to upload or share IP data
9	Total bytes transferred via Zscaler	Number of bytes transferred on web exfiltrations	This metric assesses the total bytes transferred on web applications
10	Overall Trend Analysis in the organization	Overview trend analysis of all the models and policies violated over the year or quarter	This metric will help in understanding the overall trend of policy violations by an entity, total events triggered, and the riskiest models

Table 4.1: Key risk indicators

Further, the 'Overall Trend Analysis in the organization' is the overview of the implemented models. The following are the KRI which will help in understanding the overview analysis of the implemented models w.r.t. IP data movement in the organization :

- Total number of users violating the policies - This metric accesses the overall entity in the organization violating the policies.
- Total number of events triggered over time and per country - This metric accesses the overall events triggered in the organization.
- Overview of the IP Data movement in the organization hierarchy.
- Trend analysis of the number of events triggered over the time per month within the organization hierarchy.
- Geographical analysis of the events triggered - This metric access the overall events triggered from a particular country.

The key risk indicators discussed above will be implemented in the model for developing dashboards and evaluated further in Chapters 6 and 7 respectively as per business needs. The main stakeholders who would have access to these KRIs and the models are the business people and the IRM-cybersecurity team within NXP.

4.4 Summary and Conclusion

In this chapter, we proposed an Analytics maturity model to help the NXP-IRM team understand their current state of analytics and what steps can be taken to advance cybersecurity. As a result of this research, the NXP-IRM team has advanced to the second stage of maturity and is aware of the value analytics can bring to cybersecurity. And, as discussed in Section 4.1.2, this research is the implementation of moving one level higher, to Level 2, in the analytics maturity. Following that, we proposed the data analytics pipeline architecture in Section 4.2, illustrating its components. We also discussed, in brief, the data extraction process from the FBA component and how to use the data for further data analysis. Finally, in Section 4.3, we proposed key risk indicators that will aid in developing models and achieving the organization's business goals in terms of insider risk management in NXP.

Chapter 5

DATA PREPARATION

In this chapter, we discuss the insights from preliminary data analysis. Section 5.1 discusses the detailed procedure used to extract the data sets for analysis. Section 5.2 provides a brief summary of each type of data. Finally, Section 5.3 focuses on data cleaning and transformation, assessing potential limitations and how they were overcome.

5.1 Data Extraction

The data is extracted from the forcepoint behavior analytic monitoring system, as discussed in Section 4.2. The monitoring system stores data only for the past 3 months because of the retention policy of the organization. To address this issue, jobs were scheduled to extract and augment data every week in order to have sufficient data for analytical purposes and to obtain precise insights and comparisons for data analysis.

Two types of datasets are extracted from FBA viz., entity and events datasets. The entity dataset is extracted directly from Postgres via Secure Shell Protocol (SSH). It contains all the aggregated data of employee information while the events dataset contains data of their activities/events. As discussed in Section 4.2, the events dataset is obtained by triggering queries for various models via Jenkins. Table 5.1 shows the queries used in Jenkins for data extraction for each model, based on entity activity.

Sr. No.	Model	Description	Dataset event start date	Query
1	IP File copied to removable media	Movement of IP files to removable media storage (HD, USB, Floppy, CD/DVD)	Sep 01,2022	feature.value:("FIT file copy to removable media">0) AND (feature:("IP File Extension of Interest HIGH risk V002" > 0) OR feature:("Archived file extensions HIGH Risk V002" > 0) OR feature:("Archived file extensions MEDIUM Risk V002" > 0) OR feature:("IP File Extension of Interest MEDIUM risk V002" > 0))
2	External IP data movement to external mail domains	Movement of IP files to external mail domains	Sep 01,2022	((mode="email") AND entity:("App":"fit") AND attribute:("Policies":"Analytics - Email Audit")) AND feature.value:("External Mail Domain">0) AND feature.value:("Has Attachment">0) AND NOT feature.value:("Noise Attachment">0) AND (feature:("IP File Extension of Interest HIGH risk V002" > 0) OR feature:("Archived file extensions HIGH Risk V002" > 0) OR feature:("Archived file extensions MEDIUM Risk V002" > 0) OR feature:("IP File Extension of Interest MEDIUM risk V002" > 0))
3	External IP Data Movement to web upload	IP files uploaded to external web domains	Sep 01,2022	feature.value:("FIT Web Upload">0) AND NOT feature.value:("FIT Cloud Storage Upload">0) AND NOT feature.value:("NXP Web Domain">0) AND NOT feature.value:("NXP Sanctioned Cloud Domain">0) AND NOT feature.value:("Noise Attachment">0) AND (feature:("IP File Extension of Interest HIGH risk V002" > 0) OR feature:("Archived file extensions HIGH Risk V002" > 0) OR feature:("Archived file extensions MEDIUM Risk V002" > 0) OR feature:("IP File Extension of Interest MEDIUM risk V002" > 0))
4	IP Data Movement to External Cloud Storage	Movement of IP files to external cloud storage	Sep 01,2022	feature.value:("External Cloud Storage V002">0) AND (feature:("IP File Extension of Interest HIGH risk V002" > 0) OR feature:("IP File Extension of Interest MEDIUM risk V002" > 0) OR feature:("Archived file extensions HIGH Risk V002" > 0) OR feature:("Archived file extensions MEDIUM Risk V002" > 0)) AND NOT feature.value:("Noise Attachment">0)
5	IP Data Movement to Printer	Movement of IP files to printer	Sep 01,2022	feature.value:("Print Activity">0) AND (feature:("IP File Extension of Interest HIGH risk V002" > 0) OR feature:("IP File Extension of Interest MEDIUM risk V002" > 0))
6	External IP Data Movement to Zscaler	IP files web exfiltration	Oct 01,2022	mode="zscaler" AND (feature:("IP File Extension of Interest HIGH risk V002" > 0) OR feature:("Archived file extensions HIGH Risk V002" > 0) OR feature:("Archived file extensions MEDIUM Risk V002" > 0) OR feature:("IP File Extension of Interest MEDIUM risk V002" > 0))
7	External IP Data Movement to Intercom Transcend	Movement of IP files to Transcend	Sep 01,2022	mode="transcend" and attribute:("App"="InterCom") and attribute:("Action"="Put") AND ((feature:("IP File Extension of Interest HIGH risk V002" > 0) OR feature:("IP File Extension of Interest MEDIUM risk V002" > 0)) OR feature:("Archived file extensions HIGH Risk V002" > 0) OR feature:("Archived file extensions MEDIUM Risk V002" > 0) AND NOT feature.value:("Noise Attachment">0))

Table 5.1: Data extraction queries per model

The features used in the query to extract the data are as follows :

1. Archived file extensions HIGH risk V002.
2. Archived file extensions MEDIUM risk V002
3. FIT Web Upload - IP Files uploaded to websites.
4. External Mail Domain - Mails containing an External Address.

5. External Cloud Storage V002
6. FIT file copy to removable media - Movement of IP files to removable media storage (HD, USB, Floppy, CD/DVD).
7. Has Attachment and Noise Attachment - Attachments in emails.
8. IP File Extension of Interest HIGH risk V002.
9. IP File Extension of Interest MEDIUM risk V002.
10. NXP Mail Domain - NXP internal mail domains.
11. NXP Web Domain - NXP internal web domains.

Some features are categorized into high and medium-risk file names and extensions. High-risk files include .cdk, .ao, .il, .ile, .def, etc. and medium-risk files include .lef, .sdf, .svrf, etc. The raw data is extracted and stored in **Sharepoint** which is accessible to specific people to enable further analytical capabilities.

5.2 Data Description

Entity data

An entity is a person (actor) in the organization who has access to Intellectual Property (IP) data. An entity is defined as "who they are". Currently, a total of approx 15000 entities are ingested in the monitoring system which we have employed for our analysis. The data set contains the aggregated information of all the employees. The components that can be found in the raw dataset and the ones selected for further analysis are the following :

Sr. No.	Features	Description	Column selection
1	actorid	Unique id of the entity	✓
2	key	Contains the entity attributes	✓
3	value	Contains the value of the respective entity attributes from the key column	✓

Table 5.2: Entity raw dataset (Feature selection)

Figure 5.1 depicts a snapshot of the entity raw dataset for a single entity.

actorid	key	value
User1	Status	Active
User1	Country Code	AT
User1	Monitored Entity	TRUE
User1	Risk Level	1
User1	Employee Id	13964637
User1	Email Address	user1@abc.com
User1	Title	TEMPORARY CONTRACT
User1	Cost Center Code	AT61SE0080
User1	Business Line Description	Secure Car Access
User1	Profit Center Manager	Markus
User1	Cost Center Description	R62 Sust. & Quality Engineers
User1	Business Unit Manager	Jens
User1	Manager	Marc
User1	Business Line Manager	Markus
User1	Department Manager	Bernhard
User1	Profit Center Description	Secure Car Access
User1	Profit Center Code	R62
User1	Business Line Code	BLC2
User1	Business Unit Code	BU0519
User1	Business Unit Name	Advanced Analog
User1	Full Name	User1
User1	Employee Type	Employee
User1	Risk Level	2
User1	End Date	2049-12-31T00:00:00.000Z

Figure 5.1: Example - Entity raw dataset

NXP organization hierarchy: The sequence of members in an organization is determined by the authority. It refers to the progression of employees from entry-level to high-ranking managers or executives. Organizational hierarchies typically have multiple levels, with members with greater authority occupying higher positions. Figure 5.2, shows the organizational hierarchy of NXP.

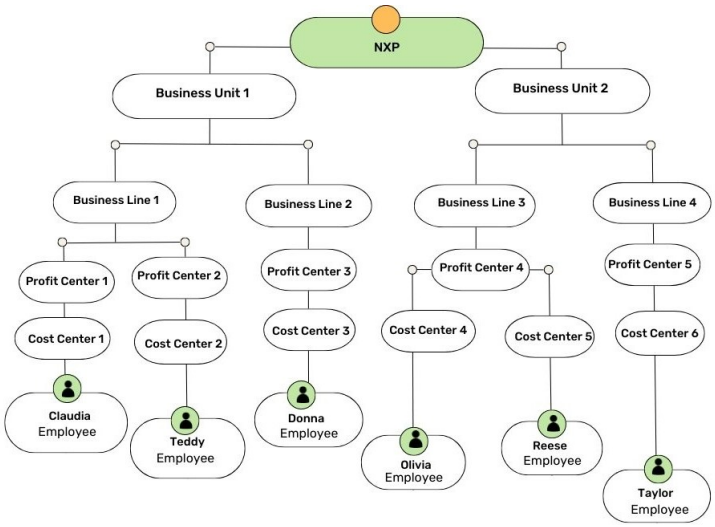


Figure 5.2: NXP organization hierarchy

Furthermore in the entity dataset the column *key* contains the following attributes for each *actorid* (entity) but not all the rows are selected for our analysis which can

be seen in Table 5.3. The reason behind the chosen features was that the analysis needs only insights into the organization and country code of the entity.

Sr. No.	*'Key'* column of entity dataset	Description	Column selection
1	Business Line Description	Business line name of the entity	✓
2	Business Line Code	Business line code of the entity	x
3	Business Line Manager	Business line manager name of the entity	x
4	Business Unit Name	Business unit name of the entity	✓
5	Business Unit Code	Business line code of the entity	x
6	Business Unit Manager	Business unit manager name of the entity	x
7	Profit Center Description	Profit center name of the entity	✓
8	Profit Center Code	Profit center code of the entity	x
9	Profit Center Manager	Profit center manager name of the entity	x
7	Cost Center Description	Cost center name of the entity	✓
8	Cost Center Code	Cost center code of the entity attributes	x
9	Department Manager name	Department manager name of the entity	x
10	Team Manager name	Team Manager name of the entity	x
11	Country Code	Country of entity work location	✓
12	Email Address	Email id of entity	x
13	Employee Id	Employee id of entity	x
14	Employee Type	Employee type of the entity e.g. Permanent, contractor	x
15	End Date	End date of an employee (Default value set to 2049-12	x
16	Full name	Full name of the entity	x
17	Status	Status of the entity if its active or not	x
18	Title	Entity title	x
19	Monitored Entity	Entities that receive a risk score. The value is a Boolean value - TRUE or FALSE	x
20	Risk Level	Risk level (on the scale of 0-5) of each entity	x

Table 5.3: Entity raw dataset - 'key' column

The event dataset consists of the data of activity performed by each entity. **One entity can have multiple events in the same or different timestamps depending on the actions triggered.** There are different models we are analyzing in our research as discussed in Section 2.4. And therefore the dataset variables vary for different models extracted for our research study. There were approx 50 columns in the raw dataset for each of the models. Table 5.4 shows the columns chosen for analysis from all of the event datasets considered for this study. The reasons for the features chosen were entirely dependent on the KRIs chosen, as discussed in Section 4.3.

Sr. No.	Features	Description	Model
1	Action	Action triggered by entity e.g. File Copied, File Moved	All
2	Bytes out	Data transferred in bytes	4
3	Category	Category of the activity performed by the entity	All
4	Destination	Location of the file	4
5	Destination Device Name	Device name of removable media	1
6	Destination Drive Letter	Drive letter of the location	1
7	Features	Features triggered for the type of activity	All
8	File extension	Extension of the file	1,3,4
9	Policies	policy violated by the entity for each event	All
10	Recipient	Recipient mail domain	2
11	Site	Site used for web upload	3
12	Time	Timestamp of the event	All
13	User	Actor id	All

Table 5.4: Event dataset (Feature selection)

5.3 Data Cleaning and Transformation

The quality of data plays an important role, to allow the decision-maker to gain better insights in order to make reliable decisions faster and better. If the data is not accurate, inferior reporting and poor business decisions that can have potentially serious consequences on the entire organization. Although, data is never clean and it may involve incomplete, inaccurate, or irrelevant data.

Data Management Association International (DAMA) [56] , identified key elements for data quality, and we would clean and check our data sets for each variable based on the following dimensions :

1. Accuracy - Is the data correct?
2. Ambiguity - Does the data lacks some information?
3. Completeness - Are there any missing values or records?
4. Consistency - Is the data stored consistently in other databases?
5. Relevance - Is the data relevant to the objective?
6. Uniqueness - Are there any duplicate values?

As part of our research, we used **Power BI**, which is a component of Microsoft Power with a primary focus on business intelligence, in the further sections for data processing and model development. The primary reason for selecting this tool was that it is the most commonly used business intelligence tool in the organization. From the Power BI apps, 'power query' is used to connect, combine, and transform data from the data sources and 'power view', a data visualization tool, is used to generate interactive charts, graphs, maps, and other visuals. However, it should be noted that any other business intelligence platform with the ability to preprocess and visualize data can be used as well.

5.3.1 Final Entity Dataset

The entity dataset consisted of approx 15000 entity/actor id. The latest data extraction for this dataset was done in December 2022. The cleaning steps done on the dataset are as follows :

1. The entity data was accurate and no ambiguous data was found.
2. Unimportant columns were discarded.
3. No duplicate values and missing values were found in the dataset. Each entry refers to a unique actor id and its attributes.
4. The dataset is relevant to the objective and consistent throughout the databases.
5. The *'Key'* column is pivoted i.e. turned the rows to columns for better understanding and analysis.
6. Transformed the *'actorid'* feature to lowercase to avoid any misinterpretation for further transformations.

Following is the list of final features and the datatype for the entity dataset:

Sr. No.	Feature	Description	Datatype
1	actorid	entity user id	string
2	Country Code	Country of entity	string
3	Business Unit Name	Business Unit of entity	string
4	Business Line Description	Business line of entity	string
5	Profit Center Description	Profit center of entity	string

Table 5.5: Final selected features of 'Entity' dataset

Figure 5.3, depicts a snapshot of one of the entities from the cleaned dataset.

actorid	Business Unit Name	Business Line Description	Country Code	Profit Center Description
User1	CTO	CTO-CTO	NL	RF Competence Center

Figure 5.3: Final 'Entity' example dataset

5.3.2 Final Event Dataset

The event dataset was extracted separately for each model via Jenkins using the queries and is stored in Sharepoint as discussed in Section 5.1. The cleaning steps done on the datasets are as follows:

1. Unimportant features were discarded from all the datasets of the respective models.
2. No duplicate values were found in the dataset.
3. Missing values were replaced with 'NA'. Each entry of the data referred to a unique actor id and its attributes.
4. The dataset is relevant to the objective and consistent throughout the databases.
5. Transformed the *'User'* feature to lowercase to avoid any misinterpretation for further transformations.

Data transformation: A left outer join is performed with each event dataset of the models with the column *'User'* and from entity dataset column *'actorid'* as seen in Figure 5.5. The new merged columns from the Entity dataset were named Sender.Business Line Description, Sender.Business Unit Name, Sender.Country Code and Sender.Profit Center Description. This is done to obtain a better analysis of the entity-triggered event and then drill down to the information regarding which level of organization has the most IP risk and from which country. Also, new features are added - 'Year', 'Month Number', 'Quarter', and 'Month' on the basis of the 'Time' feature.

Figure 5.4 depicts the data architecture and clusters of the event model datasets. It also reveals the entity's one-to-many relationships with the events dataset.

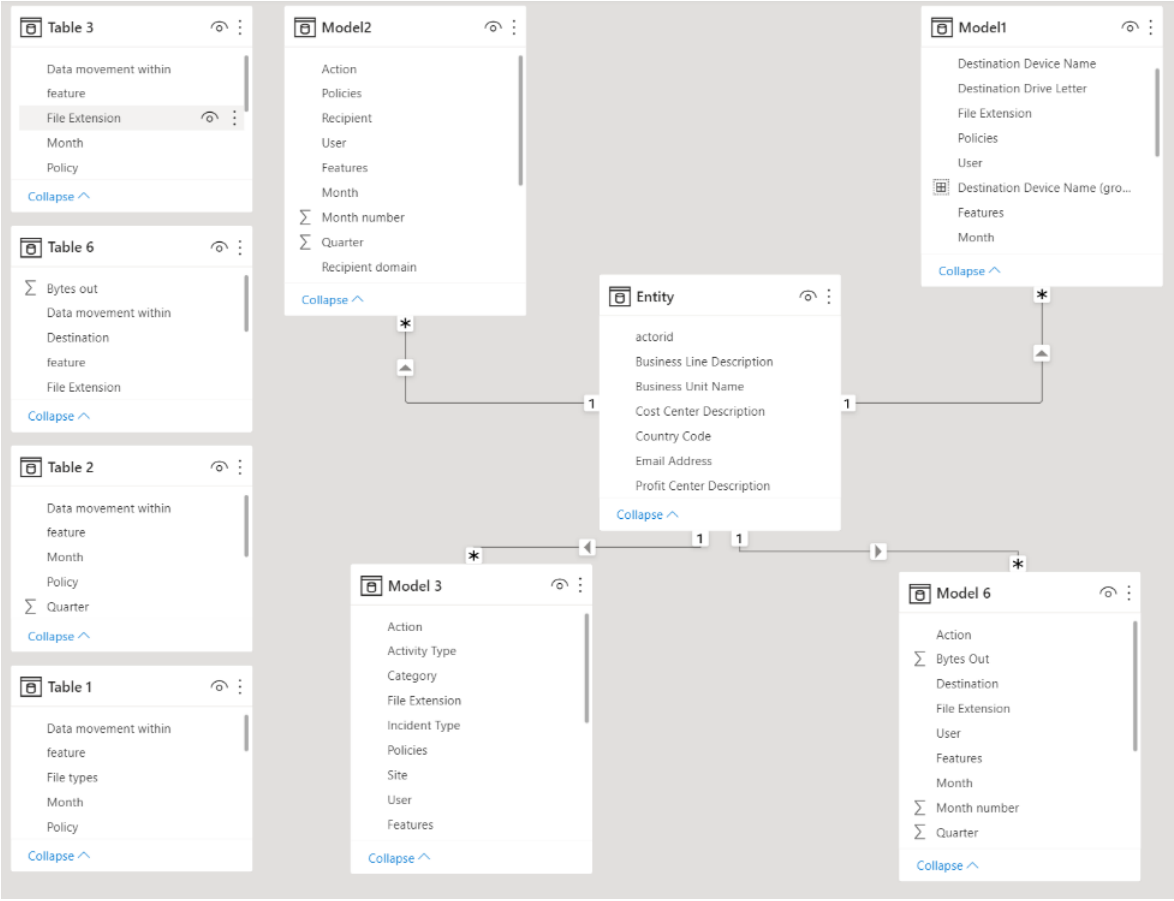


Figure 5.4: Data architecture (Star Schema)

Section 2.4 explains the reason for data extraction in different model clusters. It should be noted that additional models are possible with the current research proposal to obtain valuable insights for managing insider risk. Table 5.6 summarizes the dataset size, number of events per model dataset, and number of missing values.

Model No.	Model Name	Number of Events	Dataset size	Missing values
1	IP File copied to removable media	3.35M	1.06 GB	645
2	External IP data movement to external mail domains	1323	6.52 MB	4
3	External IP Data Movement to web upload	948	398 KB	1
4	IP Data Movement to external cloud storage	198	85 KB	0
5	IP Data Movement to printer	32	14.2 KB	0
6	External IP Data Movement to Zscaler	10989	1.16 MB	47
7	External IP Data Movement to Intercom Transcend	22	8.23 KB	0

Table 5.6: 'Event' models dataset summary

Dataset limitations

There were certain limitations in the data quality, which are mentioned below:

1. Missing values

   In some cases, entities from the event dataset had left the organization and were not present in the monitored entity dataset which led to missing values after the join. Therefore from the current analysis, such entities from the event dataset had to be removed. Although this needs to be taken into consideration in the future since IP data is anyways exfiltrated by these entities and needs to be further investigated.

2. Improper data

   - There was no proper column mentioning the specific 'cloud storage' to which IP data was transferred in the 'external IP data movement to the cloud storage' model dataset. The values were scattered across multiple columns and were formatted incorrectly. For example, an event in which data is moved to Google Drive was present in the 'Destination Device Name' column with the condition if the device name begins with "\Device\Volume," or an event in which IP data is moved to pcloud was present in a column named 'isPcloud' with the value "True". As a result, the dataset was extremely unreliable in determining the correct value of the external cloud storage.
   - There was no column mentioning the recipient of the data transferred to in the 'external IP data movement to Intercom transcend' model dataset, making this model also very unreliable.

3. Inappropriate File count and File extension

   The model with IP files sent via emails didn't have a proper file count which made it difficult to analyze the number of IP files sent. The file type for each event was given in a generic format such as (jpg, pdf, zip) therefore we couldn't figure out the actual number of files sent with the mentioned file type. Therefore, counting the number of events was taken into consideration instead of the number of files and file types per model.

The limitations listed above were either removed or were not considered for this research study. However, with the help of this study, the organization was made aware of these limitations and they are currently investigating the identified issues.

One solution is to feed data into the FIT agent in the correct format. This will help with making better decisions as well as gathering and helping us in analyzing more accurate data. Table 5.7 shows the final list of feature components used for further analysis after applying the cleaning and transformation steps to the events dataset.

Sr. No.	Feature	Description	Datatype	Model
1	Action	Action triggered by entity e.g. File Copied, File Moved	string	All
2	Bytes out	Data transferred in bytes	whole number	4
3	Category	Category of the activity performed by the entity	string	All
4	Destination	Location of the file	string	4
5	Destination Device Name	Device name of removable media	string	1
6	Destination Drive Letter	Drive letter of the location	string	1
7	Features	Features triggered for the type of activity	string	All
8	File extension	Extension of the file	string	1,3,4
9	Policies	policy violated by the entity for each event	string	All
10	Recipient	Recipient mail domain	string	2
11	Site	Site used for web upload	string	3
12	Time	Timestamp of the event	date/time	All
13	User	Actor id	string	All
14	Sender.Business Line Description	Business line of the entity	string	All
15	Sender.Business Unit Name	Business Unit of the entity	string	All
16	Sender.Country Code	Extension of the file transferred	string	All
17	Sender.Profit Center Description	Country code of the entity	string	All
18	Year	Year of the event triggered	whole number	All
19	Month Number	Month number of the event triggered	whole number	All
20	Quarter	Quarter of the event triggered	whole number	All
21	Month	Month name of the event triggered	string	All

Table 5.7: Final selected features of 'Event' dataset

Finally, Figure 5.5 shows the final cleaned and transformed dataset for one of Model.

Time	Features	User	Destination Device Name
9/1/2022 4:59:44 AM	IP File Extension of Interest MEDIUM risk V002	User1	\Device\HarddiskVolume7 ASMT ASM236X NVME 0
9/1/2022 4:59:44 AM	IP File Extension of Interest MEDIUM risk V002	User1	\Device\HarddiskVolume7 ASMT ASM236X NVME 0
9/1/2022 4:59:43 AM	IP File Extension of Interest MEDIUM risk V002	User1	\Device\HarddiskVolume7 ASMT ASM236X NVME 0

Destination Drive Letter	Action	Policies	Sender.Business Line Description
D:	File Copied	Analytics - File/Directory Copy/Move Audit	Ind and IoT Edge
D:	File Copied	Analytics - File/Directory Copy/Move Audit	Ind and IoT Edge
D:	File Copied	Analytics - File/Directory Copy/Move Audit	Ind and IoT Edge

Sender.Business Unit Name	Sender.Email Address	Sender.Country Code	Sender.Profit Center Description
Edge Processing	max.palumbo@nxp.com	US	Connectivity
Edge Processing	max.palumbo@nxp.com	US	Connectivity
Edge Processing	max.palumbo@nxp.com	US	Connectivity

Year	Month	Quarter	Month number
2022	September	3	9
2022	September	3	9
2022	September	3	9

Figure 5.5: Final 'Event' example dataset (Model 2)

5.4 Summary and Conclusion

In this chapter, we focused on the reflection of steps undertaken to retrieve, describe, and check the quality of our early data, which was the initial practical step in our study trajectory. Firstly, two different types of data sets were retrieved: entity and event data. The data was limited to the study for the period of 2022 since the data beyond that wasn't available due to the organization's retention policy. Secondly, the quality of the data has promising potential, but in some cases, it fails. An additional remark is that the events data variables for each model are different depending on the actions triggered by the entity, Hence, their sizes are not equal. Further, various steps involved in preparing the data for the modeling phase are illustrated. They are presented in a logical sequence, although multiple back-and-forth iterations were performed on them. Data selection and cleansing, for example, had to be repeated nearly every time especially when new requirements were proposed. Section 5.1 describes the data selection process and the justification behind it. Sections 5.2 and 5.3 highlight the steps followed to clean the data and their outcomes, as well as any limitations that occurred on it. To be considered, the data preparation steps were solely applied to the NXP dataset extracted from the FBA. These datasets may be also utilized for other data analysis methods and might differ from different organizations.

Chapter 6

MODEL IMPLEMENTATION

In this chapter, models are implemented and business reports are designed based on the discussion of the preceding chapters to integrate the requirements of the organization's success in terms of managing insider risks. This chapter discusses the key takeaways of the insights gained from each of the models followed by a discussion on the risk they possess with the help of a risk matrix. To be noted, the datasets and the data analysis are specific to the NXP case study and may differ from the organization's requirements along with a possibility that different analytical methods may be used.

6.1 Implementation process

In this section, we will discuss the implementation process of our research study. In Chapter 5, we discussed the data preparation process. The extracted data will be further used for our data analysis to create a business report dashboard.

A dashboard is a visual representation of the most valuable information necessary to achieve one or more goals, consolidated so that it can be supervised at a glance [57]. In this study, the decision process was to keep in mind how IP data movement occurs as well as to meet the NXP requirements. Furthermore, the majority of business leaders, according to [57] research, use dashboards for the performance of organizations. These assist users in identifying and responding to problems. As a result, dashboards are frequently designed to represent relevant information in order to monitor organizational performance and intervene as needed. In our case, this can be generalized to monitor IP data movement and the risks that insiders pose to the organization. The implemented dashboard consists of the main pages of each implemented model and one overview page.

The dashboard was created with the **Power BI** software. Power BI is simple, more powerful, and user-friendly for BI developers, data analysts, and business an-

alysts than other BI tools such as Tableau or Qlikview. Another reason is that NXP uses Power BI as a BI tool for all of its analytical tasks. Other BI platforms with the ability to connect, transform, and visualize data can also be used. For reasons of confidentiality, data related to this dashboard, such as the entity's employee id, have not been used in this thesis. In the following sub-sections, each model description and the insights it can provide are briefly explained :

6.1.1 Implementation of Intellectual property (IP) Data Movement and Analysis for Each Model

Intellectual property (IP) rights safeguard genuine business assets that can be critical to the success and profitability of your products or services [58]. As discussed in Section 2.4, we presented an IP theft tree, and one of the possibilities of IP theft that must be addressed is data exfiltration, where data exfiltration is a type of security violation that occurs when an individual's or company's data is copied, transferred, or retrieved from a computer/server without authorization of the owner. For that purpose, we will implement our model to learn more about the possible data exfiltration methods within NXP in greater detail and the significance and impact the proposed models will have in the decision-making process.

Model 1 - IP File copied to removable media

The usage of removable media, such as a flash drive, within an organization, can be risky because malware can bypass the security mechanism deployed at the network perimeter [59]. Every person has at least one flash drive stashed away in a desk drawer, and many individuals still use them on a daily basis. Despite the fact that cloud storage is slowly pushing USBs into the computer history museum, they remain among the most commonly used data exfiltration channels. Data exfiltration occurs when an entity has authorized access to IP data and transfers it to an insecure local removable media, such as USB, smartphones, cameras, computers, or other specialized devices, either by downloading files or folders from cloud services or copying the information into new files [59].

With USB drives, unplanned insider incidents are all too common. Users could lose USB drives containing sensitive company or personal data, or these USBs may fall into the hands of unauthorized individuals. USB drives are frequently used by malicious actors to distribute malware to local computers. According to Kaspersky [60] research, one in every four users worldwide is vulnerable to these types of "local" malware threats. However, any files transferred to an insecure device pose a high risk of data exfiltration.

Model 2 - External IP data movement to external mail domains

Emails are one of the most common methods to exchange information nowadays and therefore a lot of contractors and employees rely on such technologies on a daily basis to do their jobs well [61]. With the help of such channels, a large amount of sensitive information can be sent, making them particularly effective for data exfiltration. Every day, both at work and at home, we receive and send emails. However, emails are an insecure way to transfer IP data, and therefore email IP data movement cannot be trusted. Such scenarios occur when legitimate users attach files containing sensitive data to emails and send them to their personal email accounts or any other external or internal domains. They can also copy-paste sensitive information directly into the body of emails or forward confidential internal communications to their own or competitors' email addresses [61].

Furthermore, many messaging and email platforms save drafts to the cloud automatically. This means that someone with unauthorized access to their business email or another messaging service that appears to support saved drafts can use that feature to steal information. Simple mistakes may result in sensitive files being sent to the wrong audience or sensitive folders being left unprotected in the cloud, which most often goes unnoticed. In our case, we are interested in the insiders sending IP data to an external email address outside the organization's authorized mail domains.

Model 3 - External IP data movement to web upload

These scenarios occur when a user downloads sensitive data to the local network. The data is then uploaded to a third party by a user using a web browser client or other unmonitored software. A third-party service, such as a social media network, may appear to be harmless, as someone may accidentally drag and drop the wrong text or attach the wrong image.

With respect to this model NXP - IRM team has very less visibility on the web exfiltrations and for that purpose, Zscaler software is used. Zscaler is a software program that assists in identifying data offloads to the network, which are typically browsers but can be any data leaving the endpoint to the internet. With the help of Zscaler, there is a possibility to identify more exfiltrations done on the web applications such as web uploads or external cloud storage to understand more about the IP data movement to web domains. **As a result, we have defined another model, model 6.**

Model 4 - IP data movement to external cloud storage

Data can be exfiltrated from cloud storage services if data is uploaded to insecure or misconfigured resources. Cloud services introduce new exfiltration risks to be aware of, such as cases where employees or administration staff use provider features in insecure ways. Any actor with the ability to deploy code, modify virtual machine instances (VMs), or request cloud storage can exfiltrate data [62]. Furthermore, users with adequate permissions can transfer data from secure containers to insecure containers or create unauthorized services on behalf of the company.

Model 5 - IP data movement to printer

Printers are widely used in organizations to print documents, presentations, etc. However, the usage is much lesser after digitization. These scenarios occur when a user prints sensitive information. If these prints are misplaced or fall into the wrong hands, the data may be exfiltrated.

Model 6 - External IP data movement to Zscaler

Zscaler is a software program that assists in identifying data offloads to the network, which are typically browsers but can be any data leaving the endpoint to the internet. We chose to incorporate this model because the detection of web exfiltration (model 2) is currently having limitations within the NXP-IRM team. Therefore to monitor web exfiltrations, zscaler logs are used in NXP-IRM to analyze the exfiltrations over web netowrk.

Model 7 - External IP data movement to Intercom Transcend

Transcend is a website used by people in NXP to safely transfer data (IP) via the web where they upload the files to the transcend website and then can send a link to another user to download the files. However, transcend is used to transfer data between NXP users. Transcend has a website called Intercom, which is used to transfer files to external users. Therefore, with the help of this model, we can see the IP data movement to external users via Intercom transcend.

6.1.2 Main Takeaways

After analyzing and visualizing all the defined models the **main key takeaways** were as follows :

1. The number of events for model 1 was the highest when compared to other models, indicating that removable media is widely used to copy IP data. Another data exfiltration method used the most is model 2, indicating that entities within the organization transfer a large number of IP files to external mail domains.

2. The number of users who have violated the usage of removable media policy is 1354, indicating that the user-to-event-triggered ratio is high and that the users should be notified of their actions. Also for model 2, the number of users is 491 and the number of events triggered is 1323, indicating that the user-to-event-triggered ratio is less. But still, proper decisions and actions need to be taken against the users.

3. The treemap in all the models helps in determining the IP movement with the organizational hierarchy i.e. Business Unit, Business Line and Profit Center. With the help of it, proper measures can be taken on the respective organizational level regarding the usage of removable media policy violations.

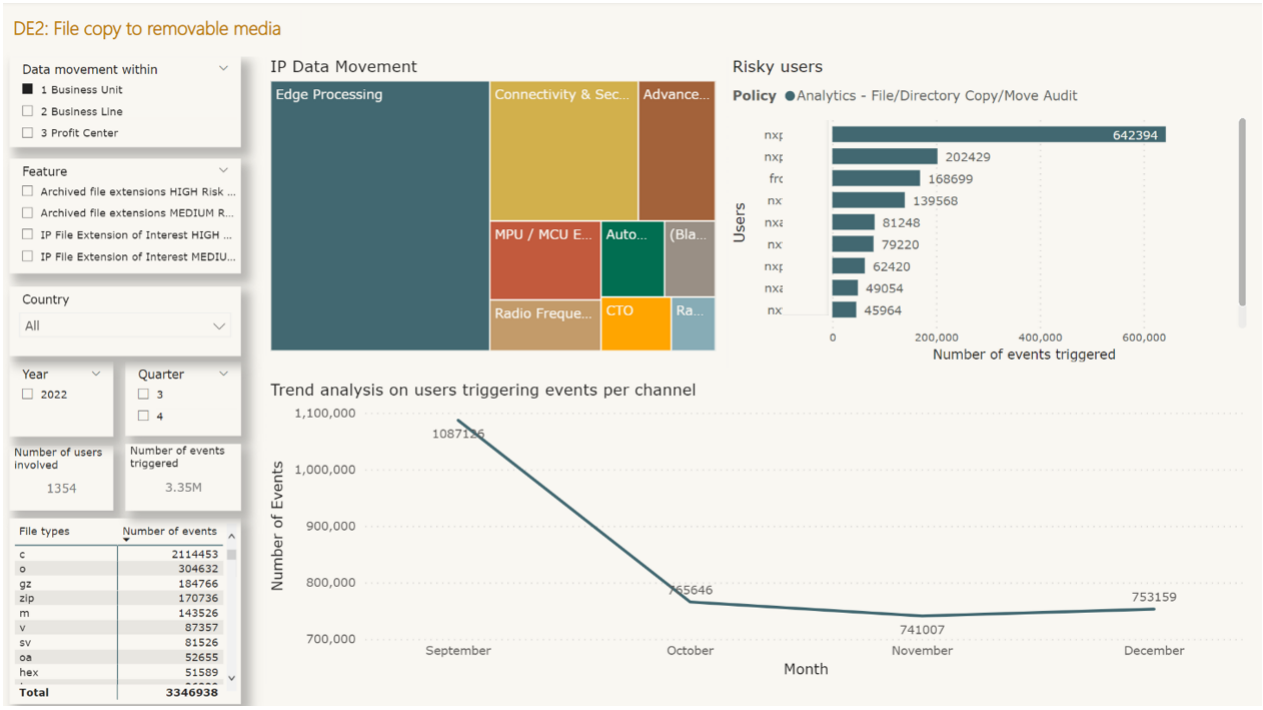
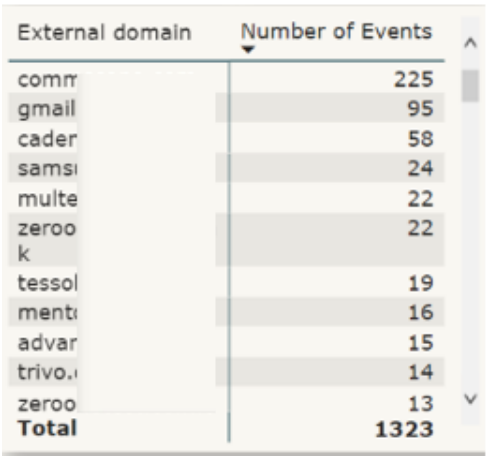


Figure 6.1: File copy to removable media business report (Model 1)

4. The top file types in model 1 are c, o, gz, m, zip, etc with the most number of events. Also, it's logical that the c file type is over represented since often it is created in large batches or part of large databases while zip and gz files are commonly used as well indicating that it can be an additional risk as we do not

know how many files it consists of. Most of the IP data movement is from the Edge processing business unit.

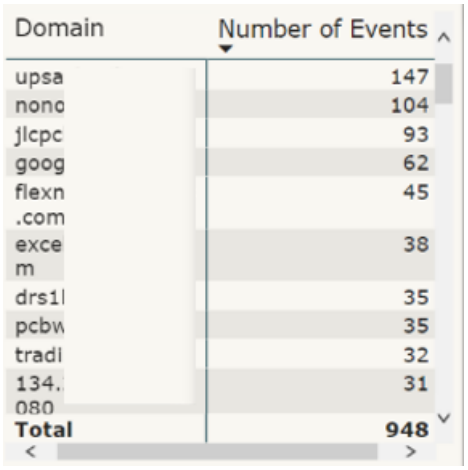
5. In model 2, the top external mail domains with the most events are comms*.com, gmail.com, and others as shown in Figure 6.2. From the analysis, usage of gmail.com indicated that most of the users are transferring IP files to their personal email addresses. And therefore, such risky behavior of transferring files to personal mail addresses and other external domains needs to be taken into consideration while making decisions.



External domain	Number of Events
comm	225
gmail	95
cader	58
samsı	24
multe	22
zeroo k	22
tessol	19
mentı	16
advar	15
trivo.(	14
zeroo	13
Total	**1323**

Figure 6.2: External mail domains (Model 2)

6. The top external web domain in model 3 with the most events is upsa*.fr, but all of the events to this web domain were triggered by a single user, indicating that a large number of files were uploaded to this web domain by one user. Other web domain extensions include google.com, pcb*.com, outlook.com, etc.



Domain	Number of Events
upsa	147
nono	104
jlcpc	93
goog	62
flexn .com	45
exce m	38
drs1l	35
pcbw	35
tradl	32
134. 080	31
Total	**948**

Figure 6.3: External web domains (Model 3)

7. The bar chart and the line chart in all the models help in recognizing the risky entities with the most policy violations and the line chart helps in identifying if the user is violating the policy is one time or on a recurring level.

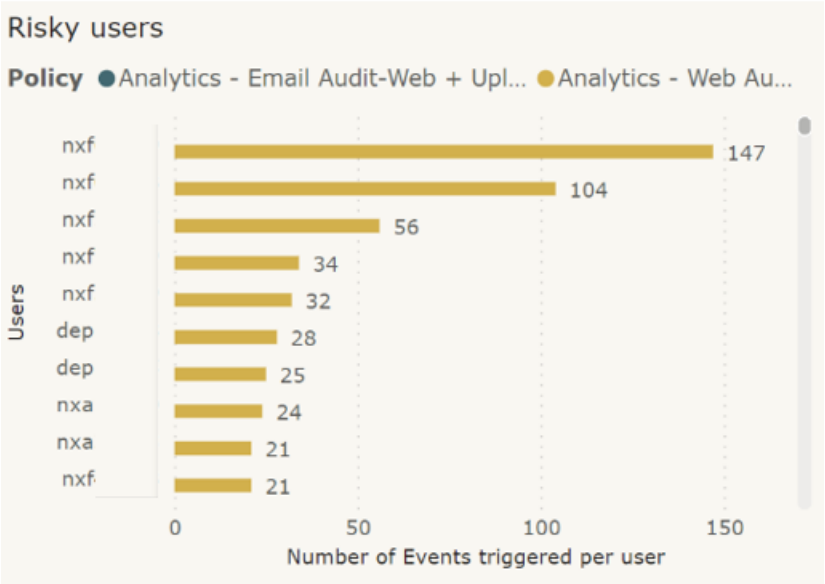


Figure 6.4: Risky users (Model 3)

Because the report is **dynamic**, clicking on any of the fields alters the report's results which makes it easier to use and get proper detailed insights. Also, it can be seen that a lot of users have recurring malicious activities in copying files to removable media. Furthermore, with the help of this insight, additional checks can be performed later to determine whether the transfer was malicious or for business purposes.

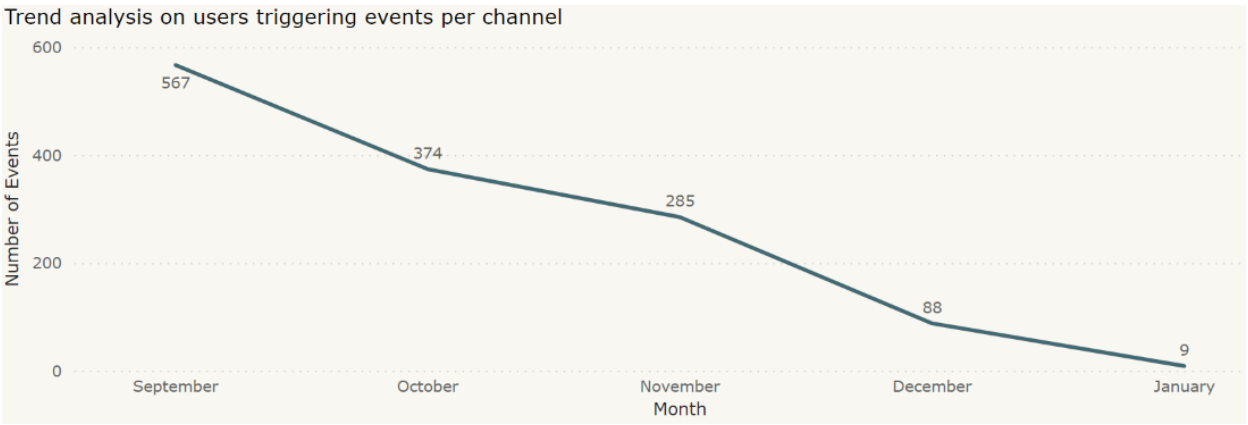


Figure 6.5: Trend analysis on users triggering events per channel (Model 2)

8. One of the limitations of visualizing the trend analysis is that the collected data is not sufficient to understand the trend on a quarterly or yearly basis. However,

this can be resolved in future work by monthly incrementally adding the data or automating the data addition.

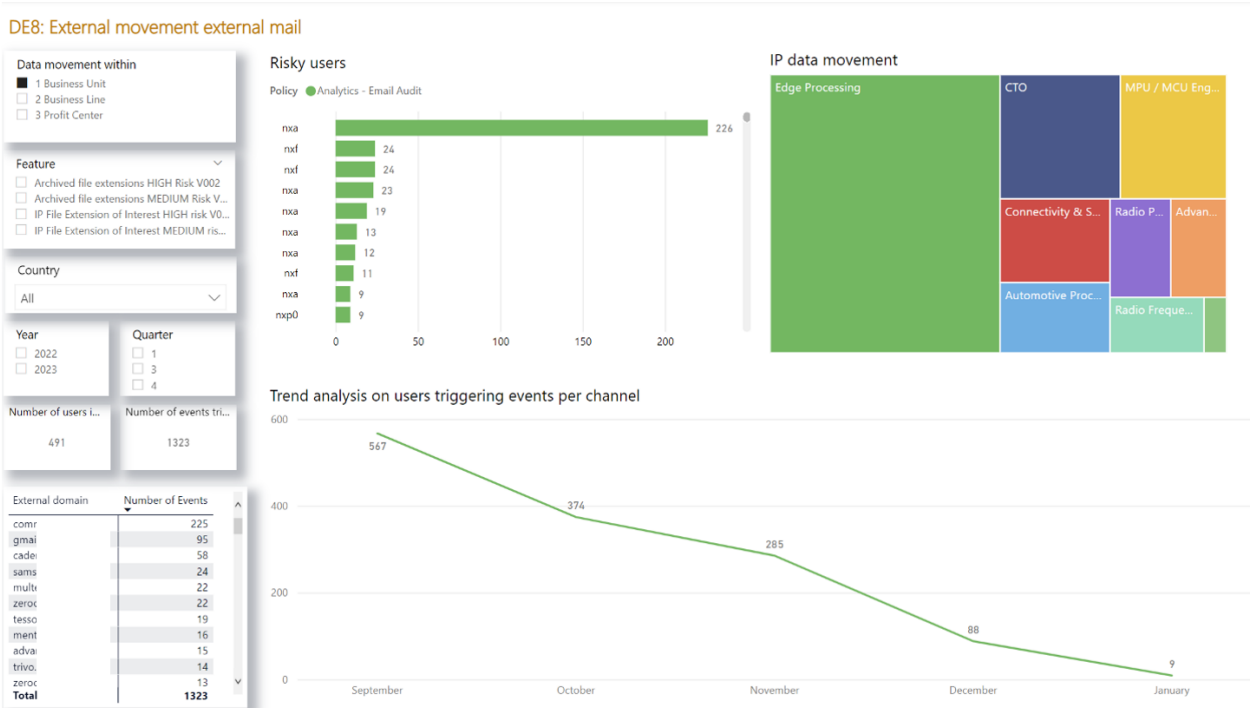
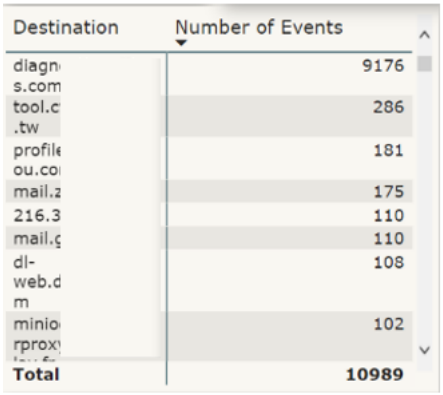


Figure 6.6: External movement external mail business report (Model 2)

9. Due to the dataset limitations for model 4,7 as discussed in Section 5.3.2 decision was taken to not implement this model but can be considered in the future work of this research study after the limitations are resolved by the NXP-IRM team. Although we couldn't get detailed insights from the data it's seen that the number of events and users involved is less. Therefore, it indicates that the risk of IP theft is low and that transcend and cloud storage is not used so much to transfer IP files in the organization.

10. As stated earlier, model 2 had less visibility and hence fewer actions were triggered on the monitoring solution therefore the Zscaler model is introduced. To monitor web exfiltrations, zscaler logs can be used to analyze the web and therefore data was extracted accordingly.



Destination	Number of Events
diagn s.com	9176
tool.c .tw	286
profil ou.co	181
mail.z	175
216.3	110
mail.g	110
dl-web.d m	108
minio rprox	102
Total	**10989**

Figure 6.7: External Web domains (Model 6)

11. The number of events was 32 and therefore the data was very limited for analysis for model 5. Therefore, the decision was taken to not implement this model but can be considered in the future work of this research study after the limitations are resolved by NXP-IRM. However, it indicates that the risk of IP theft is low and that printers are not used as a data exfiltration method to print IP files.

12. In the case of NXP, since its a semiconductor designer and manufacturer-based company they frequently rely on extremely complex software or hardware designs in the IP sector. As a result of the lack of visibility on the monitoring objective, printing does not pose the greatest risk, as printing such files is frequently not very useful or even possible. However, in other industries, such as finance, printing files can already contain a lot of valuable information, making it a much bigger indicator/data exfiltration risk. As a result, in that case, this study, model, and implementation are done similarly to other models it will be beneficial in gaining business insights and making better decisions.

13. The file types mostly used by users in model 7 are zip and gz files. As a result, we don't know how many files were copied into the zip files it's still a high risk. However, the events are fewer, and the risk of IP theft for this model is less.

Table 6.1 summarizes the final list of KRIs for the models listed above. The key risk indicators and the rationale for the metric for the above-mentioned models are as follows:

1. IP data movement within the organization hierarchy: The main motivation of the research study is to understand the IP data movement within the organization. We can use this metric to drill down throughout the organization's hierarchy and determine the riskiest business unit or business line. A greater number of events within the organization may result in data exfiltration and

harm the organization's reputation. This metric will assist in gaining insights into the movement of IP data across different organizational hierarchies and can help in taking proper actions.

2. IP data movement per country: This metric will assist in determining which countries are violating policies geographically and will aid in prioritizing countries while managing insider risk.

3. Number of events triggered: Calculating the number of events triggered is essential because the greater the number of events triggered, the more users are violating policies and putting IP at risk which may result in data exfiltration and harm the organization's reputation. This metric will assist in gaining insights into the movement of IP data across different organizational hierarchies and can help in taking proper decisions.

4. Number of users involved: The greater the number of entities, the more likely it is that insiders are not following the organization's policies. The entities are the employees with access to IP data. This metric will aid in gaining insights into users who are violating policies and putting IP at risk, as well as help in taking appropriate decisions against those users.

Sr. no.	Model name	Number of Events triggered	Number of Users involved
1	IP File copied to removable media	3.35M	1354
2	External IP data movement to external mail domains	1323	491
3	External IP Data Movement to web upload	948	203
4	IP Data Movement to external cloud storage	198	4
5	IP Data Movement to printer	32	11
6	External IP Data Movement to Zscaler	10989	377
7	External IP Data Movement to Intercom Transcend	22	13

Table 6.1: Overall results for KRI's 2 and 3

5. File types of the IP Files transferred: This metric will help us understand which file types are most commonly used by the entity to transfer IP data.

6. External mail domains to which IP Files are transferred: This metric will assist in understanding the external mail domains used the most by the entity to transfer IP data and taking appropriate actions such as restricting the most commonly used external mail domains.

7. Trend analysis on users (one-time or recurring activity/event per user per model): This metric will help in understanding if the users violating policies is one time or a recurring activity performed by them and also will help in making proper decisions against such malicious users/insiders.

8. Total bytes transferred via Zscaler: This metric assesses the total bytes transferred on web applications

Sr. No.	KRI	Model
1	IP data movement within the organization viz., Business Unit, Business Line, Profit Center	All
2	IP data movement per country	All
3	Number of events triggered	All
4	Number of users involved	All
5	File types of the IP Files transferred	1,3,5
6	External mail domains to which IP Files are transferred	2
7	Trend analysis on users - One time or Recurring activity/event per user per model	All
8	Total bytes transferred via Zscaler	6

Table 6.2: Overall KRI analysis

6.1.3 Implementation of Overview Trend Analysis of Intellectual Property (IP) Data Movement

To summarize, the dashboard overview page is displayed in Figures 6.8 and 6.12. Various parameters such as the total number of users violating policies, total events triggered, trend analysis with respect to organizational elements, and geographical overview are displayed on this overview page. The dashboard is dynamic, and therefore it provides detailed insights with the ability to drill down the organizational hierarchy in the treemap as shown in Figure 6.9.

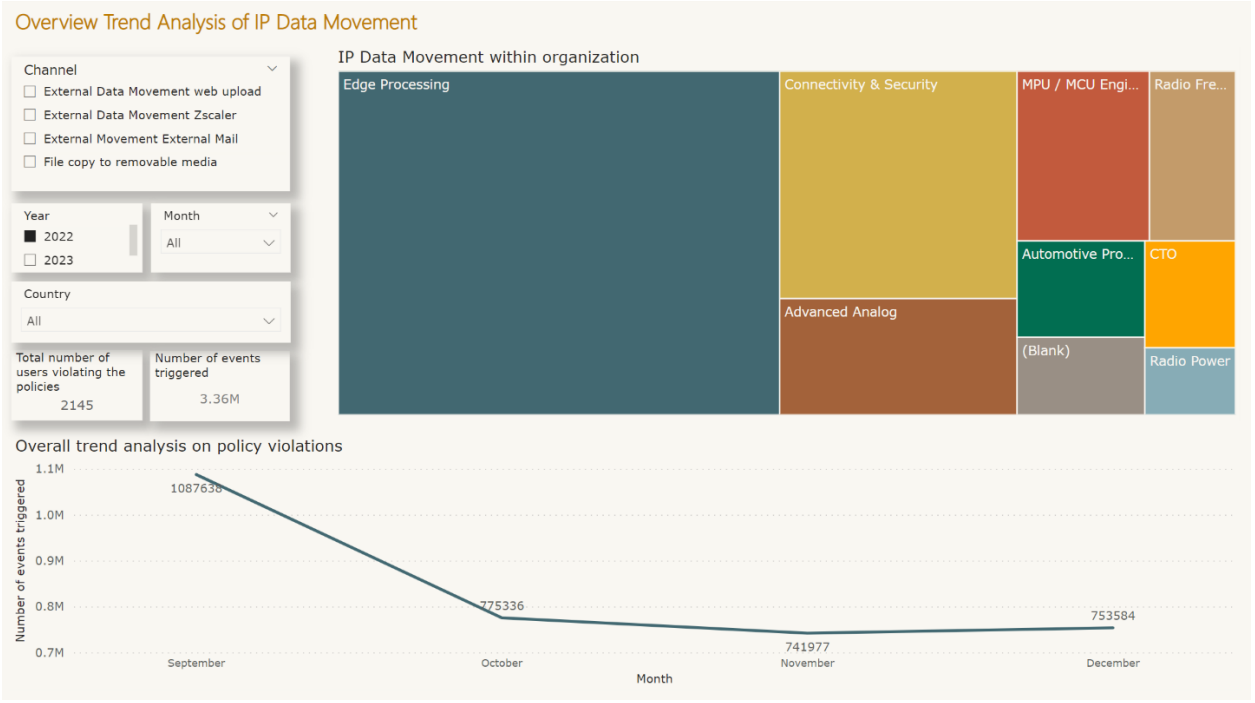


Figure 6.8: Overview Trend Analysis of IP Data Movement (Business Unit)

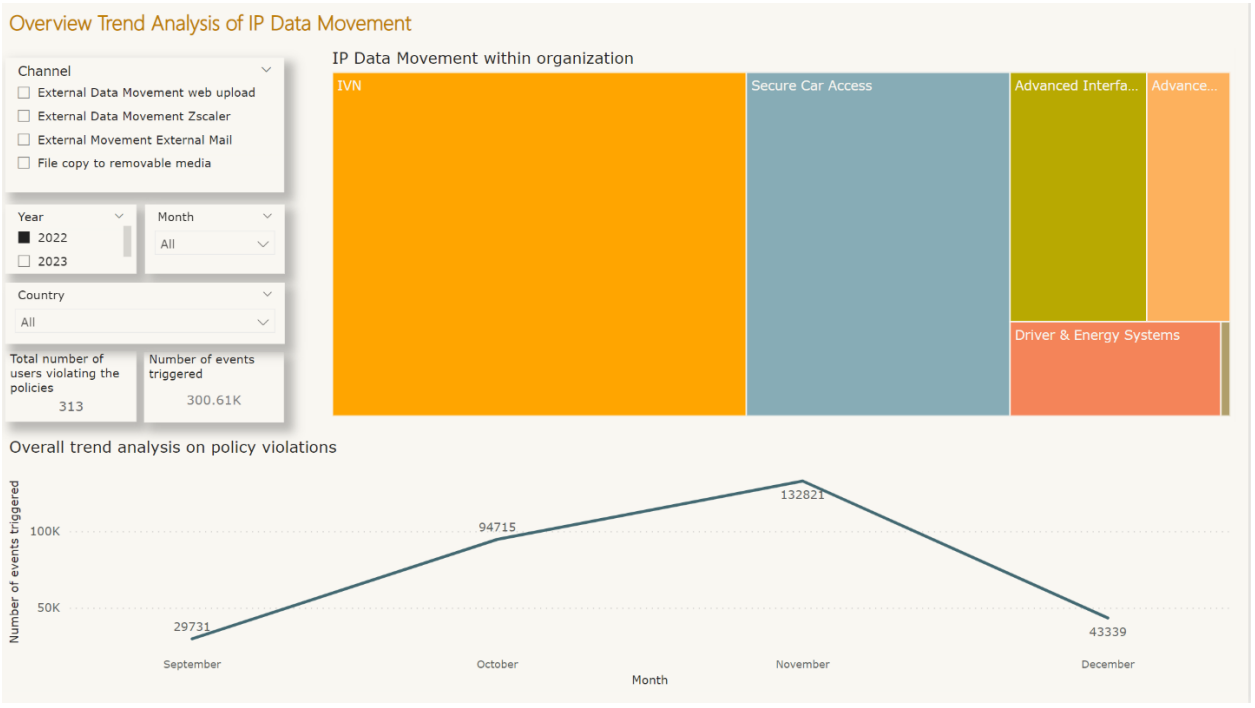


Figure 6.9: Overview Trend Analysis of IP Data Movement (Drill down to Business line)

Overall high-level business insights: Business insights are the knowledge that brings value and exists to serve, create or enhance something in business. With

this research study from the data provided, we are trying to generate intelligence and understand what is happening, why it is happening, and how it can be solved, reversed, or improved. To summarize, high-level insights gained from these implementation methods are as follows :

1. One of the challenges and limitations faced is that the dataset extracted for all of the models is of 4 months, which is a very less amount of data for trend analysis on a quarterly or yearly basis. As a part of future work, the dataset should be updated on a regular period in order to obtain the appropriate amount of data for proper trend analysis over quarters and months to make appropriate decisions to manage insider risk.

2. The most popular data exfiltration method used in NXP is the usage of removable media to copy IP files, external mail domains, and web uploads. The usage of removable media to copy or move IP files is followed very often. Proper decisions must be made for this risky behavior because the consequences can be severe.

3. Gmail.com is the most commonly used external web domain for transferring IP files. This can be an indicator that either the people in the organization are using the personal email address to transfer IP data and measures need to be taken for the same.

4. The IP file movement through the printer, transcend and cloud storage is less, so the IP risk is not as high as with other models. However, we can still see some usage, and therefore new decisions must be taken to change the way of working and reduce the use of printers to print IP documents and transcend to transfer IP files.

5. From the Zscaler model, the total number of bytes sent via web applications is 246 billion, implying that a large amount of data is exfiltrated via web applications. Although the NXP-IRM team should still manage to gain more visibility of these models.

6. Most of the IP data movement is from Business Unit - Edge processing, Business Line - Ind and IoT Edge, PC - Connectivity.

7. The overall trend shows that policy violations are decreasing month after month indicating that the risk is decreasing. However, for analysis, this 4-month data is not sufficient enough to generate a detailed trend to compare on a quarterly or yearly basis.

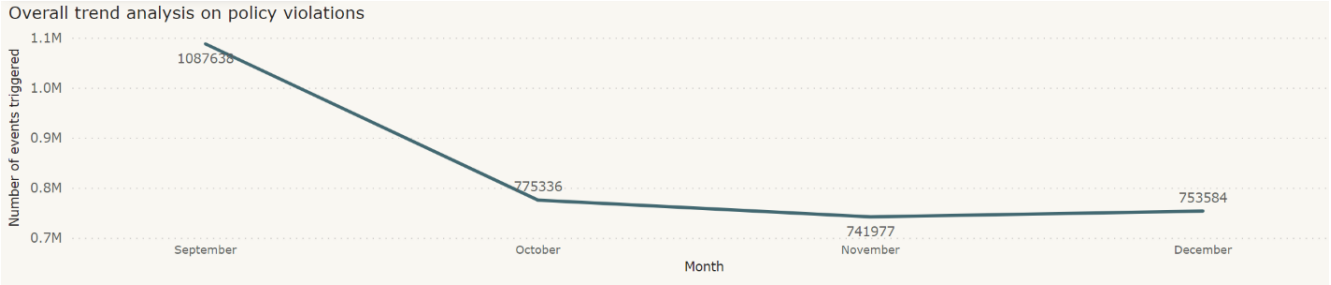


Figure 6.10: Overview Trend Analysis on policy violations (Monthly)

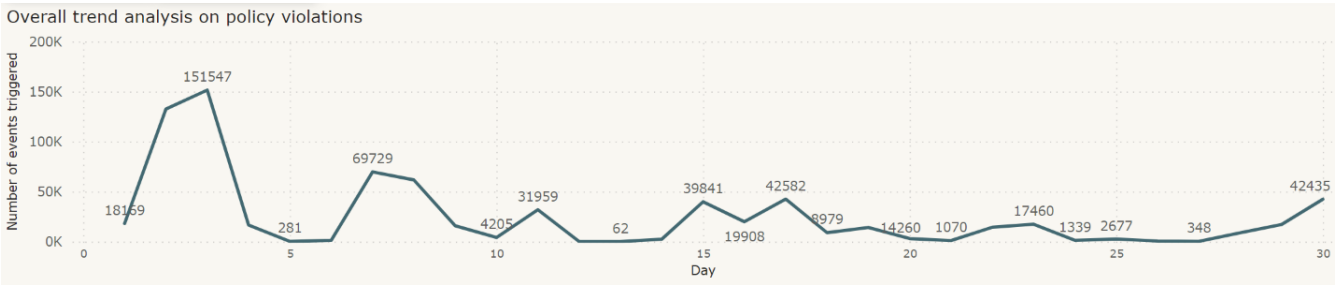


Figure 6.11: Overview Trend Analysis on policy violations (Drill Down to days)

8. The country violating most of the policies of NXP are China, India, France, and the USA.

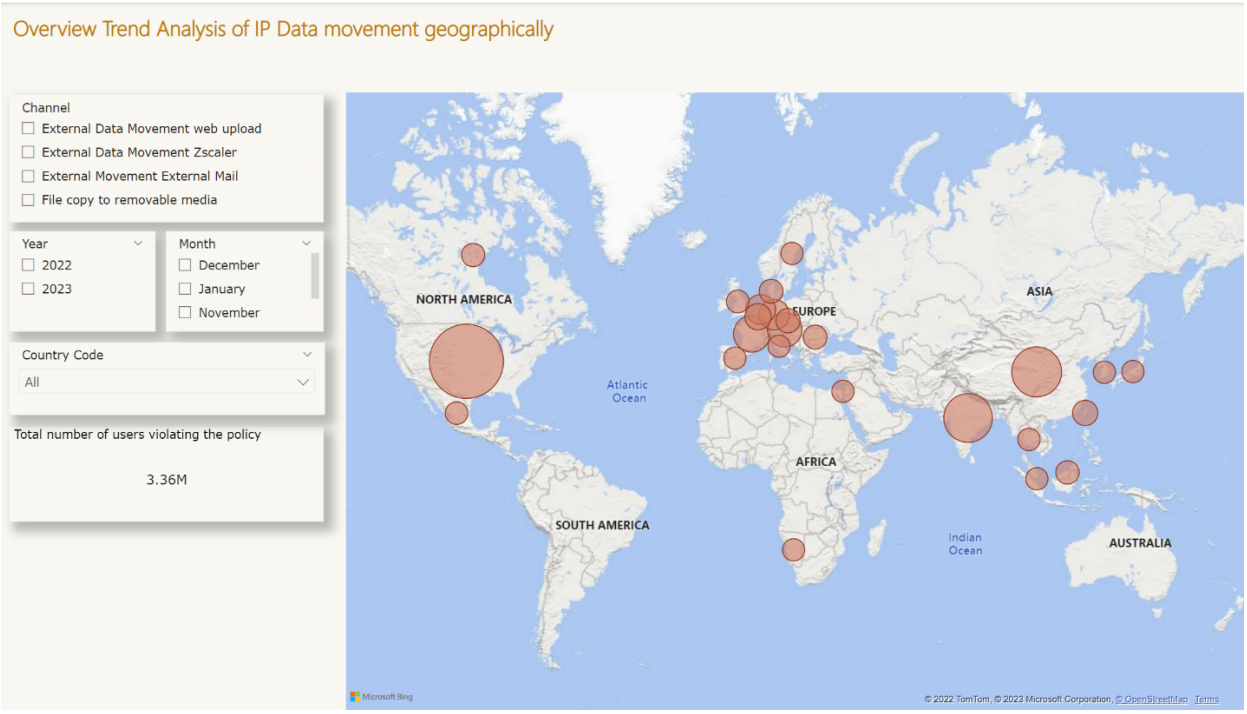


Figure 6.12: Overview trend analysis of IP Data movement geographically

6.2 Risk Matrix

In this section, we will propose a risk matrix on the basis of the models developed in the above section and the insights gained from each of the models. A risk matrix is used in risk analysis to define the risk level by balancing the categories of possibility or likelihood and consequence severity. This is a simple mechanism for increasing risk visibility and assisting management decision-making. This matrix will help in depicting the areas of risk within an organization's digital ecosystem or vendor network graphically. It will also assist in defining and categorizing various risks that an organization faces based on the significance and the impact of the risk related to it.

All of the aspects and insights gained in Section 6.1 are used to create the risk matrix. Furthermore, the risk matrix will assist NXP's high-level organizational element, considering Business Units, in determining which unit has the majority of the IP data movement. Table 6.3 lists all of the organizational Business Units that will be further mapped with the matrix.

Sr. No.	Business Unit
A	Advance Analog
B	Automotive Processing
C	Connectivity and Security
D	CTO
E	Edge Processing
F	MPU/MCU
G	RPF
H	Radio Power

Table 6.3: Business Units within NXP

The goal is to categorize the risks associated with each model in order to develop appropriate prevention measures and to create an overall risk matrix considering all the models implemented in Section 6.1.2. A control cycle must be established through continuous control and renewed risk assessment. The reports on each model's risky behavior were examined using the analytical method, and qualitative and quantitative insight analysis was obtained in order to catalog the likelihood of IP theft based on these parameters.

Two parameters are critical in assessing risks: the probability of a certain risk causing damage and the expected extent of damage when the risk occurs. Because these two parameters are interdependent, it is possible to calculate the risks. After evaluation, the following result is obtained, which can be presented in the form of a risk matrix, as shown in Figure 6.13.

Likelihood of IP Theft	Almost certain					
	Likely				1 - E,F	
	Possible		4	2,3,6 - E,G, C		
	Unlikely		5,7			
	Rare					
		Insignificant	Minor	Moderate	Major	Severe
		Impact				

Figure 6.13: Risk Matrix (Business Unit)

1 - (E,F): It is clear that many events are triggered for Model 1 - 'IP File copied to removable media,' indicating that removable media is widely used in the organization for transferring IP data. And the likelihood of IP theft is also likely to happen since a large amount of data is exfiltrated. There can be an unlikely scenario that can be maybe further unintentional where the user might happen to lose the device or it can fall into the hands of the wrong person. Therefore, proper actions need to be taken against such violations. Most of the violations are from the Business Unit - Edge processing and MPU/MCU with respect to the users violating the policies and being involved in malicious activities.

2,3,6 - (E, G, C): The next popular data exfiltration method used is IP data movement to external mail domains and web uploads, highly putting IP data at risk. There may be instances where data is sent inadvertently to the wrong email id or web, but still, in such cases, the data is also exfiltrated. This needs to be investigated and proper decisions should be made against the users sending IP data to unauthenticated external mail domains and the web.

4 - (E): Next the exfiltration method used is IP data movement to external cloud storage drives but from the analysis the risk is minor with this model indicating that a very less number of entities transfer IP data to cloud storage. Even if the events right now are less due to less visibility on the monitoring solution, one of the limitations is that while people can use cloud applications on the web in general, detection on FBA is currently again very limited. One solution for this model is that once web detection is improved, we may see more file exfiltrations via the web, as cloud solutions are frequently used on web browsers, which NXP should consider as a future task.

5,7: In comparison to other models, the risk of IP theft is minor due to rarer occurrences and relatively lower amounts of events triggered. However, proper actions need to be taken with respect to the way of working or malicious intent insiders can possess.

6.3 Summary and Conclusion

In this chapter, we put our research study's goal into action. First, business reports on the data extraction methods used in the organization are created, followed by an overall trend analysis of all the models. The use of removable media to copy IP files, external mail domains, and web uploads is the most popular data exfiltration method used in NXP, according to one of the main takeaways from this section. Proper policies and decisions must be implemented in response to these malicious activities, as well as warnings to users involved in such activities. Policy training can be given to such users to make them aware of the issues that may arise as a result of such activities. Finally, we proposed a risk matrix based on the designed models. This matrix will assist the business in understanding the risk levels that exist in the organization and will also assist them in making better decisions.

Chapter 7

EVALUATION

In this chapter, we describe our applied evaluation method, present our proposed evaluation plan, and discuss our evaluated limitations and results identified from the expert evaluation interviews. An expert panel is formed and the entire research study is presented to them to accomplish this. Following a walkthrough of the respective models, the experts are asked to complete a questionnaire. The evaluation opinions for each set can be found in Appendix C.

7.1 Evaluation Plan

The research study was carried out in a practical setting by the author within the NXP-IRM team, where the proposed prototype is implemented. *Semi-structured interviews were conducted by the author to evaluate the usefulness, feasibility, and effectiveness of the suggested proposed solution* which is accomplished by qualitatively assessing what an expert thinks of the model. The entire proposed solution from implementing models and gaining insights, as well as the proposed analytics model, data pipeline architecture, and defined key risk indicators have been evaluated. The methodology used is 'Technical action research' (TAR), which is a method of validating the artifacts in the study area. Semi-structured interviews are conducted through an 'expert opinion method' for evaluation. The model has been adjusted as a result of the feedback received during the evaluation phase. The outcomes of the evaluation interview are collected and documented in this chapter.

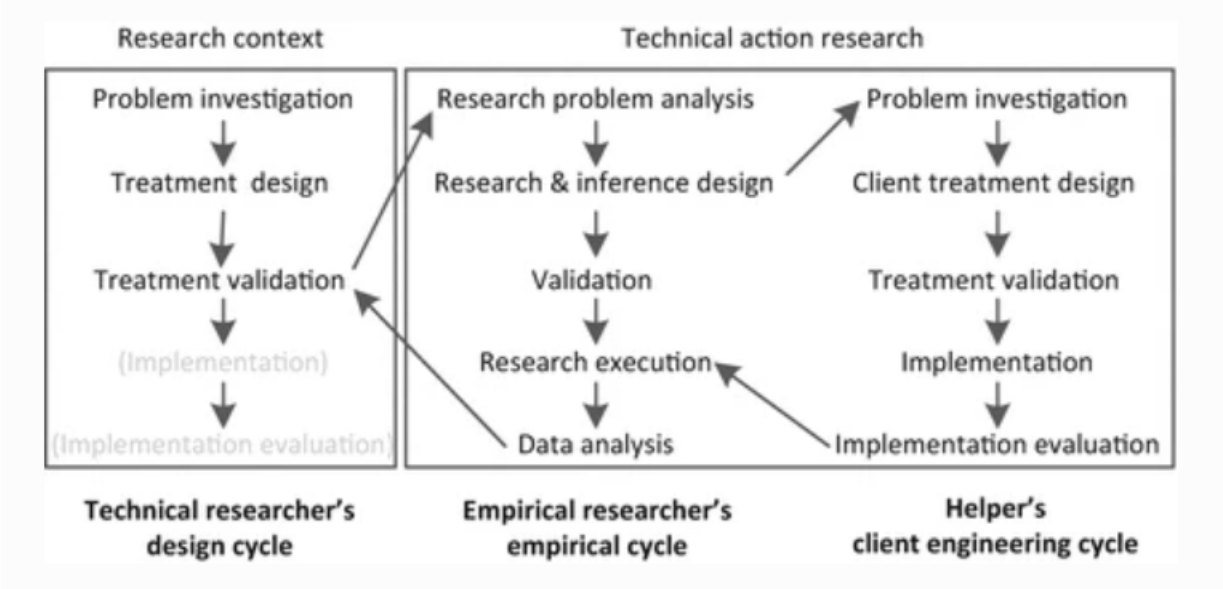


Figure 7.1: The three-level structure of TAR

The main stakeholders are the NXP Cybersecurity IRM team and the business people of NXP who will handle the solution in case of future work and the NXP as an organization will benefit from this study. The evaluation plan includes the following:

- Evaluate the design of the proposed solution and validate the models implemented in the dashboards for the NXP case with the experts in the field of interest.

Goal-oriented, process-oriented, and outcome/effectiveness evaluations were carried out as part of this thesis, and questions were formed accordingly to determine whether the goal was met, as well as to evaluate the effectiveness and impact of the proposed solution. The goal-oriented questions will aid in determining whether the defined goals and objectives were successfully met. The process-oriented questions will assess the proposed solution's strengths and weaknesses. Outcome/effectiveness evaluation questions examine broader impacts and frequently investigate what best outcome was delivered as a result of the project. These questions seek feedback from stakeholders in order to improve and critically revise the model proposed in previous chapters. The semi-structured interview for evaluating the proposed solution was conducted by the author in a hybrid manner on November 23, 2022. Following the expert interview, a questionnaire was distributed alongside. The evaluation questions were as follows:

1. How well did the problem that the project intends to solve achieved?
2. What might be the proposed solution's strengths, weaknesses, opportunities, and threats?

3. Do you think the dashboards that have been implemented are simple to understand and use?
4. Are there, in your view any limitations or ambiguities in the dashboard that you would like to improve?
5. How beneficial this solution will be? Do you think it will benefit the cybersecurity team and the organization?
6. Do you think the analytical solution can be applied in another context like HR, legal or other organizations?

7.2 Evaluation Interviews

The evaluation interviews took place in the NXP office as a live face-to-face session with the NXP - IRM cybersecurity team with 2 experts with an overall experience of more than 10 years in cybersecurity and insider risk management. We went through the whole process of our research study with the experts that were made in NXP for the last 6 months with the interviewees and evaluated the following steps from theoretical research to actual cases. The designed analytics capability was presented for each expert. Furthermore, the implementation of all the models and dashboards was explained. Afterward, a semi-structured interview session took place according to the proposed interview protocol. The questions were divided into open and closed questions.

(a) Open Questions: These types of questions were set up to know the overall impression about proposed solutions. Open-ended questions provide participants with a question prompt as well as a space in which to create their responses.

(b) Closed questions: These types of questions were designed to assess the utility of dashboard models with a limited set of possible answers.

7.2.1 Profile of the interviewees

Two experts involved during the design and development phase agreed to take part in the validation and evaluation:

(a) Expert A - Expert A is well-versed in the aspects of cybersecurity and analytics. He is a fantastic leader, a changemaker and has fulfilled multiple

different roles with an overall experience of 32 years. He has experience in areas such as enterprise solution implementation and business intelligence platform implementation. Later, he moved into cybersecurity, which is known as Risk, Security, Compliance, and Quality within IT. He is currently a Cyber Risk Manager at NXP, where he is in charge of the Insider Risk Management program, with the main goal of determining how we can create visibility inside the risks. Yeah. And with a focus on intellectual property protection. With the help of this solution, the cybersecurity insider risk management team will have a good foundation for future developments and can gain better insights and present them to the business about insider risk trends within the organization.

(b) Expert B - Expert B has extensive experience in cybersecurity. He has a strong background in intellectual property (IP) protection and is currently employed as an IP protection analyst within the IRM team. His responsibilities include developing the incident response process, coordinating with the work councils, setting up the user behavior monitoring solution, and developing models and policies for insider risks. He is knowledgeable about the Forcepoint Behaviour Monitoring solution and its limitations. This solution will assist them in filling a portion of the gap they are experiencing with FBA.

7.3 Evaluation Results

This section summarizes the proposed solution in the form of a criteria analysis based on feedback from interviewees. The rating was based on the results of the interviews.

(a) Design effectiveness: The design's effectiveness was rated as an important factor. The project's goal was achieved based on the interview results, but in order to achieve the future goal of making a full-fledged analytical solution, in that case, the goal is partially achieved. There is also a need for development with regard to NXP's internal resources. **Medium**

(b) Decision-making agility: The current case's agility implies that there is room for incremental change. If this process is repeated every quarter or year, better insights can be gained, as well as there is a possibility of switching to a new process in the future, which is also beneficial to the company. **Medium**

(c) Feasibility: One of the most important criteria was the feasibility criteria. The dashboards and risk matrix have created good feasibility within cybersecurity. With the ability to continue using the dashboards to create scenarios for how insider risk will evolve in the future within the monitoring that has been enabled with the help of this solution. Furthermore, the foundation for future dashboard expansion makes the solution feasible. **Medium**

(d) Understandability: One of the evaluation criteria was the understandability of the solution's design and development. Biweekly meetings within the team to understand and implement their changing requirements aided them also in understanding the process. **High**

(e) Usability: The usability of the implemented models and dashboards usability was one of the evaluation factors. The outcome was to gain value from dashboards, and the activity of using them in the team must be deployed. In the future, with the possibility that the dashboards will also be useful in understanding the quarterly or yearly trend analysis reports and then can inform management about how risk evolved within their organization. **High**

(f) Sustainability: Sustainability is the criterion for determining whether the benefits of the proposed solution are likely to continue. Because the IRM project is still in its early stages, it is still in the developing and improvement stage in terms of NXP's internal resources. Furthermore, the models are subject to change in the future if NXP's detection methods change. Additionally, it opens up new avenues for future development. As a result, this solution provides a solid foundation in that regard. **Medium**

This feedback and the successful implementation of the method in practice enabled the project researchers to present initial "proof-of-concept" level validation of the proposed solution. As the final step in the evaluation process, the designed and developed artifact was critically reviewed and summarized using the SWOT analysis method. Figure 7.2, shows the results of the SWOT analysis:

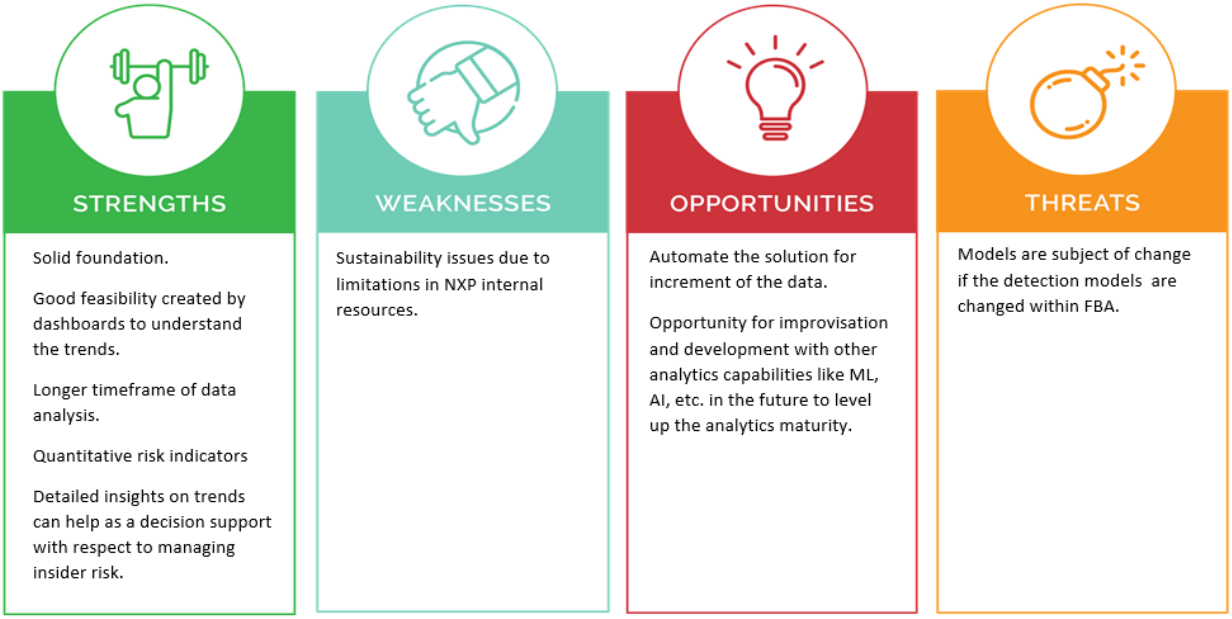


Figure 7.2: SWOT Analysis

7.4 Evaluation limitations and Threats to Validity

This section considers the potential limitations (Peffers, 2007) of our initial evaluation study [53]. First, we recognize a risk because our study only includes two practitioners. This jeopardizes the finding's generalizability, and we acknowledge that including more stakeholders in the evaluation would have been far more beneficial. However, the solution was presented in the Tech forum of the NXP cybersecurity team of 60 people which was also beneficial to evaluate our proposed solution. The feedback from the Tech forum meeting was also considered in our evaluation outcomes. Our expert participants, on the other hand, were chosen for their deep expertise as well as some commonalities, specifically:

(a) profound knowledge of the organization.

(b) expert knowledge of insider risks management and decision-making process.

To summarize, our initial evaluation of the proposed artifact provided convincing evidence that the artifact is useful for the NXP-IRM cybersecurity team and business people. The cybersecurity team sees it as useful for their company's context as decision support for managing insider risks, and they intend to use it in the future for further development.

This artifact is regarded as a strong foundation and a starting point for their journey into cybersecurity analytics. Conforming to the design science concept of iterative improvement of the proposed artifact based on repeated use in follow-up contexts within the same organization [63], the company remains optimistic about the artifact's gradual refinement and potential extensions as new information and lessons learned from its use become available.

Second, a common threat to qualitative evaluations of design research artifacts, in general, is the researcher's personal bias. However, we believe that this threat is minimal in our particular research context because the researcher is also the practitioner and is aware of the end-to-end proposed solution. Furthermore, the author demonstrated her proposed solution design and implementation to the team of five practitioners to ensure that she captured all of the information and correctly traced it back to her conclusions. This step, however, had no effect on the author's findings, reinforcing the notion that the personal researcher's bias is kept to a minimum.

Finally, the threat to validity is that the proposed solution must be deployed and used by the team in order to understand its effectiveness and gain value from the research study. Furthermore, detailed insights for annual and quarterly trend analysis will take approx 3-4 years of the dataset for proper analysis, which is currently not possible due to time constraints in the research study. An action plan for increasing data on a monthly or weekly basis must be considered for proper validation of the proposed solution.

7.5 Summary and Conclusion

In this chapter, a semi-structured interview with stakeholders (NXP-IRM experts) was used to evaluate the proposed solution. The implemented models have been adjusted and corrected in response to the feedback received, and the final version of the proposed solution and results is presented. According to the experts, the artifact has been proven to be presented in a clear and logical manner for both technical and business audiences. Our analysis of the artifact provided convincing evidence that it is useful for the NXP-IRM cybersecurity team and business people. The experts see it as useful for their company's context as decision support for managing insider risks, and they intend to use it for further development in the future. This artifact is considered a solid foundation and a starting point for their journey into cybersecurity analytics. Furthermore, the evaluation interview produced very positive results which helped to validate the quality and correctness of the proposed artifact.

Chapter 8

CONCLUSION AND DISCUSSIONS

This thesis work was done in collaboration with NXP Semiconductors N.V.'s IT cyber security department. This chapter will present all of the major findings from the research by answering the research questions raised in Section 1.4. Then, we will discuss the ethical considerations of the solution. This is followed by a discussion of how this study has influenced science and practice. Finally, the study's limitations and recommendations will be discussed.

8.1 Summary and Conclusion

8.1.1 Research Motivation

This research was initiated after observing that, organizations face a variety of consequences when it comes to insider risks. Insider threats and risks are especially damaging and difficult to manage because they are perpetrated by people who already have access to sensitive information. Because the insiders know where the data is and how to get it, insiders have a competitive advantage over external attackers. Security and risk management leaders must understand and address insider risk to protect the enterprise perimeter. As a result, to address these issues, strong tactics, techniques, and procedures should be used. This new wave of change prompted even more advancements in insider cyber security procedures, such as appropriate risk analysis, mitigation methods, business processes, new policies, and employee education.

8.1.2 Answers to Research Questions

The study's high-level goal was to design and implement data analytic capabilities as decision support in cyber security to manage insider IP risks. Therefore the main research objective set out for this thesis was: **To deliver a solution using an analytics-driven decision support approach that can be used in combination with current user behavioral monitoring systems in cybersecurity to assist organizations in making better decisions to manage and minimize the risk from insiders.** To fulfill the goal, we choose Design Science Research Methodology (DSRM) defined by Peffers et al. (2007) [53] for our thesis because it aligns with the overall objectives. This section will summarize our key findings and explain how the study addresses the research questions and objectives discussed in Section 1.4 of Chapter 1.

The main question addressed in this research: **How can organizations make better decisions by adopting an analytically driven strategy in cybersecurity to manage and minimize the risks from the insiders?**

This study successfully designed a solution in cybersecurity using an analytical approach, which can be used as decision support and will aid in making better decisions to manage insider risks. The main question is addressed in Chapters 4, 5, and 6. Chapter 4 described the primary process of designing the security analytics decision support solution. The analytics maturity model was developed to assist the NXP-IRM team in understanding the current state of analytics and what steps can be taken to take cybersecurity analytics to the next level. Moreover, data pipeline architecture and key risk indicators were developed to help organizations and businesses break down information silos, understand data movement, derive value, and achieve objectives from their data in the form of insights and analytics. The model's implementation and visualization through dashboards in Chapter 6 provide a comprehensive view of the organization's performance in terms of managing insider risks in real-time, allowing them to get a better picture of how the IP data is moving throughout the organization. This will assist businesses in making more informed decisions while managing insider risks. However, the results of a systematic literature review in Chapter 3 provide the necessary background information to begin developing the artifact. To achieve the research objective of this thesis, the **design question RQ4** was formulated which supports the development and evaluation of the final alignment model.

The systematic literature review as discussed in Chapter 3 was performed to answer **knowledge questions RQ1, RQ2, and RQ3**. The literature provided adequate knowledge to understand better how insiders and the risks and threats they posed are discussed. Insider problem has gained a lot of popularity, especially after the pandemic, and has a great impact on organizations, therefore proper steps should be taken by businesses to reorganize their security protection and IT strategies to assist reduce these occurrences. It also aided in understanding the several techniques used to tackle the security concerns posed by insiders.

Research Objective 1: To develop an effective solution and achieve the desired results, it is essential to understand and therefore conduct research on how insiders are defined, their causes, and the types of threats insiders are capable of posing. As a result, the first research question developed as follows:

RQ1: What are the security threats posed by insiders?

Insiders are more likely to break security rules and procedures because they are aware of them. The major difference between insiders and outsiders is that insiders are trustworthy. Insiders can include employees, but also contractors, consultants, workers, and outside business partners as a result of cross-company cooperation (e.g., outsourcing activities). Insider threat can be summed up as the possibility that dependable workers, temporary assistants, contractors, or consultants could gain access to sensitive information, exploit security flaws, and then break the organization's security policy whereas insider risks occur when any data exposure endangers an organization. Insider risk occurs when any data exposure endangers an organization's and its employees, customers, or partners' well-being. Insider risks focus on data first and foremost. Classifications of insider threats show that they can be categorized as sabotage, fraud, IP theft, espionage, social engineering, unintentional insider threat incident, or insider national security. These goals can range from the release of information (e.g. profit or theft) to alteration (e.g. fraud) and interference or destruction of data (e.g. sabotage).

Research Objective 2: In recent years it has seen an increase in the risk that careless or malicious insiders can pose to crucial assets and data, and the loss of important assets can inhibit an organization's recovery. Also, organizations

can face difficulties to combat and manage insider threats. As a result, the second research question developed as follows:

RQ2: Which recent cyber risks to intellectual property have been posed by insiders in an organization?

Insider attacks could have a significant impact on a company. Organizations could indeed suffer instant losses of inherent worth as well as lost revenue in addition to the lost worth of the asset that was eliminated, disclosed, or destroyed. Because internal users have valid access to critical systems, they are undetectable to traditional security solutions such as firewalls and intrusion detection systems. Also, insider threats can be difficult to deal with and manage due to budgetary constraints, lack of employees, and inadequate tools. When attempting to defend against external threats, organizations must keep in mind that insider threats also can cause harm.

Research Objective 3: Organizations have become more proactive in preventing, detecting, and responding to potential insider threats. As a result, we need to learn more about the existing organizational decision-making solutions, as well as their limitations. As a result, the third research question developed as follows:

RQ3: What are the existing decision-making analytical solutions used in an organization to manage insider risks?

As a form of large-scale data analytics, multiple proposals have been made to use machine learning and anomaly detection methodologies to make automated decisions about which insiders are behaving strangely or maliciously. The different methods and the data sources used from the reviewed literature are briefly defined in Appendix A.2. By using machine learning and behavioral analytics to analyze users, devices, and things, UEBA systems can identify insider threats, malware, and sophisticated assaults. They offer investigative insights so analysts can immediately verify and neutralize threats prior to causing more harm, as well as the knowledge to detect odd activity in real-time.

Gap in literature and in practice

To summarize, RQ1, RQ2, and RQ3 aided in answering the knowledge questions. RQ1 seeks to clarify how insiders are defined, their causes, and the

types of threats they are capable of posing. RQ2 examines recent insider cyber attacks on intellectual property, as well as the consequences and challenges that organizations face in combating and managing insider risks. Finally, RQ3 examines the existing decision-making solutions that organizations use to manage and minimize insider problems followed by a discussion.

We discovered a gap from the systematic literature review and in practice that the UEBA monitoring solution has a lot to offer but also has challenges. It lacks the ability to generate insight into general organizational behaviors, is incapable of identifying long-term sophisticated low and slow attacks, and prioritizes only the most reliable and high-risk security alerts. As a result, it has a limited ability to detect malicious behavior at an early stage.

Therefore, the design question RQ4 developed in this research will help to provide support solution to the identified gap and also address the limitations specific to data exfiltration problems within User and entity behavior analytics solutions and help to improve the feasibility of the decision-making process within the organization of the IRM cybersecurity team in NXP, for further directions and approaches that can be used to aid decision-makers.

Research Objective 4 (RO4): To enable a decision support system that can be used in combination with existing user behavioral monitoring systems in cybersecurity to help organizations make better decisions while managing and minimizing insider risk. Also, the research study should be evaluated in order to understand the added value and the types of cybersecurity decisions and strategies that can be made after analyzing the proposed solution. As a result, the following design questions and their sub-questions developed as follows:

RQ4: How to design and implement an analytic-driven solution to effectively support cybersecurity decisions in order to manage insider risks ?

Based on the aspects covered and the gaps identified from RQ1, RQ2, and RQ3, the starting point of enabling analytics in cybersecurity was to understand the current analytics maturity level. The proposed analytics maturity model, which is discussed in Chapter 4, will assist the NXP-IRM team in understanding the current state of analytics and what steps can be taken to take cybersecurity to the next level. With the help of this research, the NXP-IRM cybersecurity team is stepping into the second stage of maturity and is aware of the value analytics can bring to cybersecurity. Therefore, with respect to the proposed maturity model, this research is the implementation of moving one

level higher i.e. **Level 2** in the analytics maturity. Further, we proposed a data pipeline architecture to help organizations and businesses break down information silos, understand data movement, and derive value from their data in the form of insights and analytics. The data analytics component is introduced with the existing monitoring solution (FBA) in the pipeline. Next, 10 key risk indicators were developed in order to achieve the business objectives which are all discussed in Chapter 4.

In Chapter 5, data is extracted, cleaned, and transformed for further data analysis. Dataset limitations were also identified. Finally, in Chapter 6, during the implementation phase, four models are implemented out of seven, with some challenges and limitations. The most popular data exfiltration method used in NXP is the usage of removable media to copy IP files, external mail domains, and web uploads. New policies, warnings to risky users, and policy training can be introduced to limit external IP data movement. Finally, Chapter 6 discusses detailed insights and a risk matrix that will aid in defining and categorizing various risks that an organization faces based on the significance and impact of the risk. It is assumed that the chosen approach will aid to support IRM cybersecurity decisions to manage and minimize insider risks. Furthermore, the findings of this thesis show that the practitioner, in this case, the researcher, can adhere to the majority of the procedure.

Finally, the following sub-questions were developed to evaluate the research study to understand the added value and the kinds of cybersecurity decisions and strategies that can be made after analyzing the proposed solution:

SQ4.1: How successfully does the proposed solution provide value to the organization's business in managing insider risks?

To answer this sub-question as discussed in Chapter 7, an expert opinion method was used to assess the usefulness, feasibility, and effectiveness of the newly proposed model and generate potential improvement options. The evaluation opinions of the experts can be found in Appendix C. While the expert opinions differed, their overall assessment demonstrated that the prototype sets up a good foundation for enabling analytics in cybersecurity and can be useful for gaining better insights and trends from the dashboards which will be useful and usable for better decision-making to manage insider risks. According to the expert's opinions, the prototype is compatible with the organizational and technical infrastructure.

SQ4.2: What kinds of cybersecurity decisions and strategies can be made after analyzing the results from the proposed solution to manage insiders who pose a risk to the organization?

To answer this sub-question, after analyzing the results from the proposed solution discussed in Chapter 6 the riskiest models are the IP data movement to removable media, external mail domains, and web uploads. In order to minimize and manage insider risk efficiently, the organization can adhere to the following:

(a) Policy management: A decision can be made to implement a new policy regarding the use of removable media because it poses the greatest risk of putting IP at risk. The risk matrix will aid in identifying risky organizational units, and from the overview business report, we see that India, China, and France are high-risk countries, so decisions can be made beginning with these countries. The risky users violating the policies can be warned about their behavior and proper training can be given to them to understand the issues that can be faced due to their activities.

(b) Technical analytics advancements: The proposed maturity model assists us in determining the current maturity level of NXP-IRM, which is progressing towards level 2 with the assistance of this thesis. As a result, one strategy should involve improving cybersecurity analytics maturity by enabling advanced analytics.

(c) Continuous developments in the NXP-IRM program: Insider risk management is still in its early stages at NXP and thus has limitations which we also identified as a part of this thesis. IRM developments should be continued, and new technology involvements should be considered for improving the management of insider risks in the organization.

However, the decisions and strategies that can be developed to mitigate insider threats are inextricably linked to the recommendations resulting from this thesis work and are being presented as such in Section 8.4.

To summarize, according to the research as an NXP-IRM cybersecurity team embarks on its analytical journey, it is critical to have the proper foundation in place in order to achieve success, as it is all about evolving for the better. This research study creates a good foundation for security analytics and the design of dashboards will help as decision support for making better decisions to manage insider risk. The architecture, frameworks, models, and dashboards

developed in this study can be used to conduct additional research in this field to get support in making decisions while managing insider risk.

8.2 Ethical considerations

Data analytics and user behavioral monitoring solutions have made risk management easier. It does, however, raise some concerns. Employees can be concerned about the privacy of their work life and data. However, monitoring is legally allowed as long as it is for legitimate business purposes. In this section, we will understand data handling in an ethical way concerning our solution. Some of the key aspects and ethical considerations are taken as a part of this solution and insider risk management are as follows:

(a) Transparency is maintained within the employees: A monitoring notice is sent to all employees to notify them that their activities are being monitored, specific to their use of electronic media, and the notice includes the purpose, process to implement the monitoring and the monitoring results can be reviewed after approval from the Ethics Committee, which aids in awareness and transparency towards all the employees.

However, the type of solution used, as well as the type of monitoring performed, should be in balance with the risk to be mitigated. The work councils in NXP represent the interests of the employees to ensure that the monitoring, as well as the related processes, are balanced with both the employer's and the employee's interests. As a result, before making any decisions, the work council is consulted to understand the risk and how it should be mitigated.

(b) The monitoring solution is legally limited in what it can and cannot do within NXP leading to inefficient and ineffective decision-making processes in the incident response process. Monitoring is done primarily at the data and process levels. The IP files and their file types are stored in separate file types, allowing NXP to concentrate on the neutral file level rather than the highly intrusive personal level. And, with the help of the monitoring solution, the movement of these files is monitored, and their movement is then identified by which specific user the data is moved. And when we say that the monitoring solution is limited, our solution enters the picture to assist more with the current monitoring solution and to provide better support while managing insider risks.

(c) The automated monitoring systems are unable to distinguish between personal and professional activities. If an employee does not want their personal activities to be monitored, they should avoid using the NXP eDevice and/or NXP eMedia for personal purposes.

(d) Training is mandatory for all employees with respect to insider policies, to make them aware of their actions.

To summarize, monitoring is legal as long as it is used for legitimate business purposes. Transparency needs to be maintained, mandatory policy training and monitoring carried out by the law and in collaboration with the ethical team and work councils, resulting in an ethical approach to monitoring insider risks.

8.3 Contribution Revisited

The research is relevant from the perspective of practitioners and researchers. These contributions aim to improve the current knowledge gap in literature and practice. This study's contributions can be summarized as follows:

8.3.1 Scientific Contribution

This thesis makes three contributions of scientific relevance:

(a) An in-depth systematic literature review that presents a unique view of the current state of the art about insider risk mitigation approaches has been discussed. (Chapter 2)

(b) It contributes to knowledge by proposing a solution that includes using elements from the existing monitoring solutions and combining them in a way that evaded so far the attention of scholars. This solution was evaluated by using experts and business people opinion's regarding its usefulness and usability. (Chapters 4 and 7)

(c) An decision support analytical approach in cybersecurity to support decision-making for insider risk management. Proposed analytics maturity model, sustainable security data analytics architecture, and contribution of a case in a real-world organization and a demonstration of how the solution adds value to the business in the organization. This makes our proposed artifact more realistic. We further figure out a number of open issues and propose future research directions to motivate cybersecurity IP risk management research. (Chapters 5 and 6)

8.3.2 Practical Contribution

Insider threats and risks have gained prominence in recent years in the organization as a result of their high value in times of pandemic. Reducing business risk caused by someone with authorized access to systems and data can be a challenging issue for any firm. Therefore, the analytical-driven decision support design and approach proposed in this research will help the organization to:

(a) Proposed analytics maturity model will help the organization to understand the current state of analytics and what steps need to be taken to take analytics to a new level. (Chapter 5)

(b) Proposed a sustainable data analytics architecture based on an extension to the IRM software architecture and developed KRI which will enable cybersecurity insider risk management to leverage in the current and future efforts in managing insider risks. (Chapter 5)

(c) The proposed models will help in assisting decision makers to make better decisions, strategies, and policies in an organization to prevent such risk from occurring. Have better visibility on IP movement from the organization, and trend analysis related to insider risk. Finally, the risk matrix will help in prioritizing the risk to make better decisions. (Chapter 6)

(d) Examine the added value of analytical-driven capabilities in cybersecurity. The model implementation is evaluated by using the 'Technical action research' methodology by using the expert opinion method (Chapter 7)

To summarize, we assert that this research study will be of interest to researchers and practitioners from at least three domains viz., software engineering, cybersecurity, and data and business analytic technologies. These individuals are expected to be interested in gaining a better understanding of the architectural knowledge used for designing security data analytic systems, examining the added business value, enabling an analytical-driven approach in cybersecurity as decision support for managing insider IP risks, as well as identifying areas that require additional exploration and experimentation.

8.4 Research limitations and Future research directions

Despite the fact that the study addressed the main research question and related research objectives, as well as contributed to the scientific and practitioner communities, it has several limitations. To begin, enabling data analytics is a continuous development process rather than a one-time project. Second, the data was insufficient to grasp the trend on a quarterly or annual basis. Data needs to be incrementally added and can be automated as part of future work to clearly understand the trends of insider risks within the organization. Furthermore, due to the involvement of only two experts, our evaluation exercise has some limitations. The architecture design and models were created specifically for the NXP case study. Due to issues with NXP's internal resources, dataset limitations were encountered which is discussed in Chapter 5, and as a result, a few models were not implemented at the final stage, indicating that NXP should continue to develop and improve its IRM and detection program.

This section contains recommendations for researchers and practitioners who are following this thesis. The directions are listed as follows:

(a) With the assistance of this research study, we determined that the current state of maturity within the NXP-IRM team is level 2. Continue development, and learning is needed to improve the analytics at level 2 and to consider how to advance to the next level of analytics capability.

(b) NXP-IRM is still in the early stages of assessing insider risks. For better visibility of insider risks, continuous development of the monitoring solution and new methodologies are required.

(c) The data extraction process in our research study is done manually, and therefore further improvements for the organization can be to enable automation to reduce manual interaction.

(d) The input from the organization's business experts can also be an option for adding value to the evaluation process of the implemented models.

(e) Another future research direction can be related to the other issues with the behavioral monitoring solution that we discovered in the literature and in practice including slow and low attacks, as well as baseline issues. These issues must also be addressed, which opens the door to new additional research.

(f) Lastly, with the help of this research opportunities, open in the research and development area for enabling advanced analytics capabilities such as machine learning and artificial intelligence for managing insider risks.

These recommendations can be considered by researchers and practitioners who want to improve or evaluate the process of this study.

Appendix A

A.1 Literature review data extraction

Appendix A.1, contains a final list of literature selected and there respective authors & keywords, after following the research methodologies discussed in Section 1.2.

ID	Title	Authors	Year	Keywords
1	Behavioral Based Insider Threat Detection Using Deep Learning	RIDA NASIR, MEHREEN AFZAL , RABIA LATIF , WASEEM IQBAL	2021	Insider threat, deep learning, machine learning, user behavior, information security
2	Developing Visualisations to Enhance an Insider Threat Product: A Case Study	Martin Graham, Robert Kukla, Oleksii Mandrychenko, Darren Hart	2021	Human-centered computing—Visualization—Visualization application domains—Visual Analytics; Security and privacy—Human and societal aspects of security and privacy—Usability in security and privacy
3	Real-time Threat Detection in UEBA using Unsupervised Learning Algorithms	Joyatee Datta, Rohini Dasgupta, Sayantan Dasgupta, Karmuru Rohit Reddy	2021	UEBA, Cybersecurity, Machine Learning, Insider threats, User Data
4	Insider Threat Detection Using An Unsupervised Learning Method: COPOD	Xiaoshuang Sun, Yu Wang, Zengkai Shi	2021	insider threat detection; tree structure analysis; unsupervised learning; anomaly detection
5	User Behaviour-based Insider Threat Detection in Critical Infrastructures	Malvika Singh, BM Mehtre, S Sangeetha	2021	Cyber Security, Insider Threat Detection (ITD), User Behavior Analysis, Machine Learning, Anomaly Detection, Bidirectional, Long Short Term Memory (bi-LSTM), Support Vector Machine (SVM)
6	Role of User and Entity Behavior Analytics in Detecting Insider Attacks	Salman Khaliq, Zain ul Abideen Tariq, Ammar Masood	2020	Security Information and Event Management, Machine Learning (ML), Artificial Intelligence (AI)
7	The Visual Analytics Approach for Analyzing Trajectories of Critical Infrastructure Employers	Evgenia Novikova, Igor Kotenko and Ivan Murenin	2020	visual analytics; data mining; moving entities; route patterns; anomaly detection; self-organizing Kohonen maps; Gantt chart-based visualization; heat maps
8	Malicious Insider Attack Detection in IoTs Using Data Analytics	AHMED YAR KHAN , RABIA LATIF , SEEMAB LATIF , SHAHZAIB TAHIR , GOHAR BATOOL , AND TANZILA SABA	2019	Insider attacks, artifcial intelligence, malicious threat
9	Insider report 2019	Holger Schulze	2019	NA
10	Insider Threat Detection Based on User Behavior Modeling and Anomaly Detection Algorithms	Junhong Kim, Minsik Park, Haedong Kim, Suhyoun Cho and Pilsung Kang	2019	insider threat detection; anomaly detection; machine learning; behavioral model; latent dirichlet allocation; e-mail network
11	User Behavior Profiling using Ensemble Approach for Insider Threat Detection	Malvika Singh, BM Mehtre, S Sangeetha	2019	Cyber Security, Insider Threat Detection, User Behavior Profiling, Machine Learning, Time Series Anomaly Detection, Multistate Long Short Term Memory, Convolution Neural Network.
12	A Framework for Data-Driven Physical Security and Insider Threat Detection	Vasileios Mavroeidis, Kamer Vishi, Audun Jøsang	2018	Physical Security, Physical Security Definition, Insider Threat, Security Ontology, Digital Forensics, Data Analytics, Anomaly Detection, Attack Pattern Reconstruction, Security Provenance
13	Anomaly-based Insider Threat Detection using Deep Autoencoders	Liu Liu, Olivier De Vel, Chao Chen, Jun Zhang, Yang Xiang	2018	Insider threats, data analytics, deep autoencoder, cyber security
14	Detecting and Preventing Cyber Insider Threats: A Survey	Liu Liu, Olivier De Vel, Qing-Long Han, Jun Zhang, Yang Xiang	2018	Insider threats, data analytics, machine learning, cyber security.
15	An Insider Threat Detection Method Based on User Behavior Analysis	Wei Jiang, Yuan Tian, Weixin Liu, and Wenmao Liu	2018	Insider threat ,User behavior, Machine learning
16	Combating Insider Threats by User Profiling from Activity Logging Data	Mohamed Dahmane, Samuel Foucher	2018	Cybersecurity, insider threats, clustering, Markov chain, anomaly detection
17	Detecting Masqueraders by Profiling User Behaviors	Haohui Peng, Wei Wang	2018	insider threat; masquerader detection; intrusion detection; user behavior; network traffic
18	Insider Threat Detection Using Characterizing User Behavior	Xuebin Wang, Qingfeng Tan, Jinqiao Shi, Shen Su§ and Meiqi Wang	2018	NA
19	TOWARDS A USER AND ROLE-BASED BEHAVIOR ANALYSIS METHOD FOR INSIDER THREAT DETECTION	Qiujian Lv, Yan Wang, Leiqi Wang, Dan Wang	2018	Insider threat; isolation forest; job role; user behavior
20	Automated Insider Threat Detection System Using User and Role-Based Profile Assessment	Philip A. Legg, Oliver Buckley,Michael Goldsmith, and Sadie Creese	2017	Anomaly detection, cyber security, insider threat
21	Modeling user communities for identifying security risks in an organization	Anirban Das, Min-Yi Shen, Jisheng Wang	2017	Louvain modularity; UEBA; User and Entity Behavior Analytics; Peer grouping
22	Reducing False Positives Of User-to-Entity First-Access Alerts for User Behavior Analytics	Baoming Tang, Qiaona/Joanna Hu, Derek Lin	2017	NA
23	Visualizing Insider Threats: An Effective Interface for Security Analytics	Bar Haim, Eitan Menahem, Yaron Wolfsthal, Christopher Meenan	2017	Insider threat; user behavior analytics; anomaly detection.
24	Human-Machine Decision Support Systems for Insider Threat Detection	Phil Legg	2017	NA
25	Predict Insider Threats Using Human Behaviors	JENNIFER U.MILLS, STEVEN M. F. STUBAN, JASON DEVER	2017	Human behaviors, insider threats, linear regression, prediction interval, and systems engineering
26	A Data-Driven Evaluation for Insider Threats	Yuqing Sun, Haoran Xu, Elisa Bertino, Chao Sun1	2016	Insider threat, Audit, Behavior analysis
27	Visualizing the Insider Threat: Challenges and tools for identifying malicious user activity	Philip A. Legg	2015	Insider threat, behavioural analysis, model visualization
28	Automated Insider Threat Detection System Using User and Role-Based Profile Assessment	Philip A. Legg, Oliver Buckley, Michael Goldsmith, and Sadie Creese	2015	Anomaly detection, cyber security, insider threat
29	Methods and Metrics for Evaluating Analytic Insider Threat Tools	Frank L. Greitzer, Thomas A. Ferryman	2013	insider threat; evaluation; validation; metrics; assessment
30	Intelligence Analyses and the Insider Threat	Eugene Santos, Hien Nguyen, Member, Fei Yu, Student Member, Keum Joo Kim, Associate Member, Deqing Li, John T. Wilkinson, Adam Olson, Jacob Russell, and Brittany Clark	2012	Cognitive styles, decision-making process, insider threat, intelligence analyses
31	Exploring Adversarial Properties of Insider Threat Detection	Duc C. Le, Nur Zincir-Heywood	2020	insider threat, anomaly detection, temporal information, data representation

A.2 List of studies used in our analysis to answer RQ3

Reviewed literature contains a wide range of methodologies for detecting and preventing insider threats. Appendix A.2, includes a list of studies that were used to answer research question 3 (RQ3). The various methods used, data sources and purpose of the study has been briefly discussed.

Reference	Method	Data Source	Purpose of Study
[12]	Machine Learning Algorithms	Host, Network and Contextual data	detection method that uses a group of deep autoencoders to accomplish anomaly detection from large set of audit user data
[32]	Machine Learning Algorithms	Audit log	proposes a user behavior analysis model that combines user behavior over time, comprehensively characterizes user attributes, and detects internal attacks.
[33]	Machine Learning anomaly detection algorithms	User log data	propose methods for detecting insider threats based on user behavior modeling and anomaly detection algorithms.
[10]	Graphic Analytics Approach	Employee Login Activities	based on user profiling and derived features, demonstrated the use of visual analytics to support the identification of insider threat activity
[40]	Machine Learning Algorithms	First-access alerts	propose a user-to-entity prediction score that learns user data using a recommender system
[34]	Unsupervised learning algorithms	aggregated user activity data daily or weekly	proposed an anomaly detection method for insider threat identification based on unsupervised machine learning (ML)
[35]	Unsupervised learning anomaly detection algorithms	user activity log data	used the tree structure approach to examine user activity, create feature sequences, and integrate the COPOD method to identify anomalous users and distinguish between feature sequences
[36]	Statistical learning algorithms	user actions such as keystrokes, mouse movements, and shell command sequences during GUI interaction	This study provides a unified similarity measure approach based on probability distribution function and suggests an efficient way to extract user behavior features for creating user behavior profiles
[37]	Prediction algorithms	System logins	focused on identifying using prediction algorithms potential insider threats based on human behaviors
[38]	Machine Learning Algorithms - SVM	user's daily activities (action sequences)	Using a binary class support vector machine (SVM), users are classified as either "normal" or "malicious"
[39]	Markov Chain Models	CMU-CERT dataset which gives events of malicious activities	proposed a dynamic model based on Markov Models to identify internal risks from the network data stream
[41]	Unsupervised Learning Algorithms	user information on clickstream	proposed machine learning UEBA approach using four different unsupervised algorithms
[43]	User behavior anaytical tool	user activity	proposed a tool UBA that continuously analyzes how users interact with their organizational IT networks and efficiently illustrates the security exposures that result
[44]	User and Entity Behavior Analytics tool	user activity	proposed UEBA approaches incorporating user and role-based identification, mapping of user and entity activity, user profiling approaches, and individual risk score computations
[45]	Louvain modularity with UEBA solution	LDAP data	proposed three new algorithms to group employees based on their activity to give meaningful descriptions
[46]	Graphic Analytics Approach	Artificially generated user data and user movement data	proposed a visual analytical approach for the exploratory examination of the routes taken by the employees based on the access control system's logs

A.3 Concept Centric Matrix showing which concepts from the gathered literature are discussing which aspects of the proposed research questions.

Appendix A.3, is a concept-centric matrix which contains the literature as well as the concepts which are answering the research questions proposed in Section 1.1.

The layout of the concept matrix is straightforward. In the leftmost column, the selected literature is listed. Each column's header depicts a concept derived from the proposed research questions. An 'x' is placed in the appropriate cell if a relevant concept is discussed in a specific reference. The main takeaway from this concept matrix is that it will help in keeping track of how many selected literature deal with multiple concepts in the overall theme under investigation, as well as how many literature deal with a specific concept.

ID	Title	Definitions of Insider, Insider Risk & Insider Threat	Insider Threat Classification	Impacts on organizations due to Insider attacks	Challenges Faced by Organization	Decision-making analytical approaches/methods used and its challenges
1	Behavioral Based Insider Threat Detection Using Deep Learning	x	x			x
2	Developing Visualisations to Enhance an Insider Threat Product: A Case Study			x		x
3	Real-time Threat Detection in UEBA using Unsupervised Learning Algorithms					x
4	Insider Threat Detection Using An Unsupervised Learning Method: COPOD			x		x
5	User Behaviour based Insider Threat Detection in Critical Infrastructures					x
6	Role of User and Entity Behavior Analytics in Detecting Insider Attacks			x	x	x
7	The Visual Analytics Approach for Analyzing Trajectories of Critical Infrastructure Employers					x
8	Malicious Insider Attack Detection in IoTs Using Data Analytics	x	x			x
9	Insider report 2019	x	x	x	x	
10	Insider Threat Detection Based on User Behavior Modeling and Anomaly Detection Algorithms		x			x
11	User Behavior Profiling using Ensemble Approach for Insider Threat Detection					x
12	A Framework for Data-Driven Physical Security and Insider Threat Detection	x	x			
13	Anomaly-based Insider Threat Detection using Deep					x
14	Detecting and Preventing Cyber Insider Threats: A Survey	x	x			x
15	An Insider Threat Detection Method Based on User Behavior Analysis		x	x		x
16	Combating Insider Threats by User Profiling from Activity Logging Data	x				x
17	Detecting Masqueraders by Profiling User Behaviors			x		x
18	Insider Threat Detection Using Characterizing User Behavior			x		x
19	TOWARDS A USER AND ROLE-BASED BEHAVIOR ANALYSIS METHOD FOR INSIDER THREAT DETECTION	x		x		x
20	Automated Insider Threat Detection System Using User and Role-Based Profile Assessment			x		
21	Modeling user communities for identifying security risks in an organization					x
22	Reducing False Positives Of User-to-Entity First-Access Alerts for User Behavior Analytics					x
23	Visualizing Insider Threats: An Effective Interface for Security Analytics					x
24	Human-Machine Decision Support Systems for Insider Threat	x		x		x
25	Predict Insider Threats Using Human Behaviors		x			x
26	A Data-Driven Evaluation for Insider Threats	x		x	x	
27	Visualizing the Insider Threat: Challenges and tools for identifying malicious user activity	x		x	x	x
29	Methods and Metrics for Evaluating Analytic Insider Threat Tools					x
30	Intelligence Analyses and the Insider Threat	x				
31	Exploring Adversarial Properties of Insider Threat Detection	x	x			x

Appendix B: Interview Agenda for Evaluation

The below agenda was used to ask questions in interviews. Since it was a semi-structured interview, the actual questions posed by respondents may vary depending on the flow of the conversation, but the procedure will be followed.

(a) Introduction
   i. The interviewer introduces themselves.
   ii. The research context and purpose of the interview session are discussed.
   iii. Consent for recording the interview

(b) Background and experience of the interviewee
   i. Background experience?
   ii. Role and brief about the position of the interviewee?

(c) Presented the design and implementation of the proposed solution
   i. Explained the maturity model and data architecture.
   ii. Explained the models implemented and gave insights.
   iii. Explained the challenges and limitations.

(d) Evaluation interview
   i. Questions asked during the interview were to determine the usefulness, feasibility, and effectiveness of the proposed solution.
   ii. The opinions gathered are highlighted and briefly transcribed below.
   iii. Closing the interview.

Appendix C: Evaluation Opinions

Expert A	Background	In my 32 years, I have served in a variety of roles ranging from business consultancy within the organization to setting up enterprise solutions such as what we see today within an NXP on procurement, sales and marketing, and project management. As a peak kernel implementations within Phillips display components for 12 companies, I also serve as project manager. We also implemented what was then known as sub-business intelligence as part of that implementation. So their business intelligence is in SAP's data analytics platform. And I created the platform in this manner. And I was also a business analytics consultant to those who used that platform. So that was until 2004 when I joined NXP or Philips Semiconductors and became service delivery manager. Later, I joined cybersecurity, or what was then known as the Risk, Security, Compliance, and Quality department within IT, where I was in charge of SOC. NXP began the IP protection program in 2020 with the primary goal of determining how we can create feasibility on insider risks. And, as you have reflected in your assignment, with the goal of protecting our intellectual property. And that means that in that effort, we identified not only what DIP we want to protect, but also what solutions, and capabilities we want to enable that monitoring and, ultimately, to protect the IP, yeah, from exfiltrating from NXP. With the goal, of course, of ensuring an NXP's competitive advantage in the semiconductor market. And before that, there was no insider risk management. Until two years ago, the emphasis was on avoiding insiders as potential threat actors.
	Q1	Of course, we must consider what you have defined in terms of the data analytics maturity model or cyber security data analytics. And I believe we are still not where we want to be. But I believe the first step in creating the feasibility about the trends, how insider risk evolves, is partially addressed, and this is due to the complexity of the assignment as well as the time constraints that we have on the assignment as such, and I believe your slide on the insider risk indicator. I believe you mentioned seven or eight models, and they reflect the actual scope as well as what we have accomplished because we have only been able to understand four of those risk indicators. But it does provide a solid foundation for moving forward.
	Q2	The visibility you have created with the dashboards is, of course, the strength. And then we hope to continue using those dashboards to create scenarios for how insider risk will evolve in the future within the monitoring that we currently have enabled. For me, this is also the foundation for expanding the dashboard to include additional threat scenarios that we want to monitor in the future. The weakness or risk is sustainability, which is not related to the solution, but rather to our internal resources to ensure that it lands within the cyber security or a team to continue on the development of the data analytics solution. Moving forward, automation for data amendment poses a threat or risk. Because we now have data from the previous three months. However, in the future, we should be able to regularly update the dashboards using a predefined process, preferably an automated process. I'm not sure if you'd call that a weakness or a threat, but I think they're all opportunities. But I also believe it is an opportunity to improve the solution.

	Q3	So, to be honest, that question has not yet arrived in our organization if you understand without what we have been developing, learning is that we are regularly using it. So we still need to deploy an activity and begin using it to determine the value of the dashboard you created. Moving forward, I hope we can also use this to wrap up the monthly or quarterly reports. We met with the management team to discuss how risk has evolved within their organization. And in any development, it is important to start small. Yeah. And continue to grow. Yes, that applies here as well.
	Q4	The interpretation that we had to apply, was to make a certain risk identifiable. I like, for example, some of the decisions we had to make regarding file counts. And the dataset limitations that we faced. And there are also limitations resulting from the manner in which FBA provides the data or has ingested the data itself. So that was not what I expected, but it was something we had to come to terms with during the assignment and creation of the dashboards.
	Q5	I believe the solution will reflect where we see that decisions have changed, as well as the risk posture where we see a decrease in the strength of certain risks. Yeah, I'm not sure if it's a legitimate way to support decisions. If you understand what I mean by that. But, um, remember how they said in the meeting with the business stakeholders that if you want to make decisions, we need information about the user. And the reason is that we already own that moment saw that there isn't really a common trend if you look at the total population of 14,000 employees. Which you would be able to derive as a root cause on certain risks other than, say, USB, where we saw a lot of issues , on laptop renewals a nd also in relation to the use of USB ports like printed circuit boards or test boards. As part of the incident response, it was reported as a false positive. As part of the incident response, it was reported as a false positive. Aside from the dashboards themselves. So meaning is that we have improved our way of working to limit certain risks. It's more a result of what we've detected as false positives as part of our incident response process, but the improvements we've implemented are reflected back in the trends we see on the dashboards. Quantitative risk indicators always aid in understanding the posture and, as a result, drive decision-making processes.
	Q6	Proven the ability to support multiple types of data analytics objectives, including not only cyber security but, in this case, HR and legal. Yeah, and to the extent that solution will be scalable to those areas, and I think the if so if you look at the tools you have used and kind of dashboards which we have created, I would say no because those are too specific for inside risk management. Also the architecture, the maturity model specific to NXP-IRM I think that's very valuable to any decision.

Expert B	Background	I work as an IP protection analyst at an NXP. I started as an intern when we were just getting started with deploying the software to end users. So I was in charge of developing the incident response process, as well as coordinating with the work councils and attempting to establish the overall monitoring and policy framework. I'm quite knowledgeable about how the FBA solution works, as well as our gaps in terms of what we can see and what we do with the data. I'm primarily responsible for the incident response process, but we also want to conduct some trend analysis. So I believe there is a gap, but I am aware of it, which is why your thesis is assisting.
	Q1	In terms of creating trend analysis, it has done an excellent job, at least for the models that work properly. Some of our models aren't working properly. In that regard, I believe it has truly satisfied our desires. Yeah, in the sense that what was possible with the time you had and the, uh, you're learning the solution. I believe It's done very well in terms of trend analysis. For example, with the external mail, I was also in charge of creating the trend analysis before you started, and it was a very manual process that was also not very difficult to create trend analysis and it was not automated. So I tried to do the same thing as you, but the dashboard you've created can really help us to show on the business unit level, which I was also unable to do. How many emails are sent into which domain on a profit center level? So I think that really helps us achieve, and it will also help us connect back to the business line and tell them what kind of trend we see, as well as what we want to do with this trend, whether we want to mitigate it or not. We can also report back afterwards. So I trust your dashboards. Yes, we have achieved the fact that we now have more visibility
	Q2	So I think that the strength is really that you've identified. With our current state is uh, but you're also identified that we need some more uh trend analysis. And I think your solution really shows us that we can have a longer time frame of data that we can use to present to the business because currently our time frame was limited to 90 days reporting back on incidents. But your solution allows us to also have a longer time frame within the GDPR agreements with the Works Council. So I think that's a very good solution. As for the weaknesses, it's mainly partly also due to that NXP side because we have the data set limitations that we had, yeah. Yeah. As regards the model, well, recently I also asked you to create a model for the also filtering out of the country. There are some people in NXP working on looking into a USB and they were our first looking at East Asia. And if we can filter out for East Asia on that specific dashboard, looking at which profit centers copying to USB that would be, that would be, it's a weakness, but in the same sense it's an opportunity, right? But yeah, it's an opportunity to make it better. I would like to see that combined in a sense because the country obviously was very high level. It shows us the USB usage in one country. But then again, when we're investigating a specific country and also looking for the business units and that's something if we now look at the business unit level, we do that on a generic level without all the users. We take one business unit, but apparently some, well apparently, some teams need maybe to have a country-level few of which business units are copying to USB. I I think the threats are that that um in general to the solution is of course that we are also still, adjusting our own models, so it may be subject to change and then we need to see how well it's gonna perform. For the USB and mail, I don't see a problem. But wherever you may be going to move to another, if the Web doesn't work maybe we need to have another solution. So that could be that. But I think that's in general threats that if our detection changes the models may also change.
	Q3	No, I think from the get-go to the dashboards and the solution was already quite good and we had a couple of sessions to discuss it and then you made some adjustments. So I think they are now for our team able to understand and use it because we have some, we had so many or so many. We had several feedback moments. I think for us we really let you know what we want to see in a dashboard and you really implemented them well. So from our side, I would say it's, it's very well. I don't know how it's for the business because we have represented the business. And you also have to see if this solution now for the cyber security team. I think that's more of it's more for us and then we can use it for specific reports out and for the business. But I guess in general for us it's, it's a very well dashboard, it's easy to understand the use, OK and would you like to improve some things to it.

	Q4	It's the same opportunity I just mentioned. unless we're talking about the dashboard's interface or the dashboard's core. So, what kind of information do you see? But then I'd add a country filter, and overall, I think the dashboard is fine and the solution is very usable.
	Q5	We've already received several requests. As I previously stated, we already have a team looking into the use of USB within NXP. If I look at the mail export, I believe we can provide them with some insights into this solution as well as for us. As I previously stated, I did that manually back then, but it was a very small group, so I was able to do so. If I look at this dashboard right now, I can see how we can use it to report to the business and. For our team to see the trend, I believe that will be used in the future.
	Q6	In some ways, we as cyber security will use this in presentations to HR, legal, and other organizations. So we're already talking about using this dashboard to present to one of the managers who work on USB usage. However, in general, we can use this to generate insights for the Works Council and legal to demonstrate the type of behavior we observe. Yes, because there has been a lot of discussion about why we do this, but we can also use your dashboard to see broader trend analysis. Uh, level to demonstrate the business and also works for The legal council. Yeah, yeah. We do a lot of incident reports, so we know what kind of numbers we're talking about. but those are just incidents, and we can use this dashboard to inform them that we are still missing a lot of things because we only report on high-risk incidents. But if you look at your data, you can see how much data is flowing through the network. Also, how much data is being transferred to USB, which we do not always act on. As a result, the significance of the problem is also revealed. And I apologize if I'm harping on the USB, but that's now a real issue within NXP, and I believe your solution provides insight into the significance of that problem within NXP. So your analytics dashboard has already provided us with a very valuable insight. Yeah, that's the significance of the USB problem in that sense.

Figure C.1: Evaluation Opinions

Appendix D

Appendix D: Model Implementation - Models 3 and 6

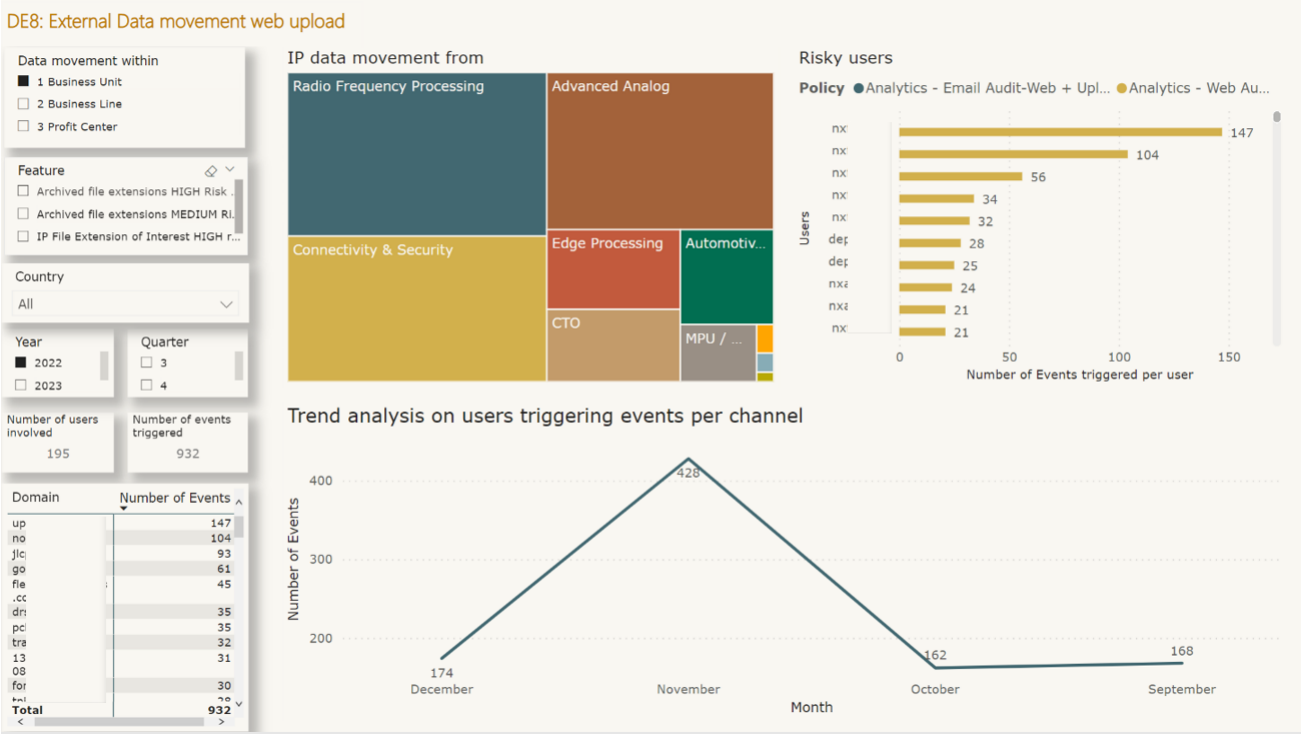


Figure D.1: External IP Data Movement to web upload business report (Model 3)

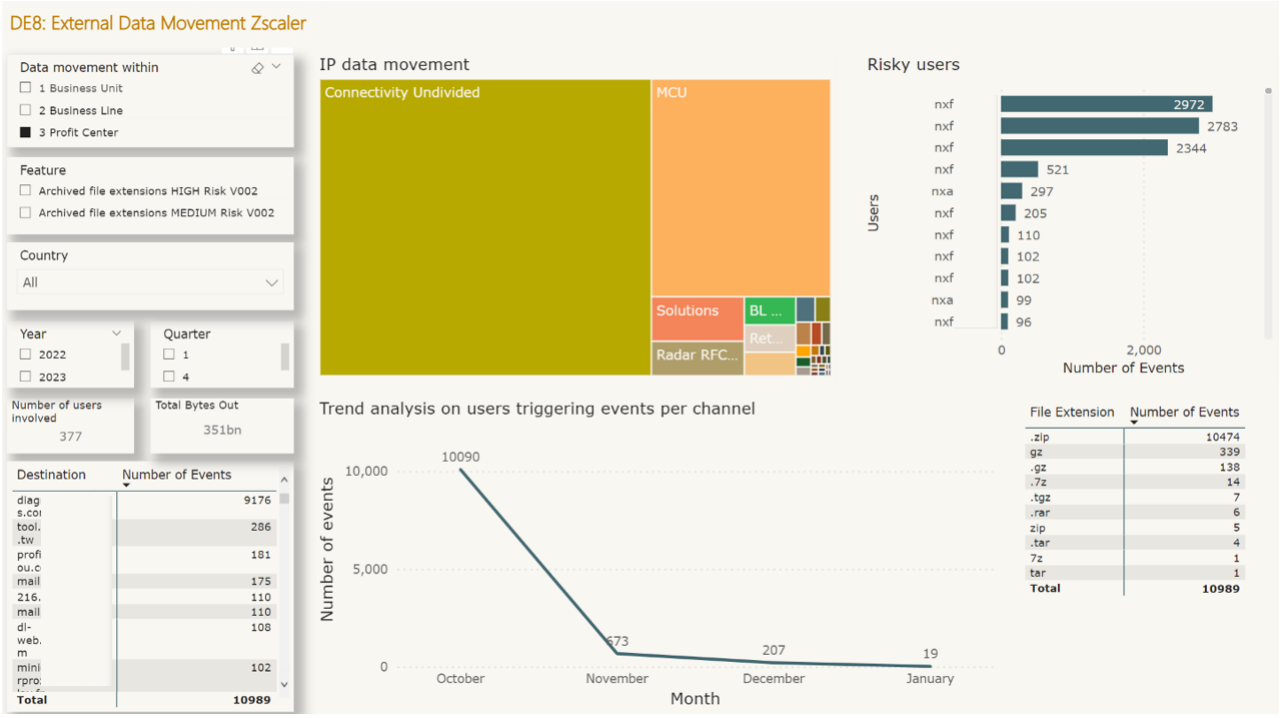


Figure D.2: External IP Data Movement to Zscaler business report (Model 6)

Bibliography

[1] D. Morrissey, “Cybersecurity budgets explained: How much do companies spend on cybersecurity?” 2021. [Online]. Available: https://cybersecurity.att.com/blogs/security-essentials/how-to-justify-your-cybersecurity-budget

[2] M. Baker, “Gartner survey reveals 82% of company leaders plan to allow employees to work remotely some of the time,” Jul 2020. [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2020-07-14-gartner-survey-reveals-82-percent-of-company-leaders-plan-to-allow-employees-to-work-remotely-some-of-the-time

[3] I. Security, “Ibm security work from home study - pr newswire,” Jun 2020. [Online]. Available: https://filecache.mediaroom.com/mr5mr_ibmnews/186506/IBM_Security_Work_From_Home_Study.pdf

[4] ——, “Cost of a data breach report 2021,” 2022. [Online]. Available: https://www.ibm.com/nl-en/security/data-breach

[5] C. company, “What is siem: Ueba,” 2018. [Online]. Available: https://www.custodian.nl/our-services/what-is-siem-ueba/

[6] F. A. Bukhsh, Z. A. Bukhsh, and M. Daneva, “A systematic literature review on requirement prioritization techniques and their empirical evaluation,” May 2022. [Online]. Available: https://research.utwente.nl/en/publications/a-systematic-literature-review-on-requirement-prioritization-tech

[7] “Defining insider threats.” [Online]. Available: https://www.cisa.gov/defining-insider-threats

[8] Q. Lv, Y. Wang, L. Wang, and D. Wang, “Towards a user and role-based behavior analysis method for insider threat detection,” *2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC)*, Aug 2018.

[9] M. Bishop and C. Gates, “Defining the insider threat,” *Proceedings of the 4th annual workshop on Cyber security and informaiton intelligence re-*

search developing strategies to meet the cyber security and information intelligence challenges ahead - CSIIRW '08, 2012.

[10] P. A. Legg, "Visualizing the insider threat: Challenges and tools for identifying malicious user activity," *2015 IEEE Symposium on Visualization for Cyber Security (VizSec)*, 2015.

[11] I. Homoliak, F. Toffalini, J. Guarnizo, Y. Elovici, and M. Ochoa, "Insight into insiders and it," *ACM Computing Surveys*, vol. 52, no. 2, p. 1–40, 2020.

[12] L. Liu, O. De Vel, Q.-L. Han, J. Zhang, and Y. Xiang, "Detecting and preventing cyber insider threats: A survey," *IEEE Communications Surveys amp; Tutorials*, vol. 20, no. 2, p. 1397–1417, Sep 2018.

[13] S. E. I. Institute, "Comparing insider it sabotage and espionage: A model-based analysis," Dec 2006. [Online]. Available: https://resources.sei.cmu.edu/asset_files/TechnicalReport/2006_005_001_14798.pdf

[14] Y. Sun, H. Xu, E. Bertino, and C. Sun, "A data-driven evaluation for insider threats," *Data Science and Engineering*, vol. 1, no. 2, p. 73–85, Apr 2016.

[15] C. insiders, "2019 insider threat report - fortinet," 2019. [Online]. Available: https://www.fortinet.com/content/dam/fortinet/assets/threat-reports/insider-threat-report.pdf?source=what-is-insider-threat

[16] M. Theis, "Common sense guide to mitigating insider threats," Dec 2018. [Online]. Available: https://resources.sei.cmu.edu/asset_files/TechnicalReport/2019_005_001_540647.pdf

[17] I. INSA, "Explanation of insa-developed insider threat definition," 2015. [Online]. Available: https://www.insaonline.org/wp-content/uploads/2018/10/INSA_InsiderThreat_definition-Flyer.pdf

[18] B. J. Wood, "An insider threat model for adversary simulation - researchgate," Dec 2013. [Online]. Available: https://www.researchgate.net/publication/228536014_An_insider_threat_model_for_adversary_simulation

[19] F. L. Greitzer and T. A. Ferryman, "Methods and metrics for evaluating analytic insider threat tools," *2013 IEEE Security and Privacy Workshops*, 2013.

[20] V. Mavroeidis, K. Vishi, and A. Josang, "A framework for data-driven physical security and insider threat detection," *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Aug 2018.

[21] N. Elmrabit, S.-H. Yang, and L. Yang, “Insider threats in information security categories and approaches,” *2015 21st International Conference on Automation and Computing (ICAC)*, Sep 2015.

[22] F. L. Greitzer, J. Strozer, S. Cohen, J. Bergey, J. Cowley, A. Moore, and D. Mundie, “Unintentional insider threat: Contributing factors, observables, and mitigation strategies,” *2014 47th Hawaii International Conference on System Sciences*, Sep 2014.

[23] C. Insiders, “2021 insider threat report [gurucul],” Jun 2021. [Online]. Available: https://www.cybersecurity-insiders.com/portfolio/2021-insider-threat-report-gurucul/

[24] A. Waldman, “Former ubiquiti engineer arrested for inside threat attack,” Dec 2021. [Online]. Available: https://www.techtarget.com/searchsecurity/news/252510411/Former-Ubiquiti-engineer-arrested-for-inside-threat-attack

[25] J. Stempel, “Pfizer sues departing employee it says stole covid-19 vaccine secrets,” Nov 2021. [Online]. Available: https://www.reuters.com/business/healthcare-pharmaceuticals/pfizer-sues-departing-employee-it-says-stole-covid-19-vaccine-secrets-2021-11-24/

[26] L. Columbus, “Dissecting the twitter hack with a cybersecurity evangelist,” Jul 2020. [Online]. Available: https://www.forbes.com/sites/louiscolumbus/2020/07/18/dissecting-the-twitter-hack-with-a-cybersecurity-evangelist/?sh=59ed226447df

[27] M. Lennon, “Insider steals data of 2 million vodafone germany customers,” Sep 2013. [Online]. Available: https://www.securityweek.com/attacker-steals-data-2-million-vodafone-germany-customers

[28] P. institute, “2022 ponemon cost of insider threats global report: Proofpoint us,” May 2022. [Online]. Available: https://www.proofpoint.com/us/resources/threat-reports/cost-of-insider-threats

[29] c. helpsystem, “New research shows security threats from insiders exploding but businesses aren’t ready to accept it,” 2015. [Online]. Available: https://www.clearswift.com/resources/press-releases/new-research-shows-security-threats-from-insiders

[30] F. IT, “Forrester insider threats drive data protection improvements (full report),” Apr 2022. [Online]. Available: https://www.imperva.com/resources/resource-library/white-papers/forrester-insider-threats-drive-data-protection-improvements-full-report/

[31] L. Whitney, T. Staff, C. R. Smith, B. Miles, S. Ingalls, C. Bohon, and F. Okeke, "How insider threats pose risks and challenges to any organization," Sep 2020. [Online]. Available: https://www.techrepublic.com/article/how-insider-threats-pose-risks-and-challenges-to-any-organization/

[32] W. Jiang, Y. Tian, W. Liu, and W. Liu, "An insider threat detection method based on user behavior analysis," *IFIP Advances in Information and Communication Technology*, p. 421–429, 2018.

[33] Kim, Park, Kim, Cho, and Kang, "Insider threat detection based on user behavior modeling and anomaly detection algorithms," *Applied Sciences*, vol. 9, no. 19, p. 4018, 2019.

[34] D. C. Le and N. Zincir-Heywood, "Exploring adversarial properties of insider threat detection," *2020 IEEE Conference on Communications and Network Security (CNS)*, 2020.

[35] X. Sun, Y. Wang, and Z. Shi, "Insider threat detection using an unsupervised learning method: Copod," *2021 International Conference on Communications, Information System and Computer Engineering (CISCE)*, 2021.

[36] X. Wang, Q. Tan, J. Shi, S. Su, and M. Wang, "Insider threat detection using characterizing user behavior," *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, 2018.

[37] J. U. Mills, S. M. Stuban, and J. Dever, "Predict insider threats using human behaviors," *IEEE Engineering Management Review*, vol. 45, no. 1, p. 39–48, 2017.

[38] M. Singh, B. Mehtre, and S. Sangeetha, "User behaviour based insider threat detection in critical infrastructures," *2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC)*, 2021.

[39] M. Dahmane and S. Foucher, "Combating insider threats by user profiling from activity logging data," *2018 1st International Conference on Data Intelligence and Security (ICDIS)*, 2018.

[40] B. Tang, Q. Hu, and D. Lin, "Reducing false positives of user-to-entity first-access alerts for user behavior analytics," *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2017.

[41] J. Datta, R. Dasgupta, S. Dasgupta, and K. R. Reddy, "Real-time threat detection in ueba using unsupervised learning algorithms," *2021 5th International Conference on Electronics, Materials Engineering amp; Nano-Technology (IEMENTech)*, 2021.

[42] P. A. Legg, “Human-machine decision support systems for insider threat detection,” *Data Analytics and Decision Support for Cybersecurity*, p. 33–53, 2017.

[43] B. Haim, E. Menahem, Y. Wolfsthal, and C. Meenan, “Visualizing insider threats,” *Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion*, 2017.

[44] S. Khaliq, Z. U. Abideen Tariq, and A. Masood, “Role of user and entity behavior analytics in detecting insider attacks,” *2020 International Conference on Cyber Warfare and Security (ICCWS)*, 2020.

[45] A. Das, M.-Y. Shen, and J. Wang, “Modeling user communities for identifying security risks in an organization,” *2017 IEEE International Conference on Big Data (Big Data)*, 2017.

[46] E. Novikova, I. Kotenko, and I. Murenin, “The visual analytics approach for analyzing trajectories of critical infrastructure employers,” *Energies*, vol. 13, no. 15, p. 3936, 2020.

[47] M. Singh, B. Mehtre, and S. Sangeetha, “User behavior profiling using ensemble approach for insider threat detection,” *2019 IEEE 5th International Conference on Identity, Security, and Behavior Analysis (ISBA)*, 2019.

[48] A. Y. Khan, R. Latif, S. Latif, S. Tahir, G. Batool, and T. Saba, “Malicious insider attack detection in iots using data analytics,” *IEEE Access*, vol. 8, p. 11743–11753, 2020.

[49] R. Nasir, M. Afzal, R. Latif, and W. Iqbal, “Behavioral based insider threat detection using deep learning,” *IEEE Access*, vol. 9, p. 143266–143274, 2021.

[50] H. Peng and W. Wang, “Detecting masqueraders by profiling user behaviors,” *2018 Eighth International Conference on Instrumentation amp; Measurement, Computer, Communication and Control (IMCCC)*, 2018.

[51] K. Król and D. Zdonek, “Analytics maturity models: An overview,” *Information*, vol. 11, no. 3, p. 142, 2020.

[52] d. science, “4 levels of data maturity every manager must know,” Jul 2020. [Online]. Available: https://computd.nl/demystification/4-levels-of-data-maturity/

[53] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A design science research methodology for information systems research,” *Journal of Management Information Systems*, vol. 24, no. 3, p. 45–77, 2007.

[54] F. BA, “Forcepoint | security simplified,” 1997. [Online]. Available: https://www.forcepoint.com/sites/default/files/resources/datasheets/ueba-platform-architecture-overview.pdf

[55] ——, 1997. [Online]. Available: https://www.forcepoint.com/

[56] D. NL, “Dimensions of data quality (ddq) - stichting dama,” 2020. [Online]. Available: https://dama-nl.org/wp-content/uploads/2020/09/DDQ-Dimensions-of-Data-Quality-Research-Paper-version-1.2-d.d.-3-Sept-2020.pdf

[57] N. N. Ramly, F. M. Nor, N. H. Ahmad, and M. H. Aziz, “Comparative analysis on data visualization for operations dashboard,” *International Journal of Information and Education Technology*, p. 287–290, 2012.

[58] n. business, “Intellectual property: The basics,” 2022. [Online]. Available: https://www.nibusinessinfo.co.uk/content/advantages-protecting-intellectual-property

[59] J. Truong, “Security best practices for removable media and devices,” Oct 2021. [Online]. Available: https://hackernoon.com/security-best-practices-for-removable-media-and-devices

[60] J. Burt, “Kaspersky,” Apr 2020. [Online]. Available: https://www.darkreading.com/security-management/usb-devices-still-a-threat-to-businesses-kaspersky-finds/a/d-id/746429

[61] P. b. Coos, “What is insider data exfiltration?” Jun 2022. [Online]. Available: https://www.endpointprotector.com/blog/what-is-insider-data-exfiltration/

[62] D. Diana, “Data exfiltration: Most common techniques and best prevention tactics,” Nov 2022. [Online]. Available: https://www.xorlab.com/en/blog/data-exfiltration-most-common-techniques-and-best-prevention-tactics

[63] R. Wieringa, “Design science as nested problem solving,” *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology - DESRIST ’09*, May 2009.